Sports Video Summarization

National Seminar on Computer Vision and Image Processing
(NaSCoVIP 2017) jointly organized by IEEE Gujarat Section,
IEEE Signal Processing Chapter and †Silver Oak College of Engineering & technology,
Ahmedabad during 8-9 September 2017

Devang S. Pandya†
Mukesh A. Zaveri, Computer Engineering Department
Sardar Vallabhbhai National Institute of Technology, Surat
mazaveri@coed.svnit.ac.in

Outline

- Introduction
- Motivation
- Literature
- Proposed Video Summarizing Algorithms
- Simulation Results
- Conclusion

Introduction

- Multimedia Content: Video, Audio, Text
- Huge amount of content
- Service based application: On demand
- Digital content analysis
- Numerous applications of multimedia processing
- Indexing, Searching, Retrieval of the content
- Summarizing for news, highlights
- Short, compact representation of digital content

Introduction: Application

- International, national events
- Movie trailer, News Summary
- Home Videos: personal, social festival and events
- Advertisement, News: abstract level representation
- Sport video summarization
- Surveillance Videos: crowd and traffic monitoring, event identification
- Crime investigation: key scene representation and identification

Motivation

- Documentary on sports, player, series
- Sport videos: cricket, football, hockey
- Training: Team work, World Cup-Football individual training
- Online decision making
- Critical analysis
- Strategic Planning

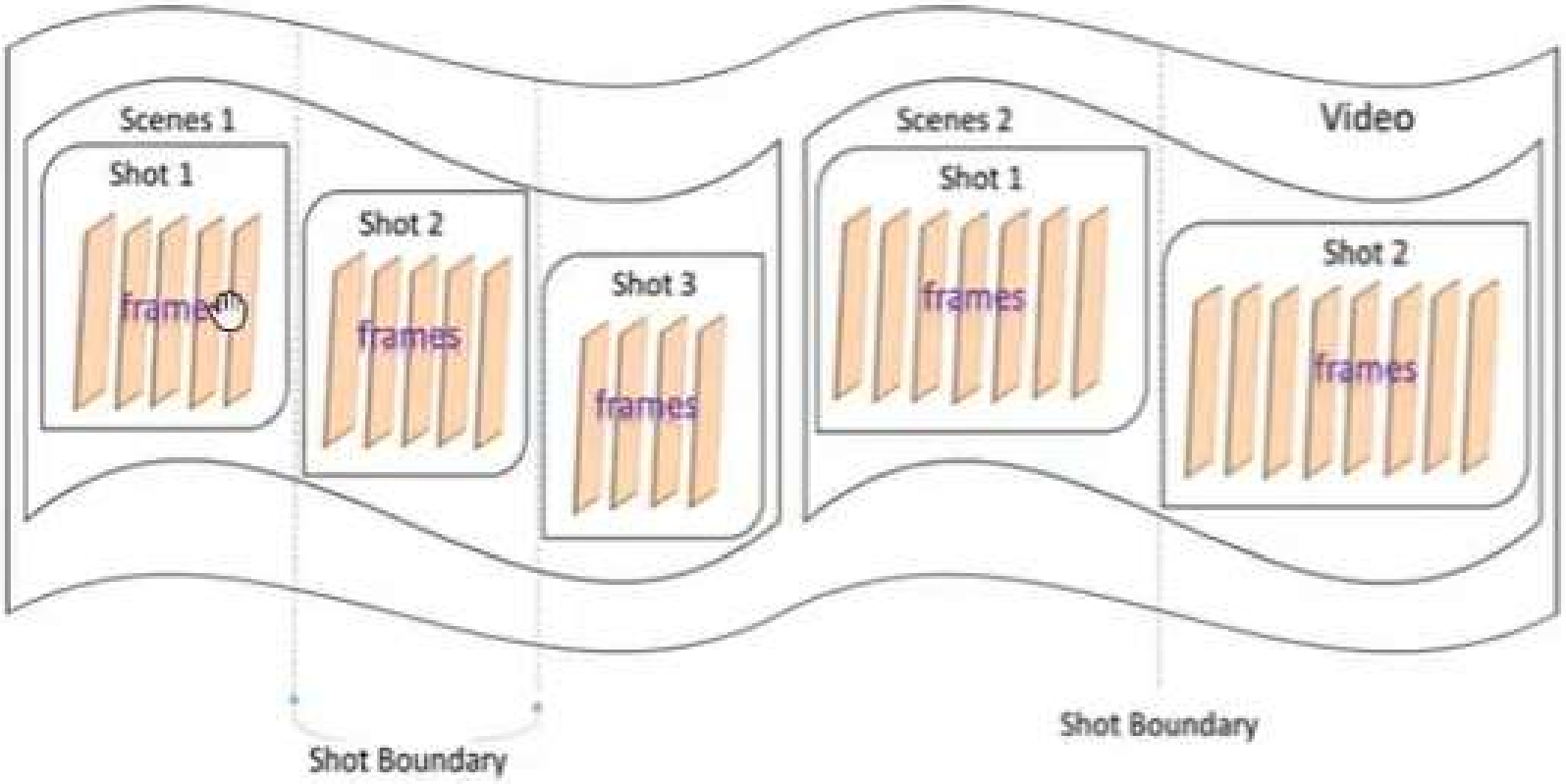
Motivation

- Video navigation
- Scanning video: detecting events
- Short summary, exciting events
- Issues:
  - Variety of videos, applications and its usage
  - Human intervention
  - Identifying common framework
  - More complex representation: video and audio combined

Video Summarization

- Abstract level vs. detailed representation
- Identifying key frames for abstract level
- Detailed representation: Events, Boundary, Significance
- Different ways for Video Summarization
  - Key frame based
  - Event based
  - Shot based
- Feature based
- Learning based
- Other approaches

Video Summarization



Structure of Video

Sport Video Summarization: Challenges

- Huge number of frames, analysis at deeper level
- Sport video: variety, broadcasters, and leagues
- Analysis per video captured stream
- Mix of all video captured streams
- Telecast: mixture of all streams
- Capturing video stream not under control
- Illumination conditions
- Sport rules varies, dress code

Sport Video Summarization: Literature

- Feature based
  - Color
    - simple and most noticeable feature
    - histogram: intensity distribution
    - major drawback: variation of intensity
    - histogram: insensitive to camera and object motion
  - Motion
    - optical flow: effective feature for identifying key frames
  - Texture
    - spatial arrangement of pixels
    - texture elements referred as texel
    - texture description: statistical and transform based
    - for example, simple texture descriptor: edge

Sport Video Summarization: Literature

- Feature based
  - Object
    - ★ target object: ball, player etc.
    - ★ object segmentation: computationally expensive
    - ★ application: human tracking, activity recognition
    - ★ limitation: not optimal for collaborative event having multiple actors
- Learning based key frame selection
  - supervised: feature extraction, training
  - unsupervised: clustering, self grouping based of features
- Other approaches: hybrid methods

Proposed Algorithms: Contribution

- Video summarization techniques
  - Key frame based methods
    - ★ histogram based
    - ★ unsupervised learning based: k-means clustering
  - Semantic based using Bayesian Network
    - ★ Dynamic Bayesian Network: Hidden Markov Model
- Goal: quick, efficient, accurate summary generation
- Solution: exploit dominant feature: color histogram and motion
- invariant features for varying illumination, ground conditions
- general framework for variety of leagues

Histogram based Key Frame Selection

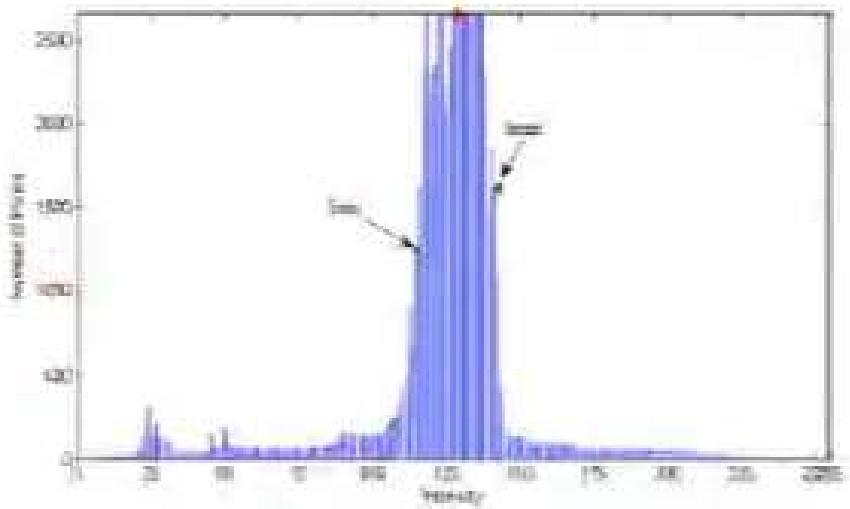
- Extract green color histogram of each frame
- Find the peak intensity of histogram
- Determine an interval which has a more concentration of pixels
- Determine intensity ranges on left and right sides of peak

$$H[i_{min}] \geq k * H[i_{peak}]$$
$$H[i_{min} - 1] < k * H[i_{peak}]$$
$$H[i_{max}] \geq k * H[i_{peak}]$$
$$H[i_{max} + 1] < k * H[i_{peak}]$$

- value of k may be set based on variance
- Mean of green color component of detected interval is computed
- Threshold for computed mean is tuned 1% to declare key frame

Histogram based Key Frame Selection

(a) Frame No:500, Flood light video

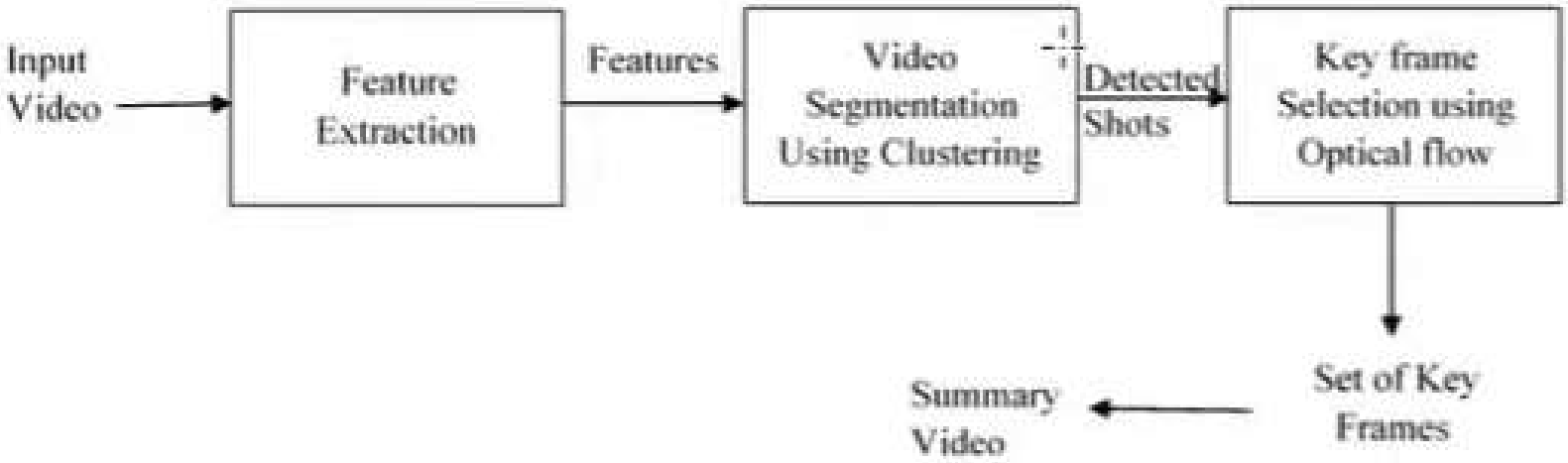


(b) Green Histogram and extracted interval

Extracted Interval

Key Frame based Video Summarization using Unsupervised Learning

- Goal: efficient video summary generation using dominant features for variety of soccer leagues
- Solution: color and change in motion
  - partition video into shots using dominant color feature
  - select key frames from each shot using change in motion



Key Frame Selection using Unsupervised Learning: Feature Extraction

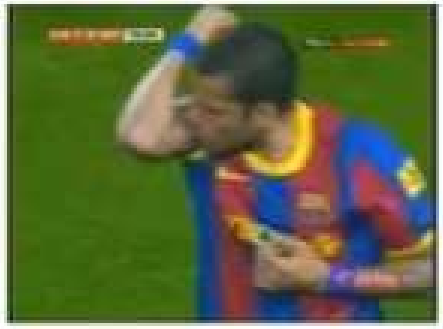
- Dominant Color
  - Soccer ground dominated by Green component
  - Hue is perceptually uniform

Frame divided into 4 parts

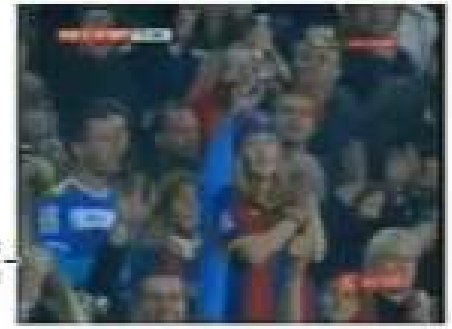
Key Frame Selection using Unsupervised Learning: Clustering

- Partitioning video into shots using unsupervised learning

Field view
Green mean:125.4
Hue mean:0.2740



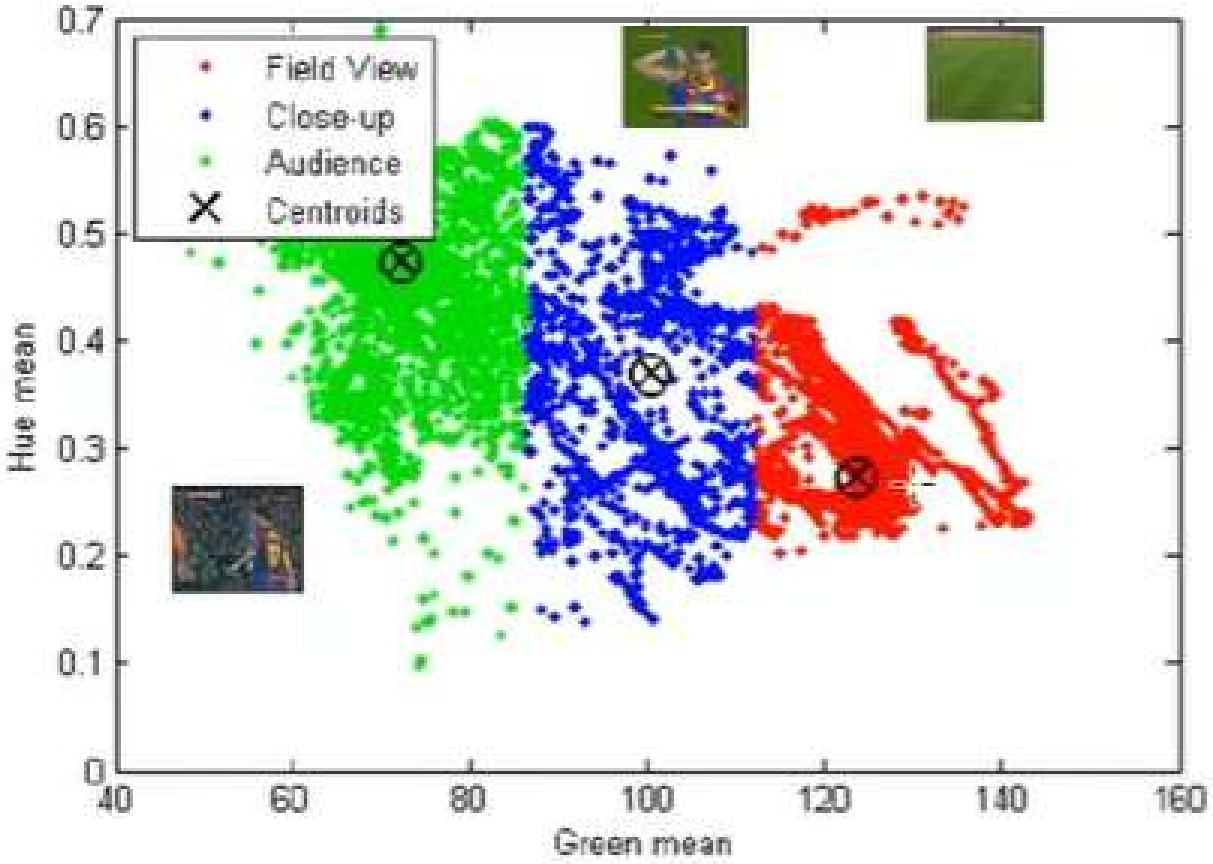
Close-up
Green mean:97.4
Hue mean:0.3498



Audience view
Green mean:69.8
Hue mean:0.4663

Field, Close-up, Audience views with Green and Hue values

Key Frame Selection using Unsupervised Learning: Clustering



Clusters formed for La Liga 2011:1

Key Frame Selection using Unsupervised Learning: Summary

- Sample output:
  - La Liga 2011
  - Serie A
- Advantage
  - Provides quick highlights
  - Effective for video which are uniform in nature
- Limitation
  - Complete event identification
  - More subjective: Key frame based summary
- **Scope for meaningful summary** in terms of
- **significant happening** during the complete match
- identifying as **Events**

Bayesian Network

- main task is knowledge modelling, (uncertain knowledge)
- limitation of domain expert
- learning structure from data - indispensable
- inference used while learning Bayesian network
- learning the parameters of Bayesian networks
- other names:
  - recursive graphical models
  - Bayesian belief networks
  - belief networks
  - causal probabilistic networks
  - influence diagrams
  - HMM - a specific type of dynamic Bayesian network
  - state space model

Event based Video Summarization

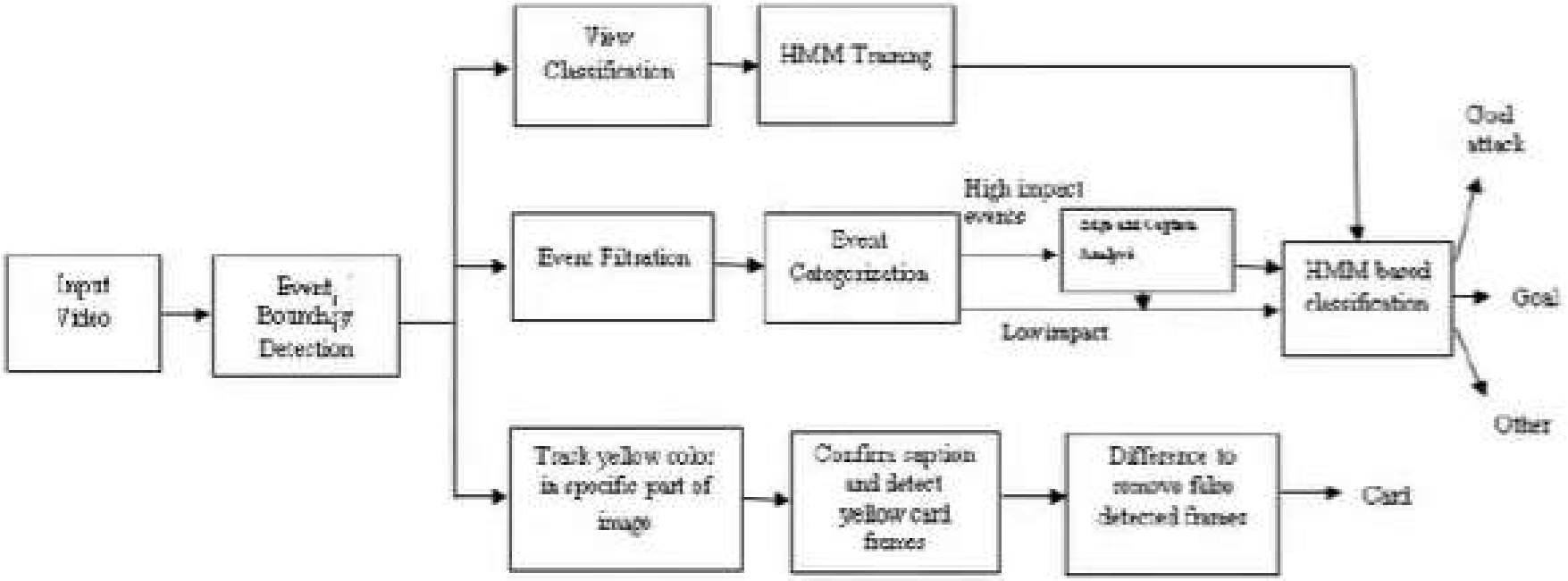
- Event identification
- State based event modeling
- Literature: Event based and Rule based approaches
- Limitation
  - lack of automatic framework, manual intervention
  - limited to few leagues
- Goal: automatic framework for variety of soccer leagues video summarization
- Solution: semantic based analysis for video summarization

Proposed Framework for Semantic based Analysis

- Semantic based event modeling
- framework consists of three phases
- event boundary detection in first phase
- second phase consists of three parallel operations
  - semantic level modeling in the second phase
    - ⋆ assign (meaning) to each frame in terms of view
    - ⋆ view based semantic for a frame of an event
    - ⋆ assign (meaning) to each event in terms of its impact
    - ⋆ impact of an event based semantic for an an event
  - extract features to define an impact of an event
  - each event defined using set of frames
  - each frame tagged based on particular view, view classification
  - event filteration: remove insignificant event detected
  - event categorized based on features to low and high impact event

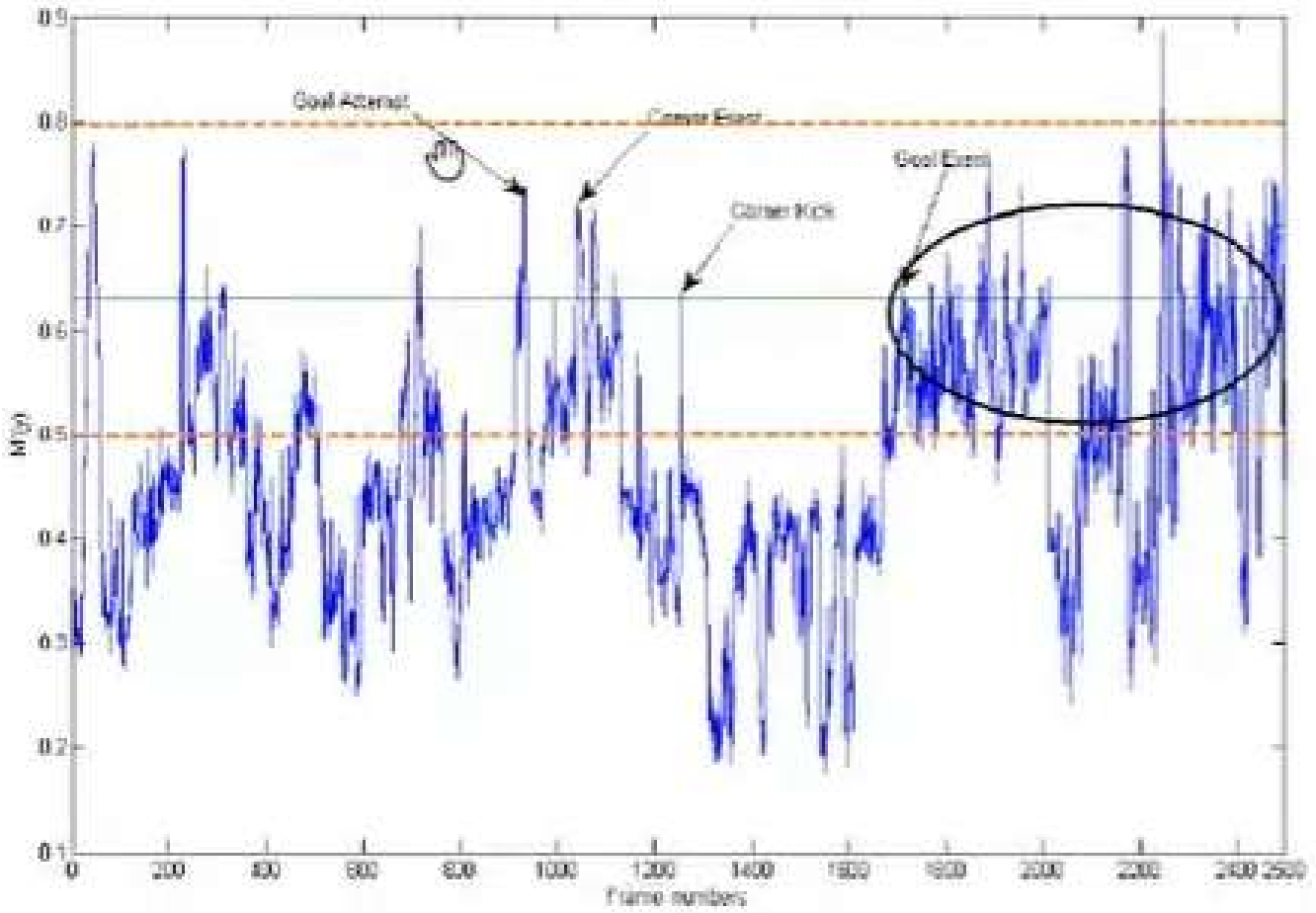
Proposed Framework for Semantic based Analysis

- event classification using Hidden Markov Model in last phase
- third parallel task yellow card event detection



Framework for semantic analysis of soccer video

Event Boundary Detection



Events with corresponding change in optical flow: La Liga

Event Boundary Detection

1063

1260

1330

1650

1701

1900

Change in optical flow and corresponding event sequences

View Classification

- Event consists of set of frames
- Depending upon the event each frame has different view
- Frame view plays significant role for event categorization
- Need frame view classification
- Frame view as semantic feature
- Field view: displays global view of the field
- Goal-post view: covers the goal post along with surrounding field
- Close-up view: Single or multiple players are present in the frame
- Audience view: Out of the ground view, generally displays the audience
- Different frame view classified as input for HMM training
- HMM models built are used for event categorization

View Classification