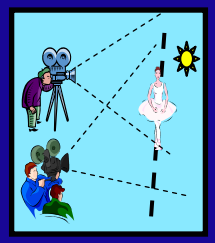
Panel: Technical and Business Challenges in Content-Based Retrieval

Shih-Fu Chang

Digital Video and Multimedia Group
Columbia University
12/14/2001

CBAIVL Workshop Panel

Consider End-to-End Content Chain



Production & post-production

text

metadata

A-V Data

feature extraction

pattern recognition

indexing

speech/text recognition

Analysis

textual search

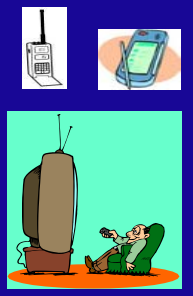
similarity matching

structure browsing

event/object search

summary

Interaction



Viewer

Production model

Media Integration

Viewer model

Types of Content Indexing

- Metadata preservation → XML-based language
  - Interoperable, extensible description tools
  - MPEG-7
- Reverse engineering
  - Production structure parsing (shot, camera)
- Feature/object extraction and matching
  - audio-visual, spatio-temporal objects and features
- Structuring, Event Recognition, Skimming
  - Structure parsing and summarization
  - Event detection and recognition

When do we need automatic tools?

- Conditions for high impact
  - Suit user's needs and tasks
  - Metadata that's not available from production
  - Large volume, low individual value
  - Work that humans are not good at
  - Time sensitive
  - Acceptable performance
- Less Promising Areas
  - Content from digital production tools with metadata
  - Prime content that can afford mega budget and manual solutions

Case 1: 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

Detecting recurrent semantic units: Serve

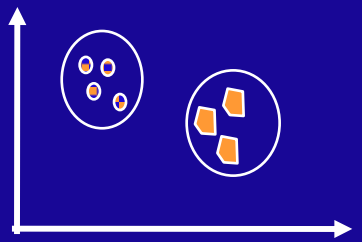


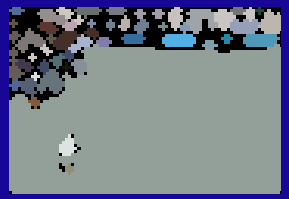
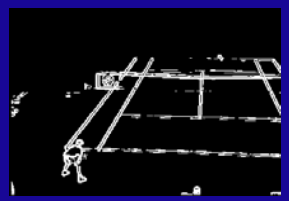

Color, lines, player

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

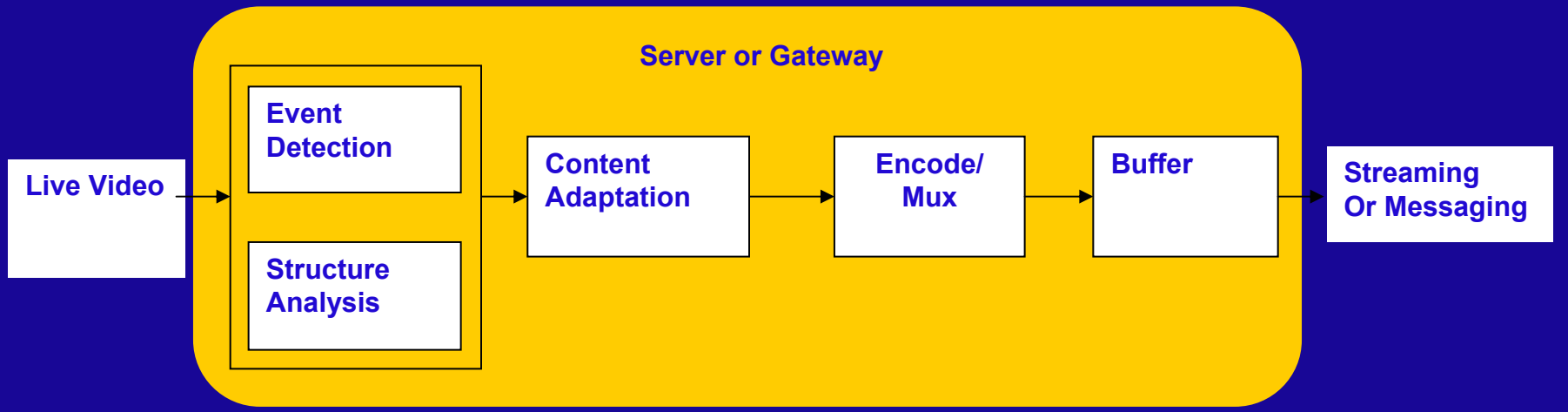
Compressed video → Shot Detection → Adaptive Color Filter → Object level verification → Canonical View Detection

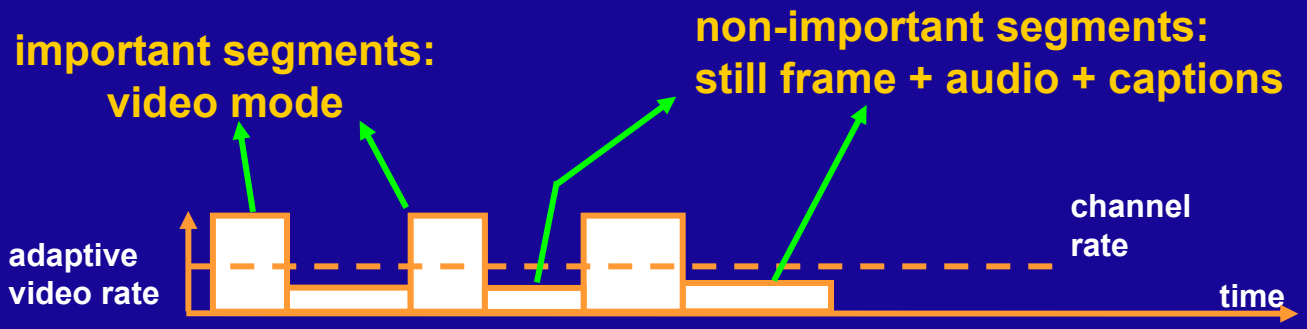
(Chang, Zhong, Kumar '00)



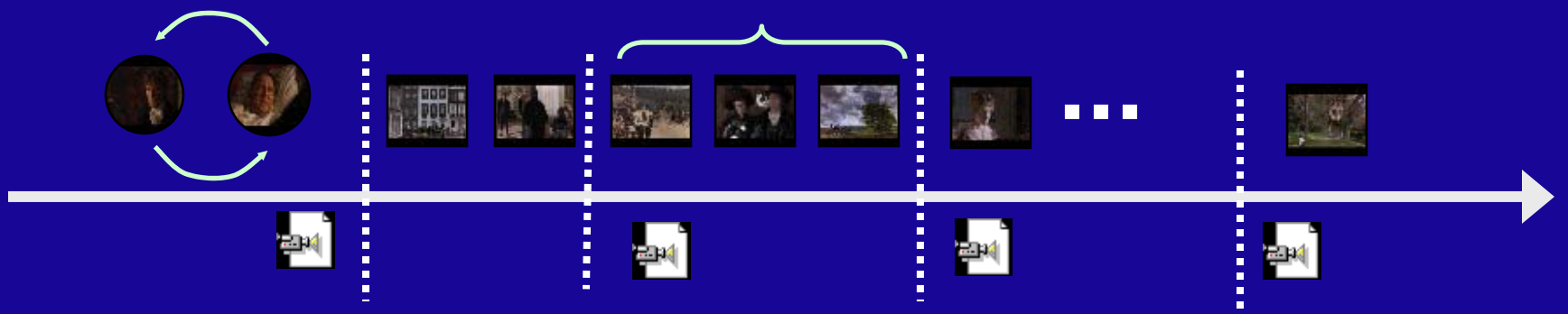



Content Adaptive Streaming/Filtering of Sports Video





Case 2: Film content structuring/summarization



- Opportunity:
  - SetTop PVR applications
  - Scene/Topic Segmentation, Browsing, Preview
- Challenge: Diverse production styles and content

Domain Consideration

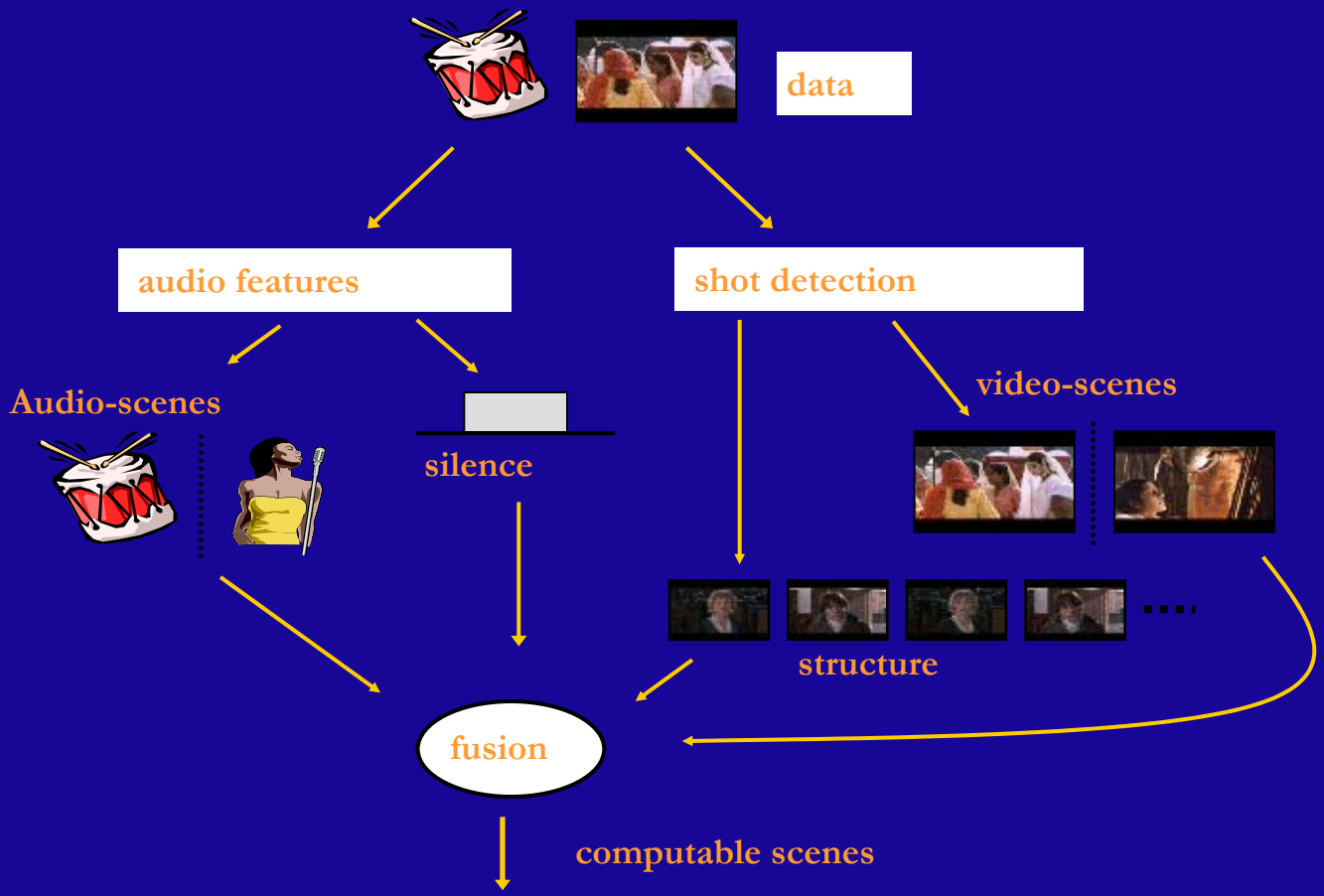
- **Film:**
  - **Does not satisfy impact conditions**
    - Mega budget production
    - Large value per content
  - **Potential impact**
    - Useful for archived collection in libraries
    - Excellent domain for intellectual/technical development

Opportunity:

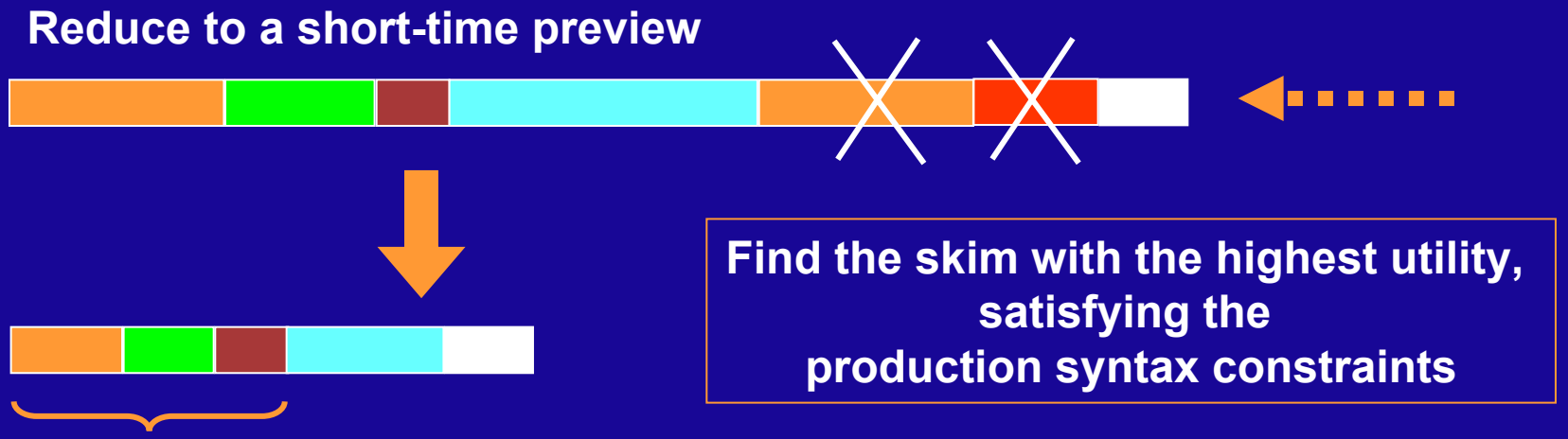
production model + multimedia integration + psychological viewer model

Scene structure analysis, summary, event detection

Computable Scene Detection Architecture

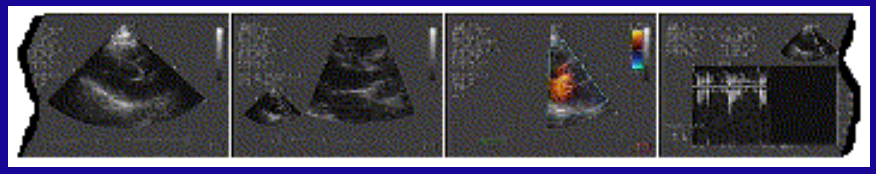


Time compression → video skims

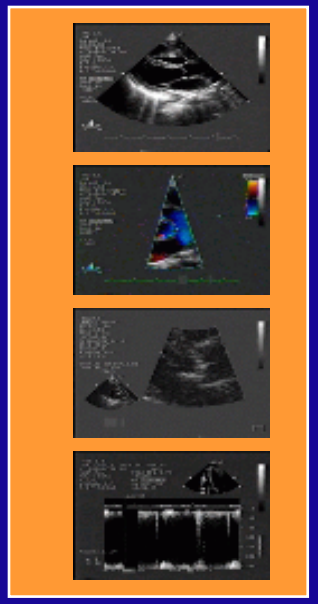


- Production Syntax → shot selection constraint
- Psychological Viewer Model → Complexity-based Utility Model
- Multimedia Integration → complementary audio-video roles

Case 3: Medical Video Library & Remote Medicine



Echo Video



Domain Knowledge

View Segmentation

KeyFrame, Heart Beat summary

View Recognizer

Diagnosis Reports

Object Annotation

Adaptive network delivery

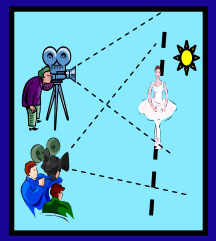
Content search

(Ebadollahi and Chang '00)

Impact Consideration

- Large volume, low individual value
- User's time extremely limited and expensive
- High annotation cost → computer-assisted annotation and summarization
- Quantitative measurement and matching of clinically significant objects objects

Challenges and Opportunities



Analysis and interaction

Production model + Media Integration + Viewer model

Systematic frameworks and methodologies

Application Success Criteria

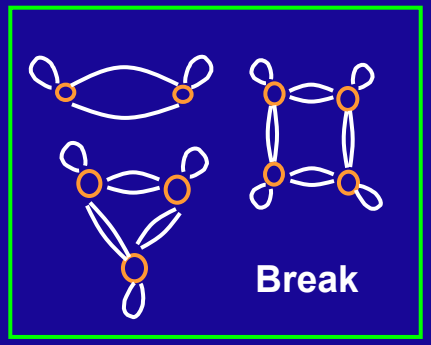
- Time-sensitive filtering
- Medical/Presentation
- Skimming
- Navigation

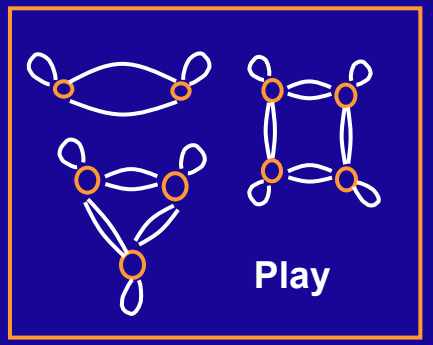
Recurrent semantic units in soccer video?

- Recovering the semantic unit structure
- PVR/SetTop box applications

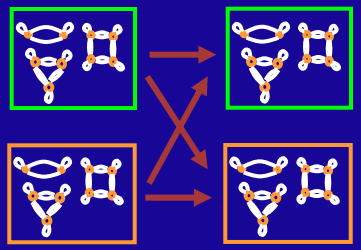
Modeling Temporal Transitions

Segment Features (color, motion, objects) → Train HMM families





Test Data → ML Detection + Linear Regression → Play-Break Transition Model (Dynamic Programming)



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