Proceedings of the Pakistan Academy of Sciences:
A. Physical and Computational Sciences: 57 (3): 19-30 (2020)

ISSN: 2518-4245 (print), 2518-4253 (online)

Pakistan Academy of Sciences

Research Article

Automatic Soccer Video Key Event Detection and Summarization Based on Hybrid Approach

Muhammad Khurram Iqbal[1], Aun Irtaza[1], and Nudrat Nida[1,2*]

[1]Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan
[2]Department of Computer Science, Air University, Islamabad, Aerospace and Aviation Campus, Kamra, Pakistan

Abstract: Sports broadcasters generate an enormous amount of video content in cyberspace due to massive viewership all over the world. Analysis and consumption of this huge repository urge the broadcasters to apply video summarization to extract the exciting segments from the entire video to capture the user's interest and reap the storage and transmission benefits. Therefore, in this study, an automatic method for key-events detection and summarization is presented for soccer videos. The proposed framework reaps benefit both from learning and non-learning methods of summarization. SVM classifier is used to classify boundary and non-boundary frames based on extracted features Histogram difference and the average motion vector. Replay detection shot view classification, and play break sequence formation are performed through efficient algorithms. Nonsubjective features such as play break duration ratio, play duration ratio, near goal duration ratio, etc. are used to perform statistical analysis which helps in devising rules for summarization. To address the shortcomings associated with key event detection and summarization algorithms and to get the best out of the merger of learning and non-learning based approaches for summarization, this research problem needed a deep inside look. The effectiveness and robustness of the proposed method are tested over an extensively huge dataset and the results are highly productive and comparative. Besides soccer, the proposed method finds its application in freestyle football and hockey but with minor modifications.

Keywords: Summarization, Rules-based approach, Machine learning, Statistical analysis

1. INTRODUCTION

Video summarization aims to produce a short and concise summary of long duration videos. Video summarization has emerged as a highly influential research area where researchers can put forth their efforts in the field of surveillance [1] (where surveillance systems develop a concise summary of events for evidence-based investigation), health [2, 3] (to produce clinical summarizes of patients' data that is easy to manage, study and evaluate) and sports video summarization.

Sports video summarization represents long-duration sports videos in concise form by capturing only significant events that occurred in a game. Summarized sports videos effectively capture the user interest, thus has the potential to generate revenue for the broadcasters. Traditionally sports video summarization requires manual involvement thus makes video summarization a laborious job. Besides, domain-specific methods for sports video summarization lot have been done concerning the development of generic frameworks that summarizes multiple games at the same time through the use of a universal feature set.

Various sports video summarization techniques and frameworks have already been proposed [4-7, 8-12]. Conventionally these methods perform feature extraction over the visual content, undergo detection of key events by applying parametric approaches or machine learning models over-extracted features to generate a summary. Thus, the sports video summarization approaches can be broadly classified into learning and non-learning

Received: April 2020; Accepted: September 2020
*Corresponding Author: Nudrat Nida <nudrat.nida@au.edu.pk>

based methods. Learning-based methods employ machine/deep learning classifiers, e.g., SVM [13], CNN's [14] to summarize sports videos and provide good performance at the expense of high computational cost. Zawbaa et al. [5], SVM, and artificial neural networks were used to detect logos, and game captions to summarize soccer videos. In [9], Javed et al. applied rule-based induction and decision trees over the game audios to summarize cricket videos. In [14] Jiang et al. used Deep Neural Network for key event detection in soccer super infusing CNN (Convolution Neural Network) and RNN (Recurrent Neural Network). CNN was used for extraction of spatial features whereas RNN was employed for mapping extracted features to soccer events. In [15] Tavassolipour et al. used Bayesian Network and Hidden Markov Model for key event detection and summarization in multiple sports (soccer, freestyle football, basketball). Besides its robust algorithms for shot boundary detection, shot view classification, mid-level feature extraction was also proposed. In [16] Namuduri et al. used Highlight Extraction Process along with the Hidden Markov Model over-extracted features for highlight generation in cricket. In [17] Godi et al. used 3D-CNN (Convolution Neural network) over-extracted visual features for key event detection in ice hockey. In [18] Midhu et al. used the Apriori algorithm to detect replays, pitch, and boundary view, players and audience, to generate concise cricket videos.

Non-learning-based methods for sports video summarization detect logos, frame transitions, player and whistle detection using parametric approaches i.e. by applying thresholds and rule-based induction. Despite being automatic, the stagnant nature of algorithms results in incorrect detections. In [19] Nguyen et.al., used bounding box size and grass pixels' ratio for shot view classification in soccer. Besides its histogram difference and contrast features were employed for replay detection. Work proposed by Tjondronegoro and Chen [20] used a rule-based model infused with game statistics for multisport highlight generation. In [21] Wang et. al., used the event annotation method for soccer by applying Hough Transform over the video frames. The use of Attack Defense Transition Analysis (ADTA) brought sufficient improvement in shot boundary detection.

Various sports video summarization methods have [9, 22, 23, 24, 25-27], combined the audio/visual features for video summarization. In [9] Bajjal et al. utilized audio features such as referee's whistle sound, commentator's excitement, crowd noise for key event detection in rugby, besides it proposed work used Mel Frequency Cepstral Coefficients (MFCC) along with Gaussian Mixture Model (GMM) for extraction and learning of features. In [22] Rui et al. utilized baseball's audio features e.g energy-related features, phoneme level features, prosodic features, and information complexity features for highlight extraction. Besides it, human speech extremities were detected through the use of E23, Etr, and MFCC algorithms. In [23] Min Xu et al. proposed a framework that utilized audio keywords for key event detection in soccer. The proposed work employed an SVM classifier for the creation of audio features from low-level features. Besides it, heuristic rules were devised for summarization through audio keywords. In [27] Raventos et al. used various audiovisual feature extraction algorithms e.g. whistle detector, logo detector, person detector, and replay detector for highlight generation. Existing sports video summarization methods have certain limitations like high computational cost, stagnant nature as well as computation inefficiency in detecting external factors like replays, logos, speed, and so on.

To get the benefit of learning and non-learning based methods, a hybrid approach for summarization of soccer videos is proposed in this paper. The proposed framework employed SVM over motion vectors and gray-level histogram difference across frames for shot boundary detection and classify the frames as boundary or non-boundary frames. Moreover, we also proposed an efficient view classification algorithm that considers important frame segments by computing game-field intensities and player bounding box sizes. To reduce misclassification of long and medium views, threshold frequencies are applied over player bounding box sizes. The proposed framework employed logo detection through statistical analysis over-extracted features from key events i-e goal, foul, shoot, and non-highlights. Likewise, a rule-based model driven on statistical data is developed to generate a concise summary of the game.

2. MATERIALS AND METHODS

Block Diagram of the proposed framework for summarization of soccer videos is presented in Figure 1. Each stage of the proposed framework is of great importance

2.1. Shot Boundary Detection

Shot boundary detection is generally considered to be an initial step, especially in video summarization applications. For this purpose, we have utilized two features, hue histogram difference as a spatial feature and resultant motion vector as a compress domain feature. Later on, these two extracted features are passed onto linear Support Vector Machine (SVM) for classification among boundary and non-boundary frames. For shot boundary detection, spatial feature-e hue histogram difference is calculated among consequent frames. It is probably a widely used approach that brings cost-effectiveness in terms of speed and accuracy.

Despite the availability of various histogram difference calculation approaches we have used block-wise histogram difference as it improves the performance of shot boundary detection. If two consecutive frames lie in the same shot their histogram difference would be small as compared to frames lying in different shots. For histogram comparison, various measures are available like Bhattacharyya, correlation, etc. we have used the Bhattacharyya comparison measure for this purpose.

$$d\,(H_1, H_2)= \sqrt{1 - \frac{1}{N\sqrt{H_1\, H_2}} \sum_i \sqrt{H_1(i) H_2(i)}} \qquad (1)$$

Here $Hj = 1/N\sum_k Hj(k)$ and N represents several histogram bins. The second feature that is a resultant motion vector is a compressed domain feature that not only improves the accuracy of the shot boundary detection process but highly cost-effective in terms of its extraction and calculation. If two consecutive frames lie in the same shot their motion vectors will show a high degree of regularity in terms of distribution and direction. Due to which the resultant motion vector's length increases Beside it if two consecutive frames lie in different shots their motion vectors show dispersion in terms of distribution and direction. Due to which the resultant motion vector's length got squeezed. Figures 2 and 3 shows motion vectors for boundary and non-boundary frames respectively. After extraction of the above-mentioned features i-e (Histogram difference and Resultant motion vector), the next step is to use an effective classifier for the classification of boundary and non-boundary frames. Here we have utilized a linear SVM (Support Vector Machine) classifier for this purpose because of its high impotence and accuracy in binary classification problems. Results associated with this stage have been discussed in the result portion of the paper.

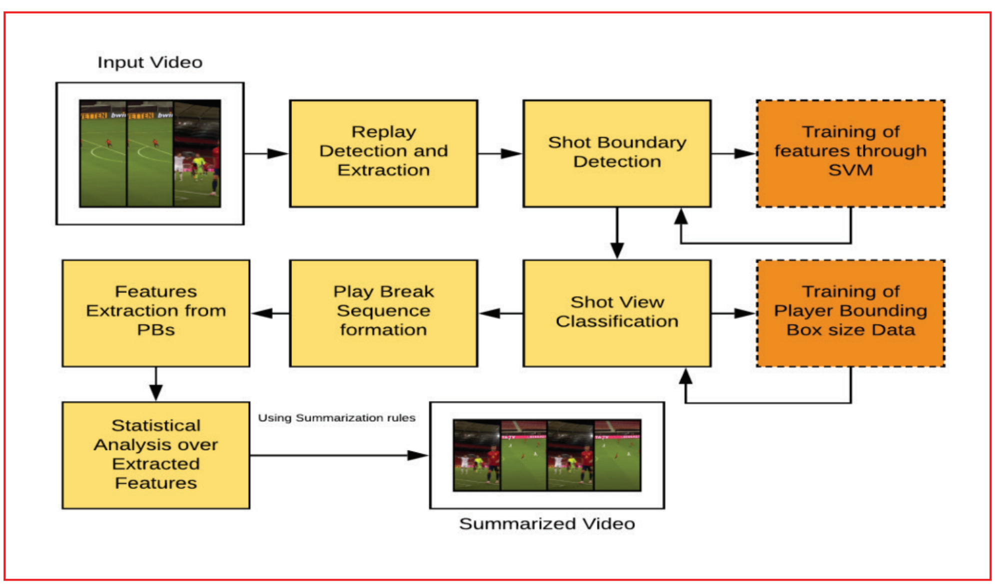

Fig.1. Block diagram of proposed framework

Fig. 2. Motion vectors for boundary frame

Fig. 3. Motion vectors for non-boundary frame

2.2. Shot View Classification

Information associated with shot views is highly productive in key event detection in sports video summarization frameworks. Shot views can lead to useful clues on the semantic content of the video. Generally, in soccer, there are three recommended views these are long, medium, and close-up views shown in figure 4. We have proposed an algorithm for shot view classification that works on green color dominance inside the frame. If there is no dominance of green color frame under consideration is classified as a close-up view. Besides it using a similar approach for the classification of long and medium views brings a lot of misclassified results. This is because both of these views are indulged with green as a dominant color. To reduce these misclassifications associated with long and medium views besides green color dominance, another feature is utilized i-e player's bounding box size. This feature is easy to compute using built-in packages of Matlab. Generally, in the long view the size of the player is small so does its bounding box size as compare to the medium view. If a frame has green as a dominant color two thresholds Tlow and Thigh associated with the player's bounding box size are employed to get a more accurate classification of long.

2.3. Replay Detection

Replay segment holds great importance in various summarization frameworks from semantic analysis prospect. They also play a key role in key event detection as the replay is the repetition of the most noticeable event using various motion techniques and camera views. Generally, a replay segment is sandwiched between two logo frames. In this paper, we have used template matching for the detection of logo frames that helps in the detection of replay segments. Histogram difference of logo frame is calculated concerning consecutive frames of video whose summary is to generate. Minimum calculation lets you know about the start and end of a replay. The proposed methodology gives a high degree of accuracy, speed, and cost-effectiveness to the overall framework. After accurate detection of replay segments, the next step is their extraction from the long-duration video.

Some researchers believe that as the replay is the repetition of the same event so its involvement might bring intensive complexity and lack of accuracy in key event detection process and might require the involvement of extra work to produce better results like work proposed in [20] utilizes replay based correction during play-break sequence formation to produce more complete sequences. We stick on with narrative and have extracted detected replay segments to bring computational cost-effectiveness as well as accuracy in the proposed framework. Figure 5 shows detected logo frames through the proposed methodology.

2.4. Play Break Sequence Formation

To be having a steady process of key event detection some researchers affirmed the highlights portion as part of the play segment. A segment happening especially in sports e.g. soccer, cricket, hockey, etc.

Fig. 4. Close-up, long and medium views

Fig. 5. Replay segment's logo frames

is a mixture of play and break as parts of the same segment. A play segment is always followed by a break and vice versa. For a better and complete understanding of any event play break sequence needed to be watched completely to conclude. Generally, it is believing that key event always falls in play segment of the play break sequence with broadcaster using a long view for its coverage. Besides it, the broken segment comprises of close-up view, indulged with replay segments as well as crowd shots.

Mostly sports broadcasters use on-screen editing effects like captions and a variety of camera views to discriminate precedent events from non-precedential ones. In soccer occurrence possibility of the key event is always high in the play segment of the play break sequence as compared to the broken segment. The break is a break of like crowd celebration, foul, etc. From the above discussion, it is pretty clear that a play break sequence is an effective container for semantic content as it contains most precedential details. The majority of automatic sports summarization frameworks extract all required features from play break sequences. In this paper, we have used the modified form of an algorithm proposed in [20] to play break sequence formation. Our proposed algorithm does not take into consideration replay based correction as proposed in [20]. The reason behind the noninvolvement of replay-based correction is that most researchers believe that replay falls in the broken segment of the play break sequence. Besides it, the noninvolvement of replay scenes gives computational cost-effectiveness to play break sequence formation stage. The proposed algorithm works over the output from shot view classification and shot boundary detection stages.

2.5. Key Event Detection and Summarization:

For sports video summarization frameworks having a universal feature set requires extensive domain-oriented knowledge. Our proposed framework uses a feature set that requires little domain knowledge which gives a high degree of flexibility during implementation. Besides it proposed framework can detect the following key events in soccer with great proficiency these are free-kick goals, penalty goals, corner goals, assist goals, shoots, and non-highlights as part of highlights.

During the training phase, statistical data associated with the following list of features from each play break segment is extracted for highlight classification:

- **Play Break sequence duration (PBseqR)**

Is the whole duration of the segment comprised of the play and break segment. The duration is in seconds. Generally, this duration is pretty long.

- **Break Duration Ratio (BDR)**

This ratio is calculated by dividing the play break segment duration with break duration.

- **Play Duration Ratio (PDR)**

This ratio is calculated by dividing play break segment duration with play duration this ratio is quite high in most games especially in world-cup games because teams are playing with a cause that is to win the world cup.

- **Near Goal Duration Ratio (NGDR)**

This ratio is calculated by Penalty Box-Duration (frames covering Penalty Box) and then dividing play break duration with this duration. This ratio is also pretty high in most soccer games due to the passing nature of teams and slowly moving towards the opponent's goal post.

- **Close-up Duration Ratio (CuDR)**

This ratio is calculated by the calculating duration of close-up views in the segment with play break duration. This ratio helps find non-highlighted portions as breaks generally lie in close-up views.

It is very important to mention over here is that this feature set is already proposed in [20]. The difference lies in the use of algorithms for their extraction and calculation. To calculate PBseqR, BDR, PDR, CuDR ratios our proposed view classification algorithm produces better results as mentioned earlier and secondly, extraction and calculation of these above-mentioned ratios are automatic for near goal ratio we have utilized an algorithm that detects three parallel lines inside the frame. Apart from this selected feature set rest of the features like the sound of the whistle, character recognition over scorecard are not included because it's quite difficult to extract these features, and secondly, there is always a greater possibility of misclassifications due to the complex nature of these features and variability in the environment for their extraction.

Statistical data associated with each highlighted segment that our proposed framework can detect is enlisted in the table. We have calculated these above-mentioned durations for each highlighted event that our framework can detect in terms of minimum, average and maximum duration as depicted in table 1. Various broadcasters' content is used for training. Some of the example from the training set is also used during performance evaluation of the proposed framework. Based on calculated statistics rules are designed to detect the most noticeable events in the game of soccer.

The generic nature of the feature set brings subjectivity and robustness to the proposed framework. Besides it with minor modifications same feature set can be implemented in various sports like freestyle football, hockey, etc Heuristic rules are constructed based on statistical data calculated from each feature over each key event. The proposed framework generates a summary of a soccer match using 3 tier algorithm stated below.

3. RESULTS AND DISCUSSION

3.1. Data Set

One of the most essential steps in the accurate and reliable evaluation of algorithms and frameworks

Highlights Classification Algorithm for Soccer (3 Tier Approach)

Step 1: Start
Step 2: Input PB sequence
Step3: If(PBseqR>=PBseqPGmin AND BseqR<=PBseqPGmax)

&& (PDR>=PDRGmin AND PDR<=PDRmax)&&

(CuDR>=CuDRGmin AND CuDR<=CuDRGmax)&&

(NGDR>=NGDRGmin && NGDR<=NGDRmax)==True)

Then Most likely to be Goal; Add serial No. of PB sequence to Summary Array else Step 4

Step4: If (PBseqR>=PBseqPFmin AND BseqR<=PBseqPFmax)

&&(PDR>=PDRFmin AND PDR<=PDRFmax)&&

(CuDR>=CuDRFmin AND CuDR<=CuDRFmax)&&

(NGDR>=NGDRFminAND NGDR<=NGDRFmax)==True)

Then Most likely to be Foul; Add serial No. of PB sequence to Summary Array else Step 5

Step5: If(PBseqR>=PBseqPSmin AND PBseqR<=PBseqPSmax)

&& (PDR>=PDRSmin ANDPDR<=PDRSmax)&&

(CuDR>=CuDRSmin AND CuDR<=CuDRSmax)&&

(NGDR>=NGDRSmin AND NGDR<=NGDRSmax)==True

Then Most likely to be Shoot; Add serial No. of PB sequence to Summary Array else
Most likely to be Non-highlights;

Add serial No. of PB sequence to Summary Array:

Pick new PB Sequence, repeat step 1 to 5Here PBseqR=Play Break sequence duration,

PDR= Play Duration Ratio, NGDR=Near Goal Duration Ratio,

CuDR= Close-up Duration Ratio. Similarly G=Goal, S=Shoot, and F= Foul

Table 1. Statistical Data Associated with Key Elements

Events	Play-break Duration (Avg; Maxi-Min)	Play Duration (Avg; Maxi-Min)	Close-up Duration (Avg; Maxi-Min)	D-Duration (Avg; Maxi-Min)
Goal	(52.31, 120, 27)	(27.8, 86, 4)	(14, 41, 4)	(11.6, 18, 4)
Foul	(41, 57, 23)	(31.7, 46, 18)	(4.9, 27, 0)	(4.9, 27, 0)
Shoot	(22.5, 7, 10)	(16.5, 31, 6)	(12.8, 31, 5)	(2.8, 8, 0)

especially involving the statistical analysis and machine learning is the availability of an appropriate dataset. Undoubtedly in soccer researchers always fall short of such appropriate datasets. Therefore, we have manually developed various datasets used at various levels for the evaluation of algorithms as well as the overall working of the proposed framework. Datasets involve soccer videos from various broadcasters and various tournaments like FIFA world cup, UEFA champion's league, etc. making 6-hour duration altogether, covering day as well as night matches with a frame rate.

The dataset, used for shot boundary detection includes 200 examples extracted manually with 100 examples for boundary frames and 100 for no-boundary frames, out of which 70% examples are sued for training and 30% for testing purposes.

Feature extraction for each example is

automatic. Extracted features are stored in an excel file later on passed onto the proposed shot boundary detector for result evaluation. The dataset used for view classification includes 200 examples each for global, medium, and close-up views respectively. Examples are extracted manually later on passed onto the proposed view classifier for result evaluation. The dataset used for evaluation of overall working efficiency of the proposed frame includes round about 630 play break sequences belonging to various categories of key events that our proposed framework can detect i-e goal, foul, shoot, and non-highlights.

These extracted play break sequences are extracted as well as categorized manually.

3.2. Results

In the proposed framework, we detected 4 key events these are the goal, foul, shoot, and non-highlights. Table 3 shows the confusion matrix for the proposed framework with the rightmost column shows the number of miss-classified examples. Out of 621 examples, 544 examples are classified correctly making accuracy number 87.6%.

The last two columns of the table show the precision and recall efficiency of our proposed framework. Fig shows a comparison of recall and precision accuracy of our proposed framework with the rule-based method and Bayesian and Copula method. In some cases, our proposed framework has better accuracy as compared to others but in some cases, it is on the lesser side. For example, precision accuracy for key event foul of our proposed framework is pretty high as compared to other methods it might be because of low recall.

Table 2. Dataset of Matches

Matches	Description
Arsenal vs Man City	UEFA Champions League
Brazil vs Argentina	Friendly Match
France vs Spain	World Cup Qualifier 2018
Germany vs England	World Cup Qualifier 2018
Man United vs Liverpool	UEFA Champions League Final 2018

Table 4 shows the performance comparison of our proposed method with commonly used classifiers such as SVM (RBF) and BN (Bayesian Network) on our dataset. Below mentioned table below listed down precision and recall rates of Bayesian Network, SVM (RBF), and our proposed method for key event detection. It is evident from the mentioned results that our proposed method has much-improved precision and recall rate for key events like Goal, Foul, and Non-highlights as compare to other methods. The reason is that our proposed method captured dependencies among various extracted features much more effectively through the use of efficient statistical analysis Algorithms. A lot of things helped in the overall improvement of results such as selection and extraction of more precise and productive features for shot boundary detection stage i-e Histogram difference across frames and motion vector calculation helped a lot in improving efficiency and preciseness of above-mentioned shot boundary detection stage which overall improved event detection methodology.

Another component of our proposed research that showed a considerable amount of improvement and been more productive in terms of results is how medium and global views were classified. By bringing additional checks over the size of the player bounding box beside the dominant color feature, helped a lot in reducing falsie detection of views. Apart from its decision over non-inclusion of replay segments during PB formation not only reduced time complexity of event detection methodology but also helped in producing more complete PB sequences. These were certain improvements that brought highly productive, effective, and more precise results as compared to other approaches of the same domain.

3.3. Shot Boundary Detection

The accuracy of the shot boundary detector of the proposed framework is shown in Figure 8 and Figure 9 in terms of the confusion matrix and scatter plot. Dataset for shot boundary detection is prepared manually, discussed in detail in the dataset portion of the paper. The accuracy of the proposed shot boundary detector is compared with, SVM with RBF kernel and Method of [28] recall and precision rate of our proposed algorithm have shown a considerable amount of improvement, this

Table 3. Confusion Matrix Depicting Recall and Precision Percentage of Proposed Framework

Key Events	Precision and Recall Matrix				
	Goal	Foul	Shoot	Non Highlights	Miss Classified
Goal	28	0	0	0	1
Foul	0	80	7	35	13
Shoot	4	1	36	2	2
NonHighlights	0	8	2	400	4
Recall %	97.05	59.25	80	96.72	*
Precision %	87.5	89.88	80	91.53	*

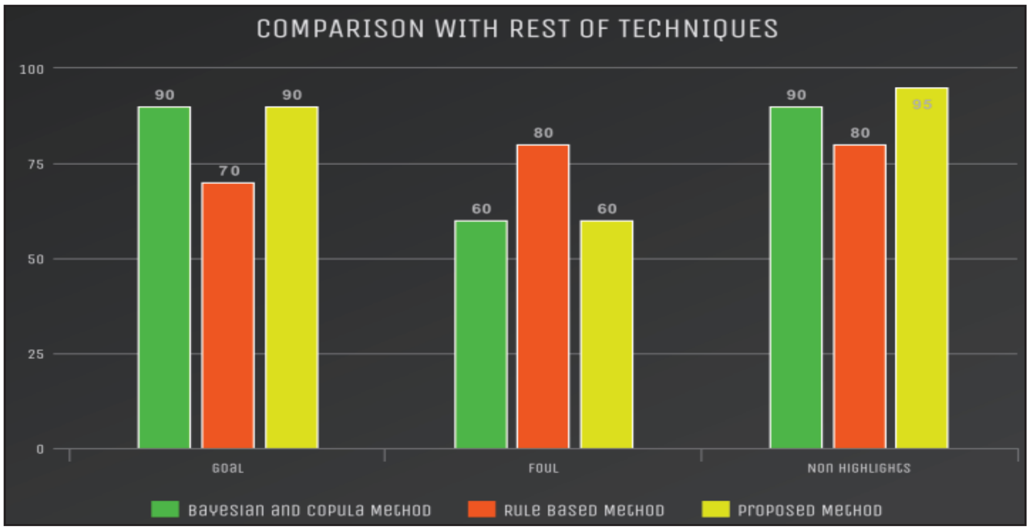

Fig. 6. Precision comparison

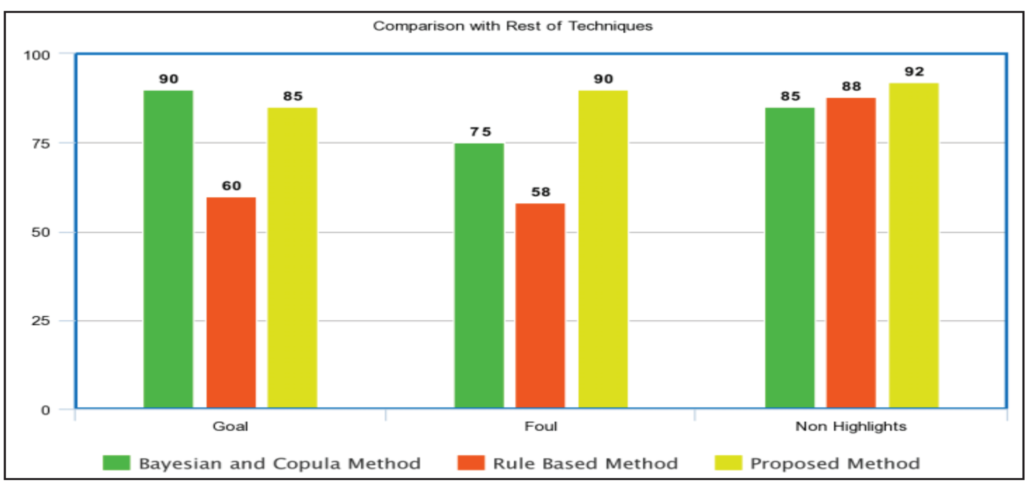

Fig. 7. Recall comparison

Table 4. Performance Comparison matrix

Events	BN		SVM (RBF)		Our Method	
	R	P	R	P	R	P
Goal	91.72	86.67	85.71	88.24	97.05	87.5
Foul	57.29	73.52	48.55	56.96	59.25	89.88
Non Highlights	90.39	85.11	89.19	79.22	96.72	91.53

because of the precise nature of extracted features as well as the methods that are used for extraction of features for shot boundary detection is highly accurate, Table 5 shows a comparison among recall and precision accuracy of the proposed methodology with rest of approaches of the same domain.

3.4. Shot View Classification

Table 6 shows the results of our proposed shot view classification method. The overall accuracy of our proposed method shows a considerable amount of improvement from the rest of the methods.

3.5. Replay Based Correction vs. Shots without Replays

The point to fonder in this portion is that though replay based correction has a greater effect on the performance and efficiency of summarization framework as proposed in [20] employing replay based correction enhances the overall cost of summarization and secondly enhances the duration of summarized content as the replay is a portion of that summarized content. In contrast to replay based correction, a sequence without replay generally brings cost-effectiveness in summarization and secondly reduces the duration of the summarized

Table 5. Shot Boundary Detector Recall and Precision

Comparison	SVM with RBF Kernel	Method [28]	Proposed Method
Recall	99.50	97.30	99.06
Precision	98.97	91.70	99.28

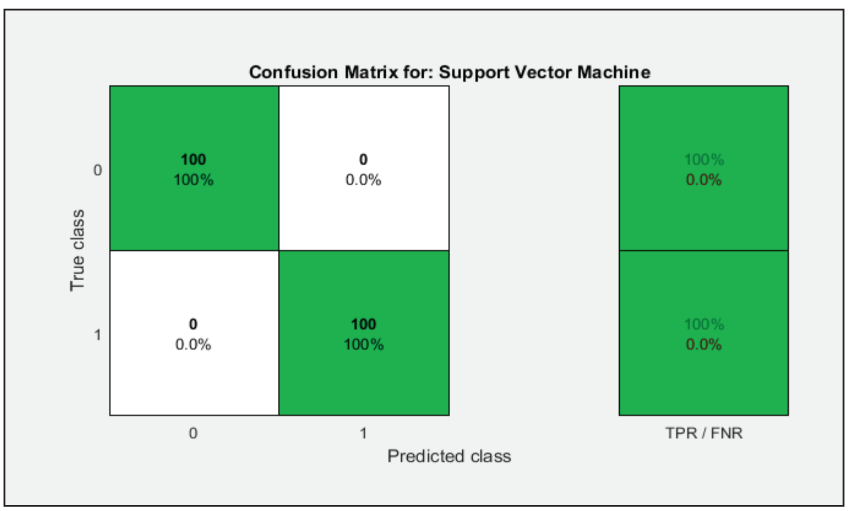

Fig. 8. Confusion matrix for boundary and non-boundary frames

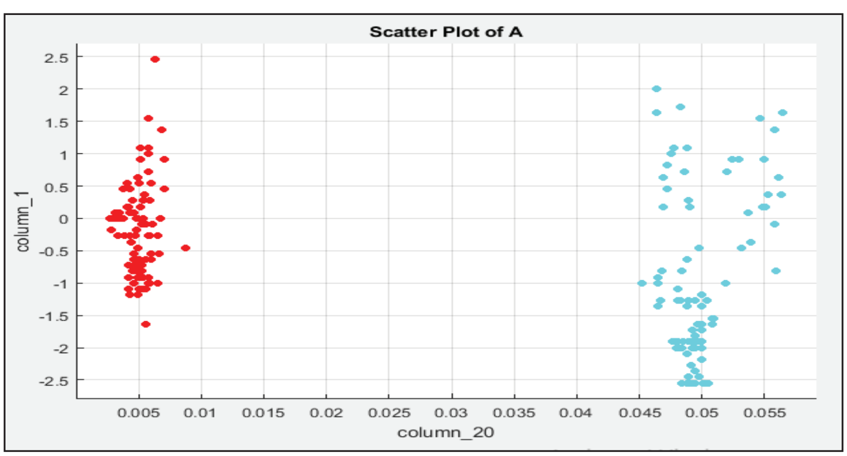

Fig. 9. Scatter plot for boundary and non-boundary frames

Table 6. Results associated with shot view classification

Stats	Global-View	Medium View	Close-Up View
Frame No.	4683-4883 France Vs Spain	14974-16974 France Vs Spain	4252-4452 France Vs Spain
Accuracy	97%	91%	98%
Examples	200	200	200
Correct	199	182	197
Incorrect	1	18	3

Table 7. Duration comparison

Matches	Duration	Summary duration with replay based correction	Summary duration without Replay segments
Arsenal vs Man City	45min	9:15 min	7:10 min
Brazil vs Argentina	90min	15:09 min	12:45 min
Germany vs England	45min	8:23 min	7:01 min

content by about 10% to 20%. Table 7 shows a reduction in duration of summary produced by our proposed framework as compared to a summary of soccer video produced through framework employing replay based correction.

4. CONCLUSION

This paper has presented a robust and computationally cost-effective framework for the summarization of soccer videos. The proposed domain-specific framework uses state of the art machine learning classifier SVM for classification of boundary and non-boundary frames over-extracted features average motion vector and histogram difference. The use of the efficient algorithm for shot view classification that takes into consideration green color dominance and threshold frequencies over player bounding box size makes this framework more effective. The proposed framework effectively utilized a view classification algorithm for the formation of play break sequences. Efficient algorithms like penalty box detection algorithm etc. are used for extraction of features that help in performing statistical analysis for key event detection keeping in view the universal nature of feature set in the future, we are looking forward to applying this narrative to other sports like freestyle football, hockey, etc. We are also looking forward to enlarging the proposed feature set so that other noticeable events that are less important can be detected with greater efficiency.

5. REFERENCES

1. P.K. Lai., M. Décombas., K. Moutet, and R. Laganièr. Video summarization of surveillance cameras. In *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. 286-294 (2016).
2. K. Muhammad., J. Ahmad., M. Sajjad., S.W. Baik. Visual saliency models for summarization of diagnostic hysteroscopy videos in healthcare systems. *SpringerPlus,* 5(1) 1495(2016).
3. I. Mehmood., M. Sajjad, and S.W. Baik. Video summarization based tele-endoscopy: a service to efficiently manage visual data generated during wireless capsule endoscopy procedure. *Journal of medical systems,* 38(9) 109 (2008).
4. M.V.M. Cirne, and H. Pedrini. VISCOM: A robust video summarization approach using color co-occurrence matrices. *Multimedia Tools and Applications,* 77(1) 857-875 (2018).
5. H.M. Zawbaa., N. El-Bendary., A.E. Hassanien, and T.H. Kim. Machine learning-based soccer video summarization system. In *International Conference on Multimedia, Computer Graphics, and Broadcasting (pp. 19-28). Springer,* Berlin, Heidelberg (2011).
6. H. Boukadida., S.A. Berrani, and P. Gros. A novel

modeling for video summarization using constraint satisfaction programming. In *International Symposium on Visual Computing* (pp. 208-219). Springer, Cham (2014).

7. T. Yao,. T. Mei, and Y. Rui. Highlight detection with pairwise deep ranking for first-person video summarization. In Proceedings of the *IEEE conference on computer vision and pattern recognition.* 982-990 (2016).
8. A. Baijal., J. Cho., W. Lee, W, and B.S. Ko. Sports highlights generation based on acoustic events detection: A rugby case study. In 2015 *IEEE International Conference on Consumer Electronics* (ICCE) 20-23 (2015).
9. A. Javed. K.B. Bajwa., H. Malik., A. Irtaza, and M.T. Mahmood. A hybrid approach for summarization of cricket videos. In 2016 *IEEE International Conference on Consumer Electronics-Asia* (ICCE-Asia) 1-4 (2016).
10. C.M. Chen, and L.H. Chen. A novel method for slow motion replay detection in broadcast basketball video. *Multimedia Tools and Applications,* 74(21) 9573-9593 (2015).
11. N. Nguyen, and A. Yoshitaka. Shot type and replay detection for soccer video parsing. In 2012 *IEEE International Symposium on Multimedia.* 344-347 (2012).
12. C.M. Chen, and I.H. Chen. Novel framework for sports video analysis: A basketball case study. In 2014 *IEEE International Conference on Image Processing* (ICIP) (pp. 961-965 (2014).
13. T. Yao., T. Mei, and Y. Rui. Highlight detection with pairwise deep ranking for first-person video summarization. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 982-990 (2016).
14. H. Jiang., Y. Lu, and J. Xue. Automatic soccer video event detection based on a deep neural network combined cnn and rnn. In *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI).* 490-494 (2016).
15. M. Tavassolipour., M. Karimian, and S. Kasaei. Event detection and summarization in soccer videos using bayesian network and copula. *IEEE Transactions on circuits and systems for video technology,* 24(2), 291-304 (2013).
16. K. Namuduri. Automatic extraction of highlights from a cricket video using MPEG-7 descriptors. In *2009 First International Communication Systems and Networks and Workshops.* 1-3 (2009).
17. M. Godi., P. Rota, and F. Setti. Indirect match highlights detection with deep convolutional neural networks. In *International Conference on Image Analysis and Processing* (pp. 87-96). Springer, Cham (2017).
18. K. Midhu, and N.A. Padmanabhan. Highlight generation of cricket match using deep learning. In *Computational Vision and Bio Inspired Computing* (pp. 925-936). Springer, Cham (2018).
19. N. Nguyen, and A. Yoshitaka. Shot type and replay detection for soccer video parsing. In *2012 IEEE International Symposium on Multimedia.* 344-347 (2012).
20. D.W. Tjondronegoro, and Y.P.P. Chen. Knowledge-discounted event detection in sports video. *Ieee transactions on systems, man, and cybernetics-part a: Systems and humans,* 40(5) 1009-1024 (2010).
21. Z. Wang., J. Yu, and Y. He. Soccer video event annotation by synchronization of attack–defense clips and match reports with coarse-grained time information. *IEEE Transactions on Circuits and Systems for Video Technology,* 27(5) 1104-1117(2016).
22. Y. Rui., A. Gupta, and A. Acero. Automatically extracting highlights for TV baseball programs. In *Proceedings of the eighth ACM international conference on Multimedia.* 105-115 (2003)
23. M. Xu., N.C. Maddage., C. Xu., M. Kankanhalli, and Q. Tian. Creating audio keywords for event detection in soccer video. In 2003 *International Conference on Multimedia and Expo. ICME'03. Proceedings (Cat. No. 03TH8698).* 2: 211-281 (2003).
24. S. Tang, and M. Zhi. Summary generation method based on audio feature. In 2015 6th *IEEE International Conference on Software Engineering and Service Science (ICSESS*): 619-623 (2015).
25. Kolekar, M. H., & Sengupta, S. (2010). Semantic concept mining in cricket videos for automated highlight generation. *Multimedia Tools and Applications,* 47(3), 545-579.
26. M.H. Kolekar, S. Sengupta. A hierarchical framework for generic sports video classification. In *Asian Conference on Computer Vision* (pp. 633-642). Springer, Berlin, Heidelberg (2003).
27. A. Raventos., R. Quijada., L. Torres, and F. Tarrés. Automatic summarization of soccer highlights using audio-visual descriptors. *Springer Plus*, 4(1),1-19 (2015)..
28. A. Ekin., A.M. Tekalp, and Mehrotra, R. Automatic soccer video analysis and summarization. *IEEE Transactions on Image processing,* 12(7), 796-807. (2003).