Video Summarization: Survey on Event Detection and Summarization in Soccer Videos

Yasmin S. Khan

ME student, Department of Computer Engineering
D. Y. Patil College of Engineering, Akurdi
Savitribai Phule Pune University
Pune, India

Soudamini Pawar

Assistant Professor, Department of Computer Engineering
D. Y. Patil College of Engineering, Akurdi
Savitribai Phule Pune University
Pune, India

Abstract—**In today's world, the rapid development of digital video and editing technology has led to fast growing of video data, creating the need for effective and advanced techniques for analysis and video retrieval, as multimedia repositories have made browsing, delivery of contents (video) and video retrieval very slow. Hence, video summarization proposes various ways for faster browsing among a large amount of data and also for content indexing. Many people spend their free time to watch or play different sports like soccer, cricket, etc. but it is not possible to watch each and every game due to the longer timing of the game. In such cases, the users may just want to view the summary of the video that is just an abstract of the original video, instead of watching the whole video that provides more information about the occurrence of various incidents in the video. It is preferable to watch just highlights of the game or just review/trailer of a movie. Apparently, summarizing a video is an important process. In this paper, video summarization approaches are discussed, that can generate static or dynamic summaries. We present different techniques for each mode in literature. We have discussed some features used for generating video summaries. As soccer is the world's most famous game played and watched, it is taken as a case study. Research done in this domain is discussed. We conclude that there is a broad perspective for further research in this field.**

Keywords—Summarization; Sports Summarization; Soccer

I. INTRODUCTION

The rapid development of digital video capture and editing technology has led to fast growing of video data that creates the need for effective and advanced techniques for video retrieval. In this era, life is becoming too busy no has free time. The users do not have enough time to watch the entire video. In such cases, the user may just want to view the summary of the video that is just an abstract of the original video, instead of watching the whole video that provides more information about the occurrence of various incidents in the video. Many people spend their free time to watch or play different sports like soccer, cricket, basketball, etc. but it is not possible to watch each and every game due to lengthy time of the game. So many people prefer to watch highlights of the game. So as per user's requirement there is a growing need for summarizing videos. Video summarization is a process that facilitates fast browsing among large video collections. It also allows more efficient content indexing. Video summarization refers to creating a summary of a video that addresses three main points. (1) The video summary should contain scenes and events not only as short as possible from the video but also the most important one. For example, in a soccer game, the summary must contain goals, fouls, shot boundaries, goal attempt, and some other important scenes. (2) The video summary should maintain a continuous connection amongst scenes. It means that the video summary should not contain video segments connected in a blind way. (3) The summarized video should not contain any redundancy. That is, the video summary should have a free repetition that is very difficult to achieve. It is necessary to detect various events in the game to generate a summary. There are two types of video summarization: Static summarization and Dynamic Summarization. As we all know soccer is the world's most famous game played and watched, the survey is discussed on soccer.

We have divided this paper into different sections. Section II describes Video Summarization, types, techniques, methods. Section III sports video summarization that takes soccer game as a case study. Section IV describes Literature Review.

II. VIDEO SUMMARIZATION

Video summarization is considered as one of the most important feature that makes the search easier and useful than before. To develop efficient indexing and search techniques to manage the huge amount of video data, new technologies need to be researched. Using this, people can use it to get the actual idea and the important events as well as scenes without watching the full original and long videos of several hours. The developed techniques in video summarization can be used for various domains, such as surveillance videos, consumer videos, movies, sports, news, etc. The summary produced can be static or dynamic i.e. it can be either KeyFrames or Video Skims. Video summarization is a tool for generating a short summary of a video, as the name implies, can either be a sequence of stationary images called key frames or moving images called video skims.

A. Static Video Summarization

These are also called representative frames also called R-frames, still-image abstracts or static storyboard. This type of video summarization can be classified in three different ways. These are as Classification based on sampling, classification based on shot segmentation, classification based on scene segmentation.

In [22], a similar method is discussed that consists of extracting the keyframes. Keyframes are extracted by pre-sampling uniformly or randomly the original video sequence. Keyframe extraction is a fundamental process in video content management that involves selecting one or multiple frames that will represent the content of the video and used for generating video summaries.

B. Dynamic Video Summarization

The idea of video skimming or dynamic summarization is generating a short video composed of informative / important scenes from the original video. The user receives an abstract view of the video story. The story is in video format [17]. For dynamic summarization also known as skimming, most techniques extract and segment video clips from the original video. Some of the techniques/ mechanisms for dynamic video summarization include applying SVD (Singular Value Decomposition), motion model [18]. In [20] and [19], method based on semantic analysis technique is applied for skimming. Compared with static summarization, there are relatively few works being addressed for dynamic video skimming. Most techniques are based mainly on visual information. Some other approaches make use of audio and linguistic information.

In [20], a dynamic video abstraction scheme for movie videos is presented. The proposed method is based on the progress of stories. The proposed approach attempts to comprehend video contents from the progress of the overall story and human semantic understanding. Here, the properties of two-dimensional histogram entropy of image pixels are adopted first, to segment a video into shots. Then, semantical meaning scenarios are obtained. It is done according to the spatio-temporal correlation among detected shots. Lastly, general rules of special scenario and common techniques of movie production are exploited. It is done to achieve the progress of a story in terms of the degree of progress between scenarios to the overall story.

C. Summarization Based On Clustering Techniques

The basic idea is clustering together similar frames/shots and then extracting some frames per cluster as key frames. These methods are different in features as color histogram, luminance, and motion vector and clustering algorithms as k-means, hierarchical [15].

A fuzzy c-means clustering algorithm is used in [23], where the original video is segmented into frames and these frames are considered as basic elements. Then color features are extracted in HSV color space. Then for grouping the frames Fuzzy c-means clustering algorithm is used. Then one frame (KeyFrame) per cluster is selected. The clustering output is a membership matrix that represents most represntable frame from each cluster. Paper [16] presents a new approach called VGRAPH. It uses keyframes extraction process as a shot-based method. For that it requires video segmentation by detecting the shot boundaries. First, the original video is pre-sampled. So that it can reduce the number of frames to be processed. Second, the pre-sampled video is segmented into shots using the color features. These features are extracted using the color histogram computed from the HSV. Next, noise frames are eliminated. And then second

frame is selected as a shot representative. At last, the keyframes are extracted using nearest neighbor graph. This graph is built from the texture features extracted from the shots R-frames using Discrete Haar Wavelet Transforms. In [25], cluster based techniques are further divided into four further classes as, techniques based on similar activity based, k-means based, partitioning based and spectral based. A disadvantage of the most of the methods that relay on clustering algorithms is making them computationally very complex for real time applications.

In [25], M. Ajmal et al have categorized video summarization techniques in six different categories based on mechanism used and overall processing. The hierarchical classification of these techniques is described by author. The six major categories discussed are based on features, events, shot selection, cluster, trajectory and mosaic. Feature based techniques have wide scope of research as it is further divided on the basis of color, motion, gesture and object based and more.

III. SUMMARIZING SPORTS VIDEOS

Sports videos are mainly contents of some fascinating events that capture attention of the user. Many people prefer for summarized version of sports video rather than to watch full lengthy videos. Full version of the video may contains many non-significant events like advertisement, unnecessary playbacks, replays etc. Even if a generic sports video summarization system is efficient and useful, the summarization technique in a domain-specific way, like soccer videos, may present much more conveniences to users. Many sports broadcasters and web sites use editing effects such as super-imposed text captions and slow-motion replay scenes to discriminate the key events. For that reason, high level semantics can be perceived by using these editing effects.

Main part of the event detection in sports video is shot boundary detection. Various methods are already purposed for this like temporal video segmentation, frame based segmentation and event detection [1][2][3]. Shot view classification contains detection of various views like Long view, close up view, medium view and out of field view. There are various techniques proposed [1][4], some of the techniques use dominant color of frames for view classification. For close up view dominant color can be skin color of the player, for long view we can say that dominant color will be green as the background of the soccer field is green. Replays are mainly played into slow motion; most of the broadcasters play replays between graphical logos [5]. Replays can be detected by identifying such logos.

IV. LITERATURE REVIEW (SOCCER GAME AS CASE STUDY)

Various applications of video summarization touch different domains such as Consumer video applications, Personal videos, Image-Video databases management, Sports videos and surveillance videos and news videos. Apparently, Media organizations and TV broadcasting companies have shown considerable interest in these applications. Hence, Sports video summarizing is a vast domain to study. In this

section various techniques/approaches are discussed in the literature of the same field.

A Dynamic summarization method is presented in [14] for the automatic extraction of summaries in soccer videos. It is based on shot detection, shot classification and Finite State Machines. Four stages discussed in the same are: playfield segmentation, shot detection using the Discrete Cosine Transform Multi-Resolution (DCT-MR) and finally, soccer video word extraction and finding out the appropriate sub-words. These sub words present summaries using the FSM and domain knowledge, where a set of rules are defined to present the semantic states in soccer game. It also explores the interesting relations between syntactic structure and the semantics of the video. Playfield segmentation is a preprocessing step. Shot detection is a step where different types of shot transition are extracted, shot classification is done in three major classes; long shot, medium shot and close-up shot using statistical method.

The proposed system in [24] is capable of detecting seven events in soccer games such as goal, foul, non-highlights, card, goal attempt, corner, offside, etc. It uses Chow-Liu Tree for structure estimation of Bayesian Network. For better pattern recognition it gives good approximation results. It has been proven that it provides a better or at least as good approximation for a discrete multivariate probability distribution.

Pattern recognition-based techniques generally extract some audio-visual features ie. Mid-level features, low level features etc. and then by using a classifier, the events or high level semantics (high level features) are detected. In [7], an automatic method is proposed that utilizes a subspace-based data mining method for feature extraction. That method is generic such that it does not use any prior knowledge in the detection process and can be considered as a domain-free method. It uses a C4.5 decision tree classifier. In [8], another method is proposed that uses a specific dimension reduction method, called mixture modality projection (MMP), to obtain high level features from low and mid-level features. Some alternative pattern recognition techniques include the use of a dynamic Bayesian network (DBN) for capturing the temporal pattern of extracted features during soccer events. For sports video highlight detection, a hybrid approach that integrates some audio-visual statistics into logical rule-based models is reported in [9]. It utilizes the play-break sequence as a semantic unit of sports videos. The method has been applied to different sports; including soccer, basketball, and Australian football.

J. Liu et al [10] had proposed a programmed player area, unsupervised naming and capable player following framework ordered broadcast soccer features. The discovery module joins background displaying and boosting acknowledgment. Naming is proficient through unsupervised player appearance learning. The outcome can be utilized for group procedures and player activity investigation, high-light distinguishing proof, et cetera. The framework can in like manner be joined with diverse applications, for instance, vision-based human-machine interaction effort. While most of players can be perceived and followed by their framework, a couple of cases,

for case in point, long obstructions, feature smear, sudden camera movement and player tangle, may direct to disappointment. They wanted to plan more gainful MCMC proposals, and improve the naming and following execution by playfield enrollment and directions incitement in future works.

J. Shen et al, in [13], exploit a subspace collection method to achieve rapid and accurate video event classification employ a subspace grit method. The technology is prepared for saving the intra-modal geometry of specimens inside an matching class and disentanglement individual classes. With the structure, feature vectors bring in miscellaneous sort of multi data can be efficiently predictable from distinct modalities and identities onto a unified subspace, on that recognition technology can be performed. In addition, the training phase is finished one time and they had a combined alteration matrix to expand miscellaneous modalities.

V.   CONCLUSION

Video summarization has attracted researchers considerably and as a result, various algorithms, mechanisms and techniques have been proposed. In this paper, a review of the research in two forms of video summarization: static summary and the video skim is carried out. Regardless of the methods used, that are static or dynamic forms, the evaluation process showed that the techniques proposed produces video summaries of high visual quality, and some approaches are suitable for real-time video processing. However, a valid evaluation method can support the field to another level. There is not any best technique for abstracting a video sequence, as video abstraction is still in the research phase largely. Also, practical applications are still limited. So, there is a scope of research in many fields such as personalized videos, consumer videos, and movie videos as well.

REFERENCES

[1] 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.

[2] 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.

[3] D. W. Tjondronegoro et al, "Knowledge-discounted event detection in sports video", IEEE Transactions vol. 40, no. 5, pp. 1009–1024, Sep. 2010.

[4] 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.

[5] A. Nagasaka, Y. Tanaka, "Automatic video indexing and full-video search for object appearances", in Visual Database Systems II, E. Knuth and L. M.Wegner, Eds. Amsterdam, The Netherlands: North-Holland, 1992, pp. 113–127.

[6] L. Xie, S. F. Chang, A. Divakaran, and H. Sun, "Structure Analysis of Soccer Video with Hidden Markov Models",ICASSP New York, USA: Columbia University, Apr. 2002

[7] M. Shyu, Z. G. Xie et al "Video semantic event/concept detection using a subspace-based multimedia data mining framework", IEEE Transactions Multimedia, vol. 10, Feb. 2008.

[8] J. Shen, D. Tao, X. Li, "Modality mixture projections for semantic video event detection", IEEE Transactions Circuits System. Video Technology., vol. 18, no. 11, pp. 1587–1596, Nov. 2008.

[9] 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, Sep. 2010

[10] J. Liu, X. Tong, W. Li, et-al "Automatic player detection, labeling and tracking in broadcast soccer video", Pattern Recognition, vol. 30, no. 2, pp. 103–113, Jan., 2009.

[11] G. Zhu, M. Yang, et al, "Detecting video events based on action recognition in complex scenes using spatio-temporal descriptor", in Proc. ACM MM, 2009, pp. 165–174.

[12] A. M. Tekalp. "Digital Video Processing", Englewood Cliffs, NJ: Prentice- Hall, 1995.

[13] J. Shen, D. Tao, and X. Li, "Modality mixture projections for semantic video event detection", IEEE Transactions. Circuits System. Video Technology., vol. 18, no. 11, pp. 1587–1596, Nov. 2008.

[14] Y. Tabii et al "A new method for soccer video summarization based on shot detection, classification and finite state machine", in SETIT 2009 5th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, 22-26 March 2009.

[15] E. Asadi and N. M. Charkari, "Video Summarization Using Fuzzy CMeans Clustering", 20th Iranian Conference on Electrical Engineering, (ICEE2012). IEEE, 15-17 May 2012.

[16] M. K.M., et-al "VGRAPH: An Effective Approach for Generating Static Video Summaries", in IEEE International Conference on Computer Vision Workshops, IEEE, 2013.

[17] 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.

[18] 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.

[19] Y. Tabii and R. O. H. Thami, "A new method for soccer video summarization based on shot detection, classification and finite state machine", SETIT 2009 5th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, 2009.

[20] 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.

[21] Zaynab El khattabi et-al, "Video Summarization: Techniques and Application", International Scholarly and Scientific Research & Innovation, 2015.

[22] 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.

[23] E. Asadi and N. M. Charkari, "Video Summarization Using Fuzzy CMeans Clustering", in 20th Iranian Conference on Electrical Engineering, (ICEE2012). IEEE, 15-17 May 2012.

[24] Mostafa Tavassolipour, Mahmood Karimian, and Shohreh Kasaei, "Event Detection and Summarization in Soccer Videos Using Bayesian Network and Copula", IEEE Transactions on circuits and systems for video technology, vol. 24, no. 2, February 2014.

[25] Muhammad Ajmal, Muhammad Ashraf et-al, "Video Summarization: Techniques and Classification", Computer Vision and Graphics, Springer Verlag, 2012.