This CVPR workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the version available on IEEE Xplore.

Automatic Cricket Highlight generation using Event-Driven and Excitement-Based features

Pushkar Shukla[1] Hemant Sadana[2,4] Apaar Bansal[2,4] Deepak Verma[2,4]
Carlos Elmadjian[3] Balasubramanian Raman [2] Matthew Turk[1]
[1] University of California, Santa Barbara [2] Indian Institute of Technology, Roorkee
[3] University of São Paulo [4] Raman Classes, Roorkee ,India

Abstract

Producing sports highlights is a labor-intensive work that requires some degree of specialization. We propose a model capable of automatically generating sports highlights with a focus on cricket. Cricket is a sport with a complex set of rules and is played for a longer time than most other sports. In this paper we propose a model that considers both event-based and excitement-based features to recognize and clip important events in a cricket match. Replays, audio intensity, player celebration, and playfield scenarios are examples of cues used to capture such events. To evaluate our framework, we conducted a set of experiments ranging from user studies to a comparison analysis between our highlights and the ones provided by the official broadcasters. The general approval by users and the significant overlap between both kinds of highlights indicate the usefulness of our model in real-life scenarios.

1. Introduction

The past few years have marked an increased interest in automatic analysis of sports content. One of the primary reasons behind this is the outburst of sports media available on the internet. With a large number of sports matches throughout the year, it is very difficult for a sports fan to keep up with all the news. Therefore, highlights serve as an important source of information to keep fans updated with the latest happenings without consuming too much of the their time. However, manual highlight generation requires professional editing skills and is a time consuming process, which limits the amount of media that can be summarized on short notice. This fuels the need for systems capable of automatically generating highlights of sports events , something that can leverage recent developments in the fields of machine learning, computer vision, and multimedia.

This work proposes an approach for automatic highlight generation of sports videos with a focus on cricket, which is a challenging problem itself due to the following reasons:

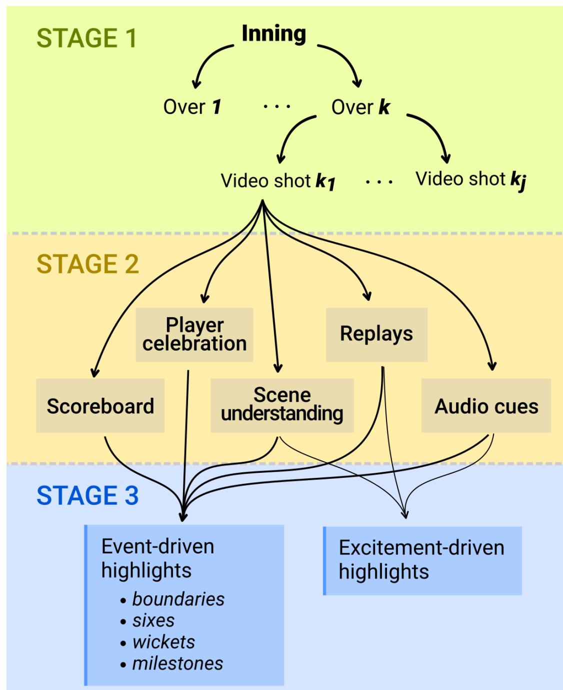

Figure 1. Block diagram representation of our model, which can be broken down into three stages. In the first one, the whole video clip is divided into a series of video shots. In the second stage, important cues are extracted from these video shots. Finally, these cues are used for generating import highlights in the last stage.

- Cricket is a complex game that has a lot of rules. Thus, a large range of events might be important depending on circumstantial factors.
- Cricket matches usually have a longer duration than most sports, with the shortest version of the game last-

ing for 3 hours and the longest lasting for 5 days.

Automatic highlight generation can either be event-driven — where specific rules of the game are considered — or excitement-based — where actions from players, crowds and commentators are used for extracting highlights. While event-based strategies are difficult to formulate, excitement-based features often suffer from a large number of false positives and are context-dependent.

We propose a novel strategy for automatic generation of cricket highlights. Event-driven features are used for extracting the four major events in a cricket match, .i.e., wickets, boundaries, sixes, and milestones, while the remaining important events are identified with the aid of excitement features. Event-driven strategies make use of OCR, playfield scenarios and replays, while the complimentary excitement-based strategy makes use of audio-based classifiers and replays to reinforce the relevance of an event in a cricket match. The major contributions of our work are:

- We propose a first-of-a-kind system capable of automatically summarizing highlights of cricket matches, with results comparable to the highlights generated by professional manual approaches.
- Our method focuses on events as well as excitement features for highlight generation. This allows us to have better quality of cricket highlights compared to previous approaches.
- We show that our system is very robust in terms of detecting important events and that users support the quality of our highlights.

2. Related Works

Highlight generation can be viewed as a sub-class of video summarization [17, 20]. Previous works have focused on generating highlights for different kinds of sports, such as soccer [8, 11, 4], cricket [10], basketball [2, 3] ,tennis[19] and golf [9, 6]. According to [13], highlight generation approaches can be divided into event-based and excitement-based approaches. The former rely on methods for detecting important events in a game, while the latter try to clip highlights using excitement cues, like audio intensity, energy, motion activity and color tracking. Considering that detecting all important events in a game might not be possible and that excitement cues may be misleading, some works [2] tried to combine both of these models for automatic highlight generation. According to [2], an automatic highlight detection model generally makes use of one or more of the following cues: replays, audio from broadcast (like audience loudness or commentator's excitement), camera motion, captions, and contextual information, such as ball position, shot segmentation and playfield scenarios.

Even though cricket is the second most popular sport in the world, automatic analysis of cricket videos is an area that has barely been researched. Most of the previous works on cricket video summarization lack robustness and focus only on a few aspects of the game. Tang et al. [18] proposed a multilevel framework that relied on Histogram of Oriented Gradients (HOG) and Color Histograms (CH) for detecting important events in a cricket match. The events, once detected, were refined using Hidden Markov Models (HMM). Namudri [10] proposed a technique that relied on MPEG-7 for highlight generation of cricket matches. The complete match was segmented into video shots (clips belonging to the same camera take). Extracted key frames were classified into one of five different classes, depending on their field view. A Hidden Markov Model (HMM) was then used to identify events from sequences of these key frames.

Kolekar and Sengupta [7] proposed a model that relied on audio-based energy for extracting important events in a cricket match. The assumption was that important events are generally surrounded by an increase in audio intensity. Once an event of relative importance was detected, the caption content was utilized to extract more important information. A different approach was proposed by [12], relying on the use of text commentary mined on-line from a third-party website. It was, however, unclear how audio-visual information and comments generated on the website were synchronized.

Compared to the literature, our work is capable of dealing with all the major aspects of a cricket match, thus being more generalizable and applicable for automatic highlight generation.

3. Methodology

3.1. Video Shot Segmentation

Video shot detection is an important aspect of sports video summarization. A complete cricket match can be broken down into a series of video shots, i.e., a set of frames that are part of the same camera take. Therefore, events in a cricket match can be considered as sequences of video shots. An example of such a sequence is illustrated in Figure 2. There were two main reasons for choosing video shots as building blocks to generate highlights:

- Since highlights are generally a combination of important events, video shots can be used for highlight generation.
- Processing video shots makes the model more efficient in terms of time. It must be taken into consideration that a cricket match might have a long duration. Therefore, applying sophisticated algorithms on individual

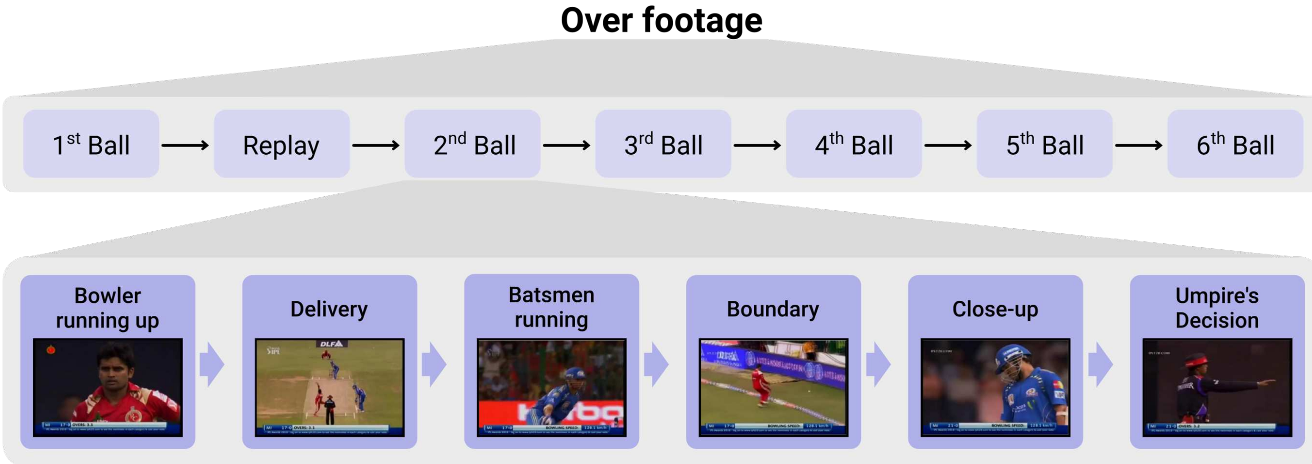

Figure 2. An example of how video shots can be thought as building blocks for a given *over*. An over in a cricket match, refers to six legal deliveries that can be bowled consecutively by one bowler. These deliveries are made up of a series of video shots. Thus, important events in a cricket match will be sequences of video shots.

frames will lead to an increase in the overall processing time of the framework. In this work, the first frame of each video shot is considered as the key frame that is used for representing the entire video shot and all the processing is carried out on that key frame.

The proposed approach for video shot detection is based on the assumption that pixels of successive video frames differ vastly at the beginning and at the end of a video shot. Therefore, the difference between the consecutive frames St and $St+k$ sampled at interval of k is calculated as shown in Eq. 1:

$$D(:,:,t) = |S(:,:,t) - S(:,:,t+k)| \quad (1)$$

The grayscale image D extracted is now thresholded and converted into a binary image B, such that

$$B(x,y) = \begin{cases} 1, & \text{if D(x,y)} \geq \tau \\ 0, & \text{if D(x,y)} < \tau \end{cases} \quad (2)$$

Here τ is a threshold parameter determined empirically. The two frames are said to be boundary frames for a given video shot if the total number of ones in the image are greater than 30% of the total number of pixels in the image. This percentage value was also chosen empirically after a few rounds of experimentation.

3.2. Replay Detection

A replay in a cricket broadcast is a video footage of one or more events in a match that are played more than once. Replays provide vital cues for important events as they are frequently used in game highlights. Furthermore, replays are particularly important in cricket because they are also used by third umpires to confirm or review a decision.

Replays can be detected by observing that *the scoreboard is absent in all replays and advertisements*. The procedure starts by cropping the area around the scoreboard in the key frame of a video shot. The cropped image is then used for classifying a video shot as an on-going event of the game or as a replay/advertisement using a Convolutional Neural Network (CNN) + Support Vector Machine (SVM) framework. This framework relies on 4096 representations extracted from the *fc7* layer of the AlexNet network, which are fed to a trained Linear Support Vector Machine that has been trained on representations of 2000 images of both classes. The training offset of the SVM was set to 0. A *one-vs-all* strategy was applied to train the model.

3.3. Scoreboard Detection

The scoreboard refers to the ball-by-ball statistics generated on screen during a cricket match. Scoreboards can be used to predict important events, such as boundaries, sixes, or the total number of wickets that have fallen. They also can tell us about the start or end of an inning.

Scoreboards can display constant as well as intermittent information about the match. For example, while the total number of runs and wickets are always displayed on the scoreboard, other stats, like the number of overs, speed of a delivery, and the current score of a batsmen, are displayed once in a while.

Accurate detection of scoreboard is important to identify key events because it changes rapidly in a cricket game. Detection of events, such as boundaries, sixes and wickets, rely heavily on the scoreboard. Thus, scoreboard recognition must be performed during the entire course of the game, meaning that the Optical Character Recognition (OCR) engine must be very effective.

Figure 3. Positive samples from a game used for training the playfied scenario mentioned in Section 3.4. Note that the key video frames when a bowler is running in a delivery are quite different from key video frames of other video shots.

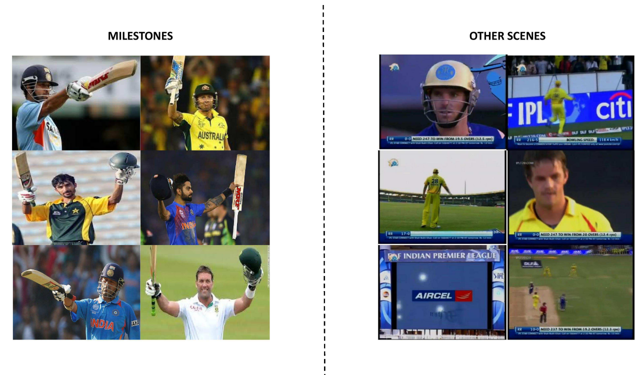

Figure 4. Samples of images from the dataset that was used for training the player celebration model. Positive samples included images of batsmen celebrating whereas negative samples were comprised of other frames sampled from cricket videos.

We employ a two-step methodology to identify runs and wickets in the scoreboard. The positions of the digits are located using an approach proposed by [5]. Once the text regions are detected, the overlapping bounding boxes are removed and each text region is allocated with a single bounding box. Only digits and the dash symbol '-' (used for distinguishing runs from wickets) are present inside these bounding boxes, so the OCR classifier was trained to recognize just this limited set of characters. The model was trained on 1100 images with 100 images for each class. Descriptors were extracted from the fc7 layer of the pretrained AlexNet model, while the dimension of the representations for each sample was 4096×1. The model was trained on a multi-class linear SVM and a *one-vs-all* strategy was used for training the model. The value of the kernel offset was set to 0 and the value of C was set to 1.0. A similar framework has previously been applied for various other domains[14, 15]

The extracted digits were then used for determining the score. The rightmost digit extracted corresponds to the total number of wickets that have fallen, whereas the remaining digits — or the digits that are to the left of the '-' symbol — are combined to calculate the total runs that have been scored.

3.4. Detecting Playfield Scenarios

Once an important event is discovered, a question arises: *How the video shots comprising an event should be clipped?*. The methodology designed to solve this problem relies on recognizing playfield scenarios and uses domain knowledge for labeling the key frame of a given video shot as the starting frame.

Almost all events in a cricket match start with the bowler running up to bowl a delivery. In such cases, the camera focuses is either on the bowler or it incorporates the bowler, the striker, the nonstriker, and the umpire, along with the pitch. Detecting these key frames would lead to capturing the entire video shot where the bowler runs and bowls a delivery. These video frames can then be used to mark the start of an event. The intuition is to develop a model that is able to distinguish key frames (starting frames) of a video shot that depict a bowler running out to bowl a delivery from other key frames that depict other aspects of the game. A sample of such key frames along with others during a cricket match is shown in Figure 2.

A CNN+SVM methodology was adopted to classify key frames as starting frames. A binary linear SVM was trained on representations extracted from the *fc7* layer of AlexNet. The model was trained on 500 images of each class to classify a given video frame as a starting frame or not.

3.5. Audio Cues

Audio intensity is an important cue to identify key moments in a cricket match. Excitement of the crowd, appeals and intense commentaries are examples of important audio cues that are often associated with important events in a match. The proposed method utilizes loudness as an audio feature for detecting important events during the game. Audio features have been used for detecting milestones and miscellaneous events, like dropped catches.

3.6. Detecting Player Celebration

Celebrations are important cues for detecting milestones in a cricket match. We trained a model to detect the celebration by a batsman. While the positive class was comprised of images in which a batsman had just completed a milestone, the negative class was trained on images that normally occur in a cricket match. The network was trained on grayscale images because the colored outfit worn by batsmen can lead to misclassification. Representations were extracted from the *fc7* layer of the AlexNet framework. The model was trained on a CNN+bilinear SVM framework that relied on representations extracted by *fc7* layer of the AlexNet model. A sample of the positive and negative classes used for training the model is shown in Figure 4.

4. Event-Driven Highlight Generation

The understanding that highlights can be termed as a collection of important events leads to a question: *What kind of event in a cricket match can be regarded as important?*

In order to answer this question, we reviewed several cricket matches and highlights to come up with a list of events that are generally considered important in such broadcasts. These events are: *boundaries, sixes, wickets, first and last deliveries of an inning, milestones, and pressure moments during a chase* .

Since the qualitative evaluation of a highlight is subjective, we conducted a user study with 15 cricket fans aged between 18 and 40 to find out how typical viewers would agree with this original list. We asked each participant to rank these events in terms of their relevance and frequency in a cricket match. The top four events were milestones, wickets, fours and sixes. In this section, we propose an event-driven strategy to extract such events.

4.1. Boundaries, Sixes and Wickets

4.1.1 Event Discovery

The approach discussed in Section 3.3 enables us to determine the total number of runs and the total number of wickets that have fallen for a given video shot. The following analogy is used for detecting boundaries, sixes or fall of wickets:

- A boundary is detected if the difference in the number of runs for two consecutive video shots is equal to four.
- A six is detected if the difference in the number of runs for two consecutive video shots is equal to six.
- A wicket is said to have fallen if the number of the wicket column increases by one for two consecutive video shots.

4.1.2 Event Segmentation

Once a boundary, a six, or a fall of wicket is detected, the next challenge lies in segmenting this event. Since our model considers events as sequences of video shots, the problem is reduced to recognizing the total video shots that make up the event. The central video is said to be the one tagging the actual event. In order to detect its starting frame, we select k video shots preceding the central video. During this step, all replays are neglected. The algorithm discussed in Section 3.4 for detecting the starting frame is then applied to the key frames of all the k video shots preceding the central video shot. The first frame detected as the positive class by the model is marked as the starting frame. In case the algorithm fails to detect a starting video shot, then the k^{th} video shot from the central video shot is marked as the starting point. Furthermore, L video shots succeeding the central video shot are clipped to complete the event. This is done in order to show the secondary events that are linked with the central event. These may include replays, celebrations from team members when a wicket falls, batsmen walking back to the pavilion, the cheer or the silence of the crowd, etc.

Skipping replays while deciding the starting frame is a crucial step as events like run-outs and stumpings involve a lot of replays before the decision is actually made. Setting the value to k without neglecting these frames would result in a faulty segmentation. It must be noted, however, that these replays are included as highlights when the event is segmented. This can be clearly demonstrated in Figure 5. The value of K was selected as 3 for all the 3 cases, Whereas the value of L was chosen to be 1 for boundaries and sixes and 3 for wickets. This was done because a larger number of secondary events are associated with the fall of wickets.

4.2. Milestones

Milestones are important events that emphasize the individual contribution of the batsmen in an inning. Batsmen often celebrate when reaching a half-century or a century. The celebration might be a gentle raise of the bat or a fierce one, in which a batsman may perform a signature move. Milestones are generally accompanied by applause from the crowd and the commentators, which translates into an increase in audio intensity. Based on this observation, all the video shots in an over are sorted in our framework according to their audio intensity and the top 5 video shots are selected for analysis. A given video shot is marked as a probable candidate for a milestone if:

- The loudness of the video shot is among the top 5 values for all video shots in the over.
- The delivery preceding the video shot has not been categorized as a boundary, a six, a wicket, or any other kind of important event.

Once the probable candidate is selected, we run through a sample of the frames composing the video shot and perform the procedure described in section 3.6 to detect player celebration in each individual frame. Milestone detection is then determined by the amount of positive occurrences in the video shot.

If a milestone is detected, the current video shot is marked as the central one. The algorithm discussed in Sec-

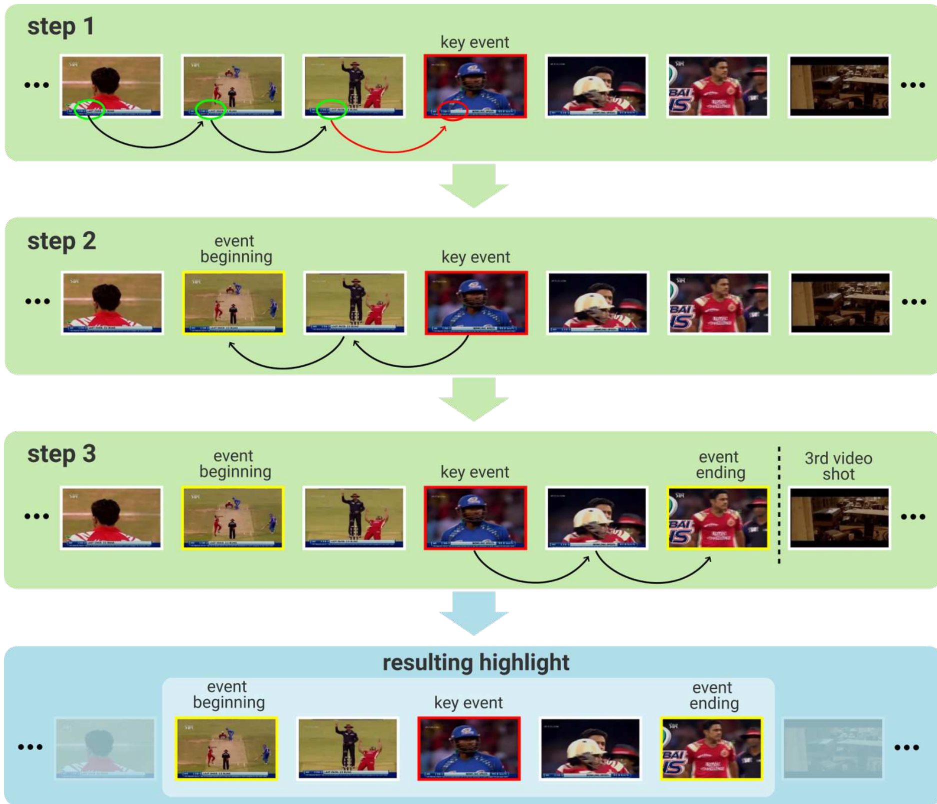

Figure 5. Detection of events in a cricket match. We sequentially scan the key frames in a video shot and look for a potential event as shown in step 1. Once an event is detected (in the figure, this happens by the change in the wicket column), the model goes backward till it detects the start of the delivery (this is done using the techniques mentioned in Section 3.4). The model now clips the video footage from the start of the delivery to the k-key frames after the central key frame.

tion 3.4 for detecting the starting frame is then applied to the key frames of all the k video shots preceding the central one. The remaining steps are the same used for segmenting highlights of boundaries.

5. Excitement-Based Highlight Generation

Excitement-based highlights are generated in a similar manner as milestones. We use replays and audio intensity to mark secondary events in a cricket match, such as dropped catches or pressure moments. A secondary event is said to be important if:

- The loudness of a video shot is among the top 5 values for all video shots in the over.
- The delivery preceding the video shot has not been categorized as a boundary, a six, a wicket, or any other kind of primary event.
- There are more than three replays associated with the event.

We employ a strategy that uses the algorithm discussed in Section 3.4: firstly, we detect the starting frame and then we perform a search over k video shots preceding the central one. Detecting consecutive starting frames gives us the duration between two deliveries. An event is clipped as a highlight if it satisfies all three conditions. The event is clipped from the start of the delivery to the time when the first replay occurs.

6. Dataset Description

A new dataset was constructed in order to evaluate the proposed framework. The dataset contains over 15 hours of video footage of T-20 matches of the 2010's Indian Premier League. In order to reduce the demand for processing power, the complete cricket match was processed over by over, with an over being comprised of six legal deliveries. The dataset was annotated manually and all the important events, such as boundaries, sixes, wickets, and milestones, were associated to corresponding video shots. Each of these videos were broken into video shots as described in Section 2. The annotations were made on video shots rather than video frames, and each video shot was labeled as a highlight, with a resolution of 640×480 pixels.

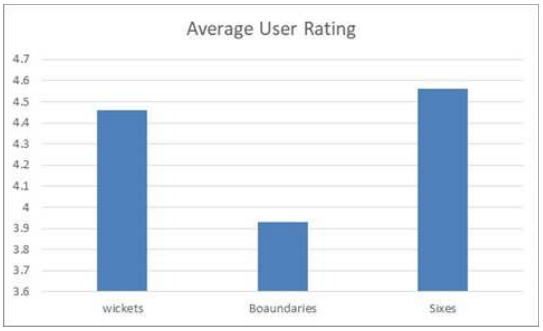

Figure 6. Average user ratings for boundaries, sixes and wickets. These ratings were calculated as a part of the user study to evaluate the performance of our algorithm.

7. Evaluation

Generating highlights is a task that presents a considerable degree of subjectivity. Hence, there is no objective ground truth when assessing the algorithm's performance. That is why we adopt a strategy similar to [2, 9] for evaluating our approach. Firstly, we performed a user study to validate the perceptual quality of our highlights. Secondly, the generated output by our algorithm was compared to the official highlights generated by the broadcasters for each game. In this section, we present the results that show a correlation between our highlights and the official ones.

7.1. User Study

We conducted a user study in order to determine the perceptual quality of the highlights generated by our approach. The procedure was similar to a user study conducted by [9]: 26 clips from a pool of 200 randomly selected highlights clips generated by our algorithm were shown to 20 different cricket fans in the age group of 20-50 years. The participants were asked to rate the clips from 0 to 5. Ratings below 2 indicated that the clips should not be considered as highlights, while ratings between 2-5 were considered relevant, with 5 designating the highlight with the highest quality possible. Since boundaries, sixes and wickets are generally considered important events, we also evaluated the specific mean average ratings for these events.

The average user ratings for boundaries, sixes and wickets is shown in Figure 6. An overall mean rating of 4.315 was obtained for all the clips that included these events as well as other important events in the match. Mean average ratings of 3.96, 4.46, and 4.57 were obtained for boundaries, wickets and sixes, respectively.

7.2. Baseline Comparison with Official Highlights

In this section we draw a comparison between the highlights generated by our model and the official highlights generated by the broadcasters. We employ the Intersection of Union (IOU) metric to capture the overlap between both highlights for a given inning [16, 1], as it is generally considered a good metric to judge the current overlap between two events that occur during the course of time.

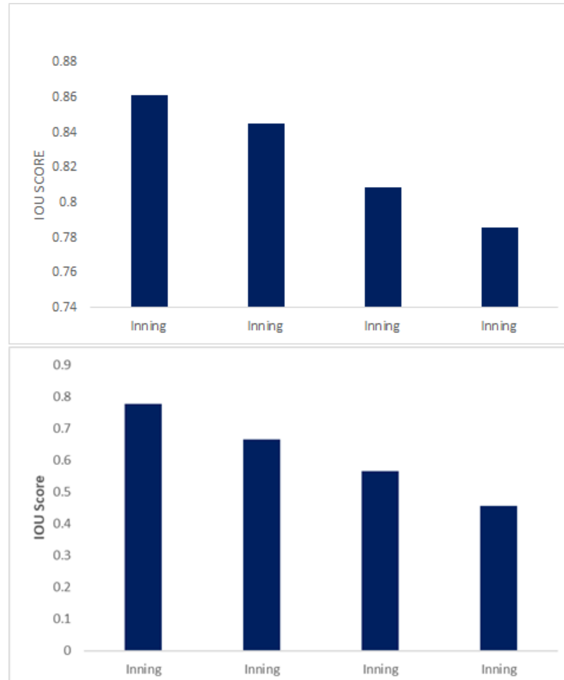

Figure 7. Innings having the highest and lowest intersection of Unions.

We were able to achieve a mean IOU of 0.731 for all innings in the dataset. The top 4 IOU's and the worst 4 IOU's obtained by model are shown in Figure 7. The highest values of IOU was 0.86 whereas the lowest value was 0.454.

7.3. Event Recognition

In this section we analyze the performance of our model to recognize important events during a cricket match. Four different events (fours, sixes, wickets and milestones) were evaluated and the Mean Average Precision (MAP) for each one of the events is shown in Table 8. A mean average precision greater than 85% was obtained for boundaries, sixes and wickets, while milestones were recognized with an MAP of 72.31%.

8. Discussion

Through a series of experiments, we sought to demonstrate the utility of our model and its applicability in real-life scenarios. Our user study resulted in an average rating of 4.315 from 26 user on a 0-5 scale, showing that highlights produced by our framework were well regarded by cricket fans.

Table 1. Mean Average Precision (MAP) calculated for the four major events that occur during a cricket match.

Event	MAP (%)
Boundaries	89.23
Sixes	86.79
Wickets	92.68
Milestones	72.31

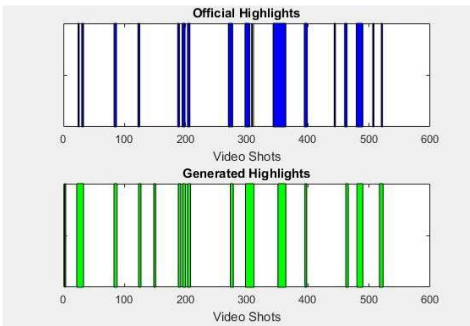

Figure 8. Comparison between the official highlights and the automatically generated highlights for an inning.

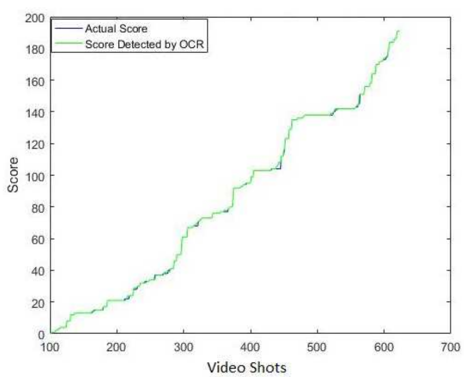

Figure 9. Comparison of the actual scoreboard for an inning and the score calculated by our OCR.

Quantitatively, our method obtained IOU values of 0.731, indicating that there was a significant overlap between our highlights and those generated by the official broadcaster. The highest value of IOU was 0.86, while the lowest was 0.45, suggesting that even though manual and automatic highlights are correlated, they still disagree on the importance of some specific events. Nevertheless, our third set of results clearly demonstrated the capability of our model to detect important events in a cricket match.

The fact that our model was able to generate highlights comparable to the official ones can be attributed to several reasons. First, using video shots instead of video frames allows the model to segment and identify smaller events more consistently. Not only do video shots provide better segmentation boundaries, they are also useful as far as labeling sub-events, such as replays and moments with high audio intensity. Thus, the underlying principal that highlights are sequences of important video shots clubbed together.

Secondly, based on the assumption that most important events on a cricket match begin when a delivery is being bowled, we managed to reduce the search space complexity of key video shots that comprise such events. Moreover, the proposed framework demonstrated to perform robustly while detecting the start of deliveries.

Finally, we used both event-driven and excitement-based strategies to extract highlights. Using both of them makes the algorithm more flexible and reliable, apart from helping detect secondary events on a cricket match. In our trials, this decision led to a perceptual improvement on overall quality of the highlights.

There are, however, certain limitations to our work that we intend to address in the future. Currently, our approach might be slightly longer and include clips that are not part of the official highlights. Besides, there are no clear metrics to evaluate false positives in highlights because false positives are intrinsically subjective in this case. Another challenging topic of investigation is detecting unusual events that can take place on a match and are of great relevance for highlights, such as extreme weather, injuries, or intense bickering between players.

9. Conclusion

We proposed a new technique to automatically generate cricket highlights, focusing on both event-driven and excitement-based features. We showed that our framework can achieve comparable results to manual highlights and that it yields acceptable results for cricket fans. Overall, we demonstrated that dividing a cricket match into video shots and cues, such as replays, audio intensity, scoreboard, player celebration, and playfield scenarios, we can create high-quality highlights without human supervision. Although there is still room for improvement, we also intend to extend this work to generate automatic captions for sports videos as well as for highlight clips of individual players.

10. Acknowledgement

We would like to extend our gratitude towards **Dr. Ankush Mittal**, Director, Raman Classes, Roorkee who motivated us to carry out the research. The authors would also like to thank **Pradeep Pant**, **Monish Sammal**, **Rohan Mishra**, **Arpit Ahuja** and **Aman Bhalla** for helping us with the project. We acknowledge partial support of this work by the Sao Paulo Research Foundation, grant 2017/06933-0

References

[1] S. Bambach, S. Lee, D. J. Crandall, and C. Yu. Lending a hand: Detecting hands and recognizing activities in complex egocentric interactions. In *Computer Vision (ICCV), 2015 IEEE International Conference on*, pages 1949–1957. IEEE, 2015.

[2] V. Bettadapura, C. Pantofaru, and I. Essa. Leveraging contextual cues for generating basketball highlights. In *Proceedings of the 2016 ACM on Multimedia Conference*, pages 908–917. ACM, 2016.

[3] D. Chauhan, N. M. Patel, and M. Joshi. Automatic summarization of basketball sport video. In *Next Generation Computing Technologies (NGCT), 2016 2nd International Conference on*, pages 670–673. IEEE, 2016.

[4] T. Decroos, V. Dzyuba, J. Van Haaren, and J. Davis. Predicting soccer highlights from spatio-temporal match event streams. In *AAAI*, pages 1302–1308, 2017.

[5] A. Gonzalez, L. M. Bergasa, J. J. Yebes, and S. Bronte. Text location in complex images. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 617–620. IEEE, 2012.

[6] D. Joshi, M. Merler, Q.-B. Nguyen, S. Hammer, J. Kent, J. R. Smith, and R. S. Feris. Ibm high-five: Highlights from intelligent video engine. In *Proceedings of the 2017 ACM on Multimedia Conference*, pages 1249–1250. ACM, 2017.

[7] M. H. Kolekar and S. Sengupta. Caption content analysis based automated cricket highlight generation. In *National Communications Conference (NCC), Mumbai*, pages 461–465, 2008.

[8] 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.

[9] 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), 2017 IEEE Conference on*, pages 57–65. IEEE, 2017.

[10] K. Namuduri. Automatic extraction of highlights from a cricket video using mpeg-7 descriptors. In *Communication Systems and Networks and Workshops, 2009. COMSNETS 2009. First International*, pages 1–3. IEEE, 2009.

[11] 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.

[12] K. P. Sankar, S. Pandey, and C. Jawahar. Text driven temporal segmentation of cricket videos. In *Computer Vision, Graphics and Image Processing*, pages 433–444. Springer, 2006.

[13] H.-C. Shih. A survey on content-aware video analysis for sports. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017.

[14] P. Shukla, I. Dua, B. Raman, and A. Mittal. A computer vision framework for detecting and preventing human-elephant collisions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2883–2890, 2017.

[15] P. Shukla, T. Gupta, A. Saini, P. Singh, and R. Balasubramanian. A deep learning frame-work for recognizing developmental disorders. In *Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on*, pages 705–714. IEEE, 2017.

[16] K. Soomro and A. R. Zamir. Action recognition in realistic sports videos. In *Computer Vision in Sports*, pages 181–208. Springer, 2014.

[17] M. Sun, A. Farhadi, and S. Seitz. Ranking domain-specific highlights by analyzing edited videos. In *European conference on computer vision*, pages 787–802. Springer, 2014.

[18] H. Tang, V. Kwatra, M. E. Sargin, and U. Gargi. Detecting highlights in sports videos: Cricket as a test case. In *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, pages 1–6. IEEE, 2011.

[19] K. Wan, J. Wang, C. Xu, and Q. Tian. Automatic sports highlights extraction with content augmentation. In *Pacific-Rim Conference on Multimedia*, pages 19–26. Springer, 2004.

[20] K. Zhang, W.-L. Chao, F. Sha, and K. Grauman. Video summarization with long short-term memory. In *European conference on computer vision*, pages 766–782. Springer, 2016.