A Survey on Sport Video Summarization

Dharmesh Tank[1]
[1]Department of Computer Engineering
[1]Assistant Professor, Aditya Silver Oak Institute of Technology, GTU, Ahmedabad,Gujarat, India

Abstract- *Over the past few years, there has been a significant increase the amount of videos. Massive growth in video content poses problem of information overload and management of content. Also it made the multimedia repositories browsing, retrieval and delivery of video contents very slow and even difficult tasks. In order to manage the growing videos on the web and also to extract an efficient and valid information from the videos, more attention has to be paid towards video and image processing technologies. Video summarization has been proposed to improve faster browsing of large video collections and more efficient content indexing and access. Video summarization provide condensed and succinct representations of the content of a video stream through a combination of still images, video segments, graphical representations and textual descriptors. Video summarization methods attempt to abstract the main occurrences, scenes, or objects in a clip in order to provide an easily interpreted synopsis. The main aim of Video summarization is to provide clear analysis of video by removing duplications and extracting key frames from the video. Video Summarization is a process of creating & presenting a meaningful abstract view of entire video within a short period of time. Mainly two types of video summarization techniques are available in the literature, i.e key frame based (static video summarization) and video skimming (dynamic video summarization)[1].This paper reviews some of the recent work on video summarization information retrieval and discusses their role in current research directions which include different techniques, browsing and search paradigms, user studies, effective computing, learning, semantic queries, new features, media types, high performance indexing. Based on the current state of the art, we also discuss the major challenges for the future.*

Keywords- Video Summarization, Key Frame, Video Skimming, Image Processing.

I. INTRODUCTION

Nowadays, we get the information from the digital media such as news, movies, T.V shows, internet etc. Due to time limitation, it is difficult for an individual to watch its full content. A large number of cameras record video around the clock, producing huge volumes. Processing these huge chunks of videos demands plenty of resources like time, man power, and hardware storage etc[1]. Video summarization plays an important role in this context. It helps in efficient storage, quick browsing, and retrieval of large collection of video data without losing important aspects. Video content can be accessed by using either a top-down approach or a bottom-up approach [2]. The top-down approach, i.e. video browsing, is useful when we need to get an "essence" of the content. The bottom-up approach, i.e. video retrieval, is useful when we know exactly what we are looking for in the content.

In video summarization, what "essence" the summary should capture depends on whether the content is scripted or not. Since scripted content, such as news, drama & movie, is carefully structured as a sequence of semantic units, one can get its essence by enabling a traversal through representative items from these semantic units. Hence, Table of Contents based video browsing caters to summarization of scripted content. video summarization, either based on Table of Contain generation or highlights extraction, deals with how to use the representation structure to provide the viewer' stop-down access using the summary for video browsing.

When the video data exceeds their limits[3], some technique is needed to summarize and extract the events from these videos and only necessary part should be given to the user as per their requirements. From the different genres of videos, Sport, Movie videos attract most of the users. Sports are basically composed of interesting events, largely called as highlights. Large number of users are interested in these highlighted events rather than the full length videos. In the same aspect audiences will attract towards the movie trailer rather than watching whole movie at first glance.

To reduce the limitation of the huge amount of stream and non-stream multimedia, we use video summarization. Video Summarization is a process of creating and presenting, an abstract view of entire video within a short period of time. This technique will generate the summaries of the videos which contain maximum information that makes individual easier.

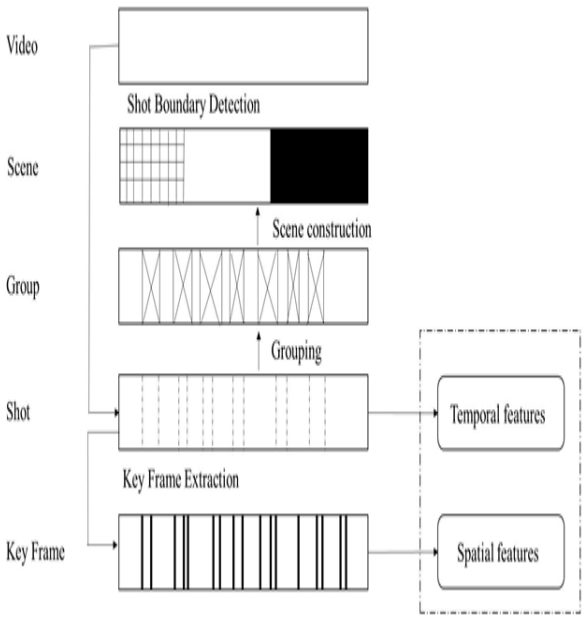

Figure 1: General classification of Video

II. TERMINOLOGY

(i) Scripted/Unscripted Content: A video that is carefully produced according to a script or plan that is later edited, compiled and distributed for consumption is referred to as scripted content. News videos, dramas & movies are examples of scripted content. Video content that is not scripted is then referred to as unscripted. In unscripted content, such as surveillance video, the events happen spontaneously. One can think of varying degrees of "scriptedness" and "unscriptedness" from movie content to surveillance content.

(ii) Video shot: is a consecutive sequence of frames recorded from a single camera. It is the building block of video streams.

(iii) Key frame: is the frame which represents the salient visual content of a shot. Depending on the complexity of the content of the shot, one or more key frames can be extracted.

(iv)Video scene: is defined as a collection of semantically related and temporally adjacent shots, depicting and conveying a high-level concept or story. While shots are marked by physical boundaries, scenes are marked by semantic boundaries.

(v) Video group: is an intermediate entity between the physical shots and semantic scenes and serves as the bridge between the two. Examples of groups are temporally adjacent shots [4]or visually similar shots.

(vi) Play and Break: is the first level of semantic segmentation in sports video and surveillance video. In sports video (e.g soccer, baseball, golf), a game is in play when the ball is in the field and the game is going on; break, or out of play, is the complement set, i.e., whenever the ball has completely crossed the goal line or touch line, whether on the ground or in the air or the game has been halted by the referee[5]. In surveillance video, a play is a period in which there is some activity in the scene.

(vii) Audio Marker: is a contiguous sequence of audio frames representing a key audio class that is indicative of the events of interest in the video. An example of an audio marker for sports. Some of the early literature in video parsing misused the phrase scene change detection for shot boundary detection. To avoid any later confusion, we will use shot boundary detection to mean the detection of physical shot boundaries while using scene boundary detection to mean the detection of semantic scene boundaries. Video can be the audience reaction sound (cheering & applause) or commentator's excited speech.

(viii) Video Marker: is a contiguous sequence of video frames containing a key video object that is indicative of the events of interest in the video. An example of a video marker for baseball videos is the video segment containing the squatting catcher at the beginning of every pitch.

III. VIDEO SUMMARIZATION

Due to the increasing volume of video content on the Web, and the human effort taken to process it, new technologies need to be researched in order to develop efficient indexing and search techniques to manage effectively and efficiently the huge amount of video data. One of the most evolving research areas is Video summarization. As the name implies, video summarization is a mechanism to produce a short summary of a video to give to the user a synthetic and useful visual abstract of video sequence, it can either be a images(key frames) or moving images (video skims). In terms of browsing and navigation, a good video abstract will enable the user to gain maximum information about the target video sequence in a specified time constraint or sufficient information in the minimum time[6].

The major task in video summarization is to segment the original video into shots and extract those video frames from the original video that would be the most informative and concessive presentation of the whole video. Such frames are referred as key frames. Basically, there two different kinds of Video Summarization techniques [7]in literature. One is static Video summarization and the other is dynamic Video Summarization. Static video storyboard summary involves a set

of key frames from original video and there is no restriction with time and sequence issue. On the other hand, it selects the most significant, small, dynamic portions of audio and video in order to generate the video summary.

3.1Method of Video Summarization

Video summarization is a mechanism for generating a short summary of a video, which can either be a sequence of stationary images (key frames) or moving images (video skims) [8]. Video can be summarized by two different ways which are as follows:

3.1.1 Key Frame Based Video Summarization

These are also called representative frames, R-frames, still-image abstracts or static storyboard, and a set consists of a collection of salient images extracted from the underlying video source[9]. Following are some of the challenges that should be taken care while implementing Key frame based algorithm: (i)Redundancy frames with minor difference are selected as key frame.(ii) When there are various changes in content it is difficult to make clustering.Key frames based video summarization can be classified in three types [10]. These are as follows:

a) Classification based on sampling

The main goal or less uniform random sampling, without considering the video content. Summary produced by this method are not for all the videos and may cause some large backup frames with similar content.

b) Classification based on scene segmentation

It extract key frames using scene detection, including the stage all parts of the semantic link in video or in the same place or at the same time. The disadvantage of this technique to produce a summary, which do not take into account the temporal position of the frame.

c) Classification based on shot segmentation

They close with altered content video key frames. First draw a picture taken from the first or key frames and the first frames and the last frames of the shot. These methods effectively for stationary and forms a small content, but they does not provide the description enough shots and movements in power.

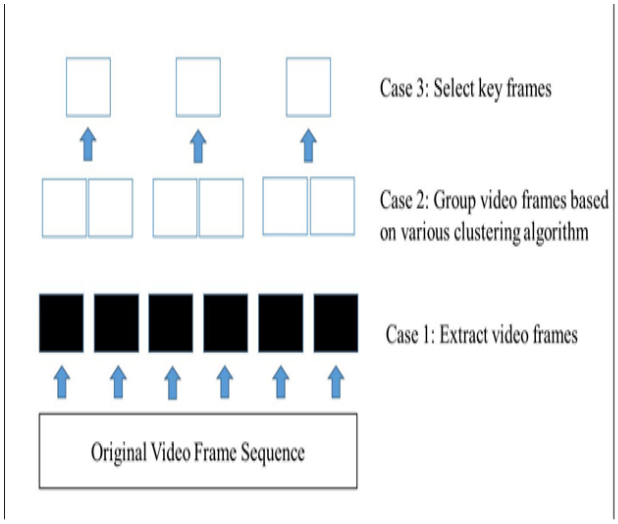

Figure 2: Key Frames based Video Summarization

3.1.2 Video Skim Based Video Summarization

This is also called a moving-image abstract, moving story board, or summary sequence [6][7]. The original video is segmented into various parts which is a video clip with shorter duration. Each segment is joined by either a cut or a gradual effect. The trailer of movie is the best example for video skimming. The simplest method for generating video skims is uniform subsampling, which extracts fixed-duration excerpts of the original video at fixed intervals. The “fast forward” method which simply increases the frame rate across the whole video content, either uniform or adaptive to the content dynamics[9].

3.1.3 Delaunay Triangulation

This method generates video summaries by capturing the visual content of the original videos in fewer frames most clustering methods require user-specified parameters or prior knowledge to produce their best results, this demands pre-processing or several trials. Both are extremely expensive and inefficient, because the best-fit parameters are not easy to get. But, video summarization technique by using Delaunay clusters generates good quality summaries with fewer frames with less redundancy Compared to many other clustering techniques. The Delaunay clustering algorithm is fully automatic with no user specified parameters and is well suited for batch processing.

An automatic video summarization technique based on Delaunay clustering is well suited for batch processing of videos without user intervention. Each cluster shows very high-quality clustering performance in terms of similarity of content. The clustering process results in clusters of different sizes based on the content represented in the original video[7].

3.1.4 K-means clustering algorithm

The K-means clustering algorithm is a simple method for estimating the mean (vectors) of a set of K-groups. It is an unsupervised clustering algorithm[12]. This is a simple and easy way to classify a given data set through a certain number of clusters fixed a priori. The main plan is to describe k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, it is better to place them as far as possible from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed. At this point we need to re-calculate k new centroids as bar centres of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. This loop may notice that the k centroids change their location step by step until no more changes are done. In this technique, video summaries are generated based on visual features and uses K-means clustering algorithm to cluster the frames. Quantitative Evaluation is based on quality of summary involving different users[13].

3.1.5 Shot Boundary Detection

In this approach, a shot is defined as the consecutive frames from the start to the end of recording in a camera. It shows a continuous action in an image sequence[14]. Deleting redundant information is achieved by segmenting the video into shots. A shot is a continuous recording of video content without breaks in a scene. Then, key frames may be extracted from each shot with different techniques based on pixel-to-pixel comparison, histogram-based comparisons, motion flow vectors, etc. This process is called Shot Boundary Detection. Shot Boundary detection is based on automatic threshold and by using reference key frame extraction videos are extracted[15].

3.1.6 Kernel video summarization

A video is partitioned into segments. A segment is a part of the video enclosed between two timestamps. A video summary is a video composed of a subset of the temporal segments of the original video. A segment can be considered important due to multiple reasons, depending on the video category and application goals: highlights of sport matches, culmination points of movies [9], influential moments of egocentric video. We make the assumption that the notion of importance can be learned from a set of videos belonging to the video category. This point of view stems from the Multimedia Event Recounting task at selecting segments containing evidence that the video belongs to a certain event category. Similarly, we define importance as a measure of relevance to the type of event. Fig. 3 shows an example video together with the importance of its segments. In this segmentation there is a video taken of a soccer match, where different semantic events take place. Major event here consider in a soccer match is a goal event. which include many classifications of the view in segmentation form as shown in Fig. 3.

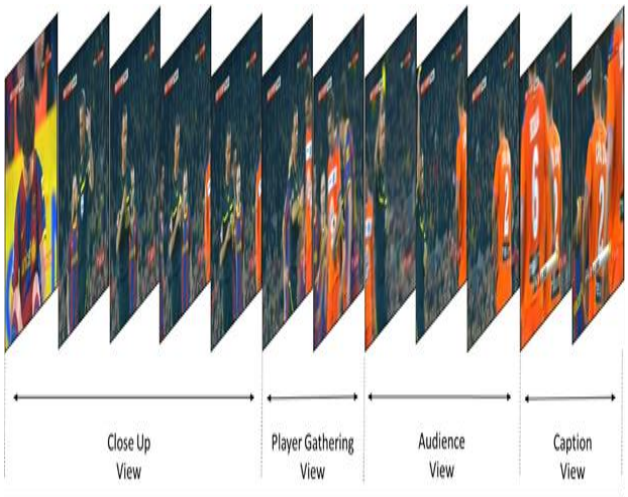

Figure 3: View segmentation in soccer video for goal event

IV. CONCLUSION

Video summarization plays important role in many video applications. The summarization viewpoint and perspective are often application-dependent. The semantic understanding and its representation are the biggest issues to be addressed for incorporating diversities in video and human perception. Depending upon the changes in contents of the video, the key frames are extracted. As the key frames need to be processed for summarization purpose, the important contents must not be missed. In this paper we have found that the key frame extraction is an important part of many video applications, like video indexing, browsing, and video retrieval. This method allows the user to generate a short relevant video from a long video. It improves the quality and also increases the efficiency of the summarized video. This survey concludes that there are several techniques used for video summarization. As per the objective we can retrieve the summary of the video by cluster based, key frames based or kernel based.

REFERENCES

[1] M. Ajmal, M. H. Ashraf, and M. Shakir, “Video Summarization : Techniques and Classification,” Comput. Vis. Graph. Springer, vol. 7594, no. 2001, pp. 8–13, 2015.

[2] Z. Xiong, Y. Rui, R. Radhakrishnan, A. Divakaran, and T. S. Huang, A Unified Framework for Video Summarization , Browsing and Retrieval. Elsevier/Aademic Press, 2006.

[3] H. Zhang, C. Y. Low, and S. W. Smoliar, "Video parsing and browsing using compressed data," » Multimed. Tools Appl., vol. 1, no. 1, pp. 89–111, 1995.

[4] D. Zhong and S.-F. Chang, "Video object model and segmentation for content-based video indexing," Proc. 1997 IEEE Int. Symp. Circuits Syst. Circuits Syst. Inf. Age ISCAS '97, vol. 2, pp. 1492–1495, 1997.

[5] L. Xie, S.-F. Chang, A. Divakaran, and H. Sun, "Structure analysis of soccer video with hidden Markov models," 2002 IEEE Int. Conf. Acoust. Speech, Signal Process., vol. 4, no. 1, pp. 4096–4099, 2002.

[6] Z. El, Y. Tabii, and A. Benkaddour, "Video Summarization : Techniques and Applications," vol. 9, no. 4, pp. 882–887, 2015.

[7] T. Sebastian, "A Survey on Video Summarization Techniques," vol. 132, no. 13, pp. 31–33.

[8] S. R. Ciocca Gianluigi, "An innovative algorithm for key frame extraction in video summarization," J. Real-Time Image Process. Springer, vol. 1, no. 1, pp. 69–88, 2006.

[9] B. T. Truong and S. Venkatesh, "Video abstraction: A Systematic Review and Classification," ACM Trans. Multimed. Comput. Commun. Appl., vol. 3, no. 1, p. 3, 2007.

[10] A. Sabbar, W.; Chergui, A.; Bekkhoucha, "Video summarization using shot segmentation and local motion estimation," Innov. Comput. Technol., p. 190,103, 2012.

[11] I. C. C. and K. Y. Cheng, "Content-selection based video summarization," IEEE Int. Conf. Consum. Electron. Las Vegas Conv. Center, USA, pp. 11–14, 2007.

[12] S. De Avila, a. Da Luz, a. De Araujo, and M. Cord, "VSUMM: An Approach for Automatic Video Summarization and Quantitative Evaluation," 2008 XXI Brazilian Symp. Comput. Graph. Image Process., 2008.

[13] D. S. Patil and S. B. Waykar, "A Survey on Event Recognition and Summarization in Football Videos," vol. 3, no. 10, pp. 2365–2367, 2014.

[14] M. H. Kolekar, K. Palaniappan, S. Sengupta, and G. Seetharaman, "Semantic Concept Mining Based On Hierarchical Event Detection for Soccer Video Indexing," J. Multimed., vol. 4, no. 5, pp. 298–312, 2009.

[15] D. Potapov, M. Douze, Z. Harchaoui, C. Schmid, D. Potapov, M. Douze, Z. Harchaoui, and C. S. Category-specific, "Category-specific video summarization To cite this version :," 2014.