

Implementing Dynamic Bayesian Network for Soccer Videos Event Detection and Summarization

Rupali S. Bongane¹, Sachin S. Bere²

^{1, 2} DGOIFOE Swami-Chincholi, Daund, Pune

Abstract: *There are expansive number of games features are accessible over the Internet. Anyhow utilizing these videos to get data is a difficult and time intensive task. In this way, extraction on occasions from the videos is needed. Semantic analysis of videos and automated event extraction assumes a key part in a few applications; utilizing content-based web search engines, indexing of video, and summarization of video. As an influential way for learning complex patterns is the Bayesian Network, this paper proposes a novel Bayesian Network (BN) based strategy automatic event recognition and summarization in Soccer videos. This system incorporates proficient shot view classification, shot boundary detection algorithms, and the related Bayesian network development. There are three fundamental stages: First is detection of the shot boundaries. The video is segmented into large and meaningful semantic segment known as play-break sequences by utilizing hidden Markov model. In second stage, play-break sequences are utilized to extract a few key events. In the third stage, the Bayesian network is utilized to obtain the high level semantic events. Developing the Bayesian network for which the structure is assessed utilizing the Chow-liu tree is the essential part of the technique. By applying group of Copula, the joint distributions of random variables of the system are displayed. Some of events that can be perceived by this technique in Soccer videos are objective, card, corners, shots on goals, offside, fouls, and missed shots. The users are more inclined to be interested in these events and not in complete and vast videos.*

Keywords: Bayesian network, video summarization, semantic video analysis.

1. Introduction

From last two decades, the rate of audio-visual data in the form of images and videos has expanded quickly. With the development of computational power and electronic storage limits, the need for huge digital image/video libraries has expanded. For example, in the Internet there exist a lot of unknown image and video information which are the bases of numerous entertainments, educational, and commercial applications. This makes the search of image/video information a difficult issue for users. A suitable computerized image/video library must give simple access to data and encourages the recovery of the content.

At the point when overall length of videos achieves a huge number of hours, users require a framework for summarizing and abstracting them to have an effective and proficient search. Albeit text-based search algorithms support users to search a particular image or a portion of long video, much of the time, the framework yields numerous unessential video to guarantee the recovery of the target video. Smart indexing and summarization frameworks are a key need in video retrieval. Among all video sorts, sports videos draw in numerous viewers and normally last for long periods of time. Sports videos, normally, are made out of some fascinating events which catch viewers' attention.

For many people, a summarized form of the sports video is more interesting than the full length video. In spite of the fact that a non specific sports video summarization framework is sufficient and valuable, the summarization framework in a domain-specific way, for example, soccer videos may provide more facilities to viewers. Most sports telecasters utilize some editing impacts, for example, slow-motion replays, and super-imposed content subtitles to recognize the key events. Hence, high level semantics can be

identified utilizing these editing effects and the audio-visual media emphasizes that are extracted. Preparing of sports videos makes it conceivable to convey sports videos over narrow-band networks, since the significant semantics usually possess just a little part of the entire content [1].

One of the challenging issues in a video event detection technique is the event boundary detection. A few systems, for example, [2], propose a frame-based algorithm for event detection, while other technique, such as, [1] and [3], utilize the temporal video segments for extracting more significant semantic units for event detection. In this strategy, same as [3], the "play-break" sequence has been utilized as a semantic unit in event detection. Every "play-break" comprises of two areas called "play" and "break." In soccer videos, the amusement is in a "play" mode when the game is going on and the "break" mode is the complement set; that is, at whatever point the game is stopped due to occurrences of an event.

2. Literature Survey

Normally, soccer video semantic analysis methods can be categorized into two main categories; pattern recognition methods [2] [3] [4] [5] and object tracking methods [6] [7] [8] [9].

A. Pattern Recognition methods

Chung-Lin Huang et al [2] presented a semantic analysis framework focused around Bayesian system (BN) and Dynamic Bayesian system (DBN). It is verified in the specific area of soccer match videos. In light of BN/DBN, it can recognize the uncommon events in soccer match, for example, goal, corner kick, penalty kick, and card. The video analyzer extricates the low-level proofs, while the semantic analyzer utilizes BN/DBN to interpret the high-

level semantics. Not the same as past shot-based semantic analysis techniques, the proposed semantic analysis is frame-based for every input frame, it gives the current semantics of the event nodes and also the hidden nodes. An alternate commitment is that the BN and DBN are consequently created by the training process rather than dictated by ad hoc. Temporal intervening network is proposed for enhancing the precision of the semantics output.

Dian W. Tjondronegoro and Yi-Ping Phoebe Chen [3] had proposed a novel methodology for detecting events in sport videos, which utilizes a definitive extent of identification and a universal set of audio/visual features. Their proposed technique is a hybrid between manual principles and statistics-based annotation to attain a base application of the domain-specific learning needed.

The methodology is generic, as the fact that domain-specific knowledge is utilized to choose or include the audio/visual features utilized; hence it is more vigorous for distinctive sports especially those that have comparable play-break structure and comparative audio/visual features. Case in point, hockey, water polo, American football, handball and rugby all contain a playing field with two goal posts, and the intriguing events.

Based on the exploratory results over different games and events, play-break scope is turned out to be efficient for detecting and classifying highlight events. It is due to that every play-break ought to contain all the obliged details from every event in sport videos. Due to this, play-break segment boundaries oblige to absolutely be recognized, and they have demonstrated that replay correction enhances the performance. The statistical-driven rules have been experimented to demonstrate its efficiency and robustness utilizing an extensive data set of sport videos, while every game is recorded from distinctive telecasters, rivalries, matches, and/or phase of rivalries.

In the study by M. Shyu et al [4], for video semantic analysis, particularly video event location, a subspace-based multimedia data mining system is presented, by tending to two essential issues, that is, semantic gap and uncommon event identification. The proposed system attains complete automation by means of multimodal content analysis and intelligent assimilation of distance-based and rule-based data mining systems. The content analysis procedure encourages the extensive video analysis by extricating middle-level and low-level features from audio/visual channels.

The assimilated data mining systems successfully attain these two fundamental problems by allaying the class imbalance problem along the procedure and by reproducing and refining the feature measurement consequently. The hopeful exploratory performance on goal event location and sports concepts extraction from soccer videos and TRECVID news accumulations exhibits the sufficiency of the proposed system. Moreover, its novel domain-free characteristic shows the extraordinary capability of extending the proposed multimedia data mining system to an extensive variety of distinctive application spaces.

Event identification is a basic amongst the most essential parts for different sorts of domain applications of video data framework. As of late, it has picked up an extensive enthusiasm of specialists and academics from diverse ranges. While identifying video event has been the focus of far reaching exploration exertions as of late, significantly less existing methodology has considered multimodal data and related effectiveness issues.

In the study, J. Shen et al [5] utilize a subspace selection procedure to attain quick and exact video event identification utilizing a subspace determination method. The methodology is equipped for separating distinctive classes and saving the intra-modal geometry of specimens inside an identical class. With the system, feature vectors introducing diverse sort of multi data can be effectively anticipated from distinctive identities and modalities onto a unified subspace, on which identification methodology can be performed. Besides, the training stage is completed once and they had a unified transformation matrix to extend diverse modalities.

Not at all like existing multimodal location frameworks, this framework functions well when a few modalities are not accessible. Trial results focused around soccer video and TRECVID news video accumulations show the efficiency and robustness of the proposed MMP for individual identification tasks in comparison with the current methodologies.

B. Object Tracking Methods

J. Liu et al [6] had proposed an automatic player location, unsupervised naming and proficient player tracking system requisitioned telecast soccer videos. The detection module joins background modeling and boosting recognition. Labeling is accomplished through unsupervised player appearance learning. MCMC information association is requisitioned tracking players. The result can be used for team strategies and player action analysis, high-light identification, and so forth. The system can likewise be connected to different applications, for example, vision-based human-machine collaboration. While the greater part of players can be recognized and tracked by their system, a few cases, for example, long impediments, video smear, unexpected camera movement and player tangle may direct to failure. They planned to design more productive MCMC recommendations, and enhance the labeling and tracking performance by playfield enlistment and trajectories induction in future works.

The framework proposed by T. D'Orazio et al [7] could give a successful help to officials much of the time: examinations showed that the utilization of different cameras with a high edge rate permitted the lessening of perspective errors and the identification of quick events that happened in distant parts of the field. The opposite cameras understood numerous cases in which the decision was troublesome by utilizing only one perspective. The tests demonstrated that the framework had the capacity perceive 35 offside events over a sum of 45. Four wrong offside or four uncertain events were produced by the framework created by the trouble in interpreting the offside rule, the compelling

imaging conditions that may happen in crowded circumstances, the complex configurations among players, the complicated interactions between the ball and the players.

M. Leo et al [8] had developed real time visual system for goal line monitoring that could regard the fundamental imperatives forced by the International Football Federation (FIFA) for the use of technologies in official matches: most importantly to be not intrusive for the field and the players, to be totally automatic and reliable, to be not reliant on human mediation, and to have real time reactions. They had introduced the results acquired from the tests acquired during the championship 2005-2006/2006-2007, with four cameras with high frame rates (200f/s) put on the two sides of the goal lines. Four computers handled the pictures procured by the cameras, recognizing the ball position in real time.

C. Wu et al [9] had presented a three-layer near-real-time event inference plan for games event recognition. First and foremost, low-level features are extricated and the framework automatically segments the frame sequences into clips as essential semantic inference units. At that point, semantic concepts of these clips are extricated to provide effective semantic description. At last, rule-based finite-state machines are intended for event inference. With the three steps, their framework can perceive sports events precisely. Exploratory results about on a live sports game video amusement feature exhibit one of its effective applications in track and field events recognition.

3. Proposed System

The basic goal of the method is capable of detecting various events in soccer videos such as: goal, card, goal attempt, corner, foul, offside, and non-highlights. The Objectives of the proposed application are as follows: 1. To detect various events 2. Large event detection is always desirable 3. When the total length of videos reaches thousands of hours; users need a system for summarizing and abstracting them in order to have an efficient and effective search. The proposed method is capable of detecting seven different events in soccer videos; namely, goal, card, goal attempt, corner, foul, offside, and non-highlights.

In our system we are using chow-Liu junction tree for structure estimation of the Dynamic Bayesian network. We are using Chow-Liu tree because its approximation is best. For better pattern recognition it gives good approximation result. It has been proven that it provides a better or at least as good approximation for a discrete multivariate probability distribution. The basic part is to construct Dynamic Bayesian network for which the structure is estimated using Chow-Liu tree. We are implementing Soccer Videos of Event Detection and Summarization Using Dynamic Bayesian Network.

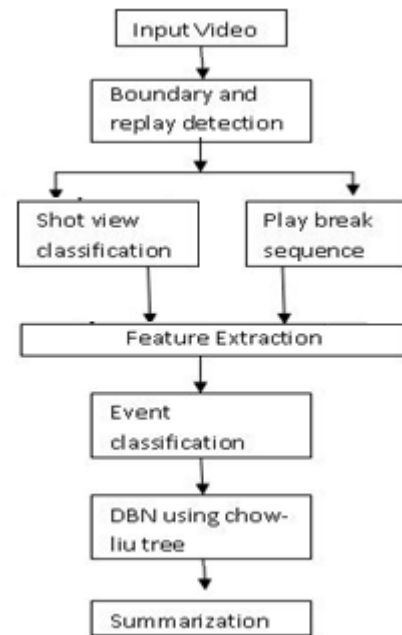


Figure1: System Architecture

4. Conclusion

In this survey, techniques for event location and summarization in video are surveyed. Also an effective technique for event location and summarization of soccer video is proposed. The Dynamic Bayesian network is utilized as a classifier for soccer event detection. Despite some existing strategies that are focused around frame or shot, the proposed strategy utilized the play-break succession as a semantic unit which prompts extraction of more genuine peculiarities from the video furthermore reduces the obliged processing cost. For calculating the joint distributions in the Bayesian network, which enabled us to use more complicated distribution models for network variables, Copula and Chow-liu tree are also utilized.

References

- [1] Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," *IEEE Trans. Image Process.*, vol. 12, no. 7, pp. 796–807, Jul. 2003.
- [2] L. Huang, H. C. Shih, and C. Y. Chao, "Semantic analysis of soccer video using dynamic Bayesian network," *IEEE Trans. Multimedia*, vol. 8, no. 4, pp. 749–760, Aug. 2006.
- [3] W. Tjondronegoro, and Y. P. Chen, "Knowledge-discounted event detection in sports video," *IEEE Trans. Syst., Man, Cybern. Part A: Syst. Humans*, vol. 40, no. 5, pp. 1009–1024, Sep. 2010.
- [4] M. Shyu, Z. G. Xie, M. Chen, and S. Chen, "Video semantic event/concept detection using a subspace-based multimedia data mining framework," *IEEE Trans. Multimedia*, vol. 10, no. 5, pp. 252–259, Feb. 2008.
- [5] J. Shen, D. Tao, and X. Li, "Modality mixture projections for semantic video event detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 11, pp. 1587–1596, Nov. 2008.
- [6] J. Liu, X. Tong, W. Li, T. Wang, Y. Zhang, and H. Wang, "Automatic player detection, labeling and

- tracking in broadcast soccer video,” *Pattern Recognit. Lett.*, vol. 30, no. 2, pp. 103–113, Jan., 2009.
- [7] T. D’Orazio, M. Leo, P. Spagnolo, P. L. Mazzeo, N. Mosca, M. Nitti, and A. Distanti, “An investigation into the feasibility of real-time soccer offside detection from a multiple camera system,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 12, pp. 1804–1818, Dec. 2009.
- [8] T. D’Orazio, M. Leo, P. Spagnolo, M. Nitti, and N. Mosca, “A visual system for real time detection of goal events during soccer matches,” *Comput. Vision Image Understanding*, vol. 113, no. 5, pp. 622–632, May, 2009.
- [9] C. Wu, Y.-F. Ma, H.-J. Zhang, and Y.-Z. Zhong, “Events recognition by semantic inference for sports video,” in *Proc. IEEE Int. Conf. Multimedia Expo*, Aug. 2002, pp. 805–808.