Event Resolution in Sports Videos: A Review

S. C. Premaratne
School of Science and Technology
Malaysia University of Science and Technology (MUST)
Malaysia.
samindap@uom.lk

K. L. Jayaratne
University of Colombo Schools of Computing
No 35, Philip Gunawardena Mawatha,
Colombo 00400.
klj@ucsc.cmb.ac.lk

P. Sellappan
School of Science and Technology
Malaysia University of Science and Technology (MUST)
Malaysia.
sell@must.edu.my

ABSTRACT

Modern digital video technologies have opened the avenues for the researchers to explore the ways for effective indexing and archiving of multimedia databases. Therefore, new concepts to mine multimedia databases have emerged, and people are doing numerous researches on how to effectively handle different types of multimedia content. Event resolution is one fundamental component in sports videos which supports effective mining and indexing because the viewers prefer to watch only the interesting or exciting portions of the videos. This paper is a review that addresses the challenges of automatically extracting the highlights from full-length ports videos. We are interested in designing a unique framework for event resolution in cricket videos by analyzing the existing research work for several different types of sports. This investigation provides room for the addition of further audio data extraction and textual data extraction for classification of a multimodal dataset.

Keywords

Video Summarization, Video Analysis, Highlight Generation, Event Resolution.

1. INTRODUCTION

Today, society is facing many problems with handling the increasing amount of sports video from TV broadcasts [1][2][3]. The essential requirement in managing video is compressing its long sequence into a more compact picture through an effective summarizing process. Video summarization will facilitate fast and efficient browsing from large video repositories[4]. It also allows more efficient content indexing. Sports video summarization need to address three key points[5][6].

(1) The video summary should identify the necessary all the critical events in the full-length video. For example, in a cricket game, the summary must contain wickets, sixes, fours, runouts, and some other important scenes.

(2) The video summary should maintain accurate connectivity between events. For example, when a player hit a six, the event should be identified as one continuous event comprising several video segments.

(3) The summary should not contain any redundant events. For example, player highlights can be shown while the game is ongoing. Therefore, sometimes the events can be recognized as two different events, but it is the same event repeated.

2. VIDEO SEGMENTATION

Video summarization is an abstract view of the original video sequence, and it facilitates an efficient way of browsing and retrieving of information in video databases[7]. It can be a collection of highlights of an original sports video represented by a collection of keyframes. To develop such indexing and search techniques to manage the massive amount of video data, new techniques need to be researched. The developed techniques in video summarization can be used for various domains, such as surveillance videos, consumer videos, movies, sports, news, etc.

The video can also be thought of as a collection of many scenes, where a scene is a collection of shots that have the same context[8]. A shot is a collection of frames. Anatomy of a video is shown in figure 1.

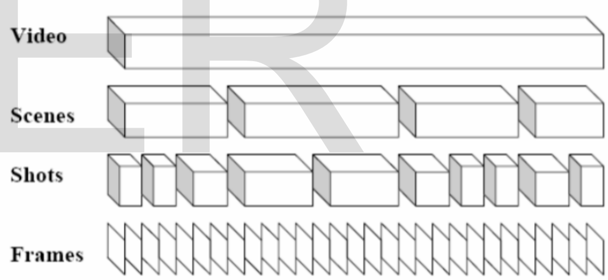

Figure 1: Anatomy of a video.

The video data should be divided into scenes for effective indexing and retrieval. A scene can be considered as continues action. For example, in a cricket video sixes, fours, runout, and no-ball can be considered as essential scenes[9]. A scene may compose of several shots which are considered as continues camera recordings. A shot represents a sequence of frames captured when the camera starts rolling and until it stops. Anatomy of a sports video by taking a cricket video as an example is shown in figure 2. The full-length cricket video is divided into individual deliveries which we consider as scenes or events. Each of the deliveries can classify into deliveries which are scored or not scored. Each delivery will have a number of shots which are from different camera recordings as well as motion changes. A number of frames will cover each shot. The first challenge of the research is to identify the scene boundaries accurately.

The challenge of segmenting a video sequence into shots is the ability to distinguish between scene breaks and regular changes which may be due to the motion of large objects or the motion of

ISSN 2229-5518

the camera[10]. There are different approaches for shot boundary detection. One such approach is histogram-based shot boundary detection. Hong Shao et al. have proposed a shot boundary detection method based on HSV color histogram and Histograms of Gradians (HOG) [4]. They deploy an HSV color histogram difference to detect the initial shot boundaries and then the shot boundaries are further verified using the HOG threshold. The experimental results show that the proposed method achieved high accuracy and is robust to brightness changes and motion.

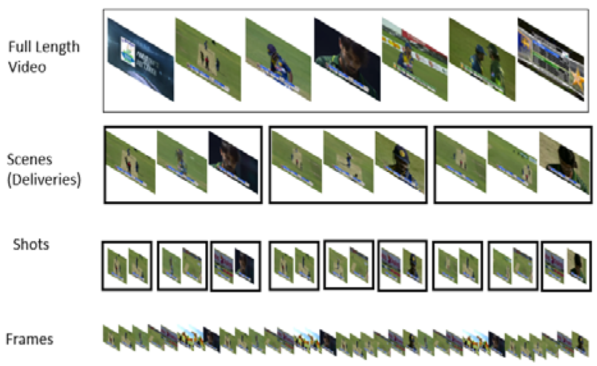

Figure 2: Anatomy of a sports video (Cricket).

A drawback of histogram-based approach is the false alarms caused in camera motion when motion analysis is done for shot boundaries [5]. When analyzing professionally produced sports videos, we observed that different sports events are frequently associated with different motion patterns. For example, in cricket videos, within an over between deliveries we can observe a similar pattern. Therefore, delivery pattern detection can be done for different types of bowlers as well. Not only within the same game but also when we compare with different games played between different countries. This may is due to the standard positions of the camera as well as common production standards. Based on this observation, Chen-Yu Chen et al. have investigated the feasibility of using motion patterns to characterize semantics and detect events in a sports video [3]. They analyze the statistical distribution of motion vectors by obtaining the entropy for each frame. The significance of the shots is determined by observing the behavior of the entropy motion value (EMV).

3. VIDEO SUMMARIZATION

Video summarization can be described based on figure 3. A critical process in content-based video indexing is feature extraction. The effectiveness of an indexing scheme depends on the effectiveness of attributes in content representation. However, it is difficult to map extractable video features (such as color, texture, shape, structure, layout, and motion) easily into semantic concepts in sports (such as goal, foul or penalty). Although the visual content is a significant source of information in a video program, an effective strategy in the video-content analysis is to use attributes extractable from multimedia sources. Much valuable information is also carried in other media components, such as text (superimposed on the images, or included as closed captions), audio, and speech that accompanies the pictorial component. A combined and cooperative analysis of these components would be far more effective in characterizing video program for both consumer and professional applications.

An essential step in the process of video structure parsing is that of segmenting the video into individual scenes. From a narrative point of view, a scene consists of a series of consecutive shots grouped because they are in the same location or because they share some thematic content. In contrast, shots are in the actual physical primary layer in the video, whose boundaries are determined by continous camera recording. When a camera is turned on/off or when the camera is switched from one to another the shot boundaries are detected. Shot boundary detection algorithms that rely only on the visual information contained in the video frames can segment the video into frames with similar visual contents. Grouping the shots into semantically meaningful segments such as events, however, usually isn't possible without incorporating information from the video program's other components. Multimodal processing algorithms involving the processing of not only the video frames, but also the text, audio, and speech components that accompany them have proven effective in achieving this goal.

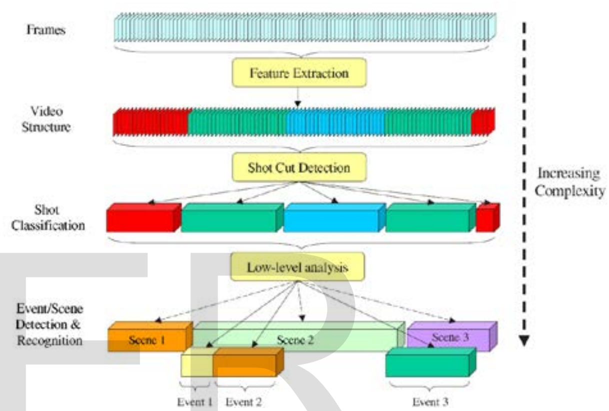

Figure 3: Hierarchy of video summarization.

Video abstraction is the process of creating a presentation of visual information about a landscape or the structure of the video, which should be much shorter than the original video. This abstraction process is similar to the extraction of keywords or summaries in text document processing. That is, we need to extract a subset of video data from the original video such as keyframes or highlights as entries for shots or scenes. Abstraction is especially important given the vast amount of data for a video program or even a few minutes' duration. The result forms the basis not only for video content representation but also for content-based video browsing. Combining the structure information extracted from video parsing and keyframes extracted in video abstraction, we can build a visual table of contents of a video program. Several terms and corresponding methods exist for abstracting video content, including skimming, highlights, and summary. A video skim is a condensed representation of the video containing keywords, frames, visual, and audio sequences. Highlights frequently involve the detection of important events in the video. A summary means that we preserved important structural and semantic information in a short version of the video represented via key audio, video, frames, and/or segments.

Video summarization can be divided into two categories they are:

1. Highlights based Summarization.
2. Table of Contents based Summarization.

In highlights-based video summarization, the primary segmentation is done based on play break scenes. This approach is potentially effective for browsing purposes because viewers will not miss any important events although they skip most of the break scenes. It is due to the fact that most highlights are contained within play scenes. But most of the time exciting events happen during the transitions between play and break. For instance, a penalty kick is how a play is resumed after stopped due to a foul which is committed inside the penalty area. Thus, a highlight should include all these play-break-play scenes to ensure that the scene contains all the necessary details.

A hierarchical structure organized for play, break and highlight scenes is shown in Figure 4. Each play, break, or combination of play-and-break can contain one to many highlight scene(s) which can be organized into a highlight collection. For example, if users are interested in storing highlight collection from team A, the corresponding highlights which belong to team A will be compiled into a highlight collection. Based on this structure, users can select to watch all play and break scenes or just the ones which have a certain number of highlights (i.e., exciting play and break scenes). Users can also refer back to the whole play or break scene if they found a highlight is not adequate.

Figure 4: Highlights-based video summarization.

Table of contents-based summarization is useful when the user wants to browse the sports video based on a particular event such as a goal in football, try in rugby or a six in cricket. On such approach is the detection of the replays within the sports video. A replay can be distinguished based on logo detection, scoreboard detection, gradual transition of frames or any of those combinations. The components of the approach are available in figure 5. Some research has been carried out for replay detection by using a combination of motion analysis and scoreboard detection. It has been observed that broadcasters omit the scoreboard in replay segments. Moreover, replay frames contain multiple gradual transitions. The research exploits these two observations for replay detection. More specifically, the proposed method uses gradual transitions and scoreboard detection s for replay detection that is then used for highlight generation. To achieve this goal, a dual-threshold based method is used for gradual transition detection. Detected gradual transition frames are used to extract candidate replay segments. Candidate replay segments are used for scoreboard detection. The detected scoreboard frames are used to discriminate between replay and live video frames. The proposed algorithm does not rely on logo template recognition for replay detection, which makes it computationally efficient.

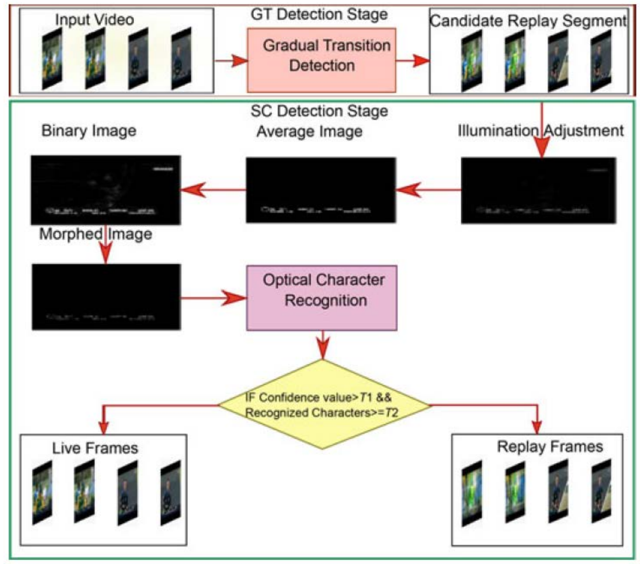

Figure 5: A hybrid technique to detect replays.

4. EVENT RESOLUTION IN DIFFERENT SPORTS.

Over recent years there has been an increasing demand for the use of analytics in sports. From players and officials to broadcasters and sports fans, the need for automated information relating to performance, accuracy and match-play tactics is becoming a pivotal asset to success. Image analysis techniques for segmentation and detection are progressing from long-established theoretical algorithms to practical, commercial installations of real-time processing systems. The outputs of such installations are often represented in the form of match statistics, or as overlay graphics tracing the dynamics of objects of interest (OOI) for specific match events.

4.1. SOCCER.

In order to detect events, the full-length video needs to be segmented into video shots. Then, the system applies support vector machine (SVM) algorithm for emphasizing important segments with logo appearance in addition to detecting the caption region providing information about the score of the game. They observe that the most exciting events occur in the goal-mouth area such as goals, shooting, penalties, direct free kicks, etc. [11]. Other non-exciting events such as dull passes in the midfield, defense, and offense or some other shots to the audiences or coaches, are not considered as exciting as the former events. Excitement event detection is based on four features:

- Scoreboard detection
- Vertical goal posts detection
- Goal net detection
- Commentator loudness detection.

Subsequently, the system uses the k-means algorithm and Hough line transform for detecting vertical goal posts and Gabor filter for detecting goal net. A goal is scored when the whole soccer ball passes the goal line between the goal posts and under the crossbar[12]. The occurrence of a goal event leads to a break in the game. Figure 6 illustrates the sequence of cinematic features after scoring a goal. Finally, the restart of the game is usually captured by a long shot. Following sequence of features are used to detect the goal event that occurs between the long shot resulting in the goal event and the long shot view that shows the restart of the game: long shot, Goal Mouth, Close-up View, Audience View, replay duration (no less than 20 and no more than 60 seconds) and Long View[13].

Figure 6: Sequence of features of a goal.

The summarized segment may contain only important events, such as goal shots, attacks, or penalty shots. The proposed system highlights the most important events. during the soccer match, such as goals and goal attempts, in order to save the viewer's time. Finally, the system highlights the most important events during the match.

4.2. SNOOKER.

Event resolution involves locating and extracting important events and conveying the information to the user in a concise manner. Accuracy is vital given the vast amount of data associated with any spots video where the most valuable semantics only occupy short time periods relative to the total duration of the video footage. The accuracy of extracting events can be improved by Motion History Images (MHI) where a synthetic representation of object motion is overlaid on a keyframe[14]. The motion is represented by a temporally graduating intensity which increases over time. Frames with recent motion are therefore represented by bright regions and earlier motion by darker regions. An illustration of MHIs for snooker is given in figure 7. In the presence of global motion, the use of MHIs for event detection is impractical unless the camera motion can be compensated. Since in a sport like snooker the global motion is minimal, this approach can be applied.

Figure 7: MHI for snooker.

The first level in the framework shown in figure 8 is to extract relevant features for sports retrieval. Playing areas in broadcast sports footage are generally well defined in terms of their color as well as geometry. Novel moment features which provide a succinct single value representation of the image surface are used. Temporal analysis is performed by exploiting the extracted geometric features and established temporal boundary detection techniques. Object tracking is performed using a color-based particle filter. A target model of the object's color distribution is created in the first appropriate frame of the footage. The likelihood of candidate models generated from particles distributed around the projected position of the region in the next frame is computed and weighted accordingly. The second and fourth levels in the framework are performed by the recognition module. Recognition is performed at both low and high levels of abstraction. The low-level analysis is performed by modeling the evolution of the moment features using Hidden Markov Model (HMM) for view classification.

Using maximum likelihood classification, the HMM was employed to classify the various view types. The tracker made use of the a-priori detected scene geometry and its relation to the real-world geometry of the delineated playing area to automatically scale the candidate regions. The spatiotemporal behavior of the objects was modeled using an HMM. It enabled six common high-level events in snooker to be detected. This research only observed the behavior of a single object under the HMM and used supplemental features such as collisions in snooker to aid in the classification. As there can be many more moving objects in view, modeling the behavior of each one could provide access to additional semantics. Significant drawbacks of this approach would include both computational expense and a need for a great deal of training material.

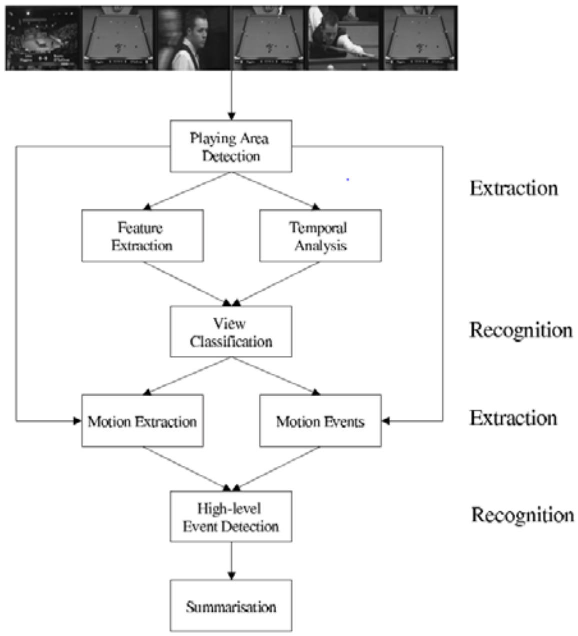

Figure 8 : Framework for event detection in snooker.

4.3. BASKETBALL.

Unlike previous works which depend on event detection, the proposed algorithm analyzes the patterns of scoring to decide semantically significant event[15][16]. A numeric character recognition method is utilized to read the score from the scene. Shot type classification is also performed to include play shots only while excluding non-play shots from the video summary (figure 9).

In the feature extraction process, play shots and non-play shots are classified by the ratio of dominant colors in a shot. To read the score of the game by the number recognition method, an initialization process is performed to localize the position of scoreboard automatically within the image frame[17]. After the feature extraction step, important shots are chosen by analyzing the variation of recognized scores. Then a set of rules are developed for this shot selection process. These rules were applied to analyze the score in determining semantically essential events.

The importance of the video scene is decided by the gap between the current score difference and the score difference in the past. Whether the shot must be included in a summary is decided by thresholding the degree of importance. The resulting highlights consist of exciting clips of scoring and can capture the flow of the basketball game.

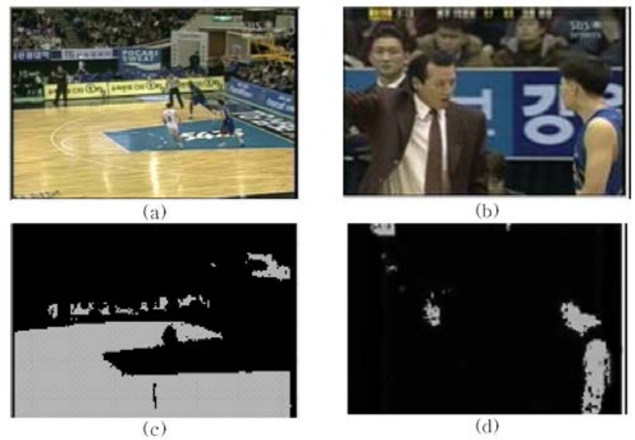

Figure 9 : (a) play shot, (b) non-play shot, (c) dominant colored pixels in play shot, and (d) dominant colored pixels in the non-play shot.

4.4. BADMINTON.

Badminton videos captured during competition and training usually contain both rally periods, in-play and break segments are alternatively concatenated[18]. In practical environments, the videos are in most cases captured by cameras on the spectator stand behind the end line. but in some cases, the game is captured by the ceiling camera in sports institutes, training centers and large-scaled gymnasiums (Figure 10). The approach to infer each rally period from the motion trajectories of players and shuttlecock has some drawbacks due to difficulties in explicitly describing all the possible contexts. In practice, the shuttlecock is not successfully detected.

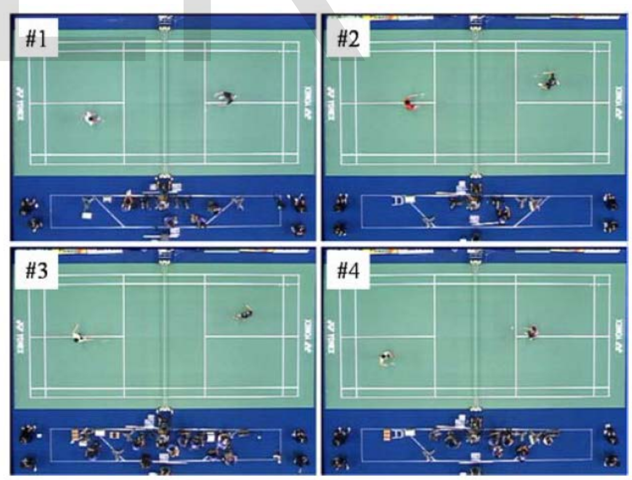

Figure 10 : Badminton game captured by the ceiling camera.

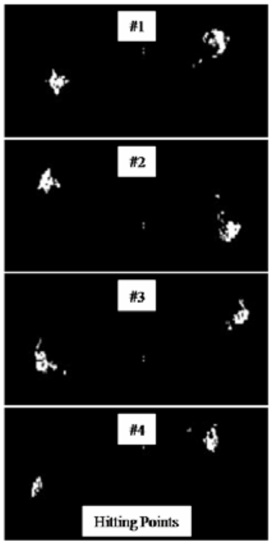

Figure 11: Image sequences in serve motions.

In the feature extraction, cubic higher-order local auto-correlation is employed which enables simultaneous extraction of spatiotemporal features from the motion images (Figure 11). The eight serves in badminton matches are used as the supervised training data. The frames in the non-training section are used as the test data. experimental results of serve detection by the proposed method. A drawback of this approach is that the method does not account for variation of serves such as forehand-long, forehand-short, backhand-short.

4.5. TENNIS.

In broadcasting tennis videos, the camera always switches to court view when two players start the game[19]. The segmentation can be according to the view changes of the camera. Some audiovisual features could be extracted to describe the characteristics of each play. The play break segmentation can be recognized in tennis videos by considering the court view shots and non-court view shots. Since a court view shot usually contains a large number of court pixels, the dominant color ratio (DCR) is utilized as the descriptor to extract court view shots.

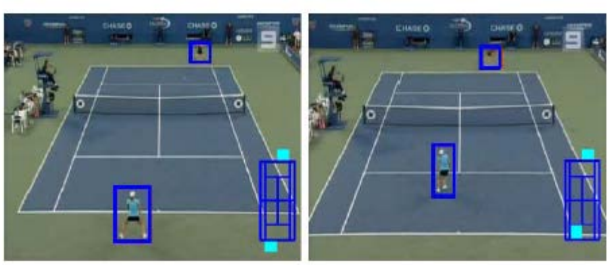

Figure 12: Player detection in tennis

According to tennis regulations, tennis events can be explicitly categorized into the following five types: Fault, Double Fault, Ace or unreturned serve, Baseline rally, and Net approach. In this approach, the following audiovisual features were considered: Moving distance of the player, a Relative position between the player and the court, Applause/cheer sound effects, Length of the play. The player detection process will be used to identify the player potion and player movement (Figure 12). A data mining algorithm is utilized to identify the frequent patterns from the extracted features. Three broadcasting videos of tennis tournament are used to evaluate the proposed methods. The evaluation data are captured from different broadcasting channels with significant variation in broadcasting styles and audio conditions.

4.6. CRICKET.

Cricket video summarization can be done by considering three basic information like Audio information, Textual information and Visual information available in the video[20]. Whenever some exciting event occurs the spectators cheer loudly, and commentator's voice also increase. Audio-based approach extracts the event based on audio level. These audio cues can be detected by measuring the loudness of the audio signal. In a text-based approach, textual information available on the video frames is used such as the scoreboard. First, the scoreboard is detected in the frame and using Optical Character Recognition (OCR), the characters are recognized to extract information like wicket fall, six, four etc. In a visual-based approach, visual features of video frames are considered from the hierarchical classification of the frame. In a cricket video, there may be two type of segments they are, real-time and replay. Replay many of times sandwiched between two logo transitions or sometimes scoreboard is absent in replay sequence. And in real time segments, the frames can be further classified to field view, i.e. pitch, long view, or boundary or non-field view, i.e. close-up of player or crowd. Further close-up can be of umpire or player of any team and crowd can be spectators or gathering of players. The sequence of these features of a cricket video can be used to extract meaningful event.

A hierarchical framework for event detection from the video is shown in figure 13. At level-1, Excitement and Non-Excitement clips are detected. Spectator's cheer and commentator's speech can help us to identify exciting events. To analyze the audio content, techniques- short-time audio energy and zero crossing rate are used.

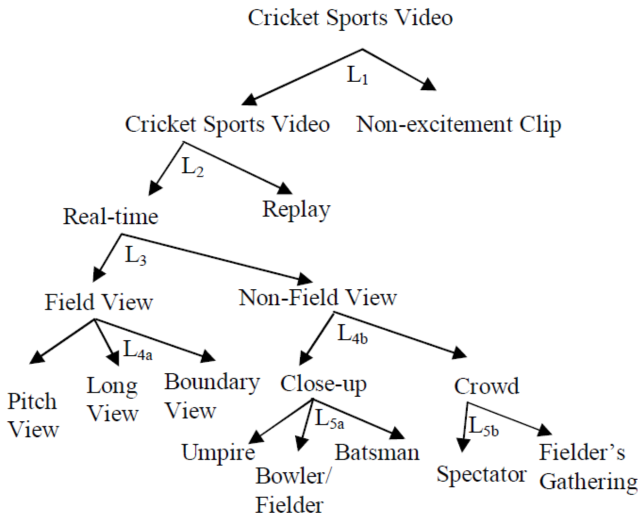

Figure 13: Hierarchical framework for event detection

At level-2, Excitement clip is identified as Real-time clip, or it can be a reply clip. The reply is identified with the help of logo transition or scoreboard appearance at the beginning and end of the reply. At level3, the Real-time clip is identified as Field-view or Non-field view. A histogram analysis can be used to classify the real-time clip into field-view and non-field view. At level-4a, Field view is separated into Pitch-view, Long view or Boundary. The long view is detected using Grass Pixel ratio and Pitch Pixel Ration to identify Long-view and Pitch-view respectively. The boundary is identified using information corresponds to motion in consecutive frames and field pixels. At level-4b, the Non-field view is divided into close-up and crowd. The close-up view is detected by identifying skin portion availability in the image. The crowd is identified by using Edge pixel to total pixel. At level-5a, Close-up is classified as Umpire, Batsman (Team-A) or Bowler/Fielder. Close-up is further classified using jersey color of the player. At level-5b, Crowd is classified as spectators or Fielder's gathering. Grass pixel ratio and jersey color are used for classification of the crowd.

In this research, they have clearly stated that cricket video analysis is a very challenging task as evaluation of results is entirely subjective. In cricket video summarization main issue is higher complexities of the game in itself. The format of the game varies as cricket is played in different manners like test, one day, tournaments, T- 20. Varying field area from venue to venue.

Another approach was introduced for event detection using the umpire gesture recognition. The systems architecture is shown in figure 14. In this approach, the entire input video is first segmented into individual scenes, each of which starts with cricket ball delivery by the bowler and contain various actions like striking the ball by the batsman, movement of the ball over the ground, catching the ball by the fielder, etc. Then the presence of the Umpire in each scene or segment is checked using threshold-based color segmentation technique since Umpires wear uniforms of a distinct color. Only those segments which contain frames with Umpire are selected and analyzed for extracting the gestures. Random Forest (RF) classifier is then trained with the intensity distributions in the frames containing Umpire that uniquely represents the different events of the game as features. RF consists of multiple decision trees, and the final classification is based on the decisions taken by a majority of the trees. Finally, the testing video is given to the trained RF. RF classifies the frames in a scene into different classes. The class labels enable the user to view the video sequence containing only exciting events in the whole game.

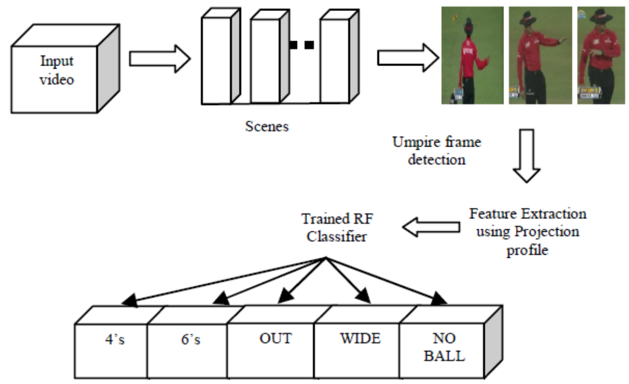

Figure 14: Umpire based event detection

Since not all OUTs are signaled by the Umpire, the system is unable to include those events in the final summary. But most of the WIDE, NO BALL, FOUR, and SIX events could be included in the final summary. Since in some cases the camera does not focus on the Umpire while showing the signal, the method cannot identify the event and hence the detection rate decreases.

5. MULTIDIMENSIONAL MODEL.

We propose a novel multidimensional model for Table of Contents based event detection shown in Figure 15. The proposed approach will process the video under three different categories. They are video, audio, and score. As an example, when a wicket is gone by taking a catch, in that case through the video processing system will identify the related video frames for that event and audio processing (commentary) it will identify relevant details such as who is the bowler, who is the batsman, who took the catch, etc. Also, in most cases there is a scoreboard on the screen, under the text processing system will catch these data and process them. For example, mentioned above, it gets the details which wicket is that and ongoing over, the number of runs, etc. After that, it will generate the table of contents for each event such as out, six, four, etc. Through the system, the user will be given the facility to browse the details related to the match events as well.

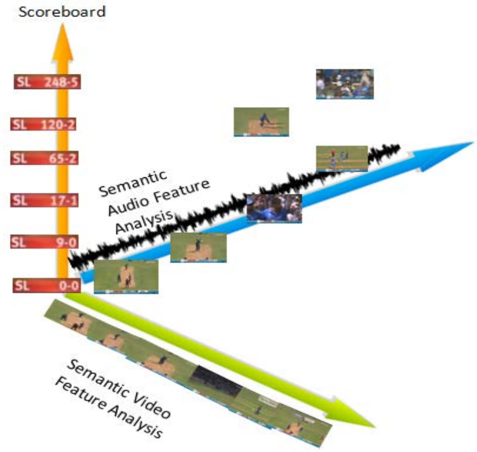

Figure 15: A multidimensional model for event mapping

Our research is an attempt to identify the events indicated in figure 16 to support automatic highlight generation of a given full-length cricket match into a video of few minutes, with selected events rather than just a video of highlights. That means the system will provide the facility of searching the highlights according to the user's wish (wickets, sixes, fours, catches, run outs, etc.).

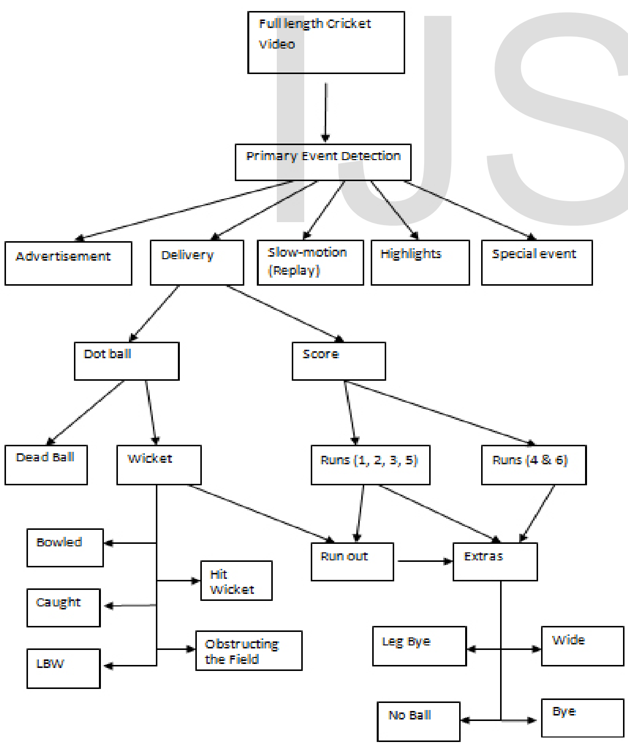

Figure 16 :Taxonomy for cricket highlight generation

When the system gets a video of a cricket match, first it will locate the location of the scoreboard since different broadcasters locate the scoreboard in different locations it. Then the video is segmented into video shots, and in each shot, we try to recognize the scoreboard. If the scoreboard is not available than we check another possibility of logo transaction of the video in the previous video segment from both the results if both are true, then we identify it as an advertisement, slow-motion reply or a highlight. Event classification will be done using a number of techniques such as motion analysis, face recognition, the sequence of event pattern recognition, etc.

6. CONCLUSION.

In this paper, we have discussed various approaches for event resolution in sports video. The techniques we have discussed will have many future applications such as automated video archiving, interesting event highlight generation, query-based sports video broadcasting, player analysis, and training. The approaches we have discussed has certain advantages and disadvantages, and the goal of each approach is to is to provide excitement clips or the necessary highlights to the viewers. By analyzing the research work done in this area, we propose a novel approach based on a multidimensional model and a new taxonomy to cricket[21]. Our approach is proven to be accurate on recognizing the primary events such as delivery starting point detection and endpoint detection which will enable to identify a delivery. The next step would be to clarify the type of delivery such as a six, four or out, etc.

ACKNOWLEDGMENT

This research received financial support from the University Grants Commission (UGC) of Sri Lanka.

7. REFERENCES

[1] C. Zhang, B. Price, S. Cohen, and R. Yang, "High-quality stereo video matching via user interaction and space-time propagation," *Proc. - 2013 Int. Conf. 3D Vision, 3DV 2013*, pp. 71–78, 2013.

[2] A. Hanjalic, "Adaptive extraction of highlights from a sport video based on excitement modeling," *IEEE Trans. Multimed.*, vol. 7, no. 6, pp. 1114–1122, 2005.

[3] C. Xiao, L. Wang, Z. Jie, and T. Chen, "A Multi-level Intelligent Selective Encryption Control Model for Multimedia Big Data Security in Sensing System with Resource Constraints," *2016 Ieee 3Rd Int. Conf. Cyber Secur. Cloud Comput.*, pp. 148–153, 2016.

[4] S. C. Chen, R. Jain, Y. Tian, and H. Wang, "Guest Editorial Multimedia: The Biggest Big Data," *IEEE Trans. Multimed.*, vol. 17, no. 9, pp. 1401–1403, 2015.

[5] N. Adami, S. Benini, and R. Leonardi, "An overview of video shot clustering and summarization techniques for mobile applications," *Proc. 2nd Int. Conf. Mob. Multimed. Commun. - MobiMedia '06*, p. 1, 2006.

[6] K. Kumar and S. Prasad, "Sports Video Summarization using Priority Curve Algorithm," *Int. J. Comput. Sci. Eng.*, vol. 02, no. 09, pp. 2996–3002, 2010.

[7] D. Zhong, "Segmentation , Index and Summarization of

Digital Video Content," pp. 1–162, 2001.
[8] S. Abburu, "Semantic Segmentation and Event Detection in Sports Video using Rule Based Approach," *Int. J. Comput. Sci. …*, vol. 10, no. 10, pp. 35–40, 2010.
[9] H. Sundaram, "Segmentation, structure detection and summarization of multimedia sequences," *Structure*, p. 331, 2002.
[10] H. Shao, Y. Qu, and W. Cui, "Shot Boundary Detection Algorithm Based on HSV Histogram and HOG Feature," *5th Int. Conf. Adv. Eng. Mater. Technol.*, no. Aemt, pp. 951–957, 2015.
[11] M. Tavassolipour, M. Karimian, and S. Kasaei, "Event Detection and Summarization in Soccer Videos Using Bayesian Network and Copula," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 2, pp. 291–304, Feb. 2014.
[12] Z. Zhao, S. Jiang, Q. Huang, and Q. Ye, "Highlight Summarization in Soccer Video based on Goalmouth Detection," pp. 1613–1616, 2006.
[13] H. M. Zawbaa, N. El-bendary, A. E. Hassanien, and T. Kim, "Event Detection Based Approach for Soccer Video Summarization Using Machine learning," *Int. J. Multimed. Ubiquitous Eng.*, vol. 7, pp. 63–80, 2012.
[14] H. Denman and A. Kokaram, "Content Based Analysis for Video from Snooker Broadcasts . Content Based analysis for video from Snooker Broadcasts," no. February, 2015.
[15] S. Liao, Y. Wang, and Y. Xin, "Research on Scoreboard Detection and Localization in Basketball Video," vol. 10, no. 11, pp. 57–68, 2015.
[16] C. Y. Chen, J. C. Wang, J. F. Wang, and Y. H. Hu, "Event-based segmentation of sports video using motion entropy," *Proc. - 9th IEEE Int. Symp. Multimedia, ISM 2007*, pp. 107–111, 2007.
[17] W. Lu, J. Ting, K. P. Murphy, and J. J. Little, "Identifying Players in Broadcast Sports Videos using Conditional Random Fields_Unknown_2010_Lu et al.pdf," no. 2, 2010.
[18] F. Yoshikawa, T. Kobayashi, K. Watanabe, and N. Otsu, "Automated Service Scene Detection for Badminton Game Analysis Using CHLAC and MRA," *Int. J. Comput. Electr. Autom. Cntrol Inf. Eng.*, vol. 4, no. 62, pp. 331–334, 2010.
[19] M. C. Tien, Y. T. Wang, C. W. Chou, K. Y. Hsieh, W. T. Chu, and J. L. Wu, "Event detection in tennis matches based on video data mining," *2008 IEEE Int. Conf. Multimed. Expo, ICME 2008 - Proc.*, no. May, pp. 1477–1480, 2008.
[20] P. K. Odedara and J. Vala, "Cricket Video Summarization Techniques," vol. 45, no. 208, pp. 87–89, 2015.
[21] S. C. Premaratne and K. L. Jayaratne, "Structural Approach for Event Resolution in Cricket Videos," *Proc. Int. Conf. Video Image Process. - ICVIP 2017*, pp. 161–166, 2017.