Mining of Statistical Temporal Patterns in Video

Shih-Fu Chang

Digital Video and Multimedia Lab
ADVENT University-Industry Consortium
Columbia University

June 20th 2003

http://www.ee.columbia.edu/dvmm
email: sfchang@ee.columbia.edu

1

Joint Work

- with Lexing Xie, Ajay Divakaran, and Huifang Sun
  - "Unsupervised Mining of Statistical Temporal Structures in Video," book chapter in *Video Mining*, edited by A. Rosenfeld, D. Doremann, and D. DeMenthon, Kluwer, 2003.
  - "Unsupervised Discovery of Multilevel Statistical Video Structures Using Hierarchical Hidden Markov Models," IEEE Intern. Conf. on Multimedia and Exhibition (ICME03), Baltimore, July 2003.
  - "Feature Selection for Unsupervised Discovery of Statistical Temporal Structures in Video, IEEE Intern. Conf. on Image Processing (ICIP03), Spain, Sept. 2003.
  - "Structure Analysis of Soccer Video with Hidden Markov Models," IEEE Intern. Conf. on Acoustic, Speech and Signal Processing, (ICASSP02), Orlando, FL, May 2002.
- Papers available at http://www.ee.columbia.edu/dvmm/publications.htm

2

Patterns Abound in Video

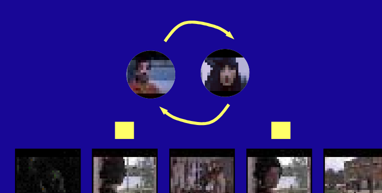
patterns exist at multiple scales
(e.g., recurrent views, plays in sports)

3

More Video Pattern Examples

News: Story

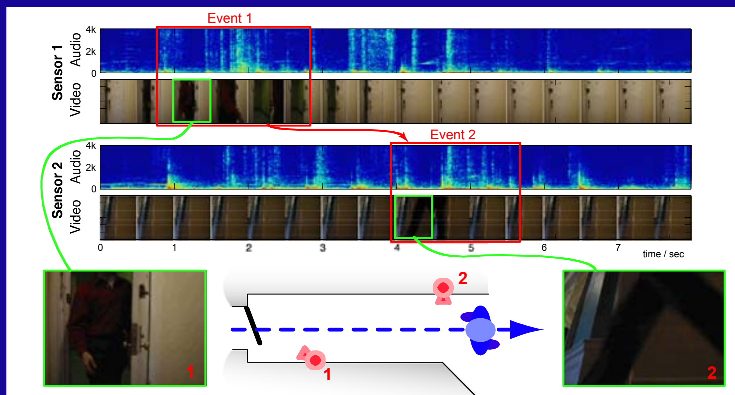
Films: Sporadic Patterns like Dialog, Point of View



4

Spatio-temporal patterns in distributed sensor networks

(Chang, Ellis, Eleftheriadis, Wang)

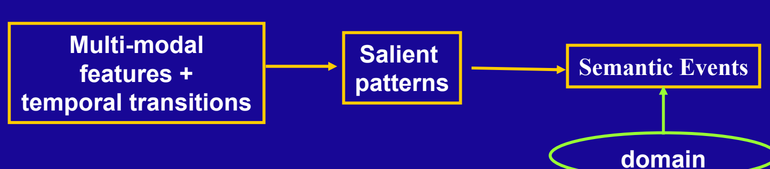


[Diagram by Ellis]

- In a distributed sensor network, patterns exist at
  - individual sensors: passing persons, footstep sounds
  - across sensors: “door opening likely followed by passing person in T seconds”
- Event breaking normal patterns → alerts

Patterns related to semantics

- Patterns correspond to semantics in specific domains



Patterns – conceptually defined

- Patterns are recurrent temporal segments in one or more sequences with predictable syntactic characteristics – MM feature and temporal transition.
  - Can be statistical or rule-based
  - Can be spare or dense, regular or irregular
- Other pattern-rich data :
  - Network traffics and transactions (for security, e-commerce)
  - Speech, language, and music
  - Biological Data (DNA or protein sequences)

7

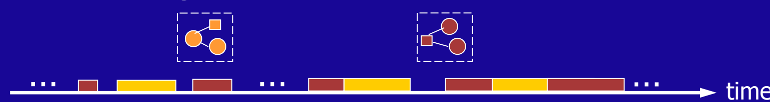
Conventional Video Indexing Tasks

- **Supervised methods**
  - **Expert-developer collaborative solutions**
  - **Given a domain**
    - identify important events, concepts, and structures
    - Develop best detectors and classifiers
  - **Problems**
    - Lack of scalability
    - Cannot address data variations (e.g., personal collection)
- **Unsupervised clustering methods**
  - **Explore feature similarity and time proximity of scenes**
  - **Rich temporal transitional characteristics unexplored**

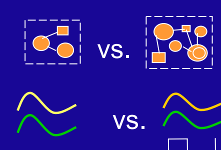
8

Challenge: Unsupervised Pattern Discovery

- Given a new domain/data, discover patterns automatically
  - E.g., Consumer, surveillance, and personal media log
- Technical Objectives:
  - Find appropriate spatio-temporal statistical models
  - Locate segments that match such models

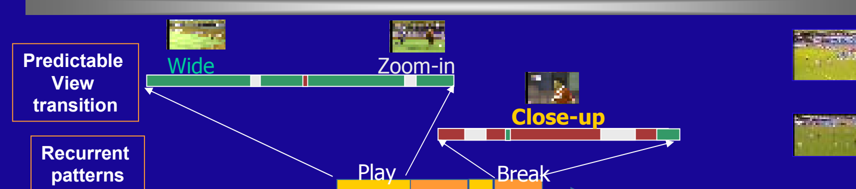


- Issues
  - What's the adequate class of models
  - How to determine model complexity
  - What's a "good" feature set



9

Simple Observations of Patterns

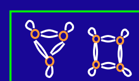
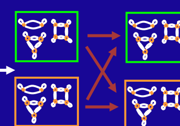


- Distinctive patterns are characterized by feature distribution and temporal transitions

10

Unsurprising evidence for Using HMM

- Prior works: HMM for temporal structures in video (*Huang, Liu, & Wang* news , *Wolf* Basketball, *Liu & Kender* documentary, *Naphade & Huang* films)
- Supervised learning of model parameters and high-level transitions
- ML classification of new videos plus temporal smoothing

(labeled & chunked Plays & Breaks)

Labeled Segments (color, motion, audio) → Train HMM families For plays/breaks → Estimate high-level transition prob. →

(testing phase)

Test Data Segment → ML Classification + Model Refinement → Play-Break sequence temporal smoothing (Dynamic Programming)

11

HMM effective for parsing soccer structures

[Xie, Chang, Divakaren, Sun 02]

- 4 test clips, 15~25 min., various countries
- Cross Validation

Test set	Training Set			
	Argentina	KoreaA	KoreaB	Espana
Argentina	87.2%	82.5%	82.5%	80.6%
KoreaA	78.1%	84.3%	84.3%	79.8%
KoreaB	79.9%	85.3%	85.3%	89.6%
Espana	79.9%	89.6%	89.6%	81.7%

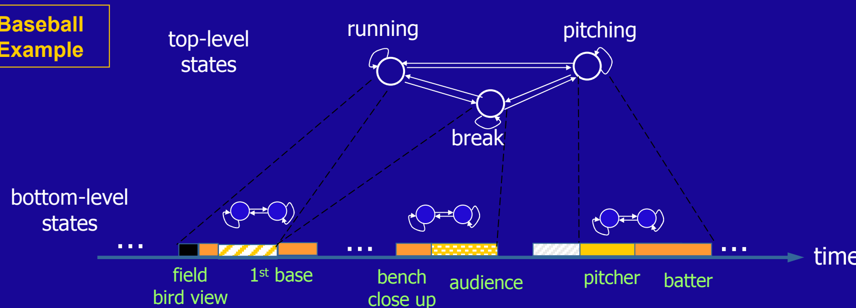
• Avg. Play-Break Classification Rate: 83.5%
vs. 60% of blind classification
• Boundary timing accuracy: 62% within 3 seconds
• The good classification provides preliminary evidence supporting HMM model and the selected features

12

Generalize to Unsupervised Discovery: Hierarchical Hidden Markov Model

- HHMM successfully used in tracking and recognition
- Intuitive Representation for Videos
  - High-level states represent distinct events
  - Presence of each event produces observations modeled by low-level HMMs

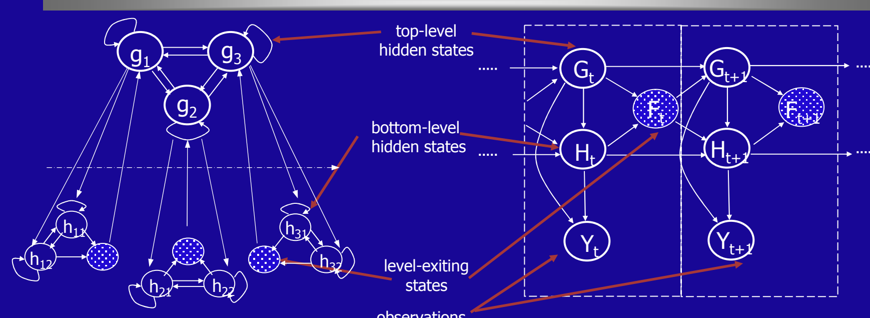
Baseball Example



13

Hierarchical HMM

[Fine, Singer, Tishby '98]
[K. Murphy, '01]



Tree-Structured representation

DBN representation, unrolled in time

- Flexible Control Structure (Bottom-up control with exit state)
- Extensible to multiple levels and distributions
- Applications to video event discovery [Clarkson & Pentland '99, Naphade & Huang '02]
  - left-right models with fixed model and features
- Given HHMM, efficient inference technique available
  - Complexity $O(D \cdot T \cdot Q^{\alpha D})$, α=1.5 to 2 [Murphy '01, Xie et al '02]

14

But Still Some Critical Issues

- No knowledge about the model structure and complexity
- No knowledge about optimal feature set

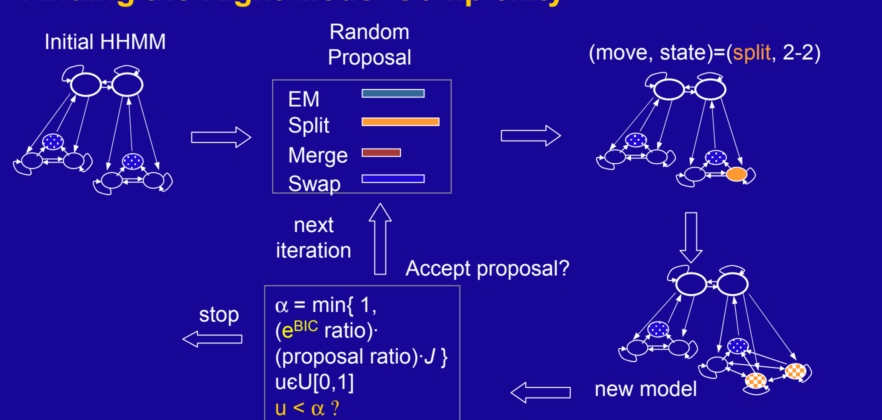
↓

- data-driven approach
  1. MCMC random model search for model adaptation
  2. Hybrid wrapper/filtering and Bayesian criteria for feature selection

15

MCMC-Bayesian Adaptation → Finding the Right Model Complexity

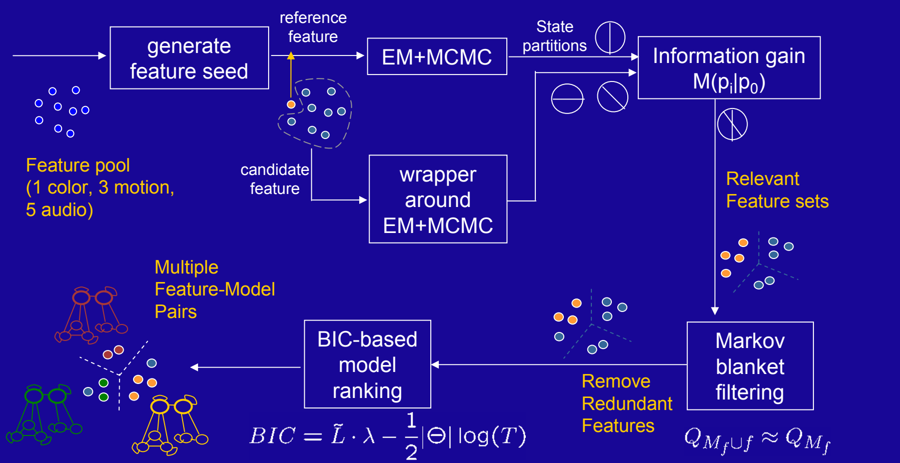
[Xie, Chang, Divakaren, Sun ICME 03]



Acceptance probability

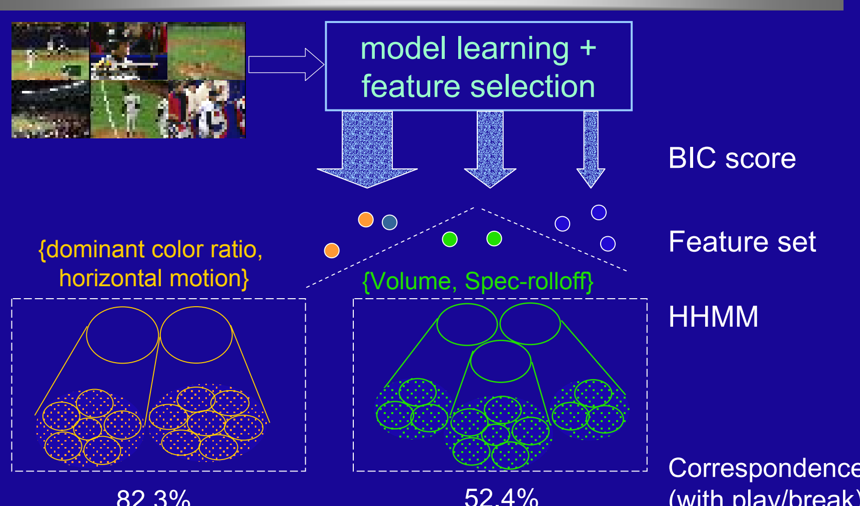
$$\alpha = \min\{1, \underbrace{\frac{P(x|\hat{m})}{P(x|m)}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{P(\hat{m})}{P(m)}}_{\text{prior ratio}} \cdot \underbrace{\frac{P(\hat{m},\hat{u})}{P(m,u)}}_{\text{proposal ratio}} \cdot \underbrace{\frac{\partial(\hat{m},\hat{u})}{\partial(m,u)}}_{\text{Jacobian}}\}$$

Feature Selection [Xie, Chang, Divakaren, Sun ICIP 03]



17

Test Case: Baseball



18

Comparing with supervised methods

Unsupervised discovery of models and features

Test clip	Feature Set	# states	Correspondence
Korea	DCR,Mx	2~4	75.2%
Spain	DCR,Volume	2~3	74.8%
Baseball	DCR,Mx	2	82.3%

Comparison with supervised approaches (soccer)

Model	Supervised?	Adaptation?	Accuracy
HHMM	N	Y	75.2%
HHMM	Y	N	75.0%
HMM**	Y	N	75.5%

* Korean soccer, MPEG-7 soccer dataset, 25 minutes long
basic feature set: dominant color ratio(DCR), motion intensity(MI)

** Trained HMM family + post smoothing with dynamic programming

- **Automatic approaches find meaningful clusters and features**
- **Unsupervised solutions achieve comparable performance in play/break classification**

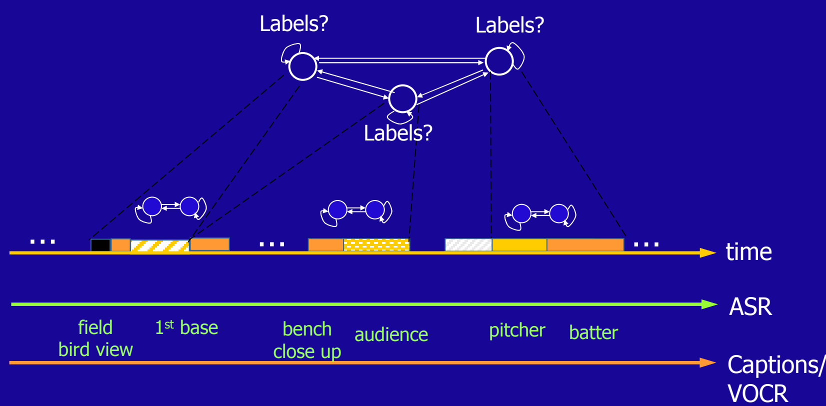
Open Issues

- **Evaluation**
  - **Is classification based on manually selected semantic classes adequate?**
    - No predefined semantic targets used in mining
  - **Miss/False rates when compared to salient patterns identified manually**
- **How to annotate discovered patterns?**
- **How to address granularity and sparseness of patterns?**
- **Low-level features vs. mid-level features (e.g., audio class, object detector)**

20

How to assign meanings to discovered patterns?

- Idea: learning statistical correspondence between patterns and synchronized text or metadata streams



21

Conclusions

- The spatio-temporal dimensions in videos offer rich patterns to be mined
- Video pattern mining facilitates scalability and personalization
- Data-driven approaches using models like HHMM and feature selection methods show interesting results
- Many issues remain
  - evaluation, semantic tagging, granularity, fusing with domain knowledge

22

More Information

- Columbia DVMM Lab
  http://www.ee.columbia.edu/dvmm

- Publications
  http://www.ee.columbia.edu/dvmm/publications.htm

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