A MULTIMODAL APPROACH FOR AUTOMATIC CRICKET VIDEO SUMMARIZATION

Aman Bhalla*, Arpit Ahuja*, Pradeep Pant† and Ankush Mittal‡
*College of Engineering Roorkee, India
†Indraprastha Institute of Technology, New-Delhi, India
‡Graphic Era University, Dehradun, India

***Abstract*—Summarizing cricket matches is a labor intensive task that demands a certain level of proficiency. The paper proposes a novel method for automatically detecting and summarizing important events in a cricket match. Our model takes into input the video recording of the entire cricket match and returns the most important clips of the game as the output. Techniques like optical character recognition, sound detection and replay detection have been used to extract important events such as boundaries, wickets and other playfield scenarios in a cricket match. These events are then clipped together to form the entire highlight for the cricket match. We performed several qualitative and quantitative experiments to evaluate our model. Our model achieves an accuracy of 89.45% to detect events like wickets, fours and sixes which indicates the usefulness of the model in real life scenarios.**

***Index Terms*—Convolution neural network, Support vector machine, Optical character recognition.**

I. Introduction

Automated sports video analysis is observed to have an outbreak attention in the field of multimedia. It led to an explosion of ideas and many research works among the various domains of media content such as images and videos over the cyberspace. With a large amount of sports video data on Internet for viewing, it is difficult for a viewer to keep a crucial glance of all media available specially in the case of sports video where the collection is enormous. Therefore highlights are the solution to the sports fans to keep them up-to-date with all important and latest scenarios. However if highlight generation is performed manually it requires a long time and even requires professional editing skills which is further a limitation to summarize the media in a short notice.

In the case of most viewed game that is cricket, we preferably need to have some summarized format of the special events taken place through the match. Therefore the automatic generation of highlights from a sports video always remained a problem for broadcasters. Hence we need a system that automatically generate highlights of sports events, something that can be made to use features of the domain like machine learning and computer vision.

This work proposes a novel approach for a computer vision based system for automatic highlight generation of cricket match videos which is a challenging problem as broadcasters usually face the main challenge when it comes to the extraction of important events from a cricket match because of the following reasons:-

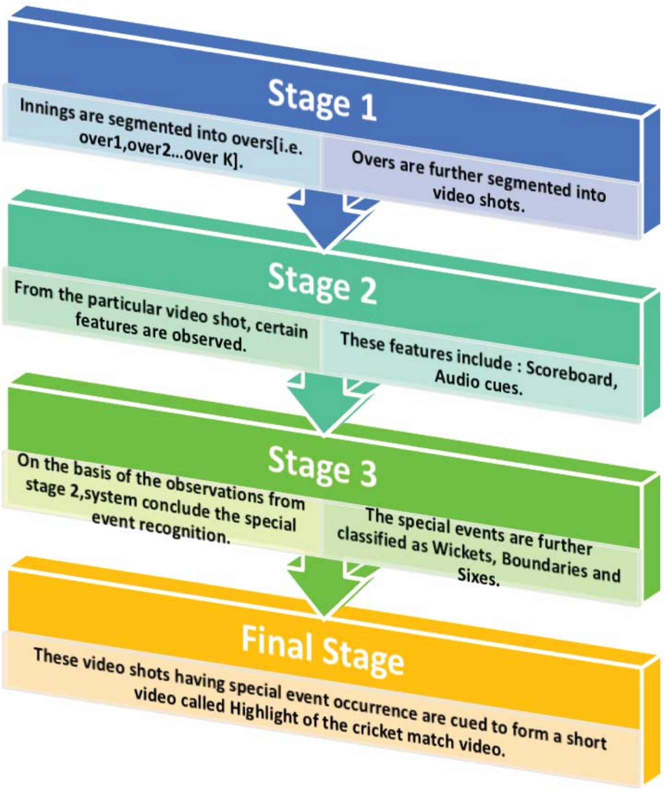
Stage 1
Innings are segmented into overs[i.e. over1,over2...over K].
Overs are further segmented into video shots.
Stage 2
From the particular video shot, certain features are observed.
These features include : Scoreboard, Audio cues.
Stage 3
On the basis of the observations from stage 2,system conclude the special event recognition.
The special events are further classified as Wickets, Boundaries and Sixes.
Final Stage
These video shots having special event occurrence are cued to form a short video called Highlight of the cricket match video.

Fig. 1. Stage-wise representation of our model.

1) Cricket constitute a set of complex rules and played in wide range of formats and on the other hand cricket matches are long lasting (up-to 4-5 days i.e. test cricket match) therefore require much more processing time.
2) The long sports video builds up a large distribution of events and classifying the events as important or unimportant remained a problem which further plays a major role in highlight extraction.

As shown in the Fig. 1, the entire architecture can be broken down into four stages. In the first stage, the video clip is broken down into series of video shot. In the second stage, important

978-1-7281-1380-7/19/$31.00 ©2019 IEEE

Authorized licensed use limited to: PES University Bengaluru. Downloaded on February 03,2024 at 03:10:00 UTC from IEEE Xplore. Restrictions apply.

cues are extracted from these video shots. Finally, these cues are used for generating highlights of the match.

II. Related Work

Previous approaches focus on generating highlights for a wide range of sports such as golf, basketball, soccer and cricket. Reference broadly reflects that the approaches can be mainly categorized into two domains : Excitement based where indexing of the frames is done according to the excitement cues such as prominent sound intensities from crowd, motion activities on field and camera saliences and another is event based which mainly rely on detection of important events in the game. As both the approaches not gives acceptable results individually, some works try to establish a model using both of the above mentioned approaches for intelligently detecting the highlights. Works of [16] mainly describe a heuristic method to extract highlights using the following criteria : audio intensity (from crowd and commentary itself) or excitement, camera motion, short moments and other playfield scenarios.

Even in case of cricket which is one of the widely played sport in the world, automatic extraction from the cricket match videos is the domain that has been more often researched. Previous works collectively project that most of the analyzing models are not up to the mark and focuses only on few phenomena of the game. Tang et al. [14] proposed an automatic detection framework which formerly used HOG (Histogram of Gradient) and CH (Color Histogram) for event detection in cricket match which serves in input to HMM (Hidden Markov Models) for categorization as important or unimportant. The same approach is adapted by Namuduri [15] with an additional segmentation technique where the match was resolved into videos shots to extract crucial key frames.

Kolekar sengupta [6] followed excitement based detection in which they used audio intensity as a basis for detection of events. The model revolves around the principle that all the major events in the match are surrounded by uplifting sound intensities and once it is categorized as important further information has been taken from the caption.

Work [17] also follows the same approach but in addition to this, a scoreboard recognition system is being used. Descriptors of scoreboard images were extracted from the fc7 layer of the pre-trained AlexNet model. The model was trained on a multi-class linear Support Vector Machine. Work [8] proposed algorithm uses running text commentary by broadcasters which may mislead due to synchronization and ambiguity of information. However, the hereby proposed framework effectively uses the crucial aspects to generate and construct a highlight series of a cricket match.

III. Implementation

A. Video shot detection

Video shot detection is an essential aspect as it reduces the processing time by dividing the full video into smaller fragments. A video shot is a sequence of frames of a video taken from a single camera where the content may vary because of the camera movement. Keeping into consideration that two successive frames having a slight change in background and objects results in a negligible absolute difference, we focus on finding the frames having an absolute difference greater than a threshold value t so that video shot can be encountered. In order to find the absolute difference between two successive frames, firstly both of them are converted to a gray-scale image and if the difference between two successive frames exceeds the threshold value t, that frame is stored and all the further processing is done only with respect to these frames. The value of t was determined empirically.

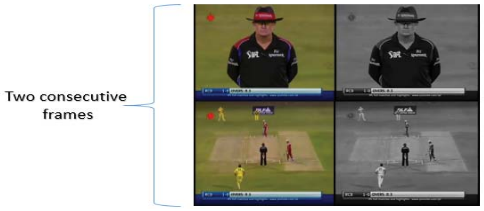

Fig. 2. absolute difference $> t$, where value of t is 60 which is determined empirically.

B. Sound detection

The key frames in a video are also determined using audio media in addition to the detection of video shot. Generally, when a batsman hits four, six or bowler takes a wicket, an increase in cheer/noise is observed. To detect the frames having higher audio levels, the number of audio samples corresponding to each video frame is calculated and stored in the module. The number of peaks in the input audio stream is determined. The frames having the highest audio levels are taken and the frame number corresponding to such video frame is noted. The audio samples are generated separately for each over because these audio samples also depend on commentator's voice apart from the crowd's noise. Since in an over, the commentary is done by a particular commentator so it is easy to find the frames with higher audio level. These frames assist in extracting the key events from the match.

C. Scoreboard Recognition

The scoreboard displays the delivery wise details of runs and wickets. OCR is applied on the scoreboard to generate the bounding boxes around each digit of the scorecard. Using the dimension of the boundary boxes generated, the individual digits is cropped for identification by the trained machine. Since '-' symbol was not recognized by OCR, therefore we made an observation that in all the matches the wickets are followed by runs in the scorecard. Therefore, in order to separate runs from the wicket, the order of location of the bounding box is used since the digit bounded by the last bounding box is of the wicket. As shown in Fig. 3, four boundary boxes are generated, first three bounding boxes determine runs and the last bounding box shows the wicket. An important thing to note is that the performance of OCR is directly dependent upon the quality of input image.

Authorized licensed use limited to: PES University Bengaluru. Downloaded on February 03,2024 at 03:10:00 UTC from IEEE Xplore. Restrictions apply.

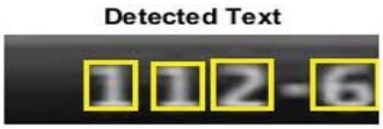

Fig. 3. Detected text

D. Highlight Generation

The runs and wickets detected by OCR are saved for each key frame which are further used in generating the highlights. The successive frames (say *k*–1 and *k*) where the difference in run is greater than or equal to four or difference in wicket is one are found. The video is played from frame number (*k*–2) to frame number (*k*+2) which was determined empirically. It is noted that if a boundary or wicket is encountered in first two frames then video is to played from first frame (not from *k*–2) and similarly when boundary or wicket is encountered in last two frames then video is played till frame number k (not till *k*+2). This step is repeated for each over of both the innings so as to play the key moments of the match. This algorithm efficiently extracts the key features from a match.

IV. Result

A. Experimental Setup

This section depicts the level wise analysis of the experiment conducted in the proposed algorithm with there respective accuracies as shown in TABLE 1.

TABLE I
Technique used and accuracies

Technique	Accuracy
GIST + OCR	76.21%
CNN + OCR	94.63%

B. Evaluation

The performance and accuracy of our implementation of automatic highlights generation completely depends upon the extend up to which the machine correctly reads the score board i.e. the correct runs and wickets. Because on the basis of the updated score board, we select the relevant frames for the highlights. Firstly, we train the machine by applying GIST features on the text detected by OCR which detects the digits with the accuracy of 74%.

When we train the machine using CNN (Convolution Neural Network), it is found that the training through the CNN lay an upper hand on the training through the GIST features as the accuracy increases from 74% to 94%.

The percentage result for the various events detected by proposed algorithm is shown in TABLE II.

The highlight of the match is generated on the basis of correctly detected boundaries, sixes and wickets. Therefore, the accuracy of the highlights of the match is determined as the average of accuracies of different events.

TABLE II
Accuracy of various events

Events	Accuracy
Boundaries	86.74%
Sixes	89.12%
wickets	92.45%

C. Qualitative Results

The graphs in Fig. 4 and Fig. 5 depict the detected and missed fours, sixes and wickets which are further used in obtaining the highlight of the match.

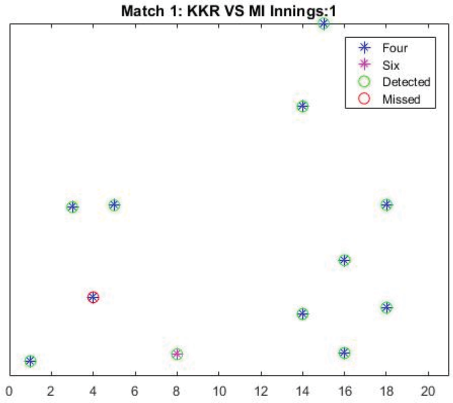

Fig. 4. Graph showing detected and missed fours and sixes

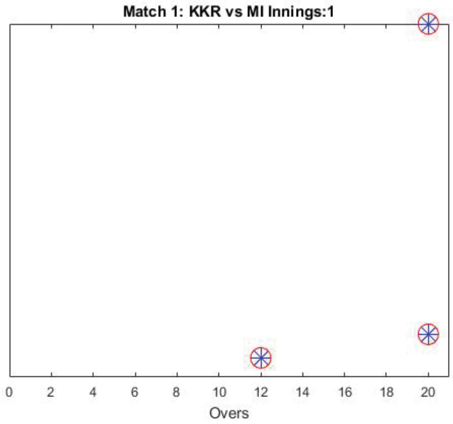

Fig. 5. Accurately detected wickets

As we can observe from the graph that the proposed algorithm is effectively detecting the fall of wickets as well as majority of fours and sixes. Although it fails to detect some of the fours and sixes that are encircled in red color but majority of four and sixes are detected that are encircled in green color. Furthermore the wicket detection is more accurate in this approach. Therefore, these results show that this approach can be effectively adapted into use of generating the series of crucial advents needed for highlight.

Authorized licensed use limited to: PES University Bengaluru. Downloaded on February 03,2024 at 03:10:00 UTC from IEEE Xplore. Restrictions apply.

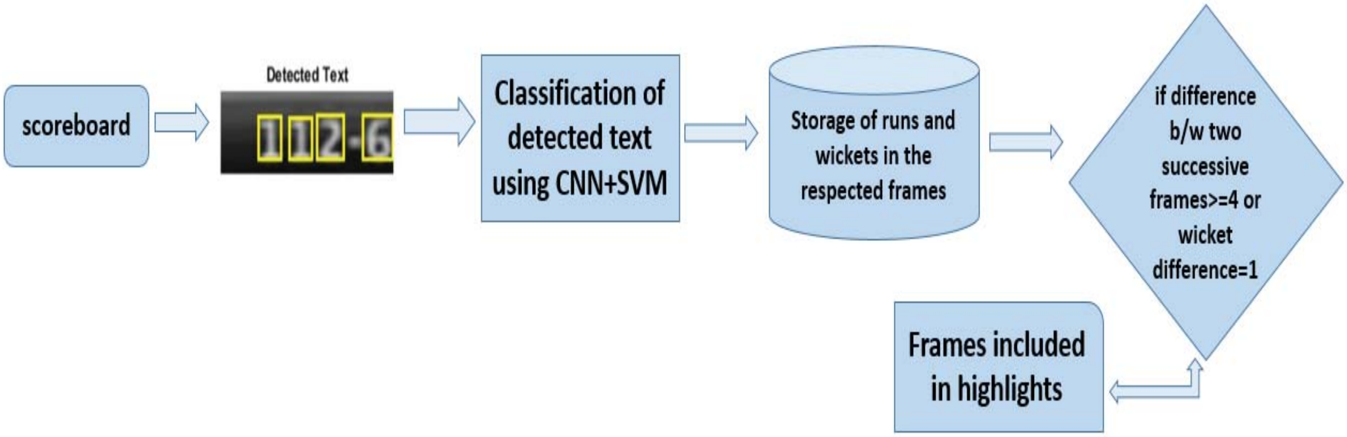
scoreboard
Detected Text
112-6
Classification of detected text using CNN+SVM
Storage of runs and wickets in the respected frames
if difference b/w two successive frames>=4 or wicket difference=1
Frames included in highlights

Fig. 6. Multilevel framework of highlight generation

V. Conclusion and future work

Generating highlights is a subjective work and there are different views about the different highlights. So we measure the accuracy of the highlights generated by our proposed algorithm by comparing it to the highlights telecast by the official broadcasters of the respective match.

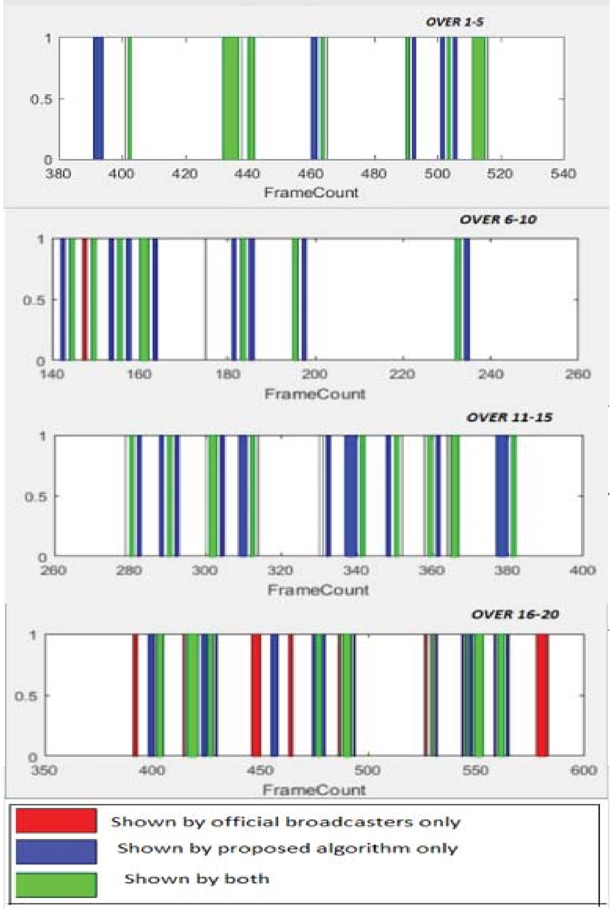

Fig. 7. Official highlights vs Generated highlight

The Fig. 7 shows the comparison between official and system generated highlights and it can be clearly observed that our proposed model is covering almost every event that should be a part of highlights of cricket match.

References

[1] Mahesh Goyani, Shreyash Dutta, Payal Raj, “Key frame detection based semantic event detection and classification using hierarchical approach for cricket Sport Video Indexing,” CCSIT, SPRINGER, LNCS in Communications in Computer and Information Science, Vol. 131, pp. 388-397, Proceedings of International Conference on Computer Science and Information Technology, Bangalore, India, 2-4 Jan ’11.

[2] Changsheng Xu, Jinjun Wang, Hanqing Lu, Yifan Zhang , “A novel framework for semantic annotation and personalized retrieval of sports video,” IEEE transactions on multimedia, vol. 10, no. 3, April 2008.

[3] R. Brunelli and O. Mitch, “Histograms Analysis for image Retrieval,” Pattern Recognition, Vol.34, No.8, pp. 1625–1637, 2001.

[4] Hu Min, Yang Shuangyuan, “Overview of content based image retrieval with high-level semantics,” 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE).

[5] Yu Xiao Hong and Xu Jinhua, “The Related Techniques of Content-based Image Retrieval,” 2008 International Symposium on Computer Science and Computational Technology.

[6] M. H. Kolekar and S. Sengupta, “Event-Importance based customized and automatic cricket highlight generation,” in Proceedings of IEEE International Conference on Multimedia and Expo, 2006.

[7] Mihai Lazarescu, Svetha Venkatesh, and Geoff West, “On the automatic indexing of cricket using camera motion parameters,” in Multimedia and Expo, 2002. ICME’02. Proceedings. 2002 IEEE International Conference on, volume 1, pp. 809–812.

[8] Yair Poleg, Ariel Ephrat, Shmuel Peleg, and Chetan Arora, “Compact cnn for indexing egocentric videos,” in Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on, pp.1–9.

[9] MinXu, Ling-Yu Duan,Chang-ShengXu,and QiTian, “Afusionscheme of visual and auditory modalities for event detection in sports video,” in Multimedia and Expo, 2003. ICME03. Proceedings, volume 1, pp. 1–333.

[10] Rashish Tandon, “Semantic analysis of a cricket broadcast video,” 2009.

[11] M. Merler, D. Joshi, Q.B. Nguyen, S. Hammer, J. Kent, J. R. Smith, and R. S. Feris, “Automatic curation of golf highlights using multimodal excitement features,” in Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 57–65.

[12] A. Rehman and T. Saba, “Features extraction for soccer video semantic analysis: current achievements and remaining issues,” Artificial Intelligence Review, 41(3):451–461, 2014.

[13] M. H. Kolekar and S. Sengupta, “Bayesian network-based customized highlight generation for broadcast soccer videos,” IEEE Transactions on Broadcasting, 61(2):195–209, 2015.

Authorized licensed use limited to: PES University Bengaluru. Downloaded on February 03,2024 at 03:10:00 UTC from IEEE Xplore. Restrictions apply.

[14] H. Tang, V. Kwatra, M. E. Sargin, and U. Gargi, "Detecting highlights in sports videos: Cricket as a testcase," in Multimedia and Expo (ICME), 2011 IEEE International Conference on 2011 july 11, pp. 1–6.

[15] K. Namuduri, "Automatic extraction of highlights from a cricket video using mpeg-7 descriptors," in Communication Systems and Networks and Workshops, 2009, pp. 1–3.

[16] V. Bettadapura, C. Pantofaru, and I. Essa, "Leveraging contextual cues for generating basketball highlights," in Proceedings of the 2016 ACM on Multimedia Conference, pp. 908–917.

[17] P.Shukla, H.Sadana, A.Bansal, D.Verma, C.Elmadjian, R.Balasubramanian, and M.Turk, "Automatic cricket highlight generation using event-driven and excitement-based features," IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2018, pp. 1800–1808.

Authorized licensed use limited to: PES University Bengaluru. Downloaded on February 03,2024 at 03:10:00 UTC from IEEE Xplore. Restrictions apply.