Implementing Dynamic Bayesian Network for Soccer Videos Event Detection and Summarization

Miss. Rupali S. Bongane
Department of Computer Engineering
DGOIFOE, Swami-Chincholi, Daund
Pune, India
bongane.r.r1@gmail.com

Ass. Prof. Sachin S. Bere
Department of Computer Engineering
DGOIFOE, Swami-Chincholi, Daund
Pune, India
Sachin.bere@gmail.com

Abstract— **Programmed events annotation is a crucial necessity for building a successful games feature synopsis. Analysts worldwide have effectively been looking for the strongest also capable answers for discover and characterize key occasions (or highlights) in distinctive games. The vast majority of the current and generally utilized methodologies have utilized decides that model the average example of varying media offers inside specific game occasions. As the Bayesian system is an influential device for learning complex examples, a Dynamic Bayesian system based strategy is proposed for programmed occasion identification and synopsis in soccer features. The proposed system incorporates productive calculations for shot limit identification, shot perspective grouping, mid-level visual peculiarity extraction, and development of the related Dynamic Bayesian network. The proposed framework applies a mixed bag of systems and exploits the fleeting excess in feature coming about. This paper presents a semantic investigation framework focused around dynamic Bayesian system (DBN). This system comprises of three stages. In the first stage, shot limits are discovered. At that point, in second stage feature is fragmented into vast and genuine semantic units, called play-break arrangements. In the following stage, a few tricks are separated from each of these units. At long last, in the last stage, with a specific end goal to accomplish abnormal state semantic tricks (occasions and ideas), the Bayesian system is utilized. DBN is generally characterized as the exceptional instance of separately associated BN particularly went for time arrangement demonstrating. The soccer videos are extracted from Youtube.com web site.**

Keywords- **Dynamic Bayesian; Network; Event Detection; Extraction; Hidden Markov Model.**

I. INTRODUCTION

In the previous decade, a lot of digital media information counting image, audio, and video, streaming video cuts, scene images, and three-dimensional (3-D) representation has been conveyed to audience. We need an adaptable and versatile way to deal with these rich media of which the digital video has been broadly acknowledged as the most available one. Several examination endeavors have been embraced by utilizing area knowledge to encourage extraction of high-level ideas straight forwardly from features. A few methodologies use stochastic techniques that frequently misuse programmed learning abilities to determine knowledge, for example, hidden Markov models (HMMs) [1]–[3]. Ekin [4] proposes a completely programmed and computationally proficient framework for sports video examination and summarization by utilizing low-level video transforming algorithms.

As of late, programmed location of the main highlights of sports video has ended up famous. Considering basic semantic ideas included in video events, a qualification can be by and large made between object oriented (alternately static-idea) events and action-oriented (or dynamic- idea) events. The events involving the ideas like Cityscape and Boatship are object-oriented as in they are basically concerned with the vicinity of specific objects in a video stream.

In the high-level feature extraction assignment of yearly TREC video recovery assessment (TRECvid) [9], the benchmark of explained video corpus is given to specialists to recognizing an extensive set of object-oriented events. Interestingly, the action oriented events, for example, People-calling-cellphone (Celltoear), People-dropping-something (Objectput) and People-pointing something (Pointing) include the semantic ideas that are solely related with particular actions performed in a video stream.

Indexing and summarization frameworks are a key need at the present time video recovery. Among all video sorts, sports videos pull in numerous viewers and ordinarily keep going for extended periods. Sports videos, when all is said in done, are made out of some fascinating events which catch clients' consideration. For a great many people, a compressed adaptation of the sports video is more alluring than the full length form. Despite the fact that a bland sports video summarization framework is sufficient and valuable, the summarization framework in an area particular way, for example, soccer videos, may offer more offices to clients.

Generally sports supporters utilize some altering impacts, for example, slow-movement replay scenes, and super-forced content subtitles to recognize the key events. Along these lines, high level semantics can be distinguished utilizing these altering impacts and the audio-visual features that are consequently separated.

One of the testing issues in a video occasion recognition technique is the occasion boundary location. A few systems, such as [2], propose a frame-based calculation for occasion discovery, while different systems, for example, [1] and [3], and utilize the temporal video sections for concentrating more genuine semantic

units for occasion recognition. In our technique, in the same way as [3], we have utilized the "play-break" grouping as a semantic unit in our occasion recognition. Each one "play-break" comprises of two areas called "play" and "break." In soccer videos, the diversion is in a "play" mode when the diversion is going on and the "break" mode is the supplement set; that is, at whatever point the diversion is ended in light of the fact that of event of an occasion.

This paper additionally shows a novel technique for sectioning the video into its play-break arrangements. From each one play-break arrangement a few features are removed which are ordered by utilizing a Bayesian network. The element Bayesian network (DBN) [20] is based on the BNs and their expansions; it tries to bind together temporal measurement with instability. DBN is a valuable apparatus for speaking to complex stochastic courses of action. Late improvements in derivation and adapting in DBN [20]–[24] have been connected to numerous genuine applications.

II. RELATED WORK

In the recent decades, numerous papers tending to player segmentation and following, player position recognition, ball distinguishment and area on the play field have been accounted for. Most of them chip away at broadcast images with the point of perceiving players' actions for video summarization [1] and virtual perspective replay [2]. A couple of works have been exhibited for constant preparing of game images managing only one of the above specified issues.

In [3] Tjondronegoro et al. (2010) present a foundation recouping calculation for soccer player segmentation has been displayed which considers the particular issue of lighting changes and the way that slow and quick motion in the scene can be considered. The issue of dispensing with shadows to get a decent segmentation for soccer player discovery has been tended to in [4]. An unsupervised learning system decides the RGB color circulations of the closer view and shadow classes.

Orazio et al. [8] proposed a three-level video-occasion discovery proposes creature chase identification in untamed life documentaries. The main level concentrates color, texture, and motion features, and catches shot limits and moving object blobs. The mid-level utilizes a neural network to focus the object class of the moving object blobs. This level likewise creates shot descriptors that join features from the first level and inductions from the mid-level. The shot descriptors are then utilized by the domain-particular deduction process at the third level to identify video portions that match the client characterized occasion model. The proposed approach has been connected to the discovery of hunts in untamed life documentaries.

In [10] Duda et al. display a novel methodology for sports video semantic occasion discovery based on analysis and arrangement of webcast content and broadcast video. Webcast content is a content broadcast channel for sports amusement which is co-delivered with the broadcast video and is effortlessly gotten from the web. They first investigate webcast content to group and recognize content events in an unsupervised way utilizing probabilistic latent semantic analysis (PLSA). Based on the recognized content occasion and video structure analysis, they utilize a conditional random field model (CRFM) to adjust content occasion and video occasion by recognizing occasion minute and occasion boundary in the video. Fuse of webcast content into sports video analysis essentially encourages sports video semantic occasion recognition. They directed probes 33 hours of soccer and basketball games for webcast analysis, broadcast video analysis and content/video semantic alignment.

In [9] Wu et al. propose a creative system for semantic video annotation through coordinated mining of visual features, discourse features and continuous semantic examples existing in the video. The proposed system essentially comprises of two primary stages: 1) Construction of four sorts of prescient annotation models, to be specific discourse affiliation, visual-affiliation, visual-successive and factual models from expounded videos, and 2) Fusion of these models for explaining un-commented videos naturally. The fundamental preference of the proposed system lies in that all of visual features, discourse features and semantic examples are considered simultaneously.

In [12], C. M. Bishop (2006) proposed a novel audio–visual feature based framework for occasion identification in broadcast video of various distinctive field sports. Features demonstrating noteworthy events are chosen and hearty indicators constructed. These features are established in attributes regular to all classes of field sports. The proof accumulated by the feature indicators is consolidated by method for a backing vector machine, which deduces the event of an occasion based on a model produced amid a preparation stage. The framework is tried nonexclusively crosswise over numerous types of field sports including soccer, rugby, hockey, and Gaelic football and the results propose that high occasion recovery and substance dismissal insights are achievable.

In [14] Szantai et al. display algorithms for parsing the structure of delivered soccer programs. The issue is essential in the connection of a customized video streaming and searching framework. While former work concentrates on the location of unique events such as objectives or corner kicks, this paper is concerned with bland structural components of the diversion. They start by characterizing two fundamentally unrelated conditions of the diversion, play and break based on the standards of soccer. They choose a domain-tuned feature set, overwhelming color proportion and motion force, based on the unique syntax and substance attributes of soccer videos.

In [7], Orazio et al. propose a powerful audiovisual feature extraction plan and content discovery and distinguishment system. Their framework gives programmed indexing of sports videos based on discourse and video analysis. They concentrate on the utilization of DBN and show how they can be successfully connected

for intertwining the confirmation acquired from distinctive media data sources.

III. IMPLEMENTATION DETAILS

A. *System Overview*

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 used planning to use Dynamic Bayesian network for structure estimation. It has been proven that it provides a better or at least as good approximation for a discrete multivariate probability distribution. The DBN is used because of its ability to capture temporal dependences among random variables. The modules of the proposed application are as follows:

- Input Video: For the training phase user give the input video to the proposed system.
- Shot Boundary Detection: A Model is based on shot boundary detection technique. Here used frame transition parameters for the technique. The previous and the next frame can be used to formulate frame estimation scheme. Frame transition parameters were based on this formulated frame estimation scheme.
- Feature Extraction: From the input frame the features are extracted. The object detection is done on this phase.
- Database: After the feature extraction, the frames are extracted. These frames are saved in the database for the detecting further frames in video.
- Dynamic Bayesian Network Construction: An algorithm to incorporate the Conditional Independence (CI) tests and ordering of nodes. The ordering of the nodes from database was generated by using the CI tests.
- Summarization: After that, text summary is generated along these all steps.

The following Figure 1. Shows the proposed system architecture. The essential goal of the technique is equipped for discovering different events in soccer videos, for example, goal, card, goal attempt, corner, foul, offside, and non-highlights. The Objectives of the proposed application are as follows:

1. The recognisation of different events.
2. Extensive occasion identification is constantly alluring
3.When the aggregate length of videos achieves thousands of hours; clients require a framework for abridging and abstracting them keeping in mind the end goal to have a proficient and successful pursuit. The proposed system is equipped for catching seven separate events in soccer videos; in particular, goal, card, goal attempt, corner, foul, offside, and non-highlights.

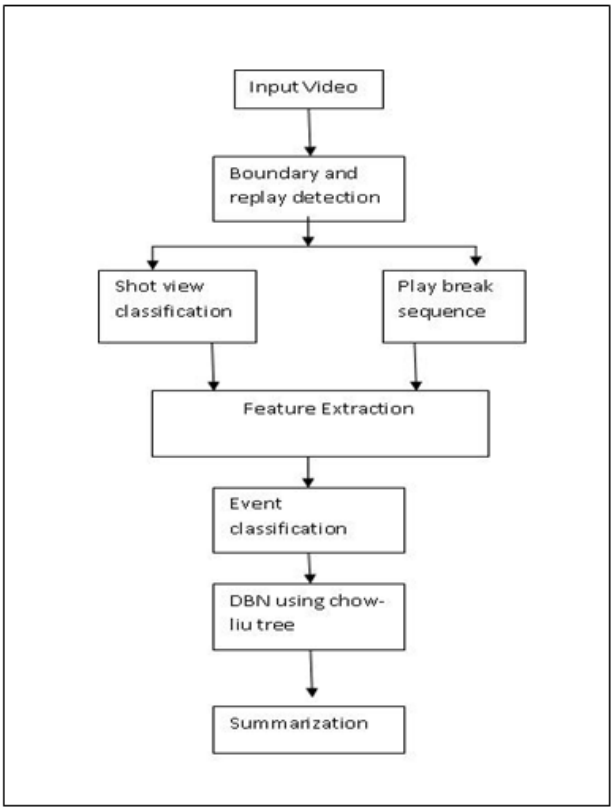

Figure1: System Architecture

B. *Mathematical Model for Proposed Work*

Let, System S is represented as: S = {D, E, C, S, N}

Boundary and Replay Detection:

Consider, D is a set of detecting shot boundary and replay

D = {d1, d2,…}

Where, d1, d2,… are number of detections.

Feature Extraction :

Let E is a set for feature extraction

E = {f1, f2, f2 …}

Where, f1,f2,are number of features.

Event Classification:

Let, C is a set for event classification

C = {e1; e2; e3 ,….}

Where, e1, e2 ,…are the number of event used for classification.

Structure Estimation:

Let, S is a set for structure estimation forming a tree,

S = {s1, s2, s3,…}

Where, s1, s2,… are number of structure form.

Video Summarization:

Let, N is a set for summarization for various shots

N = {n1, n2,…}

Where, n1,n2,..are number of summary output for various inputs.

C. *Algorithm*

Input: Soccer video
Step1: We are extracting Low level feature for given input video. (i.e. Shot boundary detection, Replay detection, Shot view classification, plays break segmentation)
Step2: Extract mid level feature of input video. (Detecting white line in grass, penalty box detection)
Step 3: To classify the events of video we are using Dynamic Bayesian network as event classifier.
Step 4: Different Important events of given soccer video are extracted
Step 5: Summary of Soccer video will be created which will show important events from given input soccer video clip. (Use 0-1 Knapsack Problem for creating summarized video)
Output: Summery of given Soccer video clip.
Length of Summery will be less than Input soccer video.

D. *Experimental Setup*

The system is built using Java framework (version jdk 6) on Windows platform. The Netbeans (version 6.9) is used as a development tool. The system doesn't require any specific hardware to run, any standard machine is capable of running the application.

IV. RESULTS AND DISCUSSION

A. *Dataset*

Here videos are extracted from Youtube.com. YouTube is a video-sharing website headquartered in San Bruno, California. The service was created by three former PayPal employees in February 2005. In November 2006, it was bought by Google for US$1.65 billion.[4] YouTube now operates as a Google subsidiary.[5] The site allows users to upload, view, and share videos, and it makes use of Adobe Flash Video and HTML5 technology to display a wide variety of user-generated and corporate media video. Available content includes video clips, TV clips, music videos, and other content such as video blogging, short original videos, and educational videos.

B. *Results*

Figure 2. shows the graph of Training time required for detecting events from videos.. X axis shows video size in Mb. and Y axis shows time in minute. The graph shows proposed system required low training time as compared to existing system for detecting events.

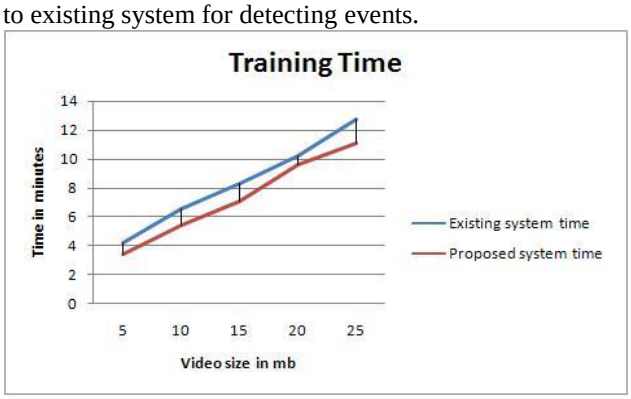

Figure.2: Graph of Training Time

Figure 3. shows the graph of Detection time required for detecting events from videos.. X axis shows video size in Mb. and Y axis shows time in minute. The graph shows proposed system required low detection time as compared to existing system for detecting events.

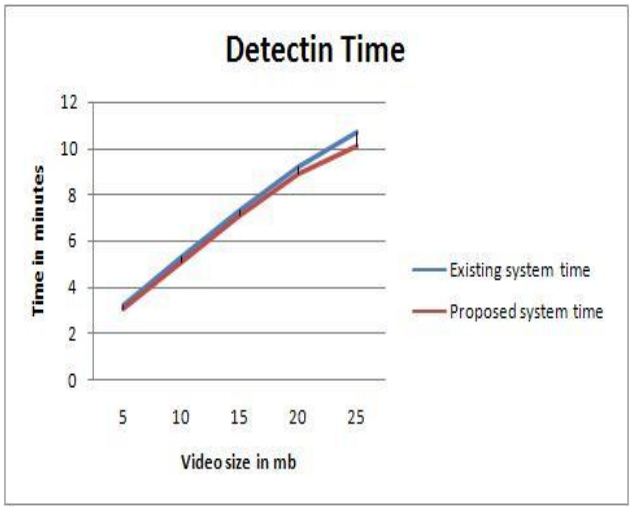

Figure.3: Graph of Detection Time

V. CONCLUSION

There are numerous video summarization and occasion distinguishment systems for videos. We are utilizing a few systems for shot boundary location to uproot the undesirable segments from the video. Essentially, utilized Shot perspective order. Dynamic Bayesian Networks are built to follow the real and complete developments of all players, the ball and the officials also. Players are expected to be followed for self-evident reasons, the development of the ball is followed for the events like Goal and shots on goal. We have proposed a video system understanding framework. Given an info succession, the framework will gather the low-level proof, and applies the surmising motor in BN/DBN to surmise high-level semantic ideas that interpret the semantic substance of video game system. The principle commitment of this paper is to add the temporal intervening network to DBN to progress the semantic interpretation precision.

ACKNOWLEDGMENT

The authors would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance. We are thankful to the authorities of Savitribai Phule University of Pune and concern members of iPGCON2015 conference, organized by , for their constant guidelines and support. We also thankful to reviewer for their valuable suggestions. We also thank the college authorities for providing the required infrastructure and support. Finally, we would like to extend a heartfelt gratitude to friends and family members.

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] C. 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] D. 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. Distante, "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.

[10] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed., New York: Wiley, 2001.

[11] C. K. Chow and C. N. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Trans. IT-14*, vol. 3, no. 3, pp. 462–467, May 1968.

[12] C. M. Bishop, *Pattern Recognition and Machine Learning*, New York, USA: Springer Science Business Media LLC, pp. 55–57, 2006.

[13] E. Kov´acs and T. Sz´antai, "On the approximation of a discrete multivariate probability distribution using the new concept of t-Cherry junction tree," *Lecture Notes in Economics and Mathematical Systems*, Berlin: Springer, 2010, pp. 39–56.

[14] T. Sz´antai and E. Kov´acs, "Hypergraphs as a mean of discovering the dependence structure of a discrete multivariate probability distribution," *Ann. Oper. Res.*, vol. 193, no. 1, pp. 71–90, 2012.

[15] R. B. Nelsen, *An Introduction to Copulas*, 2nd ed. New York: Springer, 2006, pp. 7–14.

[16] A. Sklar, "Random variables, distribution functions, and copulas—a personal Look backward and forward," in *Distributions with Fixed Marginals and Related Topics*, Hayward, CA, USA: Institute of Mathematical Statistics, 1996, pp. 1–14.