

# View Classification and Caption Analysis based Technique for Goal Event Detection

Vilas Naik\*, Vandana Athani\*\*

\*CSE, Basaveshwara Engineering College, Bagalkot

\*\* CSE, Basaveshwara Engineering College, Bagalkot

**Abstract:** - The developing era of computer and internet has led to a new ways of accessing internet, through which users will excess their interested multimedia information from different corners of the world using their electronic digital devices. Sports are one of the most interesting event that always drag the interest level of large audiences but most of the users have lack of time to watch whole of the match but in turn there is a problem of automatically extracting useful semantic information from sports video which will help users getting their requirements. Hence research in the field of sports that too especially, event detection has emerges as a important research area due to its bigger commercial requirements. One of the major tasks of sports video analysis is event detection. Event detection in sports video is a meaningful happening that has happened for example – goal, corner kick, foul etc. Event detection will assist users in many required applications such as video indexing, video summarization, enhancement in video content, analysing about opposite team and their tactics, planning a good statistics for the team, helps in referee for decision making etc. Many approaches make use of different modalities of video such as audio and video for their research and implementation technique. Few approaches will use multi-modalities for event detection. The proposed system uses grass-dominant-ratio for detecting different angles of camera positions and classifies the views as long, medium, closer and audience. This paper presents event detection method for sports video using view classes and score caption region (the caption displayed on screen for score). The proposed framework includes an algorithm for detecting dominant color of the region and detecting a shot boundary and also domain specific shot categorization. Finally, the event is detected in soccer game. The detected events are used to generate summarization.

**Keywords** - Sports video; content based ; text based; object based

## I. INTRODUCTION

An event is “something happening at that instance and drives attention”. Thus Event detection systems are built on event of interest, categorizing them as interesting or non-interesting. There are many common problems that need to be addressed as part of event detection such as modelling of an event, quality of the video taken to model the event and so on. Still a larger amount of effort has to be put in to resolve many such problems although many researchers already have contributed their work towards this. The intension of event detection in video is the need of search in large collections of video data. In order to make it possible, a system has to be made which would be able to automatically extract useful information from given data and analyse it so that it can understand what the data represents for humans. Representation for humans refers to the highlights that any

user want to watch instead of watching the whole video as they have time constraint, which is called as summary of the video.

Instead of watching the entire game, most of the audiences prefer to watch only highlights of the match which is called as summary of the game. Most of the people have the habit of watching all kind of sports such as cricket, soccer, football, basketball, volleyball etc. But they can't watch each and every match by sitting in front of the TV; rather they wish to watch most exciting occurrences of the game. So to enable this requirement of each user, there is a continuous research on getting important events of each one the above said games where at the end these detected events will be taken together and can be used in many different application such as indexing of important events, for analysing the opposite team's tactics, for improving own team's strength accordingly. Hence, proposed a model, which detects goal event in soccer video by using caption in the video which gives information about the goal and also by view classification. Here a survey has been done on different ways of detecting an event in sports video.

## II. LITERATURE SURVEY

In the literature the following type of event detection techniques are found. The survey is done for understanding good methodologies that can be implemented for detecting an event in sports video and which is more suitable method or technique for detecting an event. And the identified classes are as follows: 1) Content Based (Video based, Audio Based) 2) object Based 3) Text Based 4) Structure based event detection.

Content Based Event Detection: An object based strategy is proposed in [1], a detailed survey is done on content-aware video analysis for sports video. Different scenarios are considered for analysing the structure of content which will give the deeper understanding of content-aware analysis. The study is focused on the video content analysis techniques that are applied in sportscasts over the last decades. Content-aware analysis methods are discussed with respect to object-, event-, and context-oriented groups.

Video Based Event Detection: The videos are divided into sequence of events in [2] that is done by unsupervised event invention and detection framework. This framework totally depends on extracting the low-level visual features such as histogram of color or oriented gradient's histogram. The features are globalized to different types of games.

Here event detection is done by taking a video clip as input and getting a series of events based on the final event models obtained at the previous stage. In [6], a technique is discussed that consists of extracting the key frames from static video. By

pre-sampling Key frames are extracted uniformly or randomly in the original video sequence.

Event detection in dynamic video is the process of producing an abstract form of the original video that consists of significant scenes from the original video. It is short summary of whole video. A few of the algorithms for live video summarization include applying a method known as Singular Value Decomposition; motion model [7]. Video format summary is available in [8]. For live summarization which is known as skimming, many techniques get and fragment video shots from the original video. There are very limited works being done for live video but there are many research work are being done for static video. Many approaches use video features but few will use audio and linguistic information.

A live video summary is done for movie videos in [9]. The progress of stories has been used as core technique in this approach. This progress of stories is used to get the human semantic acceptance. To partition a video the two dimensional entropy is used.

Then, semantic meanings of the scenarios are obtained. To do this the spatio-temporal correlation among detected shots are used. Lastly, general rules of particular scenario and common methods of movie production are taken.

**Audio Based Event Detection:** In [3], an audio cue based approach is used for detecting event in sports video which will finally summarize the sports video. In this approach one of the sports character, which is audio variation is used and the events are detected based on the audio feature of the given video. There will always be variation in the audio of the sports because of audience cheer, referee whistle, and applause by audiences. These audio has been taken and are mapped on to some audio measure, if this crosses some threshold of normal audio level, then it is detected as an event. This technique detects the peak values from the audio input that is presented and also looks for the video frame for corresponding to the audio peak that has been detected.

In [4], an automatic highlights detection based on an audio classifier is proposed. The audio classifier is based on a new modelling technique of the audio spectrum called Piecewise Gaussian Modelling (PGM) and Neural Networks. In this approach, the audio stream is extracted from the video stream and the signal is down-sampled then the signal is windowed with a Hamming window overlap. The Fast Fourier Transform is applied on each window. The spectrum is then filtered using a filter-bank. It is shown that audio-based highlights detection can be effective for tennis segmentation. Goals can be detected in soccer videos using audio analysis as well.

**Object Based Event Detection:** The shot boundary detection is considered as the important part for event detection in sports video. Different methods have already proposed for this

video segmentation using like temporal feature [10], segmentation based on frame set [11] and event detection [12].

View classification contains uncovering of different views such as elongated view, middle view, and little view and beyond field view. In papers [10] and [13] various techniques proposed, where in some of those methods will use detecting color which is dominant in the field for classifying these views. For short view player's skin color may be dominant, for elongated view the dominant color may be green because the ground is green in color for soccer.

Logos are detected to identify events in [14]. These logos are detected by replays of the exciting events. Replays are always in slow motion and TV broadcasters will show these replays between logos and hence events are detected by identifying these logos.

**Text Based Event Detection:** Sports videos are mainly constituents of some mesmerizing events which will detain the concentration of the user.

Even though many times general sports summarization may give the required information for the audience, it also integrates some of the techniques which are domain specific to the respected sports. Because many TV broadcasters will use high level editing effects which includes replays in slow motion and many of the caption describing about the key events of the game.

In [5], a text based approach is used. This is called as web cast text that is broadcasted in the channel for sports. These texts are used to extract semantics and are finally used to detect the event.

Initially text is analysed to cluster and text is detected for events in an unsupervised way using semantic breakdown method. Along with detecting the text, a random field model is applied for aligning the text event and video event, which is done by generating event boundary in the video.

**Structure Based Event Detection:** In [16], the detection of video structure is done in the simplest structural elements like frame or shot but also of video scene. During news broadcast the heading aggregation should be of less time consuming hence it is always acceptable to analyse and keep the material with respect to player, scenes etc. Hence this approach is structure based event detection.

### III. PROPOSED SCHEME FOR EVENT DETECTION

The proposed model in Fig. 1 is based on event detection for emphasizing significant events during soccer game. The whole video is segmented into smaller shots initially and then classifies them according to different shot type classes. Then it will apply machine learning methods to identify caption area that describes regarding the score of the game. Later, it will detect vertical goal posts and net. As a final point the system will highlight the most vital events during the game.

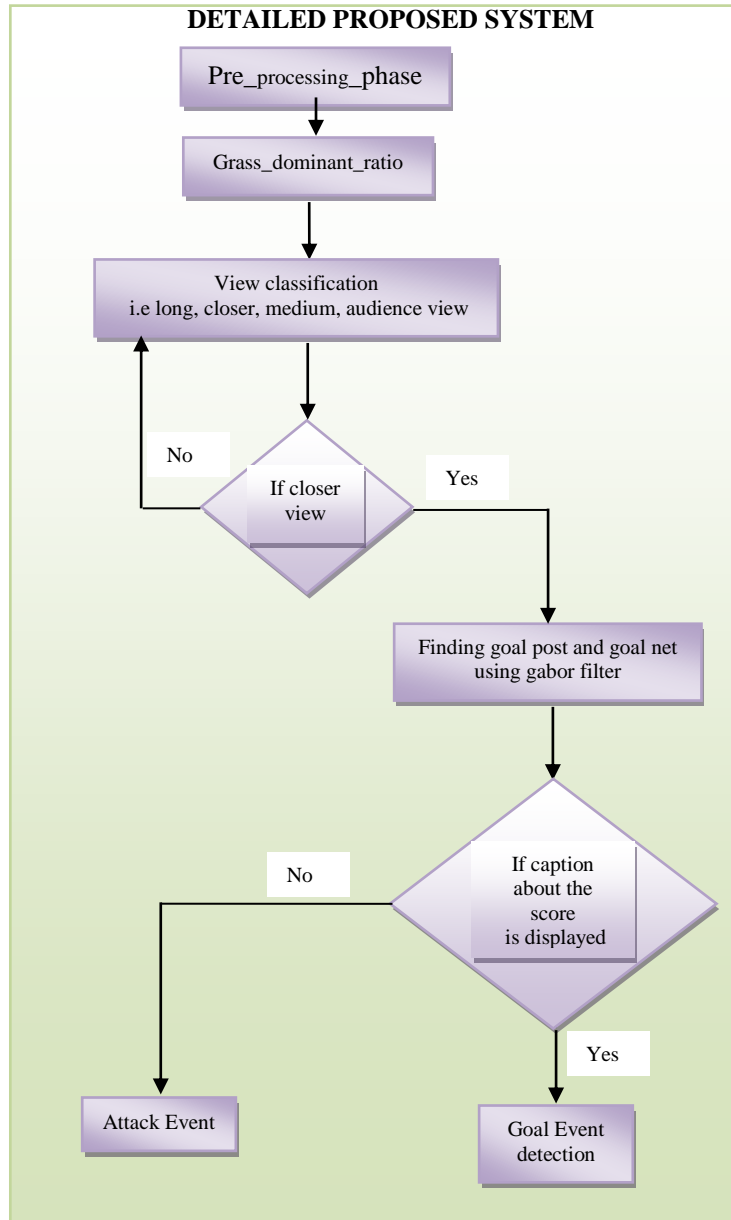


Fig. 1 block diagram for event detection

**ALGORITHM FOR GOAL EVENT DETECTION**

**Step 1:** primarily segments the given video into smaller video frames.

**Step 2:** classify camera views in each frame such as elongated, middle, nearer and audience views.

**Step 3:** If the view identified is nearer, then the system will identify goal post and net using k means and Gabor filter algorithms respectively.

**STEP 4:** Then caption area will be detected in the frame which describes regarding the score of the game. The system applies ANN algorithm for describing vital segments by detecting the caption area that describes about the score of the game.

**Step 5:** if there is a caption available, then declare it as goal event else attack event.

**Step 6:** In step 3 if there is a view other than closer then declare it as other event.

*A. Pre-processing Stage*

In this stage the given video is fragmented into smaller video shots. Initially grass dominant ratio is applied for each of the frames and then boundary detection by shot classification is done for segmentation.

*B. Dominant color extraction in the frame*

The color that covers major part of the area is called as dominant color and it is different for different sports. Since the sport chosen for this paper is soccer the major color that appears is always green because this is the color of the ground. There will be always many challenges in the process of getting dominant color in each frame such as lighting effects, shadow of the players, different resolutions of cameras and many other ecological factors.

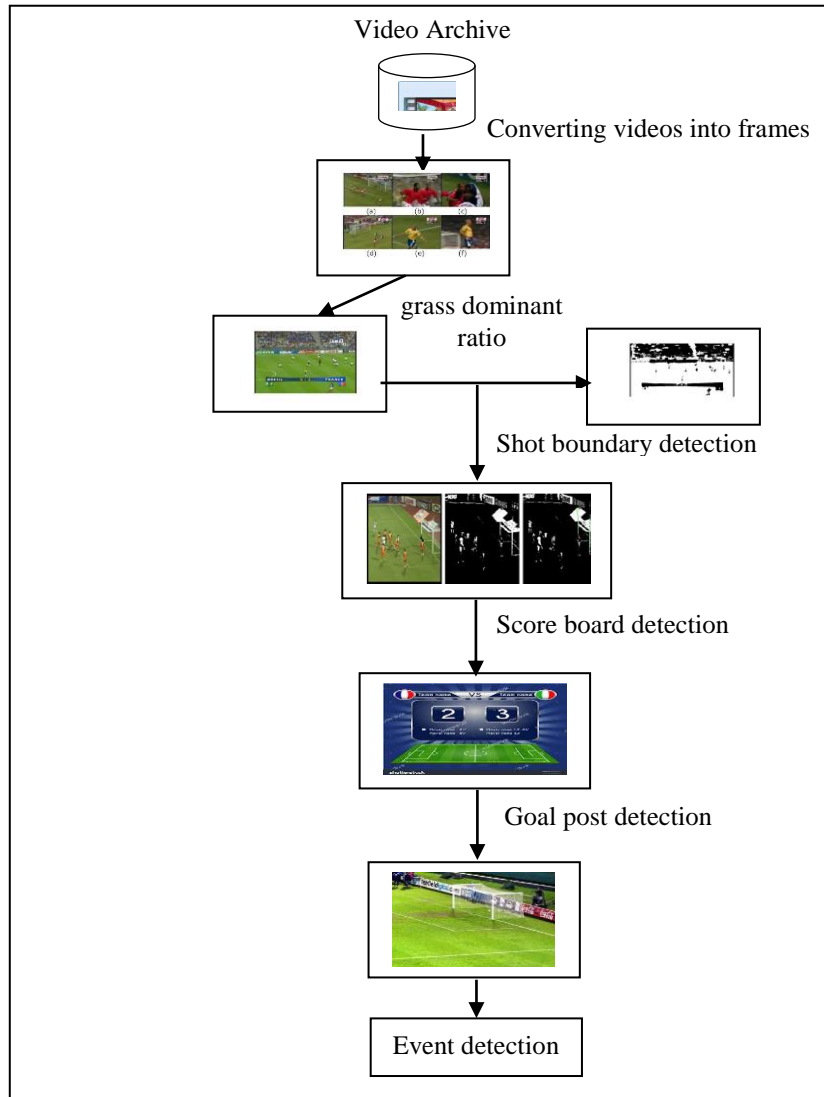


Fig. 2 pictorial representation of event detection

C. classifying views

Different views come from different cameras that are kept in different locations around the playing ground. While changing view of one camera to the view of another camera new shot will appear and that defines the boundary of new shot. Single shot is always a set of frames that are taken out of single camera in a continuous action that lies in a specified time and space. Shot transition will be of instant or gradual. Instant are more accurate than gradual shot change.

D. Event Detection stage

Soccer specific events will be more thrilling only nearer the goal mouth. Such events include goal, foul, penalty, direct kick, free kick and shooting. Thrill event detection is rely on three features;

Namely, 1) Score detection using super imposed text on the frame, 2) Detection of goal post near goal mouth, 3) Detection of goal net.

1) *Score detection using super imposed text on the frame:* The super imposed text is a caption area which is differentiated from the surrounding region, which provides details about the score of the game. The super imposed text is often called as caption and will show at the floor part of image whenever there is an occurrence of event such as goal and it will disappear soon after displaying the results within few seconds.

2) *Detection of goal post near goal mouth:* Goal posts are the vertical pillars that appear at the goal mouth. Since they are of white in color, Hough transform is applied to each frame for detecting them.

3) *Detection of goal net*: Even though the goal posts are detected, it is not enough to finalise whether the event is thrilled or not. Hence the goal net is also detected to conclude



Fig. 3 examples of captions describing about score of the game

the event is nearer to the goal mouth. Gabor filter is applied to identify goal net.

#### E. Artificial Neural Network (ANN)

In the proposed system ANN is used for training purposes. Using ANN sufficient number of training set will be given to train the system and are finally tested by using testing set.

#### IV. CONCLUSIONS

The caption score region based system proposed in this paper for event detection in soccer videos will be evaluated using videos for soccer matches. The proposed system consists of five phases; they are, pre-processing stage, view classification stage, score caption region detection phase, thrilled event detection stage, and event detection stage. The proposed system performs very well as more analysis is done. Also, it has been decided to use ANN classifier as machine learning techniques for training purpose.

#### REFERENCES

- [1] Huang-Chia Shih, Member, Ieee, "A Survey on Content-Aware Video Analysis For Sports", IEEE Transactions on Circuits And Systems For Video Technology, Vol. 99, No. 9, January 2017
- [2] Hao tang, vivek kwatra, mehmet emre sargin, ullas gargi, "detecting highlights in sports videos: cricket as a test case"
- [3] Pradeep k, "significant event detection in sports video Using audio cues", international journal of innovations in engineering and technology (IJJET)
- [4] Hadi harb, liming chen, "highlights detection in sports videos based on audio analysis I".
- [5] changsheng xu, yi-fan zhang, guangyu zhu, yong rui, hanqing lu, qingming huang, "using webcast text for semantic event detection in broadcast sports video", iee transactions on multimedia, vol. 10, no. 7, november 2008
- [6] E. Mendi, H. B. Clemente, and C. Bayrak, "Sports video summarization based on motion analysis", Computers and Electrical Engineering, vol. 39, pp. 790–796, April 2013.
- [7] Y.-F. Ma and H.-J. Zhang, "A model of motion attention for video skimming", Proceedings International Conference on Image Processing, vol. 1. IEEE, 2002, pp. 129–132.
- [8] M. S. Lew, et-al "Content-Based Multimedia Information Retrieval: State of the Art and Challenges", ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 2, February 2006.
- [9] S. Zhu, Z. Liang, and Y. Liu, "Automatic Video Abstraction via the Progress of Story", Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, vol. 6297, pp. 308–318.
- [10] A. Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization", IEEE Transactions in Image Processing, vol.12, no. 7, pp. 796–807, Jul. 2003.
- [11] C. L. Huang, H. C. Shih, and C. Y. Chao, "Semantic analysis of soccer video using dynamic Bayesian network", IEEE Transactions Multimedia, vol. 8, no. 4, pp. 749–760, Aug. 2006.
- [12] D. W. Tjondronegoro et al, "Knowledge-discounted event detection in sports video", IEEE Transactions vol. 40, no. 5, pp. 1009–1024, Sep. 2010.
- [13] D. A. Sadlier, N. E. Oconnor, "Event detection in field sports video using audio-visual features and a support vector machine", IEEE Transactions. Circuits System Video Technology vol. 15, no. 10, pp. 1225–1233, Oct. 2005.
- [14] Hossam M. Zawbaa1, Nashwa El-Bendary2, Aboul Ella Hassanien1, and Tai-hoon Kim3, "Event Detection Based Approach for Soccer Video Summarization Using Machine learning". 2010
- [15] Yasmin S. Khan, Soudamini Pawar "Video Summarization: Survey on Event Detection and Summarization in Soccer Videos", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 11, 2015
- [16] Kazimierz Choroś, "Automatic Detection of Headlines in Temporally Aggregated TV Sports News Videos", 8th International Symposium on Image and Signal Processing and Analysis (ISPA 2013)



**Vilas Naik** received BE(Electronics and Communication) from Karnataka University Dharwad and Master of Engineering in Computer Technology from Shri Guru Govind Singh College of Engineering Nanded under Sri Ramanand Teerth Marathwada University Nanded .India He is currently a research scholar registered to Visvesvaraya Technological University, Belagavi in the area of Image and Video processing working on the issues of Multimodal Video summarization and

selection. Currently he is working as Associate Professor in the Department of Computer science and Engineering, Basaveshwar Engineering College Bagalkot. His subjects of interest are Image and Video processing, Data Communications and Computer Networks, Computer Architectures and Multimedia computation and Communication



**Vandana Athani** received Bachelor's Degree in Computer Science and Engineering from BEC, Bagalkot under Visvesvaraya Technological University, Belgaum, Karnataka, India, and currently pursuing the Master Degree in Computer Science and Engineering from Basaveshwar Engineering College, Bagalkot, under Visvesvaraya Technological University, Belgaum, Karnataka, India.