Enriching Feature Extraction for Cricket Video Event Detection

[1]Ms.Deepali Bhawarthi, [2] Prof Shriniwas Gadage

[1]*PG Student, G.H.Raisoni College of Engineering and Management, Pune*
[2]*Professor, G.H.Raisoni College of Engineering and Management, Pune*

Abstract

Video is rapidly becoming one of the most popular multimedia due to its high information and entertainment capability. It consists of audio, video and text together. Finding desired information in a video clip or in a video database is a difficult and laborious task due to its semantic gap between the low-level feature and high-level video semantic concepts. A video database contains lot of semantic information. The semantic information describes what is happening in the video and also what is perceived by human users. Techniques for rapid video browsing, video retrieval become essential as video material became more and more available and subtle. Video browsing is very important functionality of multimedia systems which offer the user efficient way to view relevant information from large amount of video material. On the other hand, video retrieval enables the user to search for particular video segment based on some description. In this paper, we have presented the work for cricket video event detection as sports media is increasing day by day and all national and international news broadcasts contain specific regular segments devoted to sports. The purpose of the video event detection evolves mainly due to viewing time constraints. The goal of this paper is to outline the advance concepts for cricket highlight generation which is an effort towards summarization.

Key words: Browsing, event detection, multimedia, retrieval, semantic gap, video database.

1. Introduction

Video is the collection of continuous frames which is normally displayed at rate of 25 fps. Video itself contains a huge amount of data and complexity that makes the analysis very difficult. The huge amount of data that is produced by digitizing sports videos demands a process of data filtration and reduction. With remarkable development in multimedia systems, many sports applications came into birth. Video event analysis and recognition is a critical task in many applications such as detection of sports highlights, incident detection in surveillance video, indexing human-computer interaction [5]. Due to the long duration of the video, it is quite difficult process to index some particular event of the video. Cricket is a popular international game with a large viewer ship. Television broadcasters like Neo cricket, ESPN, Star Sports have huge databases of cricket videos. The lengths of these matches are 4-5 days and thus extracting meaningful events which are of interest to the viewer is very important. One has to go through each frame to find out some specific event from it. Because of the enormous difference in sports videos, sport Specific methods show successful results and thus constitute the majority of work. Cricket video analysis is far more challenging because of the complexities of game in itself. Cricket is a game of variable factors as compared to other famous sports such as soccer, basketball, tennis etc. Even the latest version of cricket i.e.T20-20 is played for approximately 3 hours which is greater than soccer which is played approximately for 90 minutes and hockey which is played approximately for 70 minutes. At present very few systems are implanted for cricket highlight generation as the target rating point(TRP) of the media is hiked for the channels who are able to efficiently present the news before any competitor channel produces it. The simplest way to do that is to extract one or more scalar or vector features from each frame and to define distance functions on the feature domain. Alternatively the features themselves can be used either as events for clustering the frames to view.

In this paper we have presented an approach for cricket video event detection. The system facilitates a novel technique of selecting semantic concepts and the events within the concepts, according to their degree of importance. Different importance may be assigned by different group of users. For example, a general viewer may like to have a comprehensive viewing of all important actions, whereas, specialist viewers may like to observe actions of their choice. This approach facilitates such customized highlight generation, by assigning event importance.

Event detection is thus distinct from the problem of human action recognition, where the primary goal is to classify a short video sequence of an actor performing an unknown action into one of several classes. It help us to assess the relevance or value of information within a shorter period of time while decision making. The use of only basic input as video is limiting the event retrieval and indexing capabilities of the user. The goal of this project is to expand the ways that people are able to interact with their computers. Visual features extracted from individual frames are trained to classify them into a category such as replay, crowd, etc. Namely, we wanted to enable users to interact more naturally with their computer by using simple GUI and perform various event detection actions for genre specific sports domain. In this project, a system is developed which enables to detect events in cricket video like SIXES and extract features like CLOSE-UPS, REPLAYS and CROWD which is more directly interacted by the user and can summarize all the events which can be used for specific review. This system is simple enough to run using normal personnel computers and requires little training

2. Related Work

Multi-level hierarchical framework was a successful attempt to extract semantic events from cricket video. This framework used audio-visual features to classify individual video segments [15]. On the other hand, it has been shown that camera motion parameters can also be used to classify minimal events in a cricket video. A text-based segmentation has also been proposed, where online cricket commentaries are used to annotate a cricket video [6]. In [11] Hidden markov model technique was proposed and MPEG-7 visual descriptors were used to identify cricket highlights .The semantic information of a video has two important aspects [4][8][10]. They are (a). A spatial aspect which means a semantic content presented by a video frame, such as the location, characters and objects displayed in the video frame. (b). A temporal aspect which means a semantic content presented by a sequence of video frames in time, such as character's action and object's movement presented in the sequence. To represent temporal aspects, the higher-level semantic information of video is extracted by examining the features audio, video, and superimposed text of the video [15]. The semantic information includes the detecting trigger events, determining typical and anomalous patterns of activity, generating person-centric or object-centric views of an activity, classifying activities into named categories, and clustering and determining the interactions between entities. The temporal aspect of

videos prevents the efficient browsing of these very large databases. Many efforts are conducted to extract the association between low-level visual features and high-level semantic concepts for image annotation. Video events contain rich semantic information. These are normally defined as the interesting events which capture user attentions. For example, a soccer goal event is defined as the ball passing over the goal line without touching the goal posts and the crossbar. Kolekar et al [7] proposes a method to generate highlights based on event selection and giving it an importance value based on user feedback (manual). Also the major disadvantage of real time cricket video is to generate frames as there are various video format like MPEG,AVI etc along with different frame rate, conversion rate ,frame rate,etc.So it demands for a system which is independent of the discussed parameters and need a generic approach. So in our system we have used an approach to convert any type of video into frames by fetching screenshots at run time and crop them to increase the accuracy and save them in a folder as per the user's convenient location. We have not used hierarchical approach as the retrieval time gradually increases. Hence with our approach frames are retrieved directly using the algorithms discussed in the next session

3. Proposed Method

In this paper we propose a key frame detection approach for minimizing the computation time which is crucial as the amount of data is huge. We have concentrated our work on T20-20 matches as they offer lots of events in shorter duration of time as compared to ODI's. The same work can be extended to ODIs as well as across multiple types of videos like news, movies, etc. Assuming temporal decomposition of video into structural units as clip, scenes, shots and frames forms the basis of most of the research in sports video processing. Based on single set of fixed or smoothly varying camera parameters like close-up, crowd, spectators, etc a group of sequential frames often form shot. Scene is a collection of related shots. A series of related scenes form a sequence and a part of the sequence is called as clip. A video is composed of different story units such as shots, scenes, clips, and sequences arranged according to some logical structure defined by the screen play. In our work, we extract the events in the form of clips and after analysis assign a descriptive label to each clip. The overall architecture of the system is shown in figure 1:
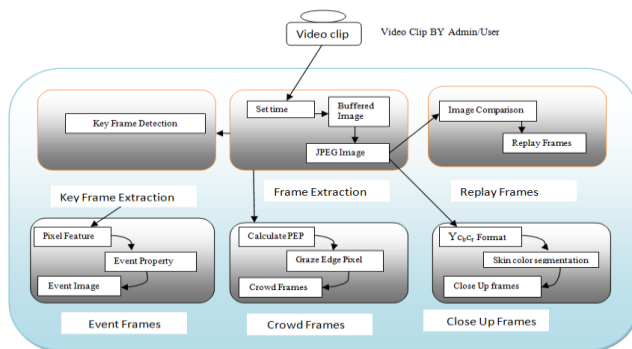
Fig.1 System Architecture

A. Conversion of Video to Frames

Instead of reading the video as it is, we converted the video into frames. Reading the video and directly processing the video is a tedious task and requires lot of memory and can work only on systems with high configurations. In a video clip each and every frame should be grabbed at a fixed frame capture rate. Now, each and every frame is an individual image and we can apply all image processing algorithms to these captured frames which is a major advantage and the size of the video does not matter here and no need of specific memory requirements. A popular method to identify frame boundaries is to compute the colour histogram of consecutive frames. Successive frames belong to the same shot if their colour histograms are similar. We have used the RGB colour histogram to estimate the similarity between two frames. The red, green and blue components of a frame are quantized into 4 (red), 4 (green) and 4 (blue) bins respectively, leading to a total of $4 \times 4 \times 4 = 64$ bins. A shot boundary is detected when the histogram difference between two successive frames crosses a threshold. This technique works well when there transitions are abrupt or hard-cut. Fig 2 shows the frames generated by this approach.

Fig 2.Key Frames

Any video frame sequences have certain properties like some frame sequences are either same or have a little difference between them. Thus using classifiers to classify all such frames is computationally inefficient. Suppose a sequence of similar (S) and dissimilar (D) frames are:

SSSDSSS

such a sequence of frames there would be a spike in the hue histogram difference (HHD) plot with a rising edge and falling edge .we have to consider only the rising edges to decide on key and non key frames.
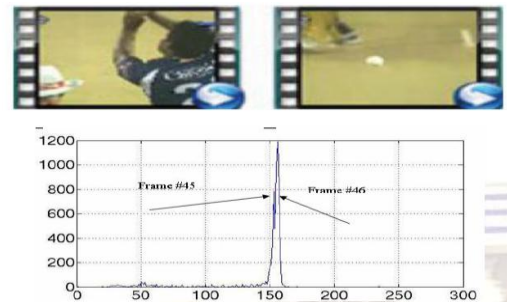

Fig 3. Hue Histogram Difference

Algorithm Key Frame Detection:
1) Convert the RGB image into HSV format.
2) Compute the frame to frame hue difference.
3)   if (diff (i)-diff (i-1) > kf_{thres}) then
        Classify frame i as key frame
    else
        Classify frame as non key frame

B. Visual Features

We now list some of the visual features used for shot categorization.

1. Grass Pixel Ratio (GPR): A pixel is identified as a grass pixel if its hue component lies between 48 and 68 (determined experimentally). We first compute a histogram for the hue component of the frame quantized into 256 bins. The GPR is the ratio of the pixel count in bins 48-68 to the total number of pixels.

2. Edge Pixel Ratio (EPR): The presence of crowd in a frame can be detected by performing canny edge detection on the given image. We therefore calculate the ratio of the number of edge pixels to the total number of pixels on a frame (EPR).

3. Skin Colour Ratio (SCR): The presence of a player/umpire on a frame can be detected by looking at the percentage of skin colour pixels in the frame. For

this, we divide the frame into 16 equally sized blocks and calculate the skin pixel ratio for each block.

C. Close-up Detection

Close-up (CU) is a photographic technique which tightly frames a person or an object. In movies, it is applied to guide audience attention and to evoke audience emotion. For close up detection we have used Haar features wavelets which use single wavelength square waves (one high interval and one low interval). The presence of a *Haar* feature is determined by subtracting the average dark region pixel value from the average light-region pixel value. If the difference is above a threshold (set during learning), that feature is said to be present. To determine the presence or absence of hundreds of *Haar* features at every image location and at several scales efficiently, integration is done. The filters at each level are trained to classify training images that passed all previous stages.

D. Crowd Detection

we see that close up or crowd frames are shown frequently whenever an exciting event occurs such as when a wicket falls, close up of batsman and bowler, then view of spectators and the player's gathering of fielding team are certainly shown. The edge detection is performed by finding the maximum gradient value of a pixel from its neighbouring pixels. If the maximum value of gradient satisfies the threshold than the pixel is classified as an edge pixel [3].The percentage of edge pixels (PEP) are used to classify the frame as crowd or close-up, since we typically observe more edge pixels for crowd frames. We applied canny edge detector and use the following ratio as the close-up detection parameter:

$$PEP = \frac{total\ number\ of\ edge\ pixels}{total\ number\ of\ pixels\ in\ the\ frame\ (EPR)} \times 100\ \%$$

The canny edge detector, first smoothes the image to eliminate noise. It then finds the image gradient to highlight regions with high spatial derivatives. The algorithm then tracks along these regions and suppresses any pixel that is not at the maximum.

Algorithm crowd detection:
1) Convert the input RGB image into YC_bC_r model.
2) Apply *canny* operator to detect the edge pixels.
3) Compute Percentage Edge Pixel (PEP) for the image.
4) Classify the image using following condition:
    if *(PEP > P$_{PEP}$)* then
            frame belongs to class crowd

        else

frame belongs to class close-up

E. Replay Detection

Motion vector [6] and replay structures [12] are used to detect replays from sports video. But these methods are not robust enough to be suitable for various kinds of sports video replay detection because replays in different sports video are various and compiled in different manners and can hardly be represented by such simple features. Therefore the recent approach is to detect the accompanying logo effect of the replays in sports videos to acquire the replay segmentations. It is been commonly observed that a replay segment is always sandwiched between two logo transitions or flying graphics which last for 8-15 frames. The following pseudo code underlines the brief concept that is used for replay detection for two consecutive frames and the results of same are shown in the fig:

Algorithm replay detection
1) For each frame i (for Image 1)
2) For Each Frame j=i+1 (for Image 2)
3) Rgbvalue1 of each Pixel (Getting RGB Value for first image)
4) Rgbvalue2 of each Pixel (Getting RGB Value for second image)
5) If(Rgbvalue1== Rgbvalue2) Then
        frame belongs to Replay Event

else

    frame belongs to non Replay Event

F. Sixer Event Detection

In our proposed system we are working on T20 cricket format videos. We have considered matches which are played in night. So when a batsman hits a sixer, it is common observation that the ball is raised up high and will be in the air for a while. At this time cameras tracks the ball as well as the background of the shot which is of course in black colour. So our system extracts the sixer frames based on the following algorithm.

Algorithm Sixer Detection
1) For each frame i
2) Convert RGB value of all pixels from binary to integer
3) Initialize blackpixel_count=0
4) if((red>=0 and red<=50) && (green>=0 and green<=50)   and (blue>=0 && blue<=50)) then
5) blackpixel_count++
6) Calculate Black Pixel Percentage (BCP)= (Blackpixel_count/total number of pixel) *100

7) If (BCP>=50) then
Frame belongs to Sixer event

4. Experimental Results

We have defined the events as scenes in the video with some semantic meaning and action associated with it and some features which are to be used as the system processes all the frames. So we have labelled features as close up, replay and crowd and highlighted event as sixes. We have implemented the system using java net beans with the intention to provide a hassle free solution for event and feature detection for cricket video.

The dataset consists of cricket video which captures frames at rate of 20 fps (variable to change) so as to get clear images to label a specific feature or event. The smallest video which was tested was for T20-20 worldcup-2006, played for duration 4 min 52 seconds. Our system extracted 292 frames with the rate of 1 frame per second. When we perform and extract the events like Close-up, Replay, crowd and sixer we found very impressive result by the system, which we plot below in a graph which matches the system extracted frame result with the human extracted frame result.
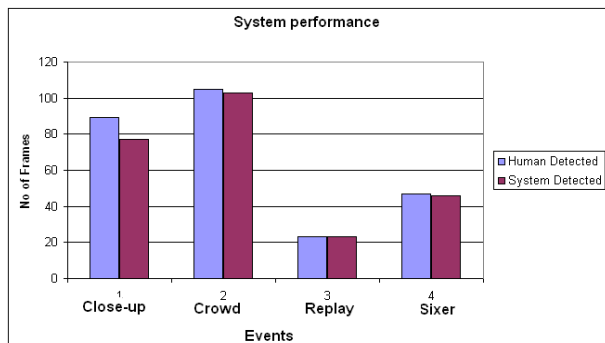
And also we perform the same kind of experiments with many videos which gives the almost gives the similar performance results. The aggregate result is shown in the below table.

TABLE I
PERFORMANCE EVALUATION

Events	Performance Percentage
Close-ups	84.34%
Crowd	97.44%
Replay	100%
sixer	95.4%

5. CONCLUSION

Our Key Frame detection based approach shows excellent detection accuracy and also results in saving of processing time. The classification exhibit better detection and classification ratio at various levels. Our algorithms are able to extract out events as shown in table I from cricket videos. Snapshots in the following figure show some of the classified frames.
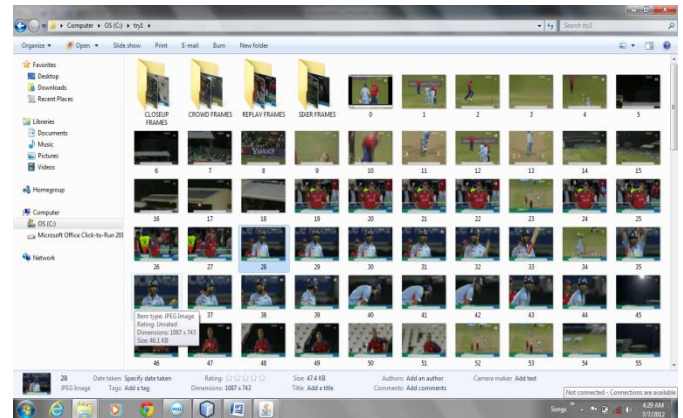
Fig 4.Resultant output folders of event detection

Fig 5.close-up Frames

Fig 6. Replay Frames

Fig 7. Crowd Frames

Fig 8.Detection of six

6. References

[1] Mahesh Kumar H. Kolekar, Kannappan Palaniappan "*Semantic Concept Mining Based on Hierarchical Event Detection for Soccer Video Indexing*", journal of multimedia, vol. 4, no. 5, October 2009.

[2] Changsheng Xu, Jinjun Wang, Hanqing Lu, Yifan Zhang , "*A Novel Framework for Semantic Annotation and Personalized Retrieval of Sports Video*", IEEE transactions on multimedia, vol. 10, no. 3, April 2008. [3] A. Hanjalic, "*Adaptive extraction of highlights from a sportvideo based on excitement modelling,*" IEEE Transactions on Multimedia, vol. 7, no. 6, pp. 1114–1122, 2005.

[4] Hu Min, Yang Shuangyuan, "*Overview of content-based image retrieval with high-level semantics*" 20IO 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE).

[5] Dr.N.Krishnan, M.Sheerin Banu and C.Callins Christiyana, "*Content Based Image Retrieval using Dominant Colour Identification Based on Foreground Objects*", International Conference on Computational Intelligence and Multimedia Applications, 2007 IEEE, DOI 10.1109/ICCIMA.2007.64

[6] M.H. Kolekar and S. Sengupta, "*Semantic concept mining in cricket videos for automated highlight generation,*" Multimedia Tools and Applications, Springer, vol. 47, no. 3, May. 2009, pp. 545-579.

[7] M.H. Kolekar, K. Palaniappan, and S. Sengupta, "*Semantic event detection and classification in cricket video sequence,*" in Proceedings Sixth of Indian Conference on Computer Vision, Graphics and Image Processing, Bhubaneshwar, India, Dec. 2008, pp. 382-389.

[8]Xiaoyun Wang, Jianfeng Zhou" *An Improvement on the Model of Ontology-Based Semantic Similarity Computation*", 2009 First International Workshop on Database Technology and Applications, 2009 IEEE, DOI 10.1109/DBTA.2009.17

[9] Yu Xiao Hong and Xu Jinhua," *The Related Techniques of Content-based Image Retrieval*", 2008 International Symposium on Computer Science and Computational Technology.

[10] A. Kokaram, N. Rea, R. Dahyot, M. Tekalp, P. Bouthemy,P. Gros, and I. Sezan, "*Browsing sports video: trends in sports-related indexing and retrieval work,*" in IEEE Signal Processing Magazine, vol. 23, no. 2, pp. 47–58, 2006.

[10] M. H. Kolekar and S. Sengupta, "*Semantic Indexing of News Video Sequences: A Multimodal Hierarchical Approach Based on Hidden Markov Model,*" in IEEE Int.Region 10 Conference (TENCON), 2005.

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

[12] B Acharya, AK Majumdar, J Mukherjee: "*Video model for dynamic objects",* Information Sciences, Elsevier, 2006.

[13]A Ekin, AM Tekalp, R Mehrotra: *Automatic soccer video analysis and summarization*. IEEE Transactions on Image Processing, 2003.

[14] KP Sankar, S Pandey, CV Jawahar: "*Text Driven Temporal Segmentation of Cricket Videos*", ICVGIP 2006.

[15] G. Evangelopoulos,A. Zlatintsi ,G. Skoumas,K.Rapantzikos:"*video event detection and summarization using audio, visual and text saliency",2009,IEEE 978-1-4244-2354-5/09*