# Automatic Highlights Extraction in Cricket

Anjani Kumar, Sumedh Masulkar
{anjanik, sumedh} @iitk.ac.in

Guide- Prof. Amitabha Mukerjee
amit@cse.iitk.ac.in

**CS365A- Artificial intelligence**
**2013-14**
**IIT Kanpur**

## ABSTRACT

AIM: Extracting highlights automatically from a sports video using audio and video analysis and removing uninteresting sequence of frames .

APPROACH:

- Divide the extraction process into multiple levels.
- Remove the uninteresting event sequences from the main video at each level.
- 5 levels of extraction for shot classification (pitch view, crowd view, field view etc.)

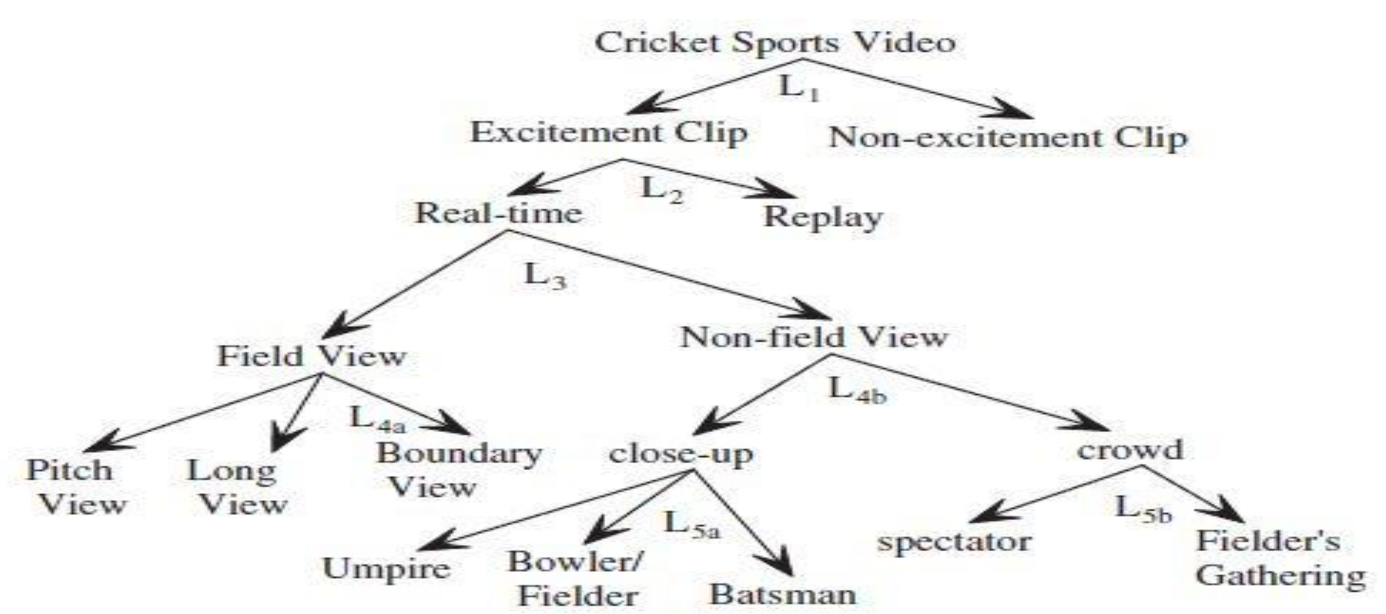



**Figure 1. Tree Diagram of Hierarchical Frame-work**
**(Image taken from [6])**

## PREVIOUS WORK

- Highlights extraction using Hidden Markov Models(HMM) in [1][2][3].
  - ❑ The states and transitions in the game were represented using HMM.
- [3] fused in audio information along with motion information for the first time
- In [4], the author proposed an unsupervised event discovery and detection framework which used color histograms(CH) or histograms of oriented gradients(HOG).
- [5] extracted event sequences from videos and classifies them into a concept using sequential association mining.
- [6] introduced a hierarchical framework for events detection and classification without shot detection and clustering.

## REFERENCES

[1] Kamesh Namuduri. “Automatic extraction of highlights from a cricket video using MPEG7 descriptors”.

[2] Jinjun Wang, Changsheng Xu, Engsiong Chng, Qi Tian.
“Sports Highlight Detection from Keyword Sequences Using HMM”, in Proceedings of the International Conference on Multimedia and Expo, 2004.

[3] ChihCheih Cheng, ChiouTing Hsu.
“Fusion of Audio and Motion Infromation on HMMBased Highlight Extraction for Baseball Games”, in Proceedings of the IEEE Transactions on Multimedia, vol. 8, no. 3, June 2006.

[4] Hao Tang, Vivek Kwatra, Mehmet Emre Sargin, Ullas Gargi.
“Detecting Highlights in Sports Videos: Cricket as a test case”, 2011.

[5] Maheshkumar H. Kolekar, Somnath Sengupta.
“Semantic concept mining in cricket videos for automated highlight generation”, 2009.

[6] M. H. Kolekar, K. Palaniappan, S. Sengupta.
“Semantic Event Detection and Classification in Cricket Video Sequence”, in Proceedings of the Indian Conference on Computer Vision, Graphics & Image Processing, 2008.

[7] Dipen Rughwani. “Shot Classification and Semantic Query Processing on Broadcast Cricket Videos”. http://cse.iitk.ac.in/~vision/dipen/.

## APPROACH IN [6]

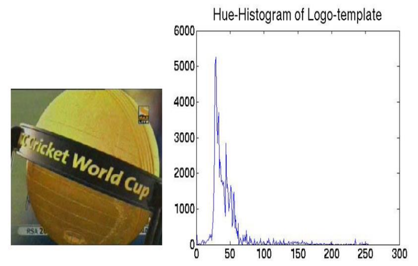
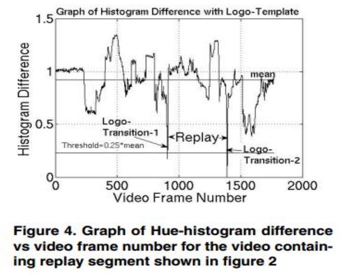
- Excitement Detection
  - ❑ Spectator’s cheer and commentator’s speech analysis.
  - ❑ Two popular content analysis techniques - Short-time audio energy($E$) and Short-time Zero Crossing Rate($Z$).
  - ❑ If $E * Z$ is greater than a given threshold, the particular frame is an excitation frame.
- Replay Detection
  - ❑ A replay is sandwiched between two logo transitions and the score bar is removed.
-   
  - ❑ Dominant Grass Pixel Ratio(DGPR) is used to classify frames.
  - ❑ DGPR = ($x_g/x$) where $x_g$ is number of pixels of grass, and x is total number of pixels.
  - ❑ For field view, DGPR values is greater than 0.07 whereas DGPR is smaller for non-field views.
- 4a - Field view classification
  - ❑ Classified as pitch view, long view or boundary view.
- 4b - Close Up view
  - ❑ RGB image is converted to $YC_bC_r$.
  - ❑ Percentage of edge pixels(EP) are calculated using *Canny* operator.
  - ❑ A threshold for EP classifies frames as close up view or crowd view.
- 5a - Close Up classification.
- 5b – Fielder gathering.


FIG: BOUNDARY VIEW


FIG: PITCH VIEW


FIG: LONG VIEW

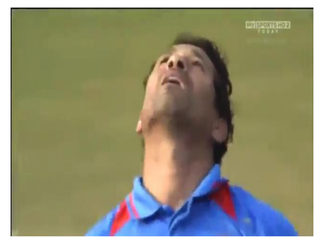


FIG: CLOSE UP


FIG: FIELDER GATHERING


FIG:SPECTATORS

## RESULTS

- corr2 computes the correlation coefficient using

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \bar{A})^2\right)\left(\sum_m \sum_n (B_{mn} - \bar{B})^2\right)}}$$

corr2 calculates the degree similarity of images. For exactly similar images, its value is 1 and 0 for two very different images. Thus, applying corr2 on a replay template and frames, we can detect action replays. After observing, for best results we have used threshold 0.65 to classify a frame as replay logo. The results, we obtained using corr2 and approach proposed by [6] were:

| | Replay detection(Our method) | Replay detection(approach in [6]) |
|---|---|---|
| Precision | 100% | 100% |
| Recall | 96.77% | 53.33% |

**Classification into Field-views and Non-field views:**

- After training svm on 4254 images, and testing on 4176 images, we observed following precision and recall -
  - Precision - 96.48%
  - Recall - 88.03%
- As compared to the following figures were produced by using approach proposed by [6],
  - Precision - 72.70%
  - Recall - 78.63%

**Classification of Field view into pitch, long, boundary views:**

| | Pitch view(Our method) | Pitch view (approach in [6]) | Long view(Our method) | Long view (approach in [6]) | Boundary view(Our method) | Boundary view (approach in [6]) |
|---|---|---|---|---|---|---|
| Precision | 98.21% | 20.66% | 96.11% | 63.82% | 97.56% | 8.47% |
| Recall | 95.21% | 69.18% | 96.60% | 28.03% | 93.69% | 27.5% |

**Classification of Non Field view into Crowd view, Close-up views:**

| | Crowd view(Our method) | Crowd view (approach in [6]) | Closeup view(Our method) | Closeup view (approach in [6]) |
|---|---|---|---|---|
| Precision | 94.29% | 44.66% | 82.42% | 98.58% |
| Recall | 98.54% | 93.70% | 52.71% | 79.11% |

**Classification of crowd view into fielder’s gathering, spectator’s crowd:**

- After training svm on 2444 images, and testing on 1184 images, we observed following precision and recall -
- Precision - 100%
- Recall - 99.42%.

Our methods have so far given promising results and shows significantly better results than the proposed approach in [6]. Although, the hierarchy used are same as suggested by authors of [6].