Automatically Extracting Excitement Clips For Cricket Match Programs

Hemlata Sinha[1], Rohit Raj Singh[2],Prashant Singh[3], Sakshi Yadav[4], Sakshi Wagh[5]
[1]Associate Professor, [2] Assistant Professor, [3,4,5] B.E Student
Department of Electronics and Telecommunication Engineering[1 2 3 45]
Shri Shankaracharya Institute Of Professional Management And Technology, Raipur, India[1 2 3 45]
(Corresponding author: Dr. Hemlata Sinha)

***Abstract*— Recordings are getting one of the significant wellsprings of amusement because of their high data capacity. And in today's world where everyone is short of time it is always better to have source of entertainment that can be consumed in shorter period of time, highlights of sports being one of the best source of entertainment for almost everyone, particularly the sports which is played for long period of time. A Sport video has a well-defined content structure and a few set of domain rules when compared to the other types. Generating sports highlights may be a labor-intensive work that needs a point of specialization. Hence Our Aim is to create an automated tool which can extract the major highlights of the Match. In this paper we divide the video into excitement and Non-excitement clips. The Excitement clips are then further divided into the real time and replays clips using Machine Learning (ML) and Digital Image Processing (DIP) Techniques.**

Keywords— Spaced Vector Machine (SVM), Hue Histogram, Image Processing, Dominant Grass Pixel Ratio (DGPR), YCbCr-Y indicates Luminance Cb & Cr indicates Luminance of Blue & Red Respectively

I. INTRODUCTION

Sports feature attracts a large population all around the world, and have turned into a critical type of multimedia content that is spilled over the web and broadcasting companies. Each and every day a large number of individuals watch sports feature of different sorts, including football, cricket, tennis, baseball, hockey, and so on [7]. Cricket is one of the popular international games which attract a large number of viewer ships. Television broadcasters like Neo cricket, ESPN, Star Sports have huge database of cricket videos. Cricket is a game of with a variable set of rules as compared to other famous sports such as soccer, basketball, tennis etc. Even some of the formats of cricket likeT20-20 is played for approximately 3 hours which is greater than the timing of soccer and hockey both together. Typically, these features (of cricket) are somewhat long, comprising of parts which are intriguing or energizing and parts which are uninteresting, moderate.

These Videos are the collection of continuous frames which is normally displayed at rate of 25 fps. Video itself contains a huge amount of data and complexity and likely "a waste of the viewer's important time."[1] The duration of some matches like 'test match' is of 4-5 days and thus extracting meaningful events which are of viewer’s interest is very important [5].

With a large count of cricket matches throughout the year, it becomes difficult for a sports fan to keep up with all the news. The huge amount of data that's produced by digitizing sports videos requires a process of data filtration and reduction. Therefore, highlights serve as an important source of information to keep fans updated with the latest happenings without consuming too much of their time. However, manual highlight generation requires professional editing skills and is a time consuming process, which limits the amount of media that can be summarized on short notice [9]. This demands the need for a system that is capable of automatically generating highlights of sports events. At present very few systems are implanted for automatic cricket highlight generation as the target.

We propose an innovative technique for automatic generation of cricket highlights. Event-driven features are used for extracting the major events in a cricket match, .i.e., wickets, boundaries, sixes, and milestones, while the remaining important events are identified with the aid of excitement features. Event-driven strategies make use of OCR, playfield scenarios and replays, while the complimentary excitement based strategy makes use of audio based classifiers and replays to reinforce the relevance of an event in a cricket match. The features of this game themselves can be used either as events for clustering the frames to view. The past few years have marked an increased interest in automatic analysis of sports content [10].

II. LITERATURE REVIEW

Jimmy Soni, Amit Bhimani,Priyanka Buch [1] proposed noteworthy or interesting area will be played with moderate movement design keeping in mind the end goal to let observers admire the subtle elements. Replay is a huge signs for highlights and routinely taken as an issue angle in occasion location. Cricket, Soccer, American Football, Basketball, Baseball, and Tennis etc. usually contains few semantic events in which viewer are interested. Games highlights can be essentially made out of intriguing occasions that may confinement the viewer's consideration. Surf highlights might be a good choice for spectators to quickly recognize the sports event therefore this review paper present various methods for generating highlights from sports video. This proposed algorithm and framework can applied on cricket.

N. Harikrishna, Sanjeev Satheesh, S. Dinesh Sriram, K.S. Easwarakumar [2] proposed method segments a cricket video into shots and identifies the visual content in them. Using sequential pattern mining and support vector machine, we classify the sequence of shots into four events, namely RUN, FOUR, SIX and OUT. The cricket video is then summarized based on user-supplied parameters. The performance of the system has been tested on a number of cricket video clips and was found to have an accuracy of 87.8%.

Namudri [3] proposed a technique that relied on MPEG-7 for highlight generation of cricket matches. The complete match was segmented into video shots (clips belonging to the same camera take). Extracted key frames were classified into one of five different classes, depending on their field view. A Hidden Markov Model (HMM) was then used to identify events from sequences of these key frames.

Tang et al. [4] proposed a multilevel framework that relied on Histogram of Oriented Gradients (HOG) and Color Histograms (CH) for detecting important events in a cricket match. The events, once detected, were refined using Hidden Markov Models (HMM).

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

III. METHODOLOGY

A method is proposed to generate the highlights based on event selection and giving it an importance value based on user feedback. Also there is a prime drawback of real time cricket video that is "to produce the frames" as there are various video formats like MPEG,AVI etc. along with different frame rate, conversion rate ,frame rate, etc [3]. So there is a need of a system which is autonomous of the demonstrated parameters and requirement of a generic approach. Features are temporarily disintegrated into a development of Events in source of an unauthorized event declaration and identification system [11]. The whole system depends uniquely upon simple to direct low level visual highlights for example, shading histogram or histogram of situated slopes, which can perhaps be summed up to different games. The proposed method of framework style breaks the entire frame feature stream into little shots of whole feature, then it classifies the game about shots into different shot-sort classes. Thus, the framework recognizes vertical objective posts.

A. Hierarchical Framework:

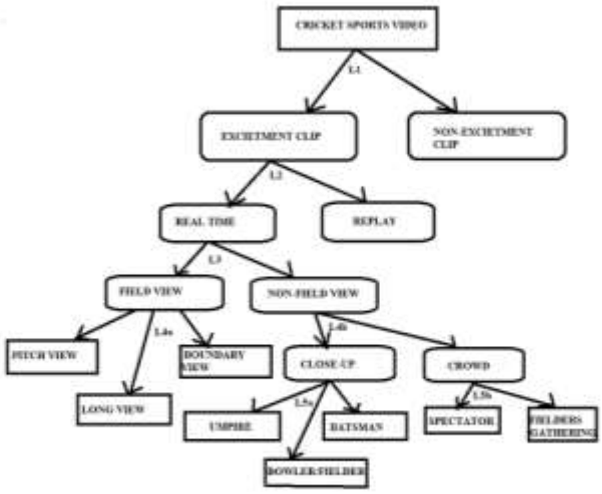

Fig. 1 Tree Diagram of Hierarchal Framework

- ➢ There are 5 levels in the hierarchy clearly shown above where each level has their own importance.
- ➢ Level 1 is the Excitement detection: Here a particular video frame is considered to be as an excitation frame if the product of its audio excitement and zero crossing rate (ZCR) exceeds a certain threshold.
- ➢ Level 2 is the Replay Detection: Here the replay segment is fit in by two logo transitions. Hence, replays can be easily recognized using Hue Histogram Difference (HDD) and removed.
- ➢ Level 3 is the Field View Detection: Dominant Grass Pixel Ratio (DGPR) is calculated for a view, which varies from 0.16 to 0.24 for the field view. Thus, a non field view can be removed.
- ➢ Level 4 is the Field View and Close Up Detection: Percentage of field pixels in regions are calculated and some thresholds are fixed, and frame can then be classified as long view, boundary view or pitch view.

In the same way, edge pixels are used to identify close views or crowd views.

- Level 5 is the Fielders gathering or crowd Detection: Crowd frames are eliminated from the video. The identification is done by computing histogram distance of hue histogram of frames [8].
- Thus, highlights for the given video are extracted.[6]

B. Level-1

Excitement level of commentators and audience can be measured and filtered by using product of Short Time Audio Energy (E) and Short Time Zero Crossing Rate(Z) and comparing that product to a particular threshold value, frames crossing that threshold value will be classified as excitement clip.

Fig. 2 Excitement Level of Commentators and Audience

1) Short Time Audio Energy (E):

In MATLAB, one of the most widely used method for extracting audio is via short time audio energy technique in which voiced and unvoiced portions are extracted in speech signal by detecting peak of energy frames if it crosses the mean that means it is voiced. Unvoiced signal has lower amplitude in comparison with audio signal, so using short time energy variation in signals of speech can be obtained.

$$E(n) = \frac{1}{V}\sum_{m=0}^{V-1} [x(m)w(n-m)]^2$$

Where

$$w(m) = \{1 \quad if\ 0 \le m \le V-1 \ 0 \quad Otherwise$$

$x(m)$ = Discrete time audio signal

V= Number of audio samples corresponding to single video frame.

2) Short Time Zero Crossing Rate(Z):

Short Time Zero Crossing Rate popularly known as STZC function, is also an energy function for detection of audio levels but this function works based on algebraic sign of frequency. It is an integer value, preferable for signals of longer duration thus making it good choice for using it highlight detection [4]. Zero valued mean is assumed for consistency.

$$Z(n) = \frac{1}{2}\sum_{m=0}^{V-1} |sgn[x(m)] - sgn[x(m-1)]|w(n-m)$$

Where,

$$sgn[x(m)] = \{1 \quad x(m) \ge 0 \ -1 \quad x(m) < 0$$

Where $w(m)$ is a rectangular window

Let,

Short Time Audio Energy = STE (E), Short Time Zero Crossing Rate=STZCR (Z) and Threshold Value=T1

So,

If

*(STE (E)*STZCR(Z))of frames>= T1*

Output: Frame

Else

Discard Frame

Level-2

At level 2, logo moves and it portrays the frames as ongoing or replay.

1) Logo Detection:

Correlation can be used for detecting logo. Popularly known as Corr2 in MATLAB, it computes correlation coefficient between two data provided or two arrays in 2-D form. Basically it computes similarity degree in images and gives output 0 for two distinct images and 1 for similar kind images. Thus checking a threshold value of coefficient for logo of frames and classifying into the replay video category.

$$r = \frac{\sum_m \sum_n (A_{mn} - \underline{A})(B_{mn} - \underline{B})}{\sqrt{\left(\left(\sum_m \sum_n (A_{mn} - \underline{A})^2\right)\left(\sum_m \sum_n (B_{mn} - \underline{B})^2\right)\right)}}$$

Where $\underline{A}$ = mean 2(A), and $\underline{B}$ = mean 2(B)

Fig. 3 Logo

Let, Threshold value for logo = T2

So,

If corr2>=T2

Output: Logo transition detected

Check for Scorecard bar

2) Scorecard Bar:

For detection of scorecard bar hue-histogram and position of scorecard bar can be used. Reference scorecard hue-histogram value can be compared with the frame at that position, dataset of frames is used for position of scorecard bar detection.

Let, Referred Scorecard= F1, frame for detection=F2, Threshold =T3 and Position of scorecard bar=P1

So,

If (HHD F2(P1) comp HDD F1 comp T3)

Output: Frame with Replay

Else

Frame can be used

Here, comp= Comparison Operator

Fig. 4 Score Card Bar

Level-3

1) Field View Detection:

Dominant Grass Pixel Ratio can be used but threshold value may vary so, other method for the same is using supervised learning by training SVM. SVM is Support Vector Machine , it is widely used for supervised learning especially for regression and classification of images[12]. So in this way SVM can be trained by images of different grounds having different values and the trained SVM can be used for classification of view as view of field or non-field view.

Fig. 5 Field View

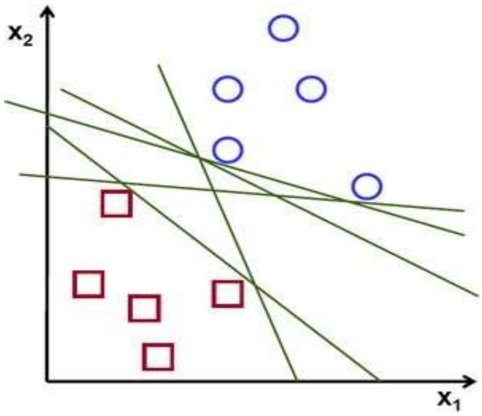

Fig. 6(A) Possible Hyper Plane

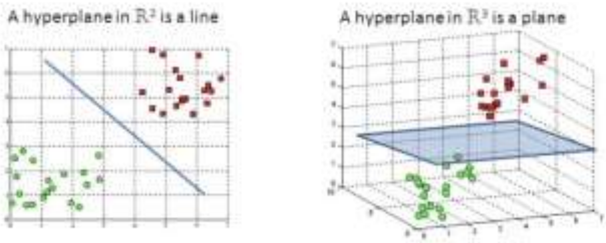

Fig. 6(B) 2D and 3D Hyper Plane Feature

2) Field View Classification:

For detection of pitch, boundary and wide view also there can be many methods like use of tensor flow i.e. using temporal variations and SVM [2]. Training SVM can be effective here as well instead of comparing and playing with threshold values, various pictures of pitch, boundary and long view can be used for training SVM.

(C) (D)

Fig.7 (A) Long View (B) Boundary View (C) Pitch View (D) Field View

Level-4

At level 4, we differentiate the close up and crowd edges through the frame. Both Close view of a person and mass gathering (Crowd View) can be classified and detected by using RGB to YC_bC_r. On conversion of RGB to YC_bC_r, it becomes easy to detect the no. of edge pixels in the pictures. So we can compute the no. of edge pixels of images using Canny edge detection. SVM training method can also use for this classification but it may not be suitable for close level detection

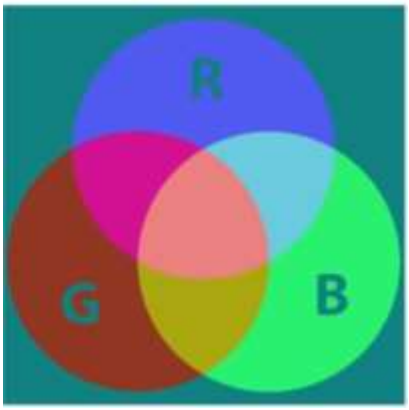

Fig. 8 YC_bC_r Color Plane

Fig. 9 Crowd View

Let, Edge Pixels=EP1, Threshold Value for Close View=T4 and Threshold Value for Crowd View=T5

So,

Convert Image into YC_bC_r

If (EP1>T4)

Output: Closed View

Elseif (EP1>T5)

Output: Crowd View

Check (For Team Gathering || Audience)

Level-5

We can again train SVM for the classification by providing images of team gathering and crowd gathering and differentiate.

If (Crowd View)

If (Crowd View== Same Color)

Output: Team Gathering

Else

Output: Crowd Gathering

Fig. 10(A) Crowd View

Fig. 10(B) Team Gathering

IV. RESULT

Here we denote true positives, false negatives, and false positives by tp, fn and fp respectively, also We know that precision and recall is calculated by;

Precision = tp/tp +fp

Recall = tp/tp+fn

We have extracted our results starting from replay detection, here we kept threshold 0.65 for corr2 and by using hue histogram difference method we calculated the values of precision and recall, these values are based on the code we developed and then tested on our dataset. The values of precision and recall when we used our approach are 100% and 97% respectively and when we used hue histogram difference technique then these values are 100% and 55% respectively.

Further on score bar detection we observed many frames for detecting frames with no score bar and as a result the values of precision and recall we got are 97% and 100% respectively.

At level where we differentiate between field view and non-field view by training SVM using about 4000 images and testing 4200 images and as a result the values of precision and recall we got are 94% and 87%, also after performing tests using due histogram method on same images the precision and recall values recoded less than the previous values.

Now the classification of whole field view is done, here we calculate the precision and recall values of pitch view, boundary view, and long view using our approach and due histogram method both. The precision and recall values using our method were approximately greater than 95% and by using due histogram method these values varied between 20 to 70 percent in all the above mentioned three views.

In the next stage classification of non-field view is done into crowd views and close up views. The precision and recall values using our approach was 94% and 98% respectively and by using due histogram method these values were 45% and 94% respectively of crowd view. Also the precision and recall values using our approach was 82% and 53% respectively and by using due histogram method these values were 98% and 79% respectively of close up views.

Now the classification of crowd view is done into fielder's gatherings or spectator's crowd. After training SVM using 1000 images and testing on 1200 images the precision and recall values came out to be 98% and 97% respectively.

V. CONCLUSION

We proposed a modern strategy to naturally produce cricket highlights, centering on both event-driven and energy-based highlights. We appeared that our framework can accomplish comparable comes about to manual highlights which it yields worthy comes about for cricket fans. Overall, we illustrated that partitioning a cricket coordinate into video shots and signals, such as replays, audio escalated, scoreboard, player celebration, and playfield scenario we will make tall-quality highlights without human supervision. In spite of the fact that there's still room for enhancement, we too proposed to amplify this work to create programmed captions for sports videos as well as for highlight clips of person players. And this paper, we propose a novel approach for identifying highlights in sports videos utilizing simple-to-extricate moo-level visual highlights.

We connected supervised strategies to perform event discovery in cricket. The outcomes were exceptionally essentially superior to the unsupervised approach. We tried our approaches for each level on a dataset of numerous outlines. For the distinctive levels within the extraction pecking order, we accomplished ~95% for nearly all levels. This appears exceptionally promising for programmed highlights extraction in sports video. Subsequently, we accomplished our objective.

VI. IMPLICATION

At display, our viewpoint may be a bit long and incorporate clips that are not portion of the official highlights. Other than, that to assess untrue positives in highlights there are no clear metrics, since wrong positives are naturally subjective in this case.

Another challenging theme of examination is identifying bizarre occasions that can take put on a coordinate and are of incredible relevance for highlights, such as extraordinary climate, wounds, or seriously quibbling between players. In future, able to work upon to progress the accuracy indeed more for the location of occasions at distinctive degree so as to extricate the highlights without much human intercession.

Another future work incorporates finding arrangements to a few than culminate, may miss out a few curiously occasions, and may instep include other sequences which may not be curiously to the watcher. This can be the most undesired cause so that manual extraction would be favored over programmed extraction of sports highlights. In this way, future work incorporates moving forward the come about by finding superior elective issues such as close up discovery.

In future, our proposed framework may too be implemented to others sports like Soccer, Tennis, American football, Table tennis, Ball, Badminton etc. So we are able to produce highlights of any sports videos by applying our proposed algorithm

VII. ACKNOWLEDGMENT

First and foremost, I would like to thank Mr. Rohit Raj Singh, our professor without his constant guidance and support, this paper would have been impossible. Also, our Head of Institute Dr. Alok Kumar Jain, his indirect support is also a blessing.

I would also like to extend my gratitude towards, family, friends, acquaintances and all other people who directly or indirectly aided the efforts during the completion of this project. Any help of any kind is greatly appreciated, without them, this would have been a rather difficult endeavour.

REFERENCES

[1] Jimmy Soni, Amit Bhimani,Priyanka Buch ".Automatic sports highlight Extraction based on logo frames and reply Algorithm" TROI 2015.

[2] N. Harikrishna, Sanjeev Satheesh, S. Dinesh Sriram, K.S. Easwarakumar. "Temporal Classification of Events in Cricket Videos", 2011.

[3] Kamesh Namuduri. “Automatic extraction of highlights from a cricket video using MPEG7 descriptors”

[4] IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 4, Issue 1, Ver. I (Jan. 2014), PP 01-05 e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197

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

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

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

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

[9] Tjondronegoro, Dian, Yi-Ping Phoebe Chen, and Binh Pham., "Integrating highlights for more complete sports video summarization", IEEE multimedia, Vol. 11, No. 4, pp. 22-37,2004.

[10] Zawbaa, Hossam M., et al., "Event detection based approach for soccer video summarization using machine learning", International Journal of Multimedia and Ubiquitous Engineering,Vol. 7, No.2, pp. 63-80,2012.

[11] BT Truong, S Venkatesh: “Video abstraction: A systematic review and classification”, ACM Transactions on Multimedia Computing, Communications, 2007.

[12] DURGESH K. SRIVASTAVA, LEKHA BHAMBHU “DATA CLASSIFICATION USING SUPPORT VECTOR MACHINE” Journal of Theoretical and Applied Information Technology Vol 12. No. 1 -- 2010