Volume 3, No. 8, August 2012
Journal of Global Research in Computer Science
ISSN-2229-371X
RESEARCH PAPER
Available Online at www.jgrcs.info

EVENT DETECTION IN CRICKET VIDEO BASED ON VISUAL AND ACOUSTIC FEATURES

Vijayakumar.V [*1], Nedunchezhian.R [2]

[*1] Research and Development Centre, Bharathiar University, Coimbatore, Tamil Nadu, India
& Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India
veluvijay20@gmail.com
[2]Department of Information Technology, Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India
rajuchezhian@gmail.com

Abstract: As digital video becomes more pervasive, efficient way of extracting the information from the video becomes an increasingly important. Video itself contains huge amount of data and complexity that makes the analysis very difficult. In this paper, we presented a video and audio features based event detection approach shown to be effective when applied to the cricket sports video. The advantage of this approach is the ability to recognize events that indicate high level of audio response and players crowd which can be correlated to key events. The proposed event detection has two steps. First, audio and visual features are extracted. Next, by defining a set of heuristic rules, the important semantic events such as wicket fall, score events are detected. The proposed framework has been tested with cricket video downloaded from the internet broadcast videos.

Keywords- Feature extraction, Heuristic model, Event detection

INTRODUCTION

Modern developments in digital video compression technologies have paved the way for extensive archiving of multimedia content. Sports video analysis has been an attractive research area in the multimedia community [2][4]. Event detection is one of the most fundamental components for sports video information and mining system because of the viewers prefer to watch only the interesting or exciting portions of the videos[5][6]. Therefore, automatic detection of interesting events in sports videos has become a fundamental problem, and is receiving increased attention from researchers in the field. Sports video is composed of temporally synchronized multimodal streams such as visual, auditory and text streams. Most of the approaches proposed recently extract some features from the information in the above mentioned three streams. Cricket videos are mainly used for entertainment purposes; the video content is usually recorded or edited using the style formats. Compared to other sports, cricket video analysis is far more challenging because of the complexities of game in itself such as various formats of matches, dynamic playing conditions (field area, pitches etc), day and day/night matches (cause illumination related problem) and duration[4].

Generally, the video content analysis is divided into two main parts called, low level feature extraction and high level concept detection. Low level feature extraction deals with video properties such as colour scheme, texture, motion, video text and audio properties such as MFCC, pitch, and zero crossing rate. High level concept detection targets to mine and describe concepts, events, scenes, locations and objects of videos such as player trajectories, crowds, audience cheering, wicket fall or score events, etc., The outline of video content analysis as shown in the Table. I.

Table I. Video Content Analysis

High level	Event detection / event classification / summarization		
Middle level	Extracting color features (color histogram, dominant color ratio, field color and player skin color)	Extracting motion vectors	Extracting audio features (MFCC,ZCR etc)
Low level	Visual features	Motion features	Audio features

The basic procedure for the existing event detection methods can be generally divided into two main steps namely, 1) Generating video content representation, where the video properties are extracted from raw sequence and 2) Event detection. This paper addresses two main problems of sports video processing such as, video feature extraction and event detection. We proposed a semantic event detection approach in a cricket video for minimizing the computation time.

The remainder of the paper is organized as follows. Related work section discusses the detailed review of literature. The proposed method section presents in detail of our framework. Experimental results section shows result of our proposed method and conclude section discuss about the benefits and future work.

RELATED WORK

Video event detection is a difficult problem in video content analysis and there are some challenges such as, bridging the semantic gap between low-level features and high-level semantic concepts.

The most popular features used by researchers in the sports video analysis are visual features such as color, motion and audio features can also contribute significantly in video event detection. Visual feature processing is executed among all representative frames to extract semantically related visual cues. Audio signals are also rich source of information in videos [1][11]. This is a challenging task, and many methods

© JGRCS 2010, All Rights Reserved

rely on detecting audio events in the video where audience cheering provides the most convincing cue for event detection. In sports videos, some special audio events, e.g., audience applause and a referee's whistle, will help to acquire some semantic cues. Generally, audience applause occurs when exciting events happen, e.g., shooting and/or a goal, and a referee's whistle may imply an interruption or another special event [10][11].

Dongqing Zhang and Dan Ellis proposed a method to detect the sound events in a basketball game with focusing on detecting cheering sound. Mel-frequency cepstral coefficient features are used to identify the cheering sounds from speeches and other confusing sounds. The MFCC features are fed into a neural network and classified into three classes such as cheering, speech and others [1]. Xingquan Zhu, Xindong Wu, Ahmed K. Elmagarmid, Zhe Feng, and Lide Wu calculated the Non-zero Pitch Ratio for detecting a cheering segment in a basket ball video database [10]. Sunitha Abburu recognized the super imposed text to detect the boundaries of the video segments and major cricket events. They proposed a hybrid multilayered approach for semantic segmentation of cricket video and automatic detection of major cricket events like score and dismissal using the rule based approach. The whole process has been done in two layers. Top layer was used the super imposed text as a primary source for the semantic extraction and the extracted semantics are passed on to the bottom layer which are combined with the rule based approach to detect the boundaries of the video segments and the major events of the cricket game [3].

N. Harikrishna, Sanjeev Satheesh, S. Dinesh Sriram, K.S. Easwarakumar presented a method to segment a cricket video into shots and identified the visual content in them. Using sequential pattern mining and support vector machine, the sequences of shots were classified into four events, namely run, four, six and out. The cricket video is then summarized based on user-supplied parameters [4]. Min Xu, Ling-Yu Duan, Changsheng Xu, Mohan Kankanhalli, Qi Tian presented a basketball event detection using multiple modalities such as visual and auditory midlevel features [6]. Kasun Wickramaratna, Min Chen, Shu-Ching Chen, Mei-Ling Shyu proposed a neural network based framework for semantic event detection in soccer videos using multi-modal processing [8].

Pradeep K. Atrey, Namunu C. Maddage and Mohan S. Kankanhalli proposed a hierarchical classification approach to assign a label to an event in a given audio frame. An audio frame to be a fixed size audio segment was extracted from the continuous audio stream. A Gaussian Mixture Model classifier was employed to classify an input audio frame, at the top level - into foreground or background, at second level - into vocal or non-vocal, and at third level - into excited events (e.g. shout/cry, door knock, running footsteps) or normal events (talk, walking footsteps) [5].

Sk.Abdul Nabi, Shaik Rasool,.P. Premchand proposed decision trees for detection and extraction of video and audio events. It consisted of three major phases namely video processing, data pre-processing and data mining. In the first phase the boundary of video events were detected and also observed some important features of video events. In the second phase, the data were cleaned i.e. it eliminated the noise data and reduced the irrelevant data from the original feature and produced as training data. In the final phase, it mined the video events from training data and finally it produced pure and relevant events [7].
Ming-Chun Tien, Yi-Tang Wang, Chen-Wei Chou,uei-Yi Hsieh, Wei-Ta Chu, and Ja-Ling Wu1, extracted real-world audiovisual features, and detect tennis events on the basis of data mining techniques[9]. P.Thirumurugan, S.Hasan Hussain detected the soccer goal event based on the combination of distance based and rule-based data mining techniques [12].

Player detection, recognition and tracking are known as important issue in the field of cricket video analysis. Moving object detection algorithms to track a player analyze the player's trajectory, and infer the actions and events associated with the player [4]. It consists of two stages: (1) an iterative motion layer detection step that is performed in individual frames, and (2) a temporal tracking process linking layers across frames and classified the linked layers into foreground moving objects vs. background[4][7].

Among all these techniques, there are primary three sources of information used as the features for event detection. The visual features include color, texture and shape. The audio features considered include timbre, rhythm and pitch. The motion features considered include pixel change ratio. In this paper, we have identified the useful events based on the visual and audio features.

PROPOSED WORK

In this section, we presented a method to facilitate video event recognition based on audio and video features and the block diagram of the proposed system is illustrated in Fig. 1. In the proposed work, visual, audio features and heuristic rule information are integrated to mine the types of events in the cricket video. It contains two basic stages namely video pre-processing which extract audio and visual feature and video event detection based on extracted features.

Video Pre-processing:

The main functionality of the video pre-processing phase is to extract visual and audio features from video streams. Initially, the cricket video is partitioned into a number of smaller segments such as key-frames based on histogram color feature metric. Later, visual, motion and audio features are extracted from the video segments. Visual feature processing is executed among all representative frames (key frames) to extract semantically related visual cues. There are some specific colors in the cricket video such as grass and non-grass (player group, audience crowd). This information is detected using the color distribution on the HSV color coordinate. The distribution of these colors contains a great amount of information for frame type classification.

© JGRCS 2010, All Rights Reserved

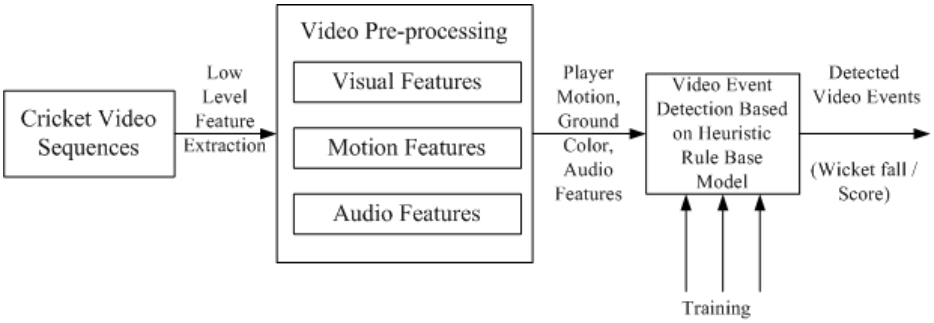

Figure 1. Block Diagram of Proposed System

Pitch is the fundamental frequency that reveals harmonic properties of audio and is an important parameter in the analysis and synthesis of speech signals. To detect audience cheering in the cricket sports video, the pitch of audio signal is used. In comparison with voice and music, the pitch value of audience applause is very small. In most cases, this value in sports videos is zero because, when cheering happens, the audio signal exhibits a constant high value noise that likely drowns out other audio signals, e.g., the voice of the anchorperson or the music.

The moving object is detected based on connected components in the key-frame image. The image region properties are measured for extracting moving object. Median filtering is a nonlinear operation used in image processing to reduce noise. It is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges. The small object is removed in the extracted binary image.

Event Detection

In the second stage, the events can be modeled either statistically or syntactically. After the video and audio features have been extracted, the event detection strategy is applied to detect the event information within video.

The heuristic rule based approach is an experience-based technique for problem solving, learning, and discovery. Where an exhaustive search is impractical, heuristic methods are used to speed up the process of finding a satisfactory solution. It pertains to the process of gaining knowledge or some desired result by intelligent guesswork rather than by following some pre-established formula.

This method was adopted to recognize the events in the cricket video based on the visual and audio-based patterns. For example, stadium supporters react to different stimuli during a match, such as a wicket fall and score. Hence, an increase in crowd response is an also important indicator for the occurrence of a key event. The heuristic rule-based approach uses domain knowledge to define rules to achieve semantic event detection. Given the extracted features, heuristic rules were applied on the features to perform further analysis. The interesting cricket events such as Wicket fall, Score were detected using heuristic rule base approach. To detect these events, some heuristic decision rules were summarized based on the percentages of pixels in the image and pitch of value of the audio respectively.

EXPERIMENTAL RESULTS AND DISCUSSION

Cricket event detection is used as the experiment test bed to validate our proposed framework. In our empirical study, five videos were collected from different Internet broadcasters. According to these heuristic rules, some interesting events were detected such as wicket fall with strong semantic meaning in cricket video. The video data (3 video) are randomly selected for training and the rest (2 video) are adopted for testing. The benefit of the proposed framework should be noted that the feature set has been reduced which brings operational benefits such as less storage requirement for multimedia database, less training time, less testing time, simplified model.

There are two parameters 'recall' and 'precision' are commonly used to evaluate the effectiveness of information retrieval techniques. The same properties can be used to the analysis the video event detection.

Recall (V) is the ratio of the number of event detected correctly over the actual number of events in a given video clip.

$$V = \frac{C}{C + M}$$

Precision (P) is the ratio of the number of events detected correctly over the total number of events detected (correctly or incorrectly).

$$P = \frac{C}{C + F}$$

F1 is a combined measure that results in high value if, and only if, both precision and recall result in high values:

$$F1 = \frac{2 * P * V}{P + V}$$

where C, the number of correctly detected events ("correct hits"), M, the number of not detected events ("missed hits") and F, the number of falsely detected events ("false hits"). All of these measures are mathematical measures, i.e. they deliver values in between 0 and 1. The basic rule is the higher the value, the better performs the algorithm.

Table II summarizes the precision and recall for wicket fall event detection. Table III evaluates the results of the evaluation of scoring event detection. The recall is less good because of some wrong detection. This has badly affected the precision in the detection of events.

Table II. The Results of Wicket fall Event Detection

S.No	*Video*	*Total Number of Events*	*Detected Events(C)*	*Missed Alarms (M)*	*False Alarms (F)*	*Recall (V)*	*Precision (P)*	*F1*

© JGRCS 2010, All Rights Reserved

1	Video- 1	15	9	3	4	0.75	0.692	0.72
2	Video -2	21	13	5	7	0.722	0.65	0.684

Table III. The Results of Scoring Event Detection

S.No	*Video*	*Total Number of Events*	*Detected Events(C)*	*Missed Alarms (M)*	*False Alarms (F)*	*Recall (V)*	*Precision (P)*	*F1*
1	Video- 1	17	14	3	2	0.824	0.875	0.849
2	Video -2	20	16	5	3	0.762	0.842	0.8

CONCLUSION

Event detection is a challenging research topic with great application potential. Audio and visual features from the broadcast are extracted to bridge the gap between the low-level features and the high-level concepts. In this paper, we the audio, visual and motion features are used to detect the important events. In particular, we would like to apply our framework to event detection in other domains. In this paper, we concentrate on wicket fall and score event in cricket video. We would like to apply our framework to the detection of other interesting events such as nice play, and appeal. The technique has a lot of domain applications including video highlight extraction, content summarization, and many more. However, there is still much scope for improvement.

ACKNOWLEDGMENT

The authors would like to thank the Management, Director, and Principal of Sri Ramakrishna Engineering College for providing laboratory resources and valuable supports.

REFERENCES

[1] Dongqing Zhang, Dan Ellis, "Detecting sound events in basketball video archive," Speech & Audio Processing class project report, 2001.

[2] J.R.Wang and N. Parameswaran, "Survey of sports video analysis: research issues and applications," in Proc. Pan-Sydney area workshop on visual information processing, Darlinghurst, Australia, June 2004, p. 8790.

[3] Sunitha Abburu, "Semantic Segmentation and Event Detection in Sports Video using Rule Based Approach," International Journal of Computer Science and Network Security, Vol.10 No.10, October 2010, pp. 35-41.

[4] N. Harikrishna, Sanjeev Satheesh, S. Dinesh Sriram, K.S. Easwarakumar, "Temporal Classification of Events in Cricket Videos," National Conference on Communications, 2011, pp. 1 – 5.

[5] Pradeep K. Atrey, Namunu C. Maddage and Mohan S. Kankanhalli, "Audio Based Event Detection For Multimedia Surveillance," in proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2006), 2006, pp. 813-816.

[6] Min Xu, Ling-Yu Duan, Changsheng Xu, Mohan Kankanhalli, Qi Tian, "Event Detection in Basketball Video Using Multiple Modalities," in Proceedings of ICICS-PCM-2003, Singapore.

[7] Sk.Abdul Nabi, Shaik Rasool,.P. Premchand, "Detection and Extraction of Videos using Decision Trees," (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 12, 2011, pp.147-151.

[8] Kasun Wickramaratna, Min Chen, Shu-Ching Chen, Mei-Ling Shyu, "Neural Network Based Framework For Goal Event Detection In Soccer Videos," in proceedings of Seventh IEEE International Symposium on multimedia 2005, pp 21-28.

[9] Ming-Chun Tien, Yi-Tang Wang, Chen-Wei Chou,uei-Yi Hsieh, Wei-Ta Chu, and Ja-Ling Wu1,,"Event Detection in Tennis Matches Based on Video Data Mining," in proceedings of ICME 2008, pp.1477-1480.

[10] Xingquan Zhu, Xindong Wu, Ahmed K. Elmagarmid, Zhe Feng, and Lide Wu, "Video Data Mining: Semantic Indexing and Event Detection from the Association Perspective," IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 5, May 2005, pp.1-18.

[11] Mark Baillie and Joemon M. Jose, "Audio-based Event Detection for Sports Video," in Proceedings of the 2nd international conference on Image and video retrieval, 2003, pp. 300-309.

[12] P.Thirumurugan, S.Hasan Hussain, "Event Detection in Videos Using Data Mining Techniques," International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012, pp.3473-3475.

© JGRCS 2010, All Rights Reserved