Event Detection and Summarization of Cricket Videos

Muhammad Haseeb Nasir
Computer Science Department, University of Management and Tech. Sialkot, Pakistan
Email: haseeb.nasir@skt.umt.edu.pk

Ali Javed
Department of Software Engineering, University of Engineering and Technology, Taxila, Pakistan
Email: ali.javed@uettaxila.edu.pk

Aun Irtaza
Department of Computer Science, University of Engg. and Technology, Taxila, Pakistan
Email: syed.irtaza@uettaxila.edu.pk

Hafiz Malik
Electrical and Computer Engineering Department, University of Michigan, Dearborn, USA
e-mail: hafizm@umich.edu

Muhammad Tariq Mahmood
School of Computer Science and Information Engineering, Korea University of Tech. and Education, Republic of Korea
Email: mtariq@koreatech.ac.kr

***Abstract*—This paper presents an efficient framework for key events detection and summarization in cricket videos. The proposed research work presents a non-learning technique based on textual features to detect three key events in cricket videos that are, boundary (4), six (6) and wicket. Image averaging is used to detect the score captions from the input video that is then analyzed to detect changes in score and wickets counters. To extract the contents of score captions, input video frame is discretized by using the mean and standard deviation. Morphological operators are applied to get rid of the noise and outliers. The extracted score caption region is passed to the optical character recognition algorithm to analyze any significant change in the score and wicket counters. The frame is marked as a key frame in case any significant change (*boundary, six, or wicket event*) is detected. A collection of video frames are selected against each key-frame to generate the summarized video. The proposed method is tested on a diverse dataset of cricket videos belonging to different tournaments. Experimental results illustrate the efficiency of the proposed method for key events detection and summarization of cricket videos.**

***Index Terms*—erosion, key events, morphology, optical character recognition, score captions, video summarization**

I. Introduction

Broadcasters generate an enormous amount of the multimedia content online at a very rapid pace. Multimedia content analysis is a taxing activity for both humans and machines. Sports videos are one of the major contributors of the available multimedia content due to its massive viewership all over the world. Sports video collections are increasing exponentially nowadays due to the increasing frequency and coverage of sports matches all over the world. This introduces a massive challenge for effective management of sports video content in the research community. The processing, storage and transmission of the available collection of sports video content is a time consuming and tedious activity. Video summarization techniques are generally used to solve these issues by generating brief synopsis of long duration videos. Effective navigation, browsing, and retrieval of relevant video chunks can be obtained by designing an effective video summarization technique that contains all key events of the original video [1].

Sports broadcasters capture and transmit massive amount of live and recorded videos of various sports which are played around the globe. Cricket is the third most popular game in the world after soccer and basketball [2]. Cricket has massive viewership in Asia, Australia, and UK. Video summarization techniques for cricket [3], [4] have not been explored much as compared to other sports such as soccer [5], tennis [6], baseball [7] and basketball [8]. The long duration and high complexity of the cricket matches are considered the main bottlenecks behind less research work reported for summarization of cricket videos.

Manuscript received January 20, 2018; revised June 22, 2018.

doi: 10.18178/joig.6.1.27-32

Existing state-of-the-arts for video summarization can be classified into learning-based [4], [9]-[12] and non-learning-based techniques [7], [13]. Learning-based techniques are computationally expensive as compared to non-learning-based techniques but offer better accuracy for classification. In [4], audio stream of the input cricket video is processed to detect the excited audio frames. A decision tree framework is trained on the corresponding excited video frames to summarize the cricket videos. Kamesh [9] proposed a learning-based method for summarization of cricket videos. Shot boundary detection is performed using histogram comparison. A feature set consisting of grass color, pitch color, audience texture, motion activity, and edges is used for shot classification. The states and transitions are represented via Hidden Markov Model (HMM). HMM is used to extract the frames containing key events, which are then used to generate the highlights for cricket videos. Kolekar et al. [5], [10]-[12] have used audio-visual features to propose a hierarchical framework for sports video summarization. Audio feature set consisting of short time energy and zero crossing rates is used to detect excited segments from the input video. Color and motion features are employed via Hidden Markov Model to identify the logo frame transitions that are then used to detect the replays. Furthermore video shots are also classified into field view and non-field view by analyzing grass color pixels. In [12], a two level abstraction approach is presented. The first level abstraction, defined as events, is presented at the micro level. The second level abstraction defined as concepts, is presented at the macro level. A hierarchical feature-based classifier is trained to detect these events in the cricket videos.

Learning-based techniques offer better accuracy but at the expense of increased computational complexity. This provides a great motivation for researchers to develop non-learning-based video summarization techniques [7], [13] that are faster in execution as compared to learning-based approaches. In [7], a non-learning-based framework is proposed to detect the replays in sports videos. Firstly, gradual transitions are detected by applying a dual threshold method to extract the candidate replay segment. Secondly, this candidate segment is processed to detect the replay/live frame based on the absence/presence of score caption in each frame. In [13], motion features are used to design a video summarization framework. Optical flow is computed to estimate the motion metrics. Subjective evaluation is used to evaluate the performance of this technique [13]. Though non-learning video summarization methods are less accurate at times but careful input data pruning results to achieve better classification accuracy.

In this research work we have proposed a computationally efficient non-learning-based technique to summarize the cricket videos. The proposed approach is tested on a diverse dataset of cricket videos of varying lengths and illumination conditions i.e. daylight, artificial lights, etc. The average accuracy of 94% at a very low computational cost illustrates that the proposed method is very efficient and precise for summarization of cricket videos.

II. Proposed Method

This paper presents an automated method to detect the key events for summarization of cricket videos. In the preprocessing phase, color frames are transformed into gray-scale images and down-sampled the sequence of frames by a factor of 20. Score captions are detected in the next stage from the input video frames. To reduce the noise and other outliers, morphological and arithmetic operators are used on the extracted region of score caption. The extracted score caption region is fed to Optical Character Recognition (OCR) method to recognize the score caption contents that are further processed to detect the key events. The frames containing significant changes in the scoreboard are marked as key frames. Each key frame represents a significant event in the input video. Few frames at the start and end of each key frame are selected to generate video skims against each key event. A summarized video consisting of each video skim arranged in a chronological order is finally generated for the input cricket video. Process flow of the proposed technique is shown in Fig. 1.

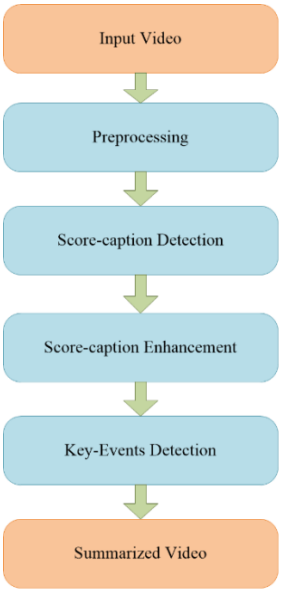

Figure 1. Process flow of proposed method

A. Preprocessing

The input color video frames are converted into gray-scale images. The sequence of grayscale frames are down-sampled by a factor of 20 in order to reduce the computational complexity of the proposed method. Shown in Eqs. (1) and (2) are the transformation of color frames into grayscale, and down-sampling by processing the 20th frame after each current frame.

$$I_i^{gray} = 0.298 * I_i^r + 0.587 * I_i^g + 0.114 * I_i^b \quad (1)$$

$$I_i^{gray} = I_{i+20}^{gray} \quad (2)$$

where I_i^r, I_i^g, and I_i^b represent the red, green, and blue components of the color image for i^{th} frame, I_i^{gray} is the

grayscale image and I_{i+20}^{gray} represents the grayscale image after 20 frames from the current frame.

B. Score Caption Detection

After watching extensive amount of cricket videos it has been observed that the score captions appear at the fixed position in the frame. Moreover, the score caption region exist in every frame while the rest of the content changes. Therefore, running image averaging is applied to extract the region of score caption from the input video. For each frame, a window of four frames are selected for averaging as shown in Eq. (3).

$$I_i^{avg} = I_{i-1}^{avg} + \frac{I_{i+1}}{n} - \frac{I_{i-1}}{n} \tag{3}$$

where I_i^{avg} and I_{i-1}^{avg} represents the average image at i^{th} and i^{th}-1 frame, I_{i+1} and I_{i-1} represent the start and end frames in the window, n is the length of sliding window that is set to four. The extracted region of score caption is shown in Fig. 2.

Figure 2. Score caption detection

C. Score Caption Enhancement

Morphological and arithmetic operators are applied to enhance the extracted region of score caption before feeding it to OCR algorithm. Morphological opening is applied on the average image of score caption with a structuring element S_1 of rectangular shape and size α=25x27as shown in Eq. (4).

$$I_i^{open} = I_i^{avg} \circ S_1 \tag{4}$$

where I_i^{open} represents the image obtained after morphological opening operation, and $\circ$ is the opening operator.

The morphed image obtained in Eq. (4) is subtracted from the average image of score caption as follows:

$$I_i^{sub} = I_i^{avg} - I_i^{open} \tag{5}$$

where I_i^{sub} represents the subtracted image for i^{th} frame. Mean and standard deviation are used to discretize the subtracted image. Two passes of morphological thinning are applied on this binary image to remove the outliers that are connected components less than 9 pixels wide. This operation is performed to remove the outliers from this binary image so that the isolated components are discarded. The resultant binary image contains the characters/numbers without any isolated pixels. Shown in Eq. (6) is the application of thinning operation on the binary image with a structuring element S_2 of square shape and size β=3x3.

$$I_i^{thin} = I_i^{bin} \otimes S_2 \tag{6}$$

where I_i^{thin} is the image obtained after applying morphological thinning for i^{th} frame, I_i^{bin} is the binary image obtained after transformation of the subtracted grayscale image, and $\otimes$ is the thinning operator.

To bridge character gaps and achieve smoothing, dilation operator with a structuring element S_3 of *square* shape and size γ=2x2 is applied on the thin image as shown in Eq. (7). This dilated image transforms the score caption contents in the form that eventually enhances the accuracy of the OCR results.

$$I_i^{dil} = I_i^{thin} \oplus S_3 \tag{7}$$

where I_i^{dil} represents the dilated image. Shown in Fig. 3 is the enhanced image of score caption.

Figure 3. Score caption enhancement

D. Key Event Detection

The processed image of score caption region is passed to the OCR method [14] to analyze the characters for recognition. The recognized characters obtained from the OCR are divided into two parts i.e. score counter (S^{count}) and wicket counter (W^{count}). The values of score and wickets are analyzed further to detect key events in the input cricket videos. The score and wicket values are compared with the values recognized in the previous frame. If there exist a significant change in the score board for score and wicket counters, then a key event is detected in the input video. For boundary event, key frames are classified as follows:

$$X = \begin{Bmatrix} Boundary\ Event, & if\ S_i^{count} - S_{i-1}^{count} = T_1 \\ No\ Event, & Otherwise \end{Bmatrix} \tag{8}$$

More specifically, if the difference in S^{count} values between the current and previous frame is equivalent to a specified threshold T_1, then a boundary event is detected. Similarly for six event, key frames are classified as follows:

$$Y = \begin{Bmatrix} Six\ Event, & if\ S_i^{count} - S_{i-1}^{count} = T_2 \\ No\ Event, & Otherwise \end{Bmatrix} \tag{9}$$

More specifically, if the difference in S^{count} between the current and previous frame is equivalent to a specified threshold T_2, then a six event is detected.

For wicket event, key frames are classified as follows:

$$Z = \begin{Bmatrix} Wicket\ Event, & if\ W_i^{count} - W_{i-1}^{count} = T_3 \\ No\ Event, & Otherwise \end{Bmatrix} \tag{10}$$

More specifically, if the difference in W^{count} values between the current and previous frame exceeds a specified threshold T_3, then a wicket event is detected.

Video skims for each key event is generated by including the frames of ten seconds prior to and five seconds after each key frame. Each of these video skims are appended in the chronological order to generate the highlights for cricket videos while preserving the temporal information.

III. Experimental Results

The proposed system is tested on a diverse collection of cricket videos. Objective metrics (i.e. *precision*, *recall*, *accuracy*, *error*, etc.) are used for performance evaluation of the proposed method.

A. Dataset

A customized dataset of thirty cricket videos is created for performance evaluation of our method. Videos are collected from various sports broadcasters namely *Fox Sports*, *Ten Sports*, *Sky Sports*, *ESPN*, *Star Sports*, and *Super Sport*. The videos recorded in the dataset has a frame resolution of 640 x 480 pixels and frame rate of 25 fps. The dataset videos are recorded in different illumination conditions i.e. daylight, artificial light, etc. Moreover video samples are recorded in different lengths i.e thirty minutes, one hour, two hours, etc. The dataset contains sample videos from 2015 test series between *Pakistan* and *Sri Lanka*, 2014 test match series between *Australia* and *Pakistan*, 2014 One Day International series between *New Zealand* and *South Africa*, and 2014 *Twenty20* cricket world cup tournament. The experiments are performed on the system parameters set to T_1=4, T_2=6, T_3>0, α=25x27 i.e. 25 rows and 27 columns, β=3x3 and γ=2x2. The reason behind using the fixed thresholds for each key-event is that these thresholds are selected according to the rules of the game. In cricket, a boundary event results in an increment of four and a six event results in an increment of six in the score counter. Moreover, a wicket event results in an increment of one or more in the wicket counter. Therefore, the thresholds for boundary, six and wicket events are selected as T_1=4, T_2=6, T_3>0 respectively. The shape and size of the structuring elements are chosen to preserve the effectiveness of the morphological operations.

B. Performance Evaluation

Thirty cricket videos are used to measure the performance of the proposed method. We have performed four different experiments to evaluate the detection performance of key events.

Precision, recall, accuracy, and error rates are computed for boundary (4), six (6), and wicket events in the first experiment. The detection results for boundary (4), six (6), and wicket events are presented in Fig. 4. Experimental results are very promising and signify the effectiveness of the proposed system for key events detection and highlights generation. The proposed method achieves an average precision, recall, accuracy, and error rates of **94.69%, 87.68%, 90.78%,** and **9.22%** for boundary (4) event, **91.53%, 82.67%**,**92.97%,** and **7.03%** for six (6) event, and **94.74%, 87.80%**,**98.04%,** and **1.96%** for wicket event.

In our second experiment, we have compared the performance of the proposed method with existing video summarization methods for cricket. Performance comparison of the proposed and existing methods [3], [7] is presented in Table I.

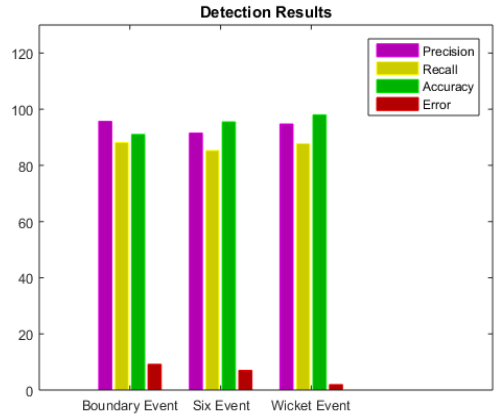

Figure 4. Detection results for boundary, six and wicket events

ROC curve analysis is performed in our third experiment to illustrate the performance of the proposed method for key events detection in *Cricket* videos. ROC curves for *Boundary*, *Six*, and *Wicket* events are plotted in Fig. 5. It can be observed from the results that the proposed method is very effective in terms of detecting the key events (*Boundary*, *Six*, and *Wicket*).

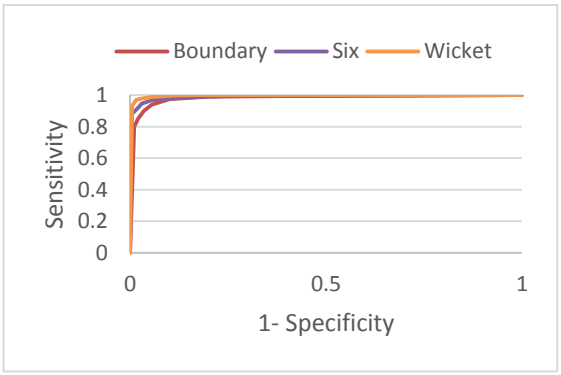

Figure 5. ROC curve analysis for boundary, six and wicket events

In our fourth experiment, a confusion matrix analysis is designed to illustrate the classification performance of the proposed technique. The classification results of boundary, six and wicket events are presented in a confusion matrix as shown in Table II. High values of True Positives and True Negatives indicate better classification of key events for video summarization.

In our last experiment we computed the computational cost of the proposed method and also compared it with our earlier work [7]. The proposed method takes 0.4 seconds to process each frame as compared to [7], which takes 1.2 seconds. The proposed method is more efficient as compared to [7] due to the following facts: (i) the grayscale frames are down sampled by a factor of 20, and

(ii) video frames are processed only for key events detection as compared to [7] where audio stream is also processed with the input cricket video.

IV. Conclusion and Future Work

The proposed work presents an efficient method based on textual features to detect key events in the input cricket videos for summarization. Score captions are analyzed to identify the significant changes in the score and wickets counter that are then used to detect the boundary, six and wicket events. Each frame containing the key event is marked as the key frame. Video skims are generated against each key frame to generate the summary. The proposed method is tested on a diverse dataset of cricket videos of various sports broadcasters. Experimental results achieves an average accuracy of 94.8% that indicates the effectiveness of the proposed method in terms of key events detection for video summarization. Currently, we are examining the performance of the proposed technique on a larger dataset. In addition, our method is independent of the game structure, therefore, it can easily be extended to adopt for multiple sports.

TABLE I. Performance Comparison of the Proposed and Existing Methods

Cricket Video Summarization methods	Dataset Details					Precision Rate	Recall Rate	Processing Time per Frame
	Frame Rate	Length	Resolution	Format	No. of Videos			
Kolekar et. al. [3]	25 fps	982 min	-	-	04	86.68%	86.93%	Not provided
Ali et al. [7]	25 fps	362 min	640 x 480	AVI	20	91.87%	89.85%	1.2 sec
Proposed Method	25 fps	600 min	640 x 480	AVI	30	93.99%	87.01%	0.4 sec

TABLE II. Confusion Matrix Analysis

Actual Class	Predicted Class						
	Classes	Positive	Negative	Positive	Negative	Positive	Negative
		Boundary		Six		Wicket	
	Positive	179	24	64	11	36	05
	Negative	8	147	05	278	02	315

References

[1] Y. J. Lee and K. Grauman, "Predicting important objects for egocentric video summarization," *International Journal of Computer Vision*, vol. 114, no. 1, pp. 38-55, 2015.

[2] Most popular sports in the world. [Online]. Available; http://mostpopularsports.net/in-the-world

[3] M. H. Kolekar and S. Sengupta, "Event-importance based customized and automatic cricket highlight generation," in *Proc. IEEE International Conference on Multimedia and Expo*, July 2006, pp. 1617-1620.

[4] A. Javed, K. B. Bajwa, H. Malik, A. Irtaza, and M. T. Mahmood, "A hybrid approach for summarization of cricket videos," in *Proc. IEEE International Conference on Consumer Electronics-Asia*, October 2016, pp. 1-4.

[5] M. H. Kolekar and S. Sengupta, "Bayesian network-based customized highlight generation for broadcast soccer videos," *IEEE Transactions on Broadcasting*, vol. 61, no. 2, pp. 195-209, 2015.

[6] G. Zhu, Q. Huang, C. Xu, L. Xing, W. Gao, and H. Yao, "Human behavior analysis for highlight ranking in broadcast racket sports video," *IEEE Transactions on Multimedia*, vol. 9, no. 6, pp. 1167-1182, 2007.

[7] A. Javed, K. B. Bajwa, H. Malik, and A. Irtaza, "An efficient framework for automatic highlights generation from sports videos," *IEEE Signal Processing Letters*, vol. 23, no. 7, pp. 954-958, 2016.

[8] V. Bettadapura, C. Pantofaru, and I. Essa, "Leveraging contextual cues for generating basketball highlights," in *Proc. ACM on Multimedia Conference*, October 2016, pp. 908-917.

[9] K. Namuduri, "Automatic extraction of highlights from a cricket video using MPEG-7 descriptors," in *Proc. First International Communication Systems and Networks and Workshops*, January 2009, pp. 1-3.

[10] M. H. Kolekar, K. Palaniappan, and S. Sengupta, "Semantic event detection and classification in cricket video sequence," in *Proc. Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, December 2008, pp. 382-389.

[11] M. H. Kolekar and S. Sengupta, "Semantic concept mining in cricket videos for automated highlight generation," *Multimedia Tools and Applications*, vol. 47, no. 3, pp. 545-579, 2010.

[12] M. H. Kolekar and S. Sengupta, "A hierarchical framework for generic sports video classification," in *Proc. Asian Conference on Computer Vision*, January 2006, pp. 633-642.

[13] E. Mendi, H. B. Clemente, and C. Bayrak, "Sports video summarization based on motion analysis," *Computers & Electrical Engineering*, vol. 39, no. 3, pp. 790-796, 2013.

[14] R. Smith, "An overview of the Tesseract OCR engine," in *Proc. Ninth International Conference on Document Analysis and Recognition*, September 2007, vol. 2, pp. 629-633.

Muhammad Haseeb Nasir (M'16) received the B.Sc. degree with honors in Software Engineering from UET Taxila, Pakistan in 2013. He received his MS degree in Software Engineering from UET Taxila, Pakistan in 2016. Currently, he is serving as a Lecturer in the Department of Software Engineering at UMT Sialkot, Pakistan.

Ali Javed (M'16) received the B.Sc. degree with honors in Software Engineering from UET Taxila, Pakistan in 2007. He got 3rd position in Software Batch-2003F. He received his MS and Ph.D. degrees in Computer Engineering from UET Taxila, Pakistan in 2010 and 2017. He received Chancellor's Gold Medal in MS Computer Engineering degree. Currently, he is serving as an Assistant Professor in the Department of Software Engineering at UET Taxila, Pakistan. Dr. Javed has served as a visiting PhD research scholar at ISSF Lab in University of Michigan, USA in 2015. He was awarded HEC scholarship to pursue his Ph.D. research work at University of Michigan, USA. His areas of interest are Digital Image Processing, Computer vision, Video Summarization, Machine Learning, Multimedia Signal Processing, Software Quality Assurance and Testing.

Aun Irtaza has completed his PhD from FAST-National University of Computers & Emerging Sciences in 2016. During his PhD he remained working as a research scientist in the Signal and image processing lab in Gwangju Institute of Science and Technology (GIST) South Korea. Currently, he is serving as Associate Professor and head of department in Computer science department at UET Taxila, Pakistan. His research interests include computer vision, pattern analysis, and big data analytics.

Hafiz M. A. Malik (S'02-M'06-SM'10) received the B.E. degree in electronics and communications engineering (with distinction) from the University of Engineering and Technology Lahore, Pakistan, in 1999 and the Ph.D. degree in electrical and computer engineering from the University of Illinois, Chicago, in 2006. After the Ph.D. degree, he joined the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, where he worked as a Postdoctoral Research Fellow. Currently, he is serving as an Associate Professor in ECE Department at University of Michigan Dearborn, MI. His research interests are in the general areas of digital content protection and digital signal processing, and the focus of current research includes information security, steganography, steganalysis, statistical signal processing, audio analysis/synthesis, and digital forensic analysis.

Muhammad Tariq Mahmood (M'2012) received the MS degree in Computer Science from Blekinge Institute of Technology, Sweden in 2006. He received the Ph.D. degree in Informatics and Mechatronics from Gwangju University of Science and Technology, Republic of Korea in 2011. Currently, he is serving as the Assistant Professor in Korea University of Technology and Education, South Korea. His research interests include Image Processing, Machine Learning and Pattern Recognition.