

# Event Detection Based Video Summarization for Sport Video

Ph.D. Synopsis

Submitted To



Gujarat Technological University

For The Degree

Of

Doctor of Philosophy

In

Computer/IT Engineering

By

**KHUSHALI R. RAVAL**

Enrollment No: 189999913039

Supervisor:

Dr. Mahesh Goyani, Assistant Professor

Government Engineering College, Modasa, Gujarat.

# Index

1 Title of the thesis and abstract .....	3
1.1 Title of the thesis .....	3
1.2 Abstract .....	3
2 A Brief description on the state of the art of the research topic. ....	3
3 Definition of the problem .....	5
4 Objective and scope of work .....	5
5 Original contributions by the thesis.....	5
6 Methodology of Research, Results / Comparisons .....	6
6.1 Shot boundary detection.....	6
6.2 Replay detection.....	9
6.3 Shot classification .....	10
6.4 Score detection.....	13
6.5 Event detection framework .....	14
6.6 Wicket fall classification.....	15
7 Results and Discussion.....	17
8 Achievements with respect to objectives .....	21
9 Conclusions .....	21
10 Publications .....	22
11 References .....	22

## **1. Title of the thesis and abstract**

### **1.1 Title of the thesis**

Event Detection Based Video Summarization for Sport Video

### **1.2 Abstract**

Sports video summarization is the process of automatically creating a shorter version of a sports game or event video, highlighting the most important and exciting moments. This technology has become increasingly important in the sports industry, as fans increasingly consume sports content on social media and mobile devices. We have proposed a method to detect and classify exciting events of cricket matches like four runs, six runs and wicket fall. First, cricket video is divided into shots, and keyframes from each shot are detected. To reduce processing time all other processes are applied to keyframes. In the following stage replay segments are identified and removed from further consideration. Keyframes are classified into different views. Finally, scoreboard text is detected to identify run and wicket differences. If the run difference is equal to 4 or 6 and the wicket difference is equal to 1, then frames between such keyframes are used to build the match's highlight. The wicket fall segment of the video is further classified into a different type. The proposed technique is modest and efficient. Experiments and results reflect the effectiveness of the proposed method.

## **2. Brief description of the state of the art of the research topic**

Sports video summarization is the process of condensing a sports video into a shorter and more concise form while retaining the key highlights and important moments of the game. It has become an important research area due to the increasing amount of sports videos available on various platforms, as well as the desire of viewers to watch condensed versions of games.

Several studies have been conducted on sports video summarization using various techniques. Existing techniques for sport video analysis are mainly classified into shot detection [1] and shot classification [2], [3], replay identification [4], [5] and event detection [6]–[8].

The primary step of semantic analysis of sports video is to divide the video into shots and play-break units. Researchers propose various ways for shot boundary detection and keyframe extraction. Aman Bhalla et al. [1] proposed a shot boundary detection method by converting RGB frames into grayscale images and calculating the absolute difference between grayscale images. If

the difference between two frames is more significant than the threshold, the frame is stored as a boundary frame. K. Midhu and N. K. Anantha Padmanabhan [9] used the hue histogram difference between the video frames and compared it with an empirically selected threshold to detect hard cuts. Static threshold methods are affected by the kind of incoming video stream. The threshold value must be determined manually to achieve an acceptable outcome.

In sports video broadcasting, key events are represented within replay segments. So the replay events can be used to construct a video summary. Slow motion replays are showing the previous significant events comprehensively. Some researchers are detecting replays and removing them from the video summary as redundant information is shown by the replays. Ali Javed et al. [4] proposed a technique for gradual transition detection based on cumulative histogram difference and dual-threshold comparison. This technique can be applied to different sports. The hue-histogram dissimilarity between each video frame and scoreboard reference frame was used to identify replay segments [9].

Certain types of shots are featured in sequential order in sports videos. Once shots are segmented, it is required to classify them for further analysis. In cricket, shots are classified into field view, boundary view, audience view, player's gathering, and close-up view. Domain-specific knowledge must be incorporated to classify the shots accurately. M. Ravinder and T. Venugopal [2] classified cricket shots using bag-of-visual features. The long view, pitch view and close-up view were detected by extracting 100 Scale-Invariant Feature Transform (SIFT) features and then clustering those using k-means. Rabia A. Minhas et al. [10] proposed a technique for classifying cricket shots into long field view, median view, non-field view and close-up view, using AlexNet CNN.

Important segments from the cricket video are recognized and indexed using different event detection techniques. Pushkar Shukla et al. [8] detected exciting events using the audio feature loudness. Hao Tang et al. [11] introduced a new method for identifying highlights in sports videos, which involves dividing the videos into a sequence of events using an unsupervised event discovery and detection framework. The framework utilizes low-level visual features such as color histograms and histograms of oriented gradients that can be easily extracted and applied to different sports. The detected events are then represented using unigram and bigram statistics to provide a concise illustration of the video.

A reliable and robust technique for the analysis of cricket videos is an important component of work for obtaining successful video summarization. Visual features have been used by several

researchers to analyze videos. Our focus is on improving automation in video analysis, which will also allow us to improve performance. This research aims to develop an efficient framework for cricket video event detection and classification of interesting events. The proposed framework effectively uses the domain knowledge for shot boundary detection, replay detection, view classification, event detection, and event classification. Once important events are identified, they are fused to generate highlights of cricket matches.

### **3. Definition of the problem**

The research title is “Event Detection Based Video Summarization for Sports Video”. This research aims to accurately detect and classify interesting events from lengthy cricket match videos. To enhance the performance of event detection, the text based approach is used. Wicket fall classification is yet not explored much. To precisely classify wicket fall into different types using a single technique text information is extracted and domain knowledge is incorporated into the proposed method.

### **4. Objective and scope of work**

- To develop an effective shot boundary detection technique.
- To detect replay segments from cricket videos efficiently.
- To accurately classify play frames into different views.
- To perform cricket event detection and classification using a robust text detection and recognition based approach.

### **5. Original contributions by the thesis**

1. In this research video dataset containing cricket videos with different playing conditions are collected as there is no standard dataset available for cricket video summarization.
2. Video is divided into frames and an efficient adaptive method for shot boundary detection using k-means clustering is proposed and applied to these frames to identify shot transitions. Keyframes are selected from each identified shot to apply further processing.
3. A replay detection technique based on scoreboard existence recognition is proposed to classify the keyframes into play frames and replay frames.
4. A modest and effective view classification technique is applied to all play frames.

5. We have proposed an efficient technique for the detection of four, six and wicket fall events and wicket fall classification based on scoreboard text detection and recognition.

## **6. Methodology of Research, Results / Comparisons**

We have proposed a system for cricket video summarization using the hierarchical framework shown in Figure 1. The figure illustrates that the cricket video is initially divided into frames. These extracted frames are used for shot boundary detection. Once shot boundaries are detected, every shot's first frame is preserved as a representative frame of that shot and identified as a keyframe. All further processing is applied to these keyframes. After shot boundary detection, replay and play frames are classified.

As replay segments of the video show redundant information in slow motion, we have detected and removed replay segments from further consideration. After replay removal, all play frames are further classified as field view and non-field view frames. In cricket, at the start of balling, pitch views are shown. So it is necessary to detect pitch as a mark of delivery of a new ball. The pitch view detection technique is applied to all field view frames and from that, frames showing pitch are extracted. All non-field view frames are classified as close-up or crowd frames. Finally, the score caption detection technique is applied for event detection. The highlight is built for that segment of the video where a significant change in run score or wicket score is identified. The detailed methods for the event detection framework are described in the following subsections.

### **6.1 Shot Boundary Detection**

A video is a collection of shots. Shots are a set of continuous frames representing semantically similar concepts. The video is divided into shots in the first stage of the event detection framework. To do such segmentation, shot boundaries need to be detected. Usually, two different types of shot transitions are observed in sports videos. 1. Cut and 2. Gradual transition. In the cut type of transition, changes between the frames are abrupt at the boundary, while in gradual transition the last few frames of the former shot are progressively substituted by the first few frames of the present shot. The discontinuity between shots is shown in Figure 2.

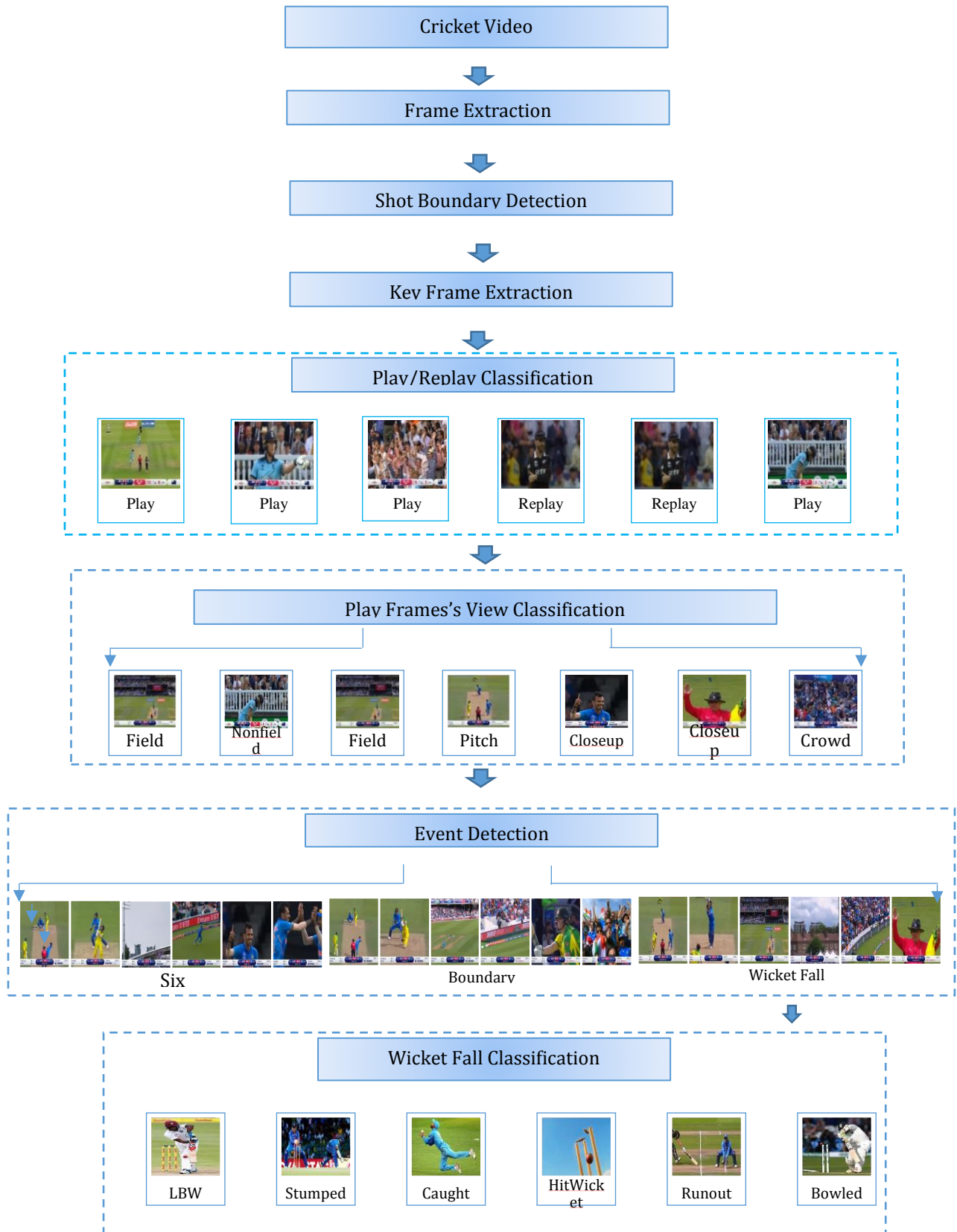


Figure 1. Event detection framework

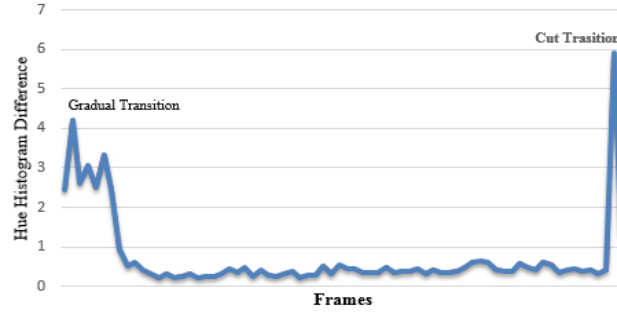


Figure 2. Discontinuity between shots

We have proposed a method for shot boundary detection and keyframe extraction using a clustering-based technique. As frames in shots usually cover similar concepts, other than processing each video frame, representative frames from every shot are identified. Then, each further processing is applied to such representative frames to reduce processing time. The detailed algorithm is shown below in Algorithm 1.

**Algorithm 1. Shot boundary detection**

$N \leftarrow \text{length}(\text{InputVideo})$

**for**  $i \leftarrow 1$  to  $N$

$I_{RGB} \leftarrow \text{ReadFrame}(\text{InputVideo})$

$I_{HSV} \leftarrow \text{RGBtoHSV}(I_{RGB})$

$$\text{HueHistogramDifference}(n) = \frac{1}{r \times c} \sum_{i=1}^{256} |(HH_{I_{HSV(n)}}(i) - HH_{I_{HSV(n+1)}}(i)|$$

**end**

$(idx, C) \leftarrow \text{kmeans}(\text{HueHistogramDifference}, 2)$

$$\text{ShotBoundaryCandidate} = \bigcup_{i=1}^N I_{RGB(i)} (idx_i == 1)$$

$lsbc \leftarrow \text{length}(\text{ShotBoundaryCandidate})$

$\text{ShotBoundaryFrames}$

$$= \bigcup_{i=1}^{lsbc} (\text{ShotBoundaryCandidate}_{i+1} - \text{ShotBoundaryCandidate}_i) \neq 1$$

## 6.2 Replay Detection

In sports videos, exciting events are followed by replays. As shown in Figure 3. replays are generally shown between two gradual logo transitions and the scoreboard is absent in the replay video segment. The replay represents interesting events in slow motion. Hence contents of the replay are redundant. To improve the accuracy and avoid redundant processing, we have to filter out replays.



Figure 3. Replay segment of video

The main drawback of replay detection techniques using logo transition detection is it needs to learn the logo template each time. In the proposed method, we have tested the nonexistence of a scoreboard to identify replay segments. The algorithm for replay detection is shown below.

### **Algorithm 2. Replay detection**

```

 $I_{AVG} = \sum_{i=1}^{150} I_{RGB(i)} / 150$ 
 $I_{BW} \leftarrow BW(I_{AVG})$ 
 $I_{SBR} \leftarrow BlackWhiteAreaFilter(I_{BW})$ 
 $SCX_{MIN} \leftarrow MinX(I_{SBR})$ 
 $SCX_{MAX} \leftarrow MaxX(I_{SBR})$ 
 $SCY_{MIN} \leftarrow MinY(I_{SBR})$ 
 $SCY_{MAX} \leftarrow MaxY(I_{SBR})$ 
 $SCR_{RefFrame} \leftarrow Region(SCX_{MIN}, SCY_{MIN}, SCX_{MAX}, SCY_{MAX})$ 
for  $i \leftarrow 1$  to  $length(KeyFrame)$ 
     $SCR_{KeyFrame_i} \leftarrow Region(SCX_{MIN}, SCY_{MIN}, SCX_{MAX}, SCY_{MAX})$ 
     $Correlation = CorrelationCoefficient(SCR_{RefFrame} - SCR_{KeyFrame_i})$ 
    If  $Correlation > Th_{Replay}$  than  $PlayFrame \leftarrow KeyFrame_i$ 
    Else  $ReplayFrame \leftarrow KeyFrame_i$ 
end

```

Figure 4. shows the outcome of temporal image averaging and scoreboard region detection process.

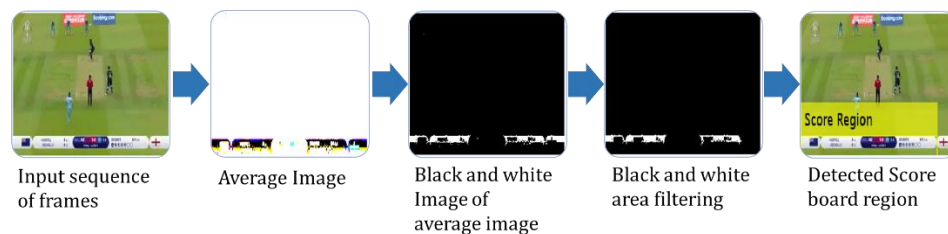


Figure 4. Scoreboard detection process

### 6.3 Shot Classification

Sports video events cover different shots in sequential order. Shot view classification is an elementary requisite for high-level analysis of sports video [12]. To classify various shots usually, visual cues are used. All keyframes belonging to play segments are fed into the shot classification framework. Initially, our work classified cricket views into field view and non-field view. Pitch views are extracted from field view frames and non-field view frames are further classified into close-up and crowd views. In the following subsection, cricket view classification algorithms are discussed.

#### 6.3.1 Field view detection

Generally, the field view covers a large green area. The hue histograms of the field view and non field view are shown in Figure 5. The field view and the non-field view are classified by examining the green pixel ratio similar to [13].

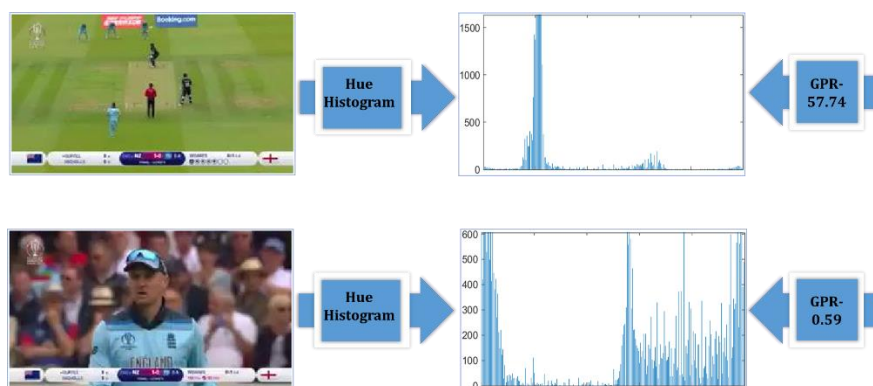


Figure 5. Hue histogram of field view and non field view

Algorithm 3 shows steps for field view detection. RGB frames are converted into HSV images. As the field color is different in each cricket video, we need to check the hue range for the green color. Once the hue range is decided, the peak value of green is observed and a number of pixels having a hue value nearby the green peak are summed and used to identify the Green Pixel Ratio (GPR) [14]. If the GPR value is high, the keyframe is classified as a field view frame otherwise; it is classified as a non-field view frame. The algorithm for field view detection is shown below.

**Algorithm 3. Field view detection**

$I_{RGB} \leftarrow \text{PlayFrame}$

$I_{HSV} \leftarrow \text{RGBtoHSV}(I_{RGB})$

$\text{GreenWindow} \leftarrow \text{InputGreenRange}$

$G_{\text{Peak}} \leftarrow \text{Max}(\text{GreenWindow}_{\text{Start}} : \text{GreenWindow}_{\text{End}})$

$$\text{GPR} = \frac{\text{Total number of pixel having hue} = G_{\text{peak}} \pm 2}{\text{Total number of pixels in the frame}} \times 100$$

If  $\text{GPR} > Th_{\text{Field}}$  then  $\text{FieldFrame} \leftarrow \text{PlayFrame}$

Else  $\text{NonFieldFrame} \leftarrow \text{PlayFrame}$

### 6.3.2 Pitch view detection

The pitch view covers a large pitch area, so the pitch color ratio is more significant than other colors. For the detection of pitch view, RGB frames are converted into HSV images. As the pitch color is dissimilar in each cricket video, we need to check the hue range for the soil color. Figure 6 displays pitch views and the corresponding hue histogram of it.

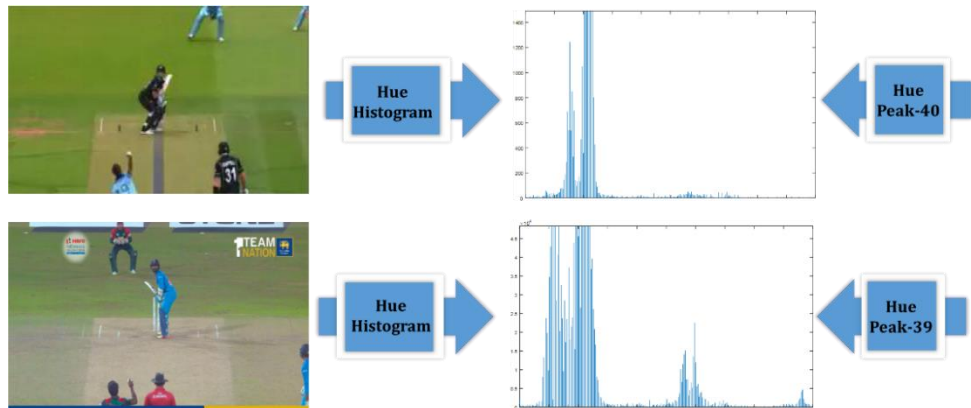


Figure 6. Hue histogram of pitch view

Steps for pitch view discovery are given in Algorithm 4. Once the hue range for pitch is decided, the peak value of the soil is observed, and a number of pixels having a hue value equal to the soil peak are summed and used to identify the Pitch Pixel Ratio (PPR) [14]. If the PPR value is high, the keyframe is classified as a pitch view frame. Otherwise, it is classified as a non-pitch view frame. The algorithm for pitch view detection is shown below.

**Algorithm 4. Pitch view detection**

```

 $I_{RGB} \leftarrow \text{FieldFrame}$ 
 $I_{HSV} \leftarrow \text{RGBtoHSV}(I_{RGB})$ 
 $\text{PitchWindow} \leftarrow \text{InputPitchRange}$ 
 $P_{\text{Peak}} \leftarrow \text{Max}(\text{PitchWindow}_{\text{Start}}:\text{PitchWindow}_{\text{End}})$ 

$$\text{PPR} = \frac{\text{Total number of pixel having hue} = P_{\text{peak}}}{\text{Total number of pixels in the frame}} \times 100$$

If  $\text{PPR} > Th_{\text{pitch}}$  then  $\text{PitchFrame} \leftarrow \text{FieldFrame}$ 
Else  $\text{NonPitchFrame} \leftarrow \text{FieldFrame}$ 

```

### 6.3.3 Non-field view classification into Close-up and Crowd

It is perceived that all non-field views cover either a crowd or a close-up view of any player. Crowd shots generally display the audience or player's gathering and a close-up view spectacles zoomed-in view of a player [12]. As shown in Figure 7 edge density is high in the crowd view compare to close up view. For the classification of non-field views into close-up and crowd views, we have used the edge detection method. Algorithm 5 displays the process for non field view classification.

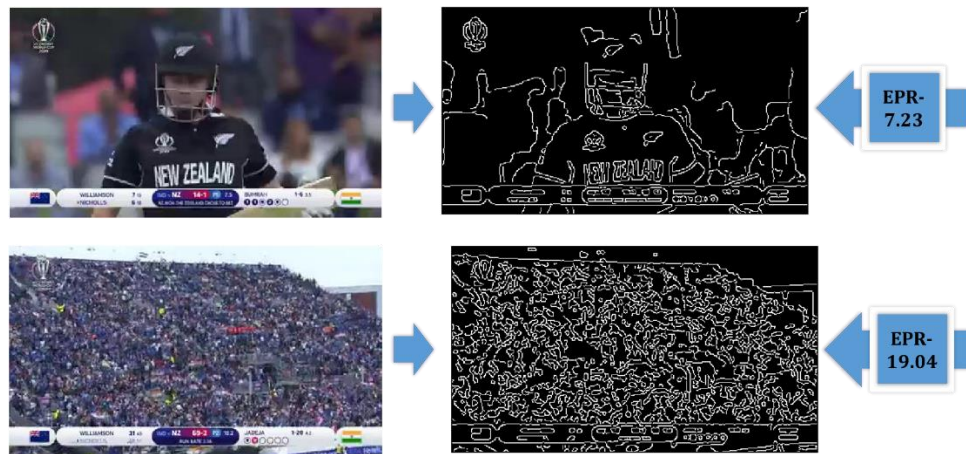


Figure 7. Edge detection of a crowd and close-up views

First, non-field view RGB frames are converted into Grayscale images. The canny edge detector is applied to a grayscale image. As close-up views have less edge concentration than crowd views, EPR is lesser for close-up views.

**Algorithm 5. Non field view classification**

```

 $I_{RGB} \leftarrow NonFieldFrame$ 
 $I_{Gray} \leftarrow RGBtoGray(I_{RGB})$ 
 $I_{BW} \leftarrow CannyEdgeDetection(I_{Gray})$ 

$$EPR = \frac{Total\ number\ of\ edge\ pixels}{Total\ number\ of\ pixels\ in\ the\ frame} \times 100$$

If  $EPR > Th_{Crowd}$  then  $CrowdFrame \leftarrow NonFieldFrame$ 
Else  $CloseupFrame \leftarrow NonFieldFrame$ 

```

#### 6.4 Score detection

The text data presented in the scoreboard encompasses fruitful information for the event discovery. Events in cricket like four runs, six runs and wicket fall can be mined by checking the difference in run and wicket mentioned on the scoreboard. The following steps are usually taken to extract the information from the scoreboard. 1. Locate the interesting region of the scoreboard 2. Recognize the information of that region [15]. The style, dimension and position of the scoreboard are generally static for every frame of sports video. In the proposed work as illustrated in Algorithm 6 score and wicket caption regions are first located and segmented. As the size of this region is too small to perform the caption recognition task, the cropped image is enlarged. After enlarging the image, top-hat filter is applied. The image is then converted into black and white. Finally, OCR is applied to recognize the score and wicket caption.

**Algorithm 6. Score Detection**

```

 $I_{RGB} \leftarrow KeyFrame$ 
 $I_{SCR} \leftarrow ScoreCaptionRegion(I_{RGB})$ 
 $I_{SCR} \leftarrow Enlarge(I_{SCR})$ 
 $I_{TH} \leftarrow TopHatFilter(I_{SCR})$ 
 $I_{BW} \leftarrow Binarization(I_{TH})$ 
 $(Run, Separator, Wicket) \leftarrow OCR(I_{BW})$ 

```

Figure 8 illustrates the outcomes of the score detection process. Top hat filter followed by opening operation increases the accuracy of text recognition.

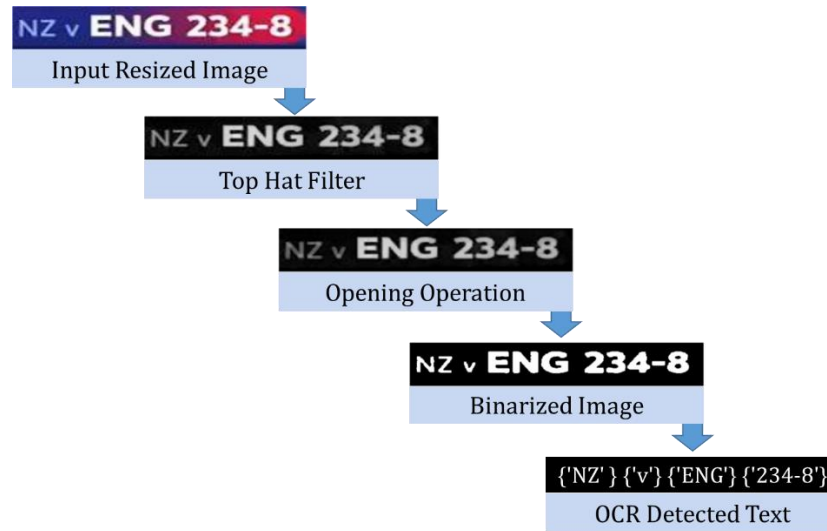


Figure 8. Score detection process

## 6.5 Event detection framework

Once run and wicket information is extracted, we save the consecutive keyframes for which the run difference is 4 or 6 and the wicket difference is 1. Then we search for the pitch view from that keyframe in the reverse direction. Once we get a keyframe that shows a pitch view, we save that keyframe. The highlight is generated that starts from a pitch view and ends with either a close up or a crowd view. This is how the highlights of exciting events are generated.

### **Algorithm 7. Event Detection**

*Load run and wicket information as RC and WC*

*If  $(RC(i+1)-RC(i) == 4 \parallel 6)$  or  $(WC(i+1)-WC(i) == 1)$  then Last\_frame=playframe(i+1)*

*For j=i+1; keyframe(j) != pitch view; j=j-1;*

*First\_frame=playframe(j)*

*Generate highlight from First\_frame to Last\_frame*

The change in score is observed at keyframe 1805 in Figure 9. So that is marked as the end of the highlight segment and for checking the starting of bowl delivery, the pitch view is searched in the previous direction.

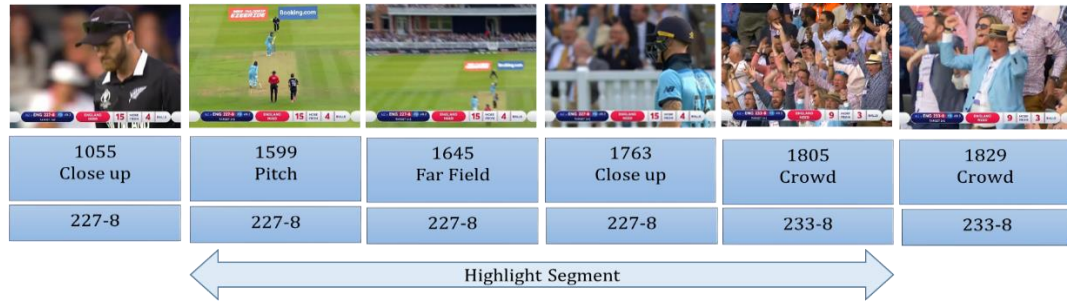


Figure 9. Highlight segment

## 6.6 Wicket fall classification

In cricket, the following types of wicket falls are mainly observed: Caught, Bowled, Leg Before Wicket, Stumped and Run Out. We have proposed a method for the classification of wicket fall using text detection and recognition. It is observed that the type of wicket fall is displayed on the score bar after each wicket fall event. If we can extract that information automatically, then the classification of wicket fall type can be given without any training phase. Generally, researchers are adopting a specific approach to identify each kind of wicket fall. The proposed method can detect each type of wicket using a single process. Table 1 is showing the text displayed on the scoreboard for every kind of wicket fall.

Table 1. Scoreboard text indicating types of wicket fall

Wicket Fall Type	Text on Scoreboard
Caught	<b>c</b> PlayerName <b>b</b> PlayerName
Bowled	<b>b</b> PlayerName
Leg Before Wicket	<b>lbw</b> <b>b</b> PlayerName
Stumped	<b>st</b> PlayerName <b>b</b> PlayerName
Run out	<b>run out</b> PlayerName
Hit Wicket	<b>hit wicket</b> PlayerName

Steps for the wicket fall classification are shown below in Algorithm 8.

### **Algorithm 8. Wicket fall classification**

*If  $(WC_{(i+1)} - WC_{(i)} == 1)$  then First\_frame = playframe(i+1)*

*For j = playframe(i+1) ; j = j+50*

*Read frame(j), Crop ROI, Resize ROI, Black and White conversion, Apply OCR*

*If OCR text contains any of pre-specified wicket types, assign that label to **wickettype**.*

*Assign wicket type label to generate highlight*

As shown in Figure 10. From the keyframe where the difference between scores is observed, we start searching the frame for wicket fall information in the forward direction. In the proposed work region showing the detailed information of wicket fall is located and segmented. For text recognition from this region, the size of the region must be enlarged. Next, morphological operation opening is applied to the wicket information image. After applying the opening operation, the resulting image is converted into a black and white image. This binary image is fed to OCR for character recognition. Finally, string matching is applied to check whether OCR's output contains the information on wicket fall. If the OCR detected text matches any of the wicket fall types, the previously generated highlight segment is labeled with that particular type.



Figure 10. Search segment for wicket fall detail

## 7. Results and Discussion

For cricket video summarization, there is no standard database available. The diverse dataset consists of 11 different cricket videos with various broadcasters and is used to evaluate the performance of the proposed event detection system. The details of the datasets are presented in Table I.

**Table I. Description of dataset**

No.	ID	Name of the Match	No. of Frames	Length (minutes)	Resolution	Frame Rate
1	Test19_WvE	West Indies vs England Test 2019	2836	1.34	1280×720	25 fps
2	NatWest11_IVE	India vs England NatWest 2011	2350	1.41	1280×720	25 fps
3	ODI14_IvN	India vs New Zealand One Day International 2014	4422	2.27	1280×720	25 fps
4	T2019_SvN	Sri Lanka vs New Zealand T20 2019	5875	3.55	1920×1080	25 fps
5	BBL11_HvS	Big Bash League 2011 Hobart Hurricanes and Melbourne Stars	5551	3.55	1920×1080	25 fps
6	ICCWC19_NvE	New Zealand vs England ICC Cricket World Cup 2019	5983	3.59	1280×720	25 fps
7	ODI2022_IVB	India vs Bangladesh One Day International 2022	7500	5.00	1280×720	25 fps
8	AC22_IvS	India vs Shrilanka Asia Cup 2022	10218	5.4	1280×720	30 fps
9	Nidahas18_IvB	India vs Bangladesh Nidahas Trophy 2018	10241	6.49	1280×720	25 fps
10	RSWS_IvB	India vs Bangladesh RoadSafety World Series	15225	10.15	1280×720	25 fps
11	AC20_IvP	India vs Pakistan Asia Cup 2020	30969	17.12	1280×720	30 fps

For measuring the performance of proposed techniques, the following evaluation metrics are used.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Where TP is true positive, FP is false positive and FN is false negative.

### 7.1 Performance of shot boundary detection

For shot boundary detection, the proposed method was tested for eight videos. As shown in Table II our approach attains an average of 99% precision and 96% recall. The performance drop of recall for some of the videos is due to sudden brightness variations between consecutive frames. This is mainly observed in night matches where illumination change is perceived by night light. Because of this, the difference between the frames increases and is detected as a boundary candidate. Due to diverse patterns and varied lengths of gradual transition, some of the intermediate frames of gradual transitions are misclassified as shot transitions

Table II. Results of shot segmentation

<i>Dataset</i>	<b>Actual Shot Boundaries</b>	<i>TP</i>	<i>FP</i>	<i>FN</i>	<i>Precision</i>	<i>Recall</i>
TEST19_WvE	20	16	0	4	1.00	0.80
ODI14_IvN	37	36	0	1	1.00	0.97
T2019_SVN	57	54	0	3	1.00	0.95
BBL11_HVS	61	61	0	0	1.00	1.00
ICCWC19_EvN	62	61	0	1	1.00	0.98
ODI22_IvB	80	79	0	1	1.00	0.99
Nidahas18_IvB	84	84	0	0	1.00	1.00
AC22_IvS	108	103	5	0	0.95	1.00
Average					0.99	0.96

A performance comparison of the proposed shot boundary detection technique with existing methods is shown in Figure 11. It is perceived from the comparison that the proposed method attains greater performance in precision and recall.

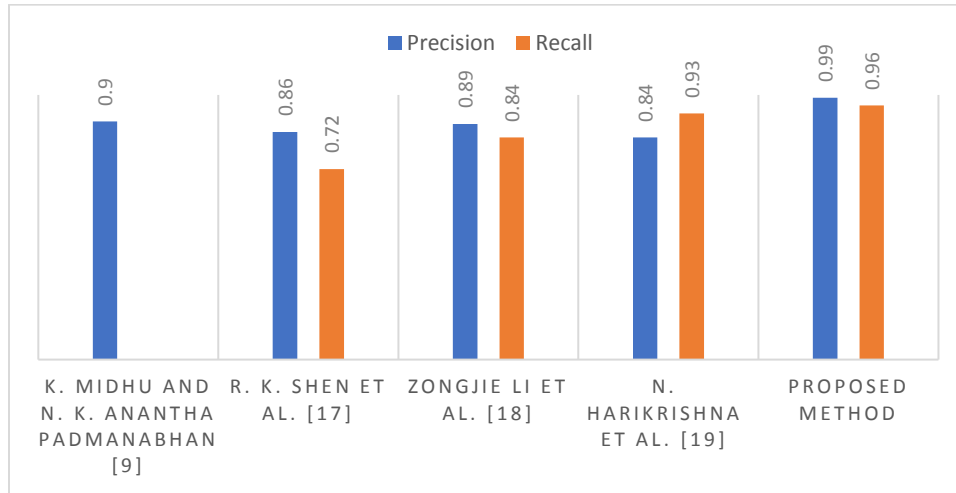


Figure 11. Performance comparison of shot boundary detection

## 7.1 Performance of replay detection

The performance of the replay detection technique is shown in Table III. The results indicate that our approach achieves 96% precision and 97% recall. The misclassification of replay segments is due to variations in scoreboard style for some frames. For example, suppose scoreboard style and dimensions vary to show specific types of information, in that case, the algorithm may get a low correlation value and lead to the misclassification of the play frame into the replay frame.

Table III. Results of replay detection

Dataset	TP	FP	FN	TN	Precision	Recall
TEST19_WvE	15	1	0	4	0.93	1
ODI22_IvB	20	20	0	0	1	1
AC22_IvS	24	23	0	1	1	0.96
ODI14_IvN	31	3	1	2	0.91	0.96
BBL11_HVS	38	2	1	20	0.95	0.97
ICCWC19_NvE	49	3	1	8	0.94	0.98
T2019_SVN	49	0	2	6	1	0.96
Nidahas18_IvB	56	5	4	19	0.91	0.93
Average					0.96	0.97

Figure 12. illustrates a comparison of the proposed system with state of art techniques. The proposed method achieves better performance in precision. For the recall metric our system achieves higher performance compared to two existing methods. In addition, our method can apply to different types of sports videos. This method does not need to recognize the logo template for each video. The correlation coefficient is proved effective to classify play and replay frames of video.

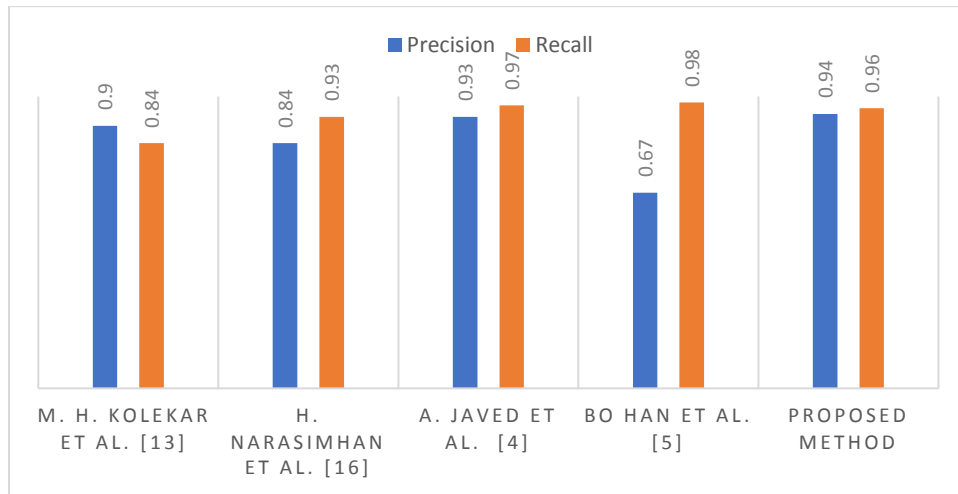


Figure 12. Performance comparison of replay detection

## 7.2 Performance of event detection and classification

Table IV shows the results of event detection and wicket fall classification method. From the outcome, it is observed that key events are identified effectively by the proposed method.

Table IV. Results of Event Detection

Type of Event	No. of Events	Detected Events				
		TP	FP	FN	Precision	Recall
Boundary	12	12	0	0	1.00	1.00
Six	12	11	0	1	1.00	0.92
Wicket	16	16	0	0	1.00	1.00
Wicket Classification	14	13	0	1	1.00	0.93

The proposed system's performance is compared with present event detection techniques and shown in Table V.

Table V. Comparison of proposed method with existing event detection techniques

Method	Boundary		Six		Wicket	
	Precision(%)	Recall(%)	Precision(%)	Recall(%)	Precision(%)	Recall(%)
Ali Javed et al. [20]	94.79	89.65	90.41	88.00	88.09	90.24
Pushkar Shukla et al. [8]	89.23		86.79		92.68	
N. Harikrishna et al. [19]	92.00	90.00	25.00	25.00	92.00	85.00
A. Bhall et al. [1]	86.74		89.12		92.45	
Proposed Method	100	100	100	92.00	100	100

Wicket fall classification is yet to get much attention from researchers. So we have compared the proposed system's performance with available wicket fall classification techniques and shown in Figure 14. The approach presented by Qamber Abbas and Youmeng Li [21] is working with image sequence and using HOG and LBP features and SVM classifier to predict wicket fall type. It can identify four different categories Bowled, Caught Behind, Catch Out and LBW. Whereas our approach is modest text detection and recognition based approach. It needs to localize the region of the score bar where wicket fall information is displayed. Before applying OCR preprocessing to score bar image is essentially required. Our approach can accurately classify Caught, Bowled, Leg Before Wicket, Stumped, Run out and Hit Wicket types of wicket.

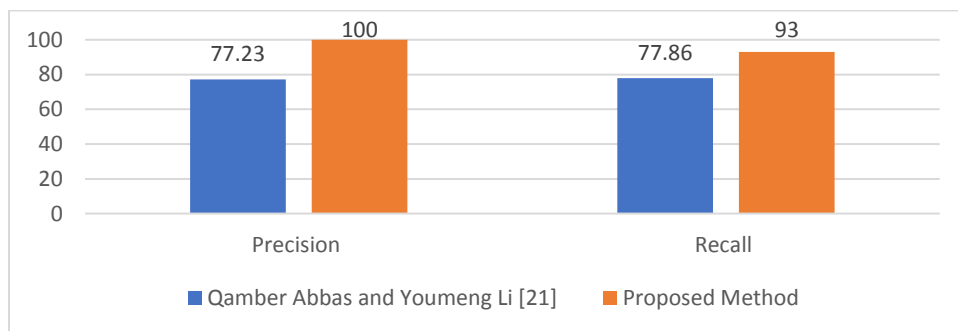


Figure 14. Performance comparison of wicket fall classification

Once interesting events are detected using run and wicket difference value, highlight is built using the method shown in Algorithm 5. According to this method, the highlight segment is started from the pitch view which is the mark of delivery of the new ball. If the pitch view corresponding to the current event is classified, then our approach will search for the pitch view in a backward direction and may include the previous ball delivery segment of the video in the current interesting event video summary. Which probably reduces the attention of the viewer. The other possibility is if some other view is missed and classified as a pitch view then the summary will start from that wrongly classified view. So the accuracy of shot boundary detection, replay detection and view classification affects the outcome of video summary preparation.

## **8. Achievements with respect to objectives**

- Successfully developed efficient shot boundary detection technique which can detect cut and gradual transitions with an average precision of 0.99 and recall of 0.96.
- Developed an effective replay detection method that detects replay segments with an average precision of 0.96 and recall of 0.97. This method is not sport genre specific, it can be applied to other sports as well.
- The event detection and classification technique can perform better for the detection of interesting events from cricket videos. We have effectively developed a wicket type classification method that classifies 6 different types of wickets in cricket using a single approach.

## **9. Conclusions**

An efficient framework for cricket video event detection and categorization is presented in this paper. It includes visual features and text features in an effective way. Shot boundary detection based on K-means clustering avoids setting a static threshold. We have applied processing on keyframes rather than each video frame, reducing processing time. The automatic replay detection approach is proven efficient as it escapes the prerequisite of learning logo templates. It is not dependent on sport structure and can be applied to other sports genres. The shot view classification is based on traditional feature extraction methods. Shots are classified into field view, pitch view, close-up view and audience view. The event detection and classification method can detect four, six and wicket fall events. The wicket classification technique using text detection and recognition is a novel contribution that achieves 100% precision and 93% recall. The results show the

effectiveness of our approach.

## 10. Publications

1. Title: A survey on event detection based video summarization for cricket

Journal: Multimedia Tools and Applications, Springer

Status: Accepted

Indexing: SCI, SCOPUS

2. Title: Shot segmentation and replay detection for cricket video summarization

Conference: International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), 2023

Status: Accepted

Scheduled: 28<sup>th</sup> - 30<sup>th</sup> April, 2023

Published in: e-Proceedings in IEEE Xplore

## 11. References

- [1] A. Bhalla, A. Ahuja, P. Pant and A. Mittal, "A multimodal approach for automatic cricket video summarization," 6th International Conference on Signal Processing and Integrated Networks: 146-150, 2019
- [2] M. Ravinder and T. Venugopal, "Content-Based Cricket Video Shot Classification Using Bag-of-Visual-Features," Artificial Intelligence and Evolutionary Computations in Engineering Systems. Advances in Intelligent Systems and Computing: 599–606, 2016
- [3] D. Karmaker, A. Z. M. E. Chowdhury, M. S. U. Miah, M. A. Imran, and M. H. Rahman, "Cricket shot classification using motion vector", 2nd International Conference on Computing Technology and Information Management: 125–129, 2015
- [4] A. Javed, A. Irtaza, Y. Khaliq, H. Malik, and M. T. Mahmood, "Replay and key-events detection for sports video summarization using confined elliptical local ternary patterns and extreme learning machine," Applied. Intelligence 49: 2899–2917, 2019
- [5] B. Han, Y. Yan, Z. Chen, C. Liu, and W. Wu, "A general framework for automatic on-line replay detection in sports video," ACM international conference on Multimedia: 501–504, 2009
- [6] S. C. Premaratne, "Structural Approach for Event Resolution in Cricket Videos," Proceedings of the International Conference on Video and Image Processing, pp. 161–166, 2017
- [7] M. H. Kolekar and S. Sengupta (2006) Event-Importance Based Customized and Automatic Cricket Highlight Generation. IEEE International Conference on Multimedia and Expo: 1617-1620
- [8] P. Shukla et al. , "Automatic cricket highlight generation using event-driven and excitement-based features," IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops: 1881–1889, 2018
- [9] K. Midhu and N. K. Anantha Padmanabhan, "Highlight generation of cricket match using deep learning," Lecture Notes Comput. Vis. Biomech. 28: 925–936, 2018
- [10] R. A. Minhas, A. Javed, A. Irtaza, M. T. Mahmood, and Y. B. Joo, "Shot classification of field sports videos using AlexNet Convolutional Neural Network," Applied Sciences 9, 2019
- [11] H. Tang, V. Kwatra, M. E. Sargin, and U. Gargi, "Detecting highlights in sports videos: Cricket as a test case," Proceedings - IEEE International Conference on Multimedia and Expo: 1-6, 2011

- [12] K. R. Raval, M. M. Goyani, "A Survey on Event Detection based Video Summarization for Cricket", *Multimedia Tools and Applications*, 81, 29253–29281, 2022
- [13] M. H. Kolekar and S. Sengupta, "Semantic concept mining in cricket videos for automated highlight generation". *Multimedia Tools and Applications* vol. 47, 545–579, 2010.
- [14] M. M. Goyani, S. K. Dutta, and P. Raj, "Key frame detection based semantic event detection and classification using heirarchical approach for cricket sport video indexing," *Commun. Comput. Inf. Sci.*, vol. 131, pp. 388–397, 2011.
- [15] M. Hung, C. Hsieh, and S. Member, "Event Detection of Broadcast Baseball Videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 12, pp. 1713–1726, 2008.
- [16] H. Narasimhan, S. Satheesh, and D. Sriram, "Automatic summarization of cricket video events using genetic algorithm," *Proc. 12th Annu. Genet. Evol. Comput. Conf. GECCO '10 - Companion Publ.*, pp. 2051–2054, 2010
- [17] R. K. Shen, Y. N. Lin, T. T. Y. Juang, V. R. L. Shen, and S. Y. Lim, "Automatic Detection of Video Shot Boundary in Social Media Using a Hybrid Approach of HLFPN and Keypoint Matching", *IEEE Trans. Comput. Soc. Syst.*, vol. 5, no. 1, pp. 210–219, 2018.
- [18] Z. Li, X. Liu, and S. Zhang, "Shot Boundary Detection based on Multilevel Difference of Colour Histograms," *1st Int. Conf. Multimed. Image Process. ICMIP*, pp. 15–22, 2016.
- [19] N. Harikrishna, S. Satheesh, S. D. Sriram and K. S. Easwarakumar, "Temporal classification of events in cricket videos", *National Conference on Communications*, pp. 14-18 2011
- [20] A. Javed, K. B. Bajwa, H. Malik, A. Irtaza, and M. T. Malunood, "A hybrid approach for summarization of cricket videos,". *IEEE International Conference on Consumer Electronics-Asia*: 1-4, 2016
- [21] Q. Abbas and Y. Li, "Cricket Video Events Recognition using HOG, LBP and Multi-class SVM," *Journal of Physics: Conference Series*, vol. 1732, October 2021