Review: Highlight Extraction and Classification of Sports Video

[1]Vishwa Parmar, [2]Devangi Kotak
[1]Student, [2]Asst. Prof.
[1,2]Computer Engineering Department
[1,2]V.V.P. Engineering College, Rajkot, India

***Abstract* – In this paper we review all the methods which can be helpful to extract highlights from sports video. There are several methods which work on replay detection, caption detection, event detection, logo detection etc. All methods have theirs pros and cons. We have briefly discussed about this all methods and try to analyze best matching method according to our application. Normally highlight extraction is giving good results but to reduce the length of extracted video we can use classification in it which can improve the outcome.**

***Index Terms* - sports highlights, event detection, feature extraction, video summarization, classification**

I. INTRODUCTION

In recent years, sports videos have drawn increasing attention in automatic video analysis, since they are globally widespread and attract large audiences. In addition, as more convenient digital equipment emerges, it is much easier to record or further archive digital video data for general users at home. The distribution of sports video over the Internet further increases necessities for automatic video analysis.

Normally, sports videos are rather long, consisting of portions which are interesting or exciting and portions which are boring, bland, and likely "a waste of the viewer's valuable time."

Cricket is the most popular sport after soccer. It is the most favourite sport of Indians. Research work reported less in cricket comparison to other sports. The reason behind that is cricket is lengthy sport comparison to soccer, basketball, tennis, badminton. Addition to that it is also complex game and has various formats like ODI, test cricket and recently added T20.

The sports video analysis techniques adopted so far can be broadly classified as event-based or excitement model-based. The event-based video abstraction techniques are mostly based on feature-based event detection. Excitement-model based highlight generation techniques look for excitement through features such as motion activity, cut density, audio energy, colour tracking.

Event-based highlights use more semantically meaningful content than the excitement-based highlights and its success depends upon the richness of the semantic concepts. Excitement-model based techniques are generic in nature and require less domain specific knowledge, but their performance changes from game to game.

Classification also used with the extraction. It is used to classify the frames. Frames may be classified into field-view, non-field view, close-up, crowd, etc. Classification emphasizes the extraction video. It also reduces the size of the video and removes the unrelated frames.

II. LITERATURE SURVEY

Various extraction methods are: play and break event, object detection, crowd cheer (auditory), keyword spotting detection, replay detection.

The aim is to analyse cricket pictures that are to appear on TV and to extract captions [1]. It automatically generates highlights of game sequences so that selections of events can be located and played back. Excitements levels are gathered from the audio energy and short time zero crossing.Consider fig. 1. Caption recognition is carried out using sum of absolute difference based caption recognition model. Method reduces manual processing, enables the generation of personalized highlight and also can be used for Content Based Video Retrieval. The approach seems effective and around 80-85% accurate in practical tests.It is necessary to give complete set of characters of the channel and prior knowledge of the caption location is required.

A novel hierarchical framework and effective algorithms for cricket event detection and classification is given [2]. The proposed scheme performs top down video event detection and classification using hierarchical tree which avoids shot detection and clustering. In the hierarchy, at level-1, audio features are used to extract excitement clips from the cricket video. At level-2, excitement clips are classified into real-time and replay segments. At level-3, the segments are partitioned into field view and non-field view based on dominant grass colour ratio. At level- 4a, field view is classified into pitch-view, long-view, and boundary view using motion-mask. At level-4b, non-field view is classified into close-up and crowd using edge density feature. At level-5a, close-ups are classified into batsman, bowler/fielder, umpire using jersey colour feature. At level-5b, crowd segment is classified into spectator and players' gathering using colour feature. Consider fig. 2.

Fig. 1. Caption style in cricket video[1]

Fig. 2. Replay detction frames[2]

The proposed system in [3] detects exciting clips based on audio features and then classifies the individual scenes within the clip into events such as replay, player, referee, spectator, and players gathering. A probabilistic Bayesian belief network based on observed events is used to assign semantic concept-labels to the exciting clips, such as goals, saves, yellow-cards, red-cards, and kicks in soccer video sequences. The labelled clips are selected according to their degree of importance to include in the highlights. The proposed system selects 100% premium concept which gives the good result. The proposed approach not only be extended to other kind of sports, but also to other types of videos such as news, movies for video summarization applications. Close up and crowd frames are shown in fig. 3.

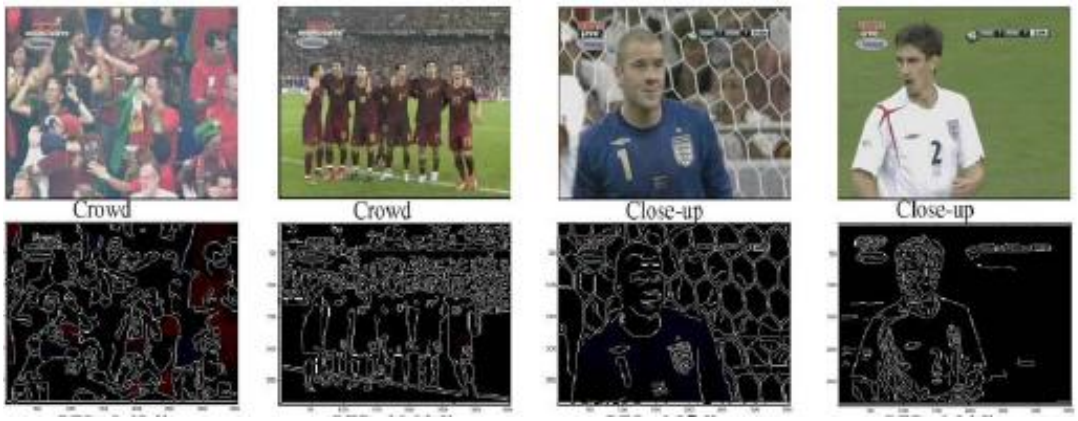

Fig. 3. Close up and crowd frames.[3]

A generic method for sports video highlight selection is presented in [4]. Processing begins where the video is divided into short segments and several multi-modal features are extracted from each video segment. Excitability is computed based on the likelihood of the features lying in certain regions of their probability density functions that are exciting and rare. The proposed measure is used to rank order the partitioned segment stream to compress the overall video sequence and produce a contiguous set of highlights. The video is first segmented into small blocks for feature extraction. Several features (scalar parameters) are extracted from each segment that is modelled to be generally proportional to the excitement level of the given segment. The multimodal events/features used for excitability measure: (1) slow motion replay, (2) camera motion activity, (3) scene cut density, (4) commentators' speech in high and (5) low excitement levels, and (6) audio energy.

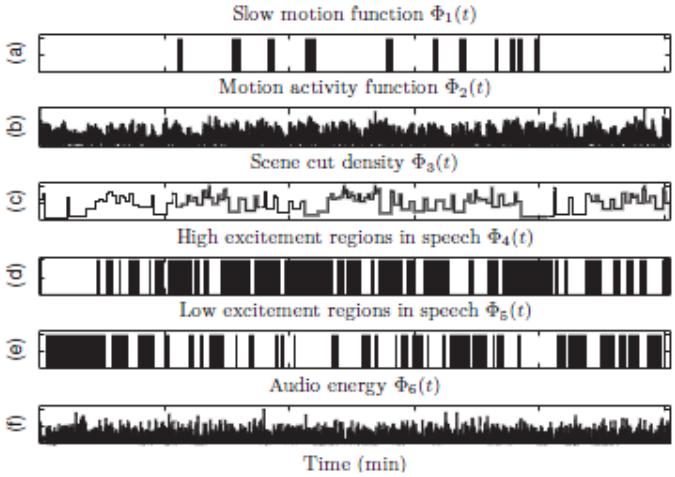

Fig. 4. A time-line view of the detected events/feature functions [4]

In [5] the efficacious method for event detection in soccer game broadcasted video and comprehending aspects which have been proposed to detect event and classify them in order to generate highlights is proposed. The aim is to propose a method to minimize the amount of manual supervision in considering which set of features are related to an event and to provide a flexible system being able to tackle sequences exclude a regular pattern. Logo transition frames are shown in fig. 5. All frames from the video are analysed for calculating the threshold value. Threshold value v is the smallest detectable sensation on calculating the brightness of the frame. This value v is the primary measure to evaluate whether the frame is logo frame or any normal frame. It is difficult to extract distinguish motion features to represent slow-motion pattern. Analysis of intra-shot and inter-shot is not used.

Fig. 5. Logo transition frames[5]

An algorithm to detect semantic concepts from cricket video is proposed in [6]. The proposed scheme works in two parts. In first part a top-down event detection and classification is performed using hierarchical tree. In second part, higher level concept is identified by applying A-Priori algorithm. High level concept mining from given video is two levels hierarchical process. At first level, meaningful events are identified associated with video using low level features. And at second level, higher level concepts are recognized based on previous level results. For cricket, events like, real time video, reply, pitch view, field view, close up classification, crowd classification are extracted in first level. Collections of such events are used later on for extracting the concepts like wicket fall (balled out, catch out, run out, stumped out etc.), boundary (fours and sixes), run milestones (Half century or century) etc. There is always a semantic gap between low level feature and high level concepts which needs to be filled up. The proposed approach is tested on T20 matches as they offer lots of events in short duration as compared to ODI's.

A novel semantics-based content analysis system for reliable media highlight extraction using Dynamic Bayesian Network (DBN) is proposed in [7]. It extracts the low-level evidences and then converts the input video to high-level semantic meaning. Specific domains contain rich spatial and temporal transitional structures that help the transformation process. A robust audio-visual low-level evidence extraction scheme is introduced. Based on DBNs, soccer events such as goal event, corner kick, penalty kick event, and card event can be found. Given a video in specific domain, the proposed system can extract the low-level evidence and interpret the input video in terms of high-level semantic. The low features like close-up view, camera motion, audience region, audio, gate, replay, board, referee. Training can be categorized into two kinds: qualitative (structural training) and quantitative training (parameter training). Qualitative training concerns the network structure of the model and quantitative training determines the specific conditional probabilities.

In [8], a reliable logo and replay detecting approach is proposed. It contains two main stages: first, a logo transition template is unsupervised learned, a key frame (K-frame) and a set of pixels that describes logo object (logo pixels, L-pixels) accurately are also extracted; second, the learned information are used jointly to detect logos and replays in the video. A logo transition usually contains 10-30 frames, describes a flying or varying object(s). In this paper, a novel unsupervised learning and detection approach for logos and replays cross different sports videos is proposed. It first learns a logo template based on the motion characteristics; then learns a colour representation (Lpixels) of a key frame of the logo. In the detection stage, the colour representation is used to filter out frames that not lie in a logo; and the logo template is used to verify the true logos .Finally pair these found logos to get the replay clips. It uses a sequence template to model the transition of the logo object. Motion feature is used. The whole detection procedure is totally automatically. The approach is not applicable in case when multiple types of logo object occur in a same video.

In [9], a novel approach for detecting highlights in sports videos is proposed. The videos are temporally decomposed into a series of events based on an unsupervised event discovery and detection framework. The framework solely depends on easy-to-extract low-level visual features such as colour histogram (CH) or histogram of oriented gradients (HOG), which can potentially be generalized to different sports. The unigram and bigram statistics of the detected events are then used to provide a compact representation of the video. The effectiveness of the proposed representation is demonstrated on cricket video classification: Highlight *vs.* Non-Highlight for individual video clips (7000 training and 7000 test instances). Fig. 6 shows the overview of proposed approach. Feature extraction is used to pre-process all videos; event discovery and model refinement stages construct the event vocabulary; event detection and video clip representation transcribe the video; and finally feed into the training of the highlights detection classifier. A low equal error rate of 12.1% using event statistics based on CH and HOG features is achieved.

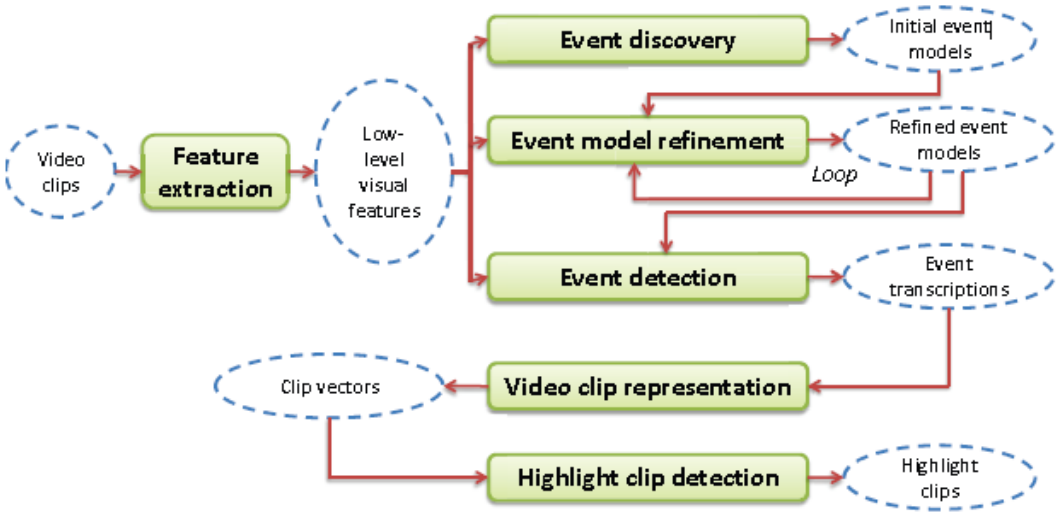

Fig. 6 overview diagram of proposed framework.[9]

A novel hierarchical framework for soccer (football) video classification is proposed in [10]. The proposed scheme performs a top-down video scene classification which avoids shot clustering. This improves the classification accuracy and also maintains the temporal order of shots. The hierarchy, at level-I, audio features are used, to extract potentially interesting clips from the video. At level2, it is classified into field view and non-field view using feature of dominant grass colour ratio. At level-3a, it classifies field view into three kinds of views using motion-mask. At level-3b, it classifies non-field view into close-up and crowd using skin colour information. At level-4,it classifies close-ups into the four frequently occurring classes such as player of team-A, player of team-B, goalkeeper of team-A, goalkeeper of team-B using jersey colour information. The proposed hierarchical semantic framework for event classification can be readily generalized to other sports domains as well as other types of video.

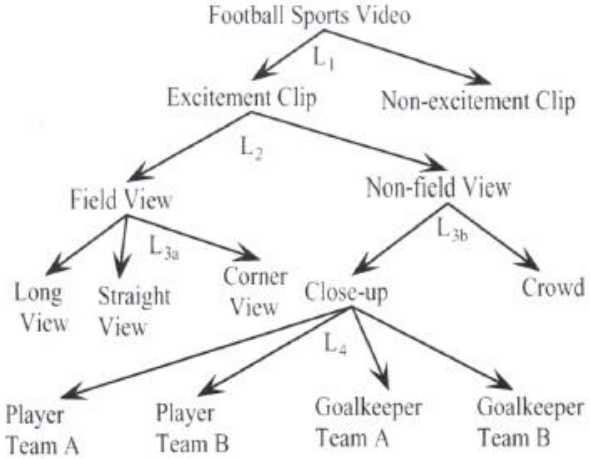

Fig. 7 tree diagram for hierarchical framework.[10]

In [11], a visual content based algorithms to automate the extraction of video frames with the cricket pitch in focus. As a pre-processing step, first select a subset of frames with a view of the cricket field, of which the cricket pitch forms a part. This filtering process reduces the search space by eliminating frames that contain a view of the audience, close-up shots of specific players, advertisements, etc. The subset of frames containing the cricket field is then subject to statistical modelling of the grayscale (brightness) histogram (SMoG). Since SMoG does not utilize colour or domain specific information such as the region in the frame where the pitch is expected to be located, an alternative algorithm: component quantization based region of interest extraction (CQRE) for the extraction of pitch frames is proposed. The SMoG-CQRE combination for pitch frame classification yields an average accuracy of 98:6% in the best case (a high resolution video with good contrast) and an average accuracy of 87:9% in the worst case (a low resolution video with poor contrast). Since, the extraction of pitch frames forms the first step in analysing the important events in a match; a post processing step, viz., an algorithm to detect players in the extracted pitch frames is proposed. The method presented in this paper to extract frames with a view of the cricket pitch is carried out in three phases. The first phase comprises a pre-processing step, which acts as a coarse filter. The field-frames include both pitch frames and extraneous information such as a view of the field near the boundary. In the second phase of the method, field-frames are further processed to separate frames with a view of the pitch ("pitch frames") from non-pitch frames using the following pitch detection algorithms: SMoG, CQRE and a combination of SMoG and CQRE methods. This second phase forms the crux of the proposed work. Finally, the third phase comprises a post-processing step, which uses the pitch frames to localize key players in the field of view. Fig. 8 shows the histograms difference of pitch and non-pitch image.

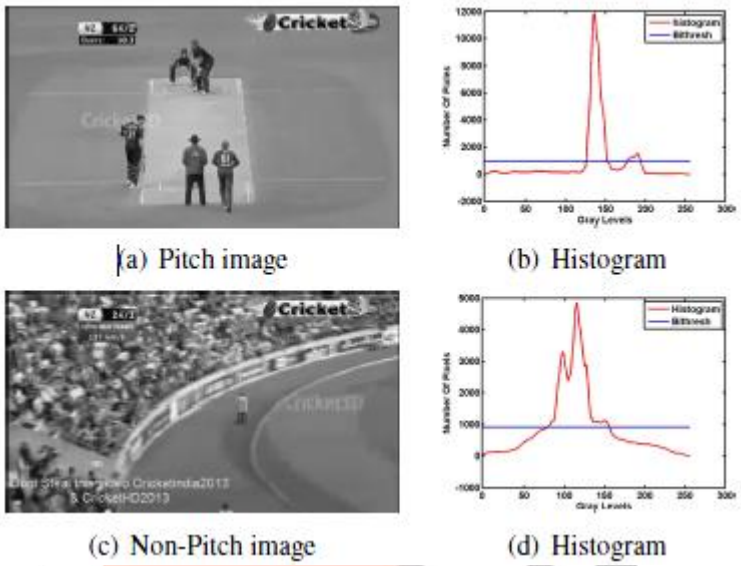

Fig. 8 pitch and non-pitch image with histograms[11]

In [12] a composite feature combining Optical flow analysis along with camera view analysis to model the type of shots played is presented. The work first presents an improved camera shot analysis based on learning parameters from a small supervision set. This splits the broadcast video into shots which are combined into balls and, the segment where the batsman is playing the stroke is identified. The approach works at three levels: a) classifying the type of view b) detecting the shots where effective play happens, and c) identifying the direction in which stroke is played. Work can be combined with these to develop a fully automated cricket commentary system, which can provide a detailed analysis on a ball by-ball basis. The work has potential to be integrated with other requirements - such as video summarization for mobile devices, or identifying commercial break insertion points. The general approach of combining optical flow with other features already available may also be useful in other spatially coherent games such as baseball. After that optical flow analysis is used to determine the direction of the stroke with an accuracy of 80 percent. Fig 9 and 10 shows the cut and change in frames.

Fig. 9. Cut between two frames [12]

Fig. 10 Gradual/fade change in frames [12]

In [13] a novel approach towards customized and automated generation of sports highlights from its extracted events and semantic concepts. A recorded sports video is first divided into slots, based on the game progress and for each slot, an importance-based concept and event selection is proposed to include those in the highlights. Our approach uses two levels of abstractions- one at the micro level, called events and the other at the macro level, called concepts. This approach can not only be extended to other sports, but also to other type of videos such as news, movies, etc. for video summarization applications.

In [14] a new framework meant for replay frames detection in cricket video is proposed. Framework is based on a block of score bar which is present in non-replay frame and not present in replay frame. Correlation is a signal matching technique used for obtaining similarity between two signals. Support Vector Machine (SVM) is a supervised machine learning technique meant for binary classification.. A training set of thousand frames containing five hundred non replay frames and five hundred replay frames used for training SVM. Feature vectors are calculated using correlation coefficient between templates and the training frames in the training set. Results have achieved an average recall of 96% and precision of 99%.

III. CONCLUSION

As reviewed, there are different methods of highlight extraction in sports videos and many techniques in classification of the extracted frames. Each methods and techniques have their merits and demerits. From these methods we can compare best among these methods according to our application. We can also combine two methods for better extraction. From above discussion we can conclude that while using classification with highlight generation video can be reduced and frames which are not related like group gathering, cheering of crowd can be excluded from highlight. We can further work on more sports video or any other genre of video.

REFERENCES

[1] Kolekar, Maheshkumar H., and S. Sengupta. "Caption Content Analysis Based Automated Cricket Highlight Generation." *National Communications Conference (NCC), Mumbai*. 2008.

[2] Kolekar, Maheshkumar H., KannappanPalaniappan, and SomnathSengupta. "Semantic event detection and classification in cricket video sequence."*Computer Vision, Graphics & Image Processing, 2008. ICVGIP'08. Sixth Indian Conference on*. IEEE, 2008.

[3] Kolekar, Maheshkumar H., and SomnathSengupta. "Bayesian Network-Based Customized Highlight Generation for Broadcast Soccer Videos."

[4] Hasan, Taufiq, et al. "A multi-modal highlight extraction scheme for sports videos using an information-theoretic excitability measure." *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012.

[5] Arbat, Shivani, Shashi Kumari Sinha, and Beena Khade Shikha. "Event-Detection-In-Broadcast-Soccer-Video-By-Detecting-Replays." *international journal of scientific & technology research* 3.5 (2014): 282-285.

[6] Goyani, Mahesh, et al. "Wicket fall concept mining from cricket video using a-priori algorithm." *Proc Int J Multimedia Appl (IJMA)* 3.1 (2011).

[7] Chao, Chung-Yuan, Huang-Chia Shih, and Chung-Lin Huang. "Semantics-based highlight extraction of soccer program using DBN." *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*. Vol. 2. IEEE, 2005.

[8] Huang, Qiao, et al. "A reliable logo and replay detector for sports video."*Multimedia and Expo, 2007 IEEE International Conference on*. IEEE, 2007.

[9] Tang, Hao, et al. "Detecting highlights in sports videos: Cricket as a test case."*Multimedia and Expo (ICME), 2011 IEEE International Conference on*. IEEE, 2011.

[10] Kolekar, M. H., and K. Palaniappan. "A hierarchical framework for semantic scene classification in soccer sports video." *TENCON 2008-2008 IEEE Region 10 Conference*. IEEE, 2008.

[11] Jayanth, Sandesh Bananki, and GowriSrinivasa. "Automated classification of cricket pitch frames in cricket video." *Electronic Letters on Computer Vision and Image Analysis* 13.1 (2014).

[12] Kumar, Ashok, Javesh Garg, and AmitabhaMukerjee. "Cricket activity detection." *Image Processing, Applications and Systems Conference (IPAS), 2014 First International*. IEEE, 2014.

[13] Kolekar, Maheshkumar H., and SomnathSengupta. "Event-importance based customized and automatic cricket highlight generation." *Multimedia and Expo, 2006 IEEE International Conference on*. IEEE, 2006.

[14] Ravinder, M., and T. Venu Gopal. "Replay Frames Classification in a Cricket Video Using Correlation Features and SVM." *European Journal of Applied Sciences* 7.2 (2015): 92-97.