Real-Time Content-Based Adaptive Streaming of Sports Videos

Shih-Fu Chang, Di Zhong, and Raj Kumar

Digital Video and Multimedia Group
ADVENT University/Industry Consortium
Columbia University

December 14, 2001, CBAVIL

Live Sports Video Filtering and Navigation

Linear/Passive Video

Interactive Video

- Time sensitive interest
- Massive production and audience
- Time compressibility
- Temporal structure and production rules

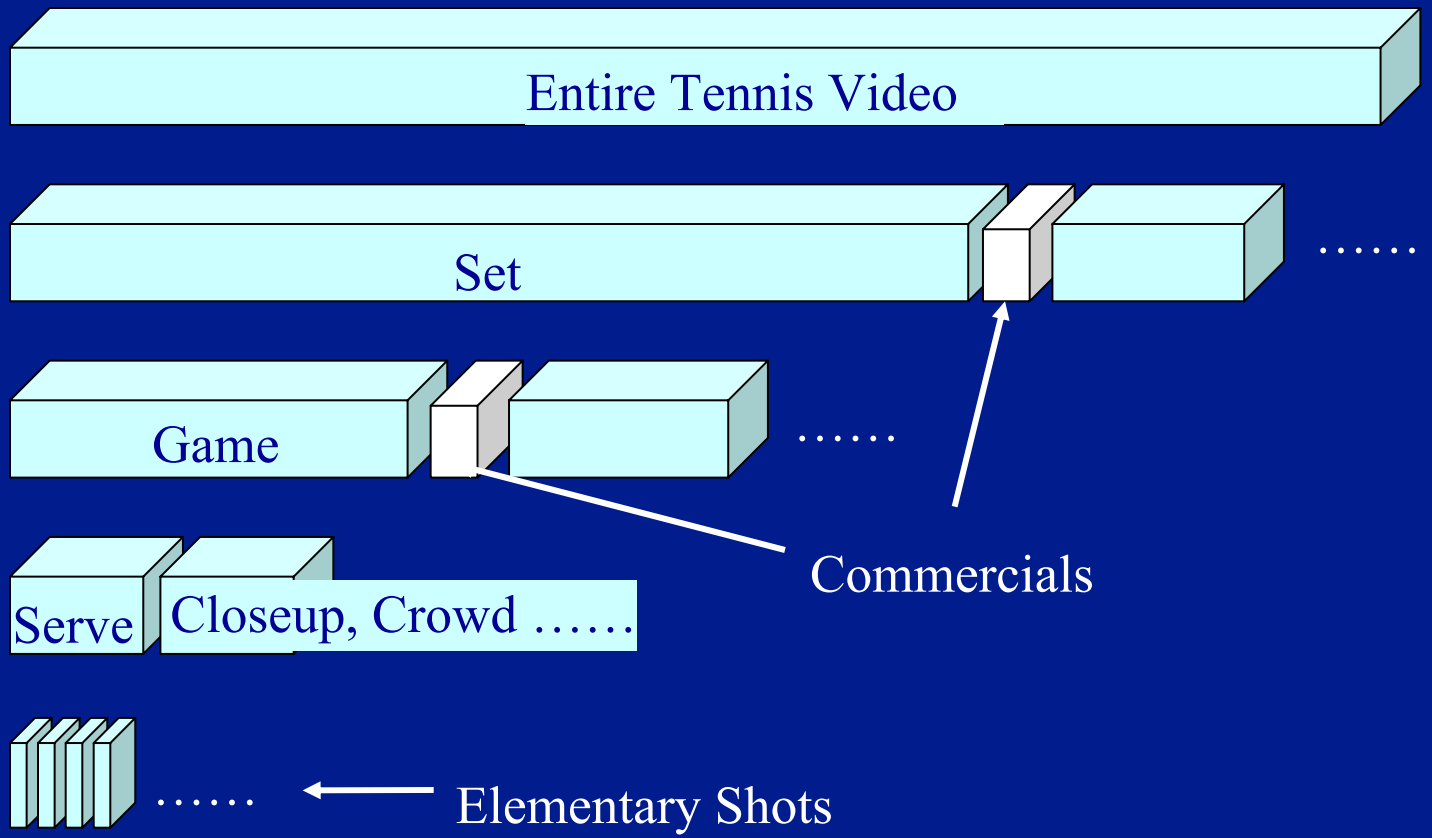
Beyond Ringer and Clip Download

Services:

- Messaging
- Localized information
- Media/game download
- Multi-person games
- TV phone
- On-site purchase

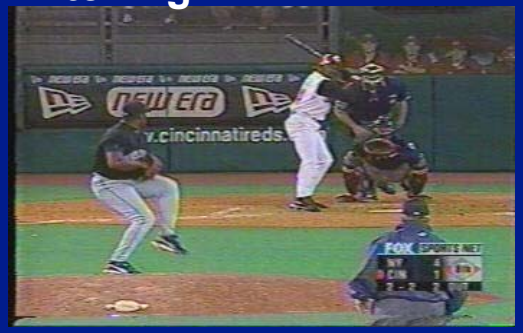
Time-sensitive short video messages suiting personal needs

Regular Structure and Views in Sports Video



Regular set of views

Pitching

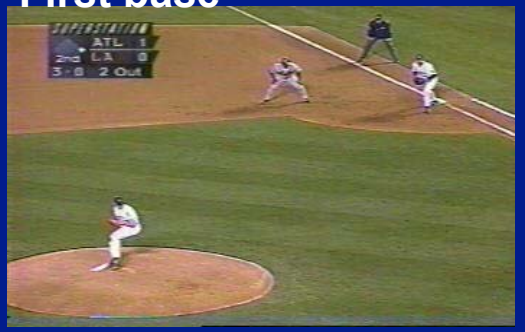


Close-up pitcher

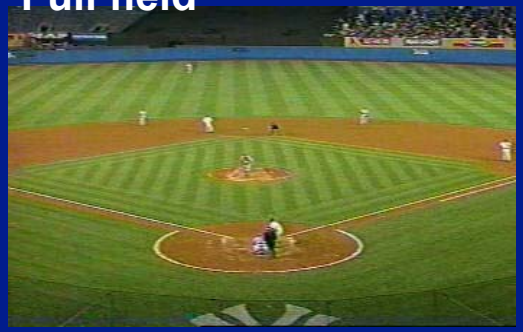
Catcher

Base hit

First base



Full field



- **Important issues:**
  canonical view ←→ beginning of event
  view transition pattern ←→ types of events

Prior work

- Basketball: Saur, Tan *et al* '97
  - Use camera motion to detect events (wide angle, close up, fast break)
- Tennis: Sudhir *et al*, '98
  - Detect play events (rally, net play *etc*) by detecting the court lines and players
- Soccer: Gong *et al* '95
  - Classify location and event by detecting lines, motion, and players
- Baseball: Rui et al, '00
  - Detecting highlight from audio track
- Sports Recurrent Semantic Event Extraction
  - Zhong/Chang '01: Pitching, Serving, etc
  - Xie/Chang '01: play/break in soccer

Review of RSU Extraction

- Framework for automatic structure parsing (canonical view detection)
- Real-time processing → simple features and compressed-domain
- Combine learning tools and systematic rule acquisition
- Combine frame-level and object-level analysis

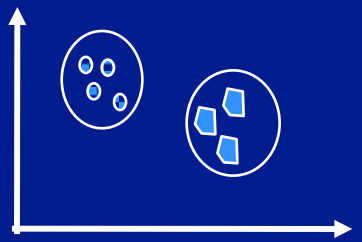
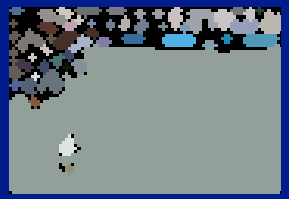
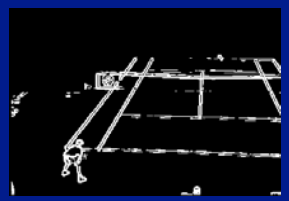
System for "serve" view detection



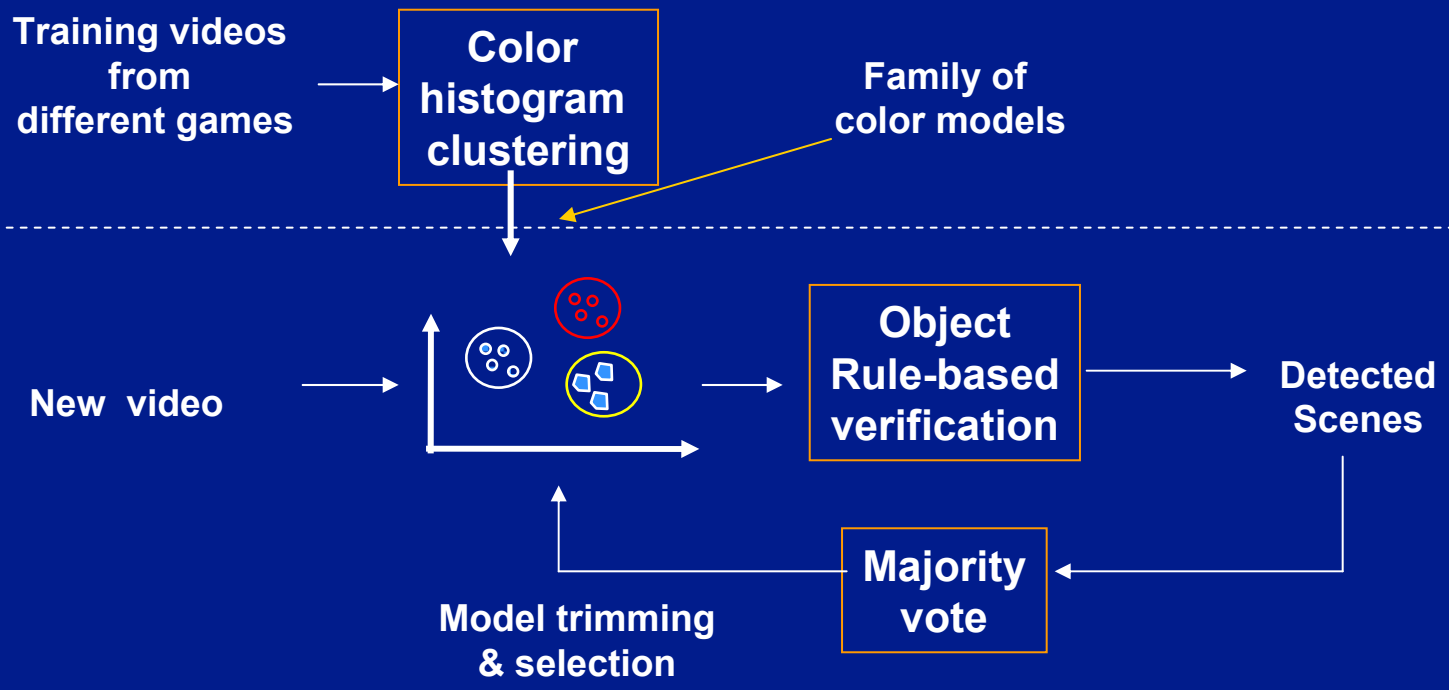

Color, lines, player

Every I/P frames | Key frames/shot | Down-sampled I/P frames

Compressed video → Shot Detection → Adaptive Color Filter → Object level verification → Canonical View Detection

Adaptive color filtering



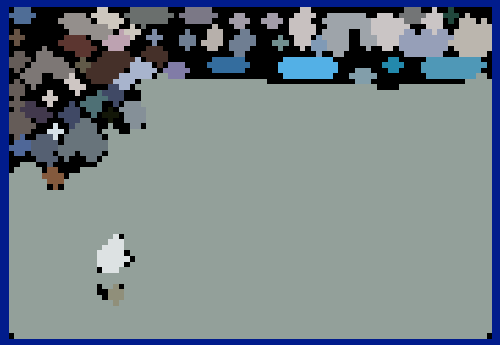
- Improve robustness over source variations
- Improve speed

Object rule verification

Segmentation based verification

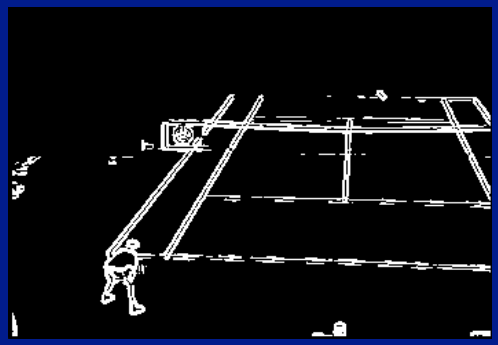
For each key frame in a shot

- **automatic region segmentation based on color, edge, and motion projection**
- **detect court region based on color consistency, size, and location**
- **region tracking and background layer modeling to detect moving objects**
- **verify the size and position of court and player**



Object-level rules

Edge based Verification

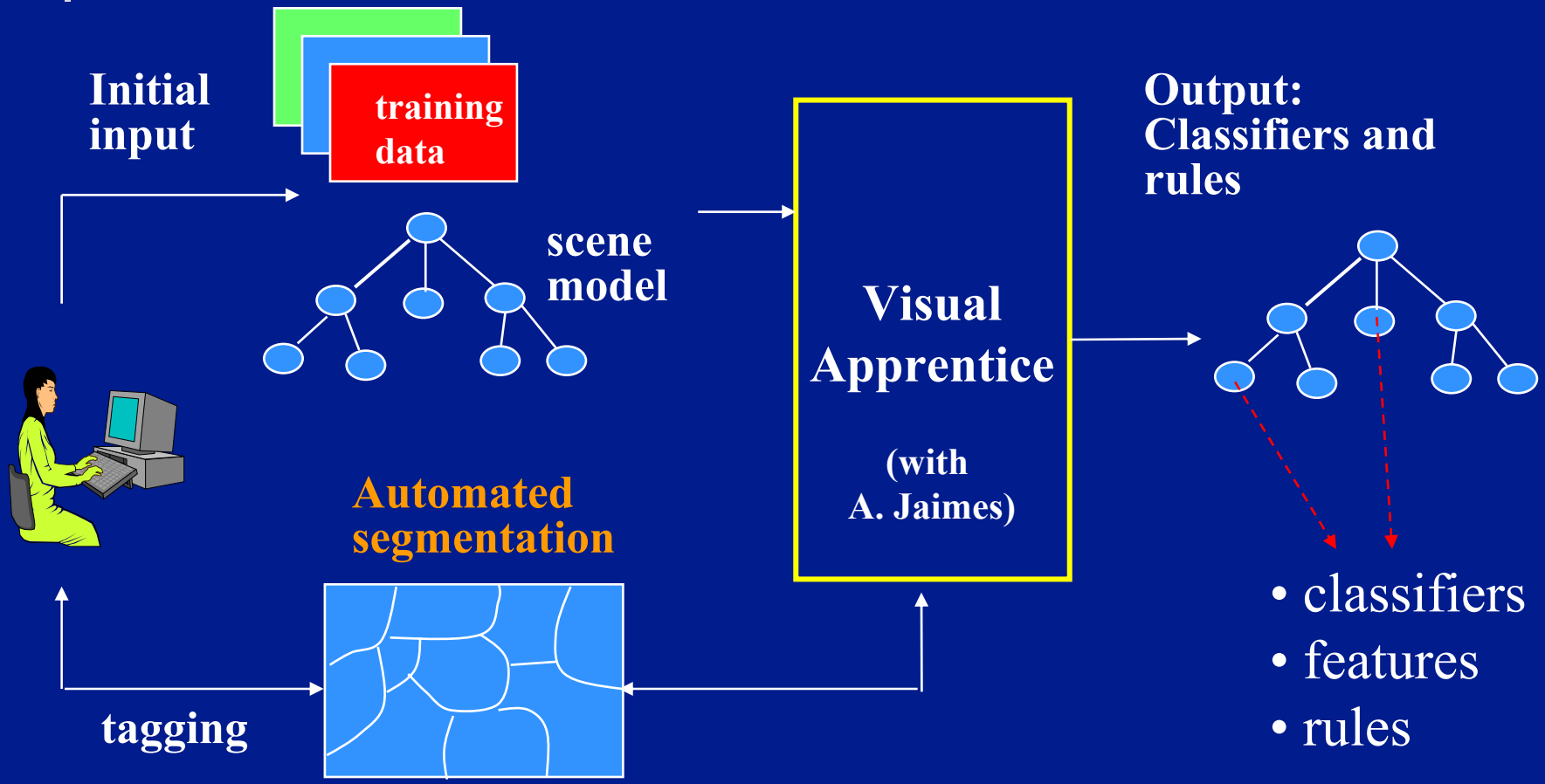


- Using Hough transform to detect court lines within the court field

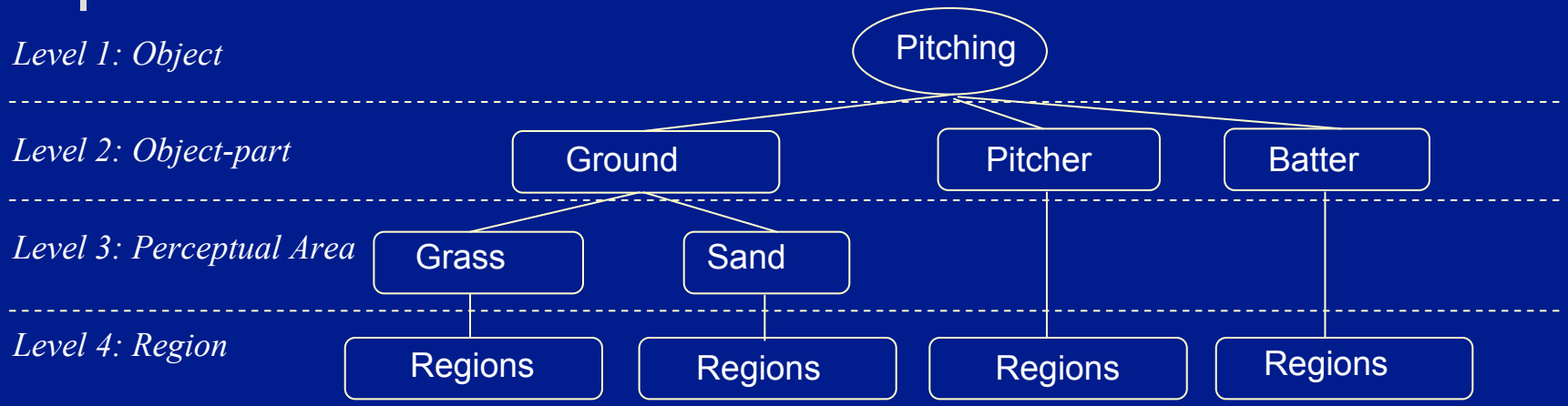
- Verify number of horizontal and vertical court lines

- Verify the locations of the lines

Systematic learning of object rules



Extension to other types: learning new rules



Experiment results

Training data – 3-4 different games (10 mins each)

Testing data – 1 hour tennis and 1 hour baseball

	Ground truth	# of Hits	# of False
Tennis (Serve)	89	82 (92%)	2
Baseball (Pitch)	93	90 (97%)	4

Case 1: initial segment of testing data included in training set

Experiment Results – cont.

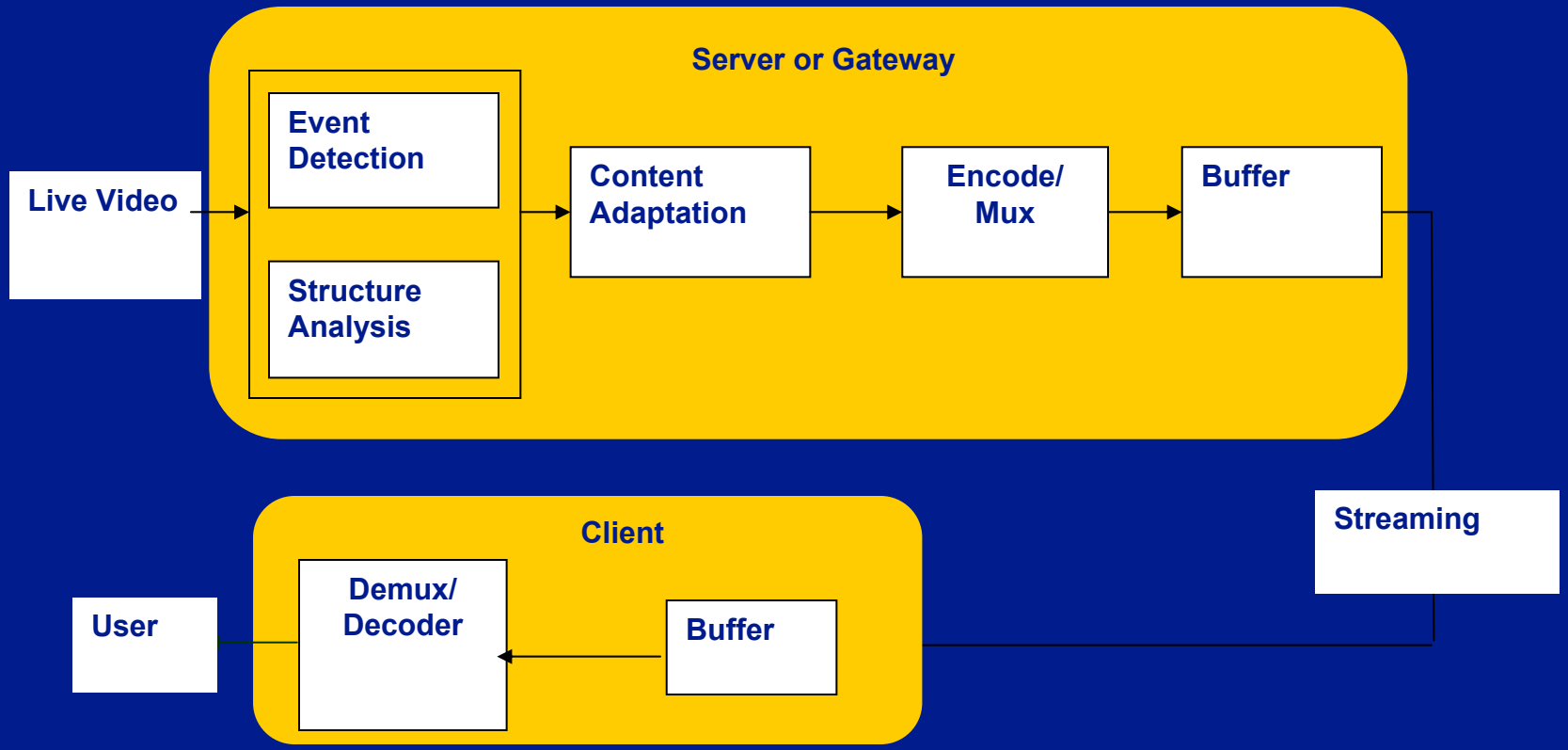
Case 2: initial segment of testing data excluded from training set

	Ground truth	# of Hit	# of False
Tennis (Serve)	74	68 (92%)	1
Baseball (Pitch)	57	56 (98%)	5

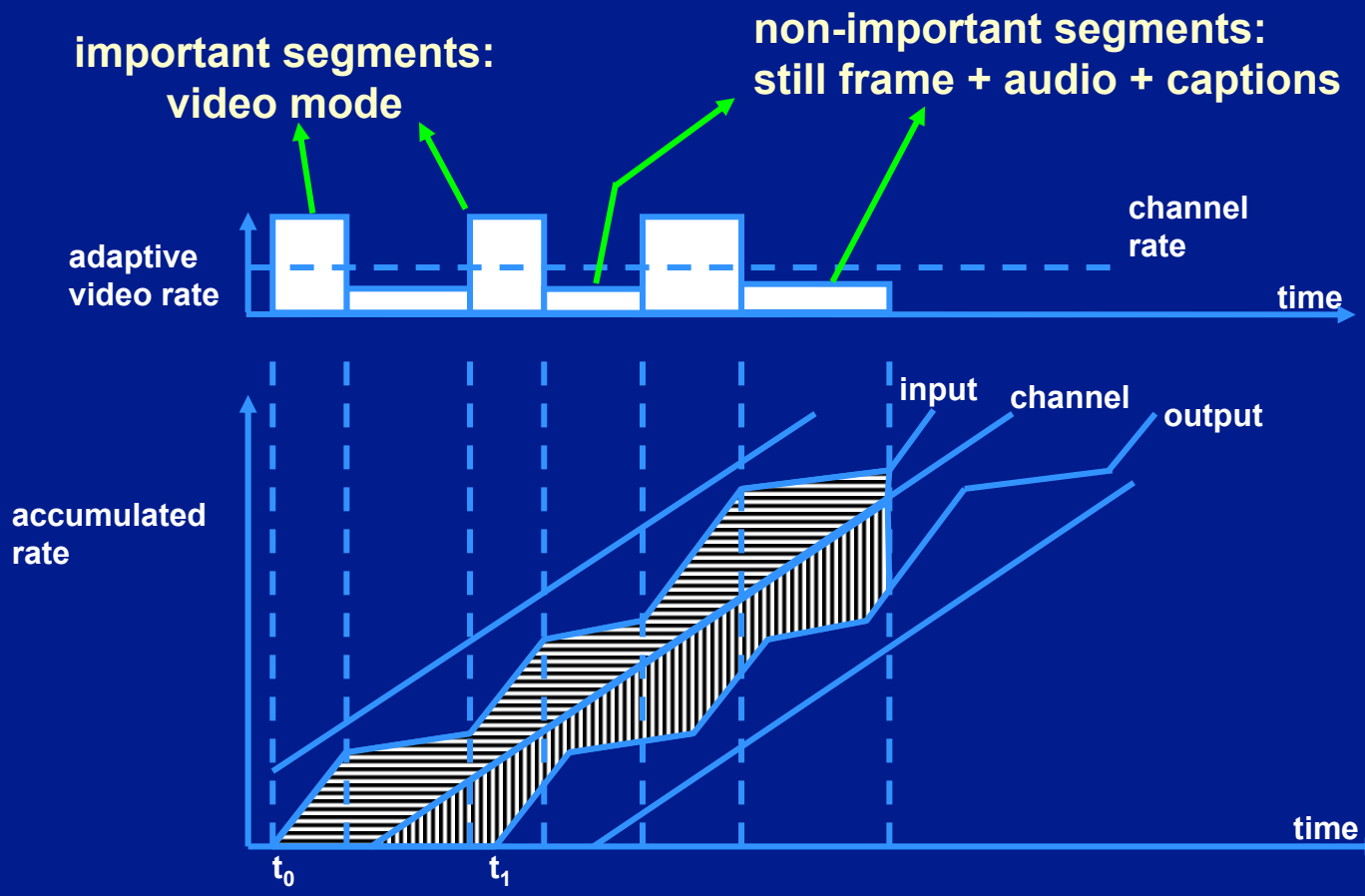
- **Key factors:**
  - **Comprehensive color models + adaptive selection**
  - **Error correction by object-level verification**

Application: Time-Sensitive Mobile Video

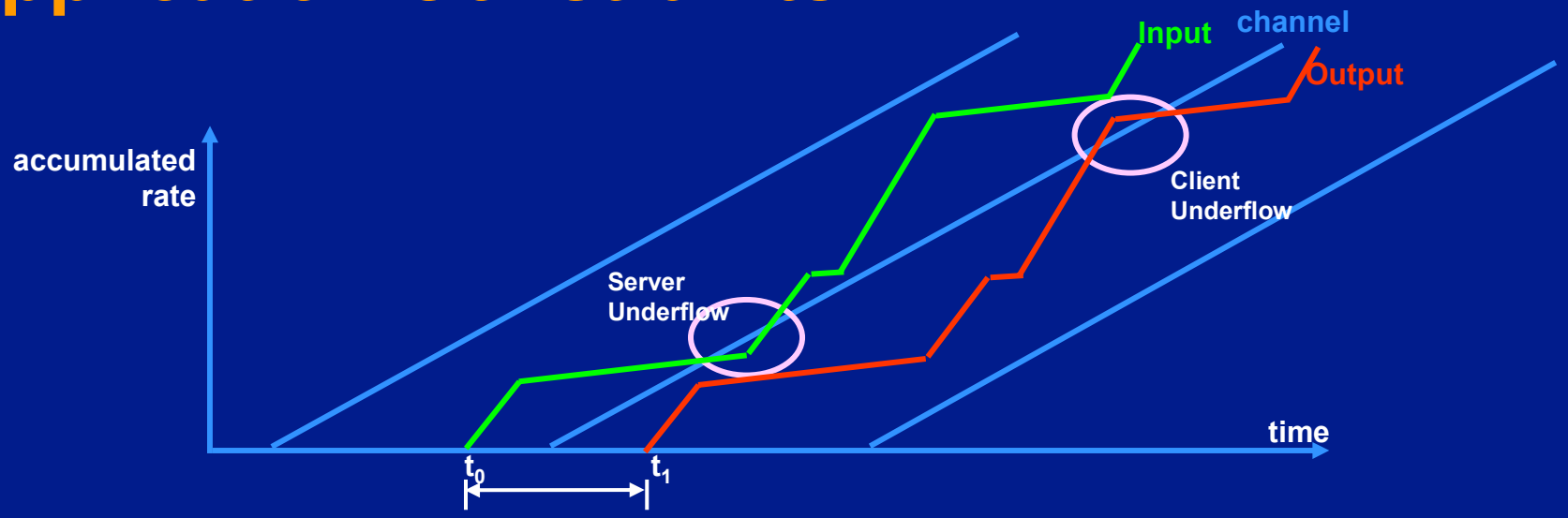
Content-based Adaptive Streaming



Content Adaptive Streaming



Application Constraints



- Buffer size: not a constraint
  - 8MB buffer can store 2000 sec (32Kbps) - 250 sec (256Kbps)
- Playback delay is the main constraint
  - E.g., real-time event delivered in 60 sec
- Client error more serious than server error
  - Different handling policies
    - Degrade to Channel Rate vs. Freeze and resume

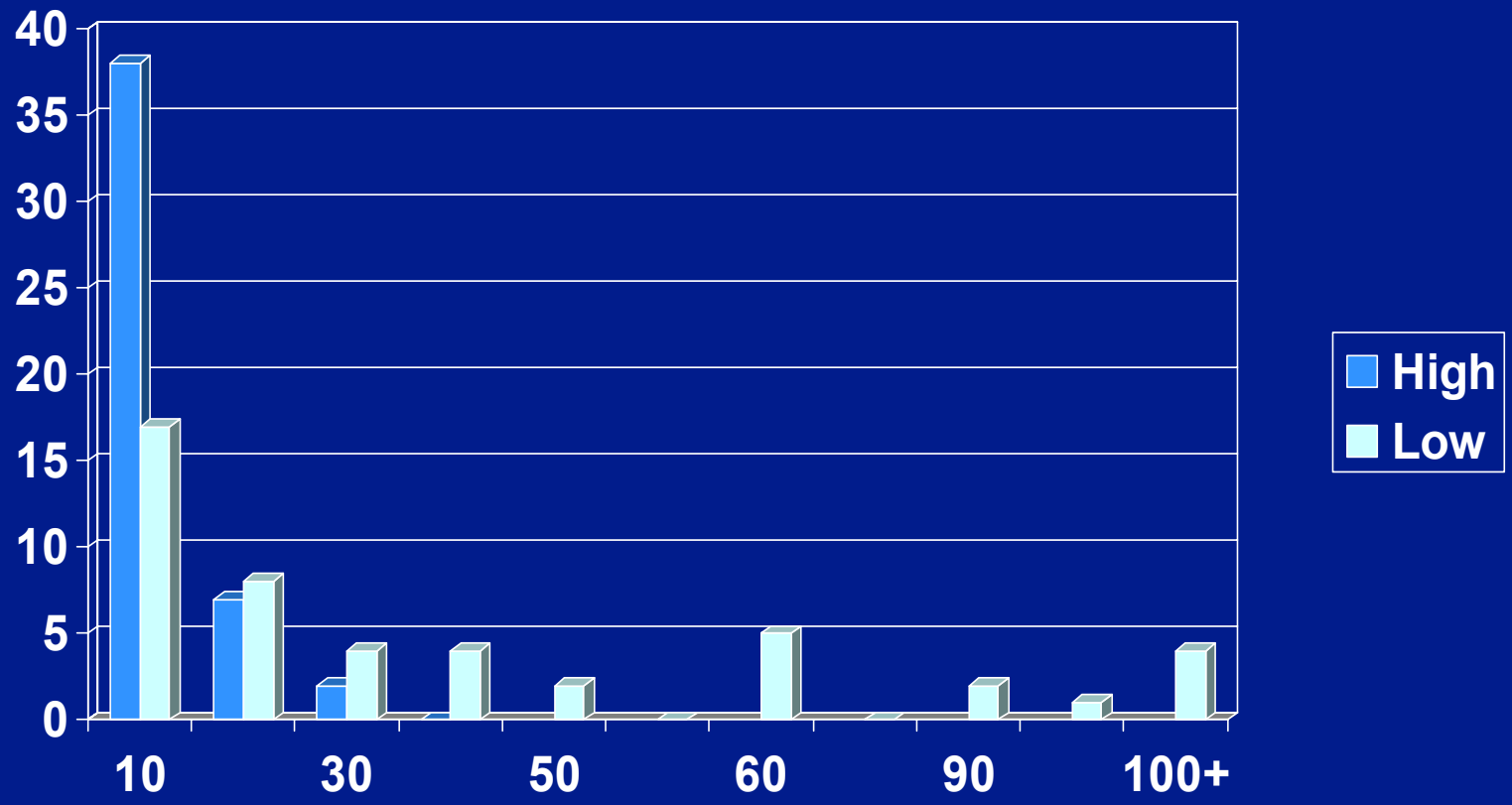
Simulation Setup

- 2 full innings, concatenalted
- No commercials
- 94 segments (47 h+l)
- Means(h,l)=7.05, 35.7
- Var(h,l)=1.44,2.4
- MaxL(h,l)=23,208

- Low rate = 0.2 * channel rate, with a min. of 10 kbps
- Avg. rate = 0.9 * channel rate

  i.e. utilization efficiency = 90%

Content Distribution



Server Underflow Rate

delay / avg rate	*30 sec*	*60 sec*	*120 sec*	*240 sec*
300 kbps	6%	6%	6%	4%
128 kbps	6%	6%	6%	4%
64 kbps	6%	6%	6%	4%
32 kbps	6%	6%	6%	3%
16 kbps	13%	6%	4%	1%

Client Underflow Rate

delay / avg rate	*30*	*60*	*120*	*240*
300	11	9	8	2
128	11	9	8	2
64	11	9	8	2
32	11	9	7	0
16	19	7	0	0

Recurrent semantic units in soccer video?

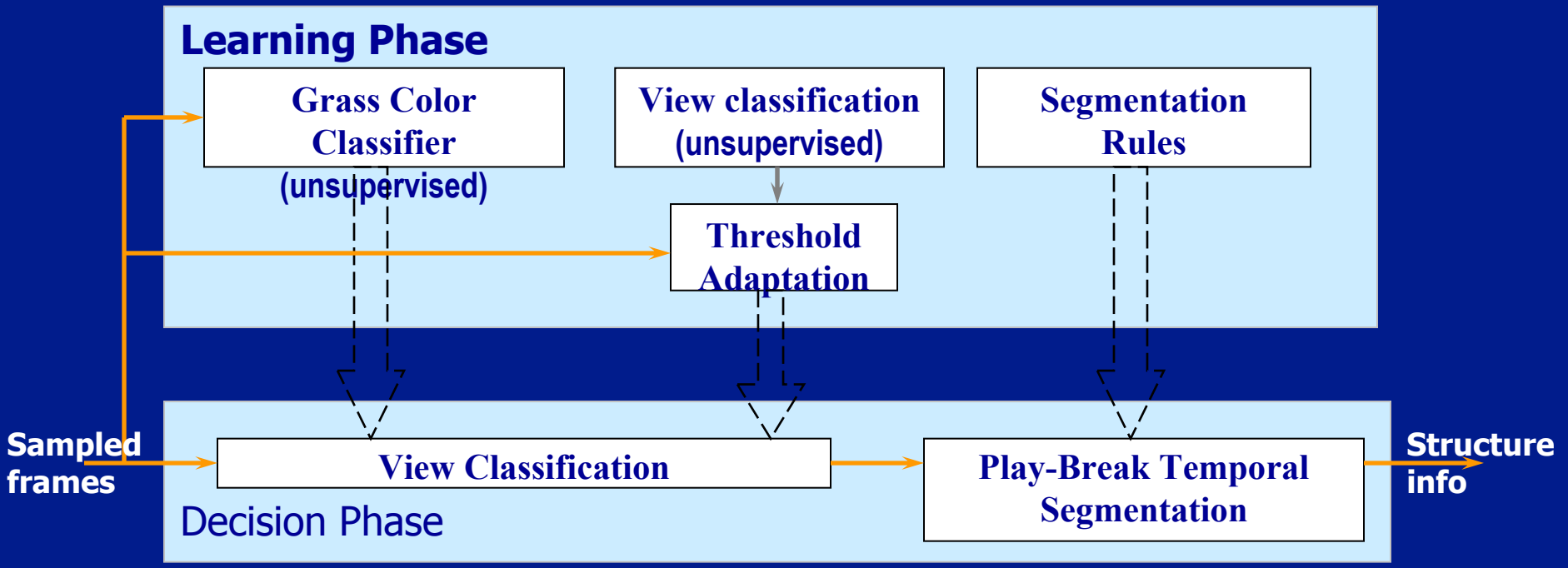
- Game → a sequence of play and break segments
- Play/break → No canonical views/events
- Sporadic events (start and end)
- Shot boundary ≠ play boundary

Heuristic Model: Features→ Views → Structure

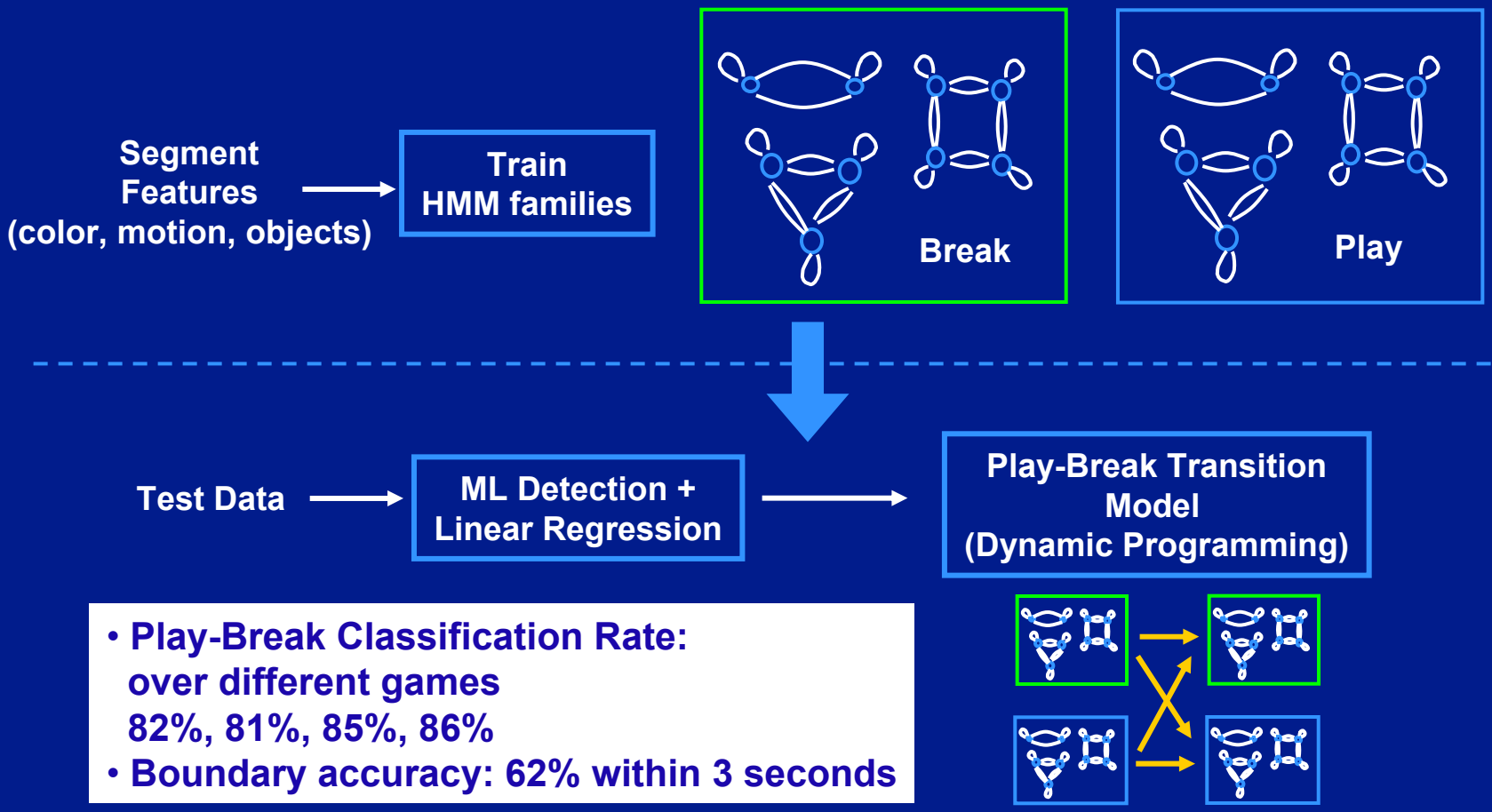
global

zoom-in

close-up



Modeling Transitions with HMM



- Play-Break Classification Rate:
  over different games
  82%, 81%, 85%, 86%
- Boundary accuracy: 62% within 3 seconds

Simple Event Detection

- **Baseball**

  - simple constraints to detect runs/hits after pitch

  - detection of camera motion and field

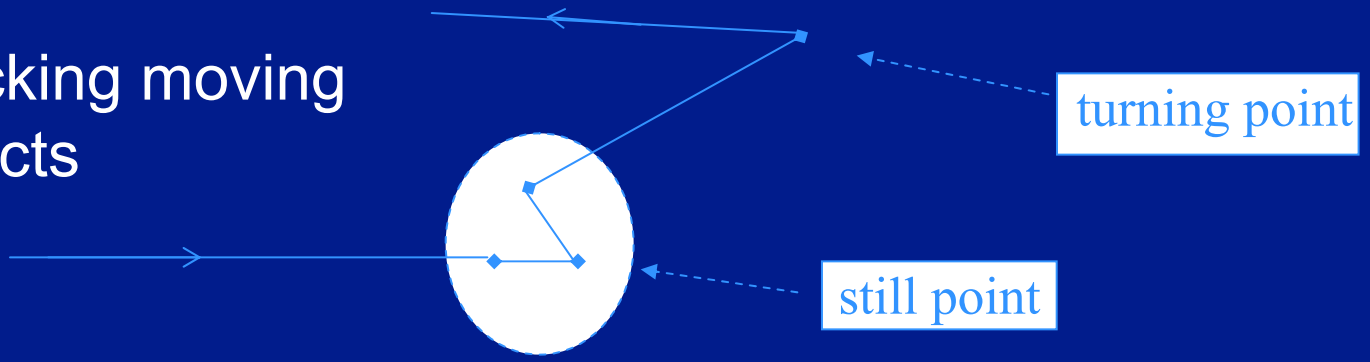
- **Tennis**

  - detect play types (net, base) and number of strokes

  - Extract and analyze player trajectory

Detecting Player Behaviors

Tracking moving objects



	# of Net Plays	# of Strokes
Ground Truth	12	221
Correct Detection	11 (92%)	216 (97%)
False Detection	7	81

Conclusions

- A real time system for view detection and temporal structure parsing
  - multi-stage, multi-time-scale, frame + object level
  - compressed domain processing
  - systematic learning of domain rules
- Systematic method for extracting Fundamental Semantic Units (FSU)
- Promising application in “content-based adaptive streaming and real-time messaging”

Current work

- Integrate visual analysis with
  - Video text recognition
  - Audio cue analysis
  - Speech and transcript analysis
- Event Detection
  - E.g., player, runs, goals

END