Sound Approach for AI Generated Video Highlights

Prof. B. B. Gite
Assistant Professor and Head of Department
of Computer Engineering,
Sinhagad Academy of Engineering.
bb.gite@sindhagad.edu

Gaurav Kolhe
Department of Computer Engineering,
Sinhagad Academy of Engineering.
kolhegaurav123@gmail.com

Pranav Kadam
Department of Computer Engineering,
Sinhagad Academy of Engineering
kadamps295@gmail.com

Mohit Patil
Department of Computer Engineering,
Sinhagad Academy of Engineering.
mohitpatil1999@gmail.com

Suhas Surse
Department of Computer Engineering,
Sinhagad Academy of Engineering.
suhasassurse8421@gmail.com

Abstract: A large number of cameras record video around the clock, producing huge quantities of data. Processing or scrutinizing such huge chunks of videos and extracting salient events from the large video databases manually is a quite laborious and time-consuming task. Again such video data requires huge memory space to store. Due to such an increase in video data, there is a need for effective techniques for video retrieval and analysis as no one has enough time to watch the entire video.

Generating highlights from the video is called as video summarization and this technique produce a short, informative summary of a long video without losing the important data frames. This technique is useful in the indexing and classification of such videos in the video database. This paper provides a review of different approaches given by various researchers in the field of video summarization.

I. Introduction

The rapid advancement in the field of digital video technology makes it very easy to record a huge volume of videos of news, movies, sports, and documentaries. Again due to the increase in demand for security there are Thousands of video cameras found at public places, banks, ATMs and airports, etc. resulting in a large amount of information that is difficult to process in real-time i.e the user may not have enough time to watch the entire video like the user may just want to view the summary of the video to get an actual idea and the important events about the occurrence of various incidents in the original video(such as any suspicious activity in surveillance camera OR Important catches, goals or wickets for sports ), instead of watching the whole video.

Video Shortening has been a field of active research for a long time. However, the main focus was on either minimizing storage usage by compressing or removing redundant frames without loss of actual content. 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 the video sequence, it can either image either called static
(sequence of keyframes) video summarization or moving images either called as dynamic (video skims i.e. collection of dynamically-composed audio-video sub-clips) [1].

**Static Video Summarization:** Static video summaries are composed of a set of keyframes extracted from the original video.

**Dynamic Video Summarization:** It is a short video composed of informative scenes from original video presented to the user. The original video is condensed with the important content of a video in short time. It also preserves the motion information. The ability to include audio and motion elements that potentially enhance both the expressiveness and the amount of information conveyed by the summary.

**Classification of video summarization techniques**

Generate a video summary from the wide video contents can be done based on their properties and characteristics. On that basis they are classified mainly into four categories.

A. **Feature based Video summarization** [2]: In this the summarization is based on the certain features like voice, event, color, and motion. Different summarization techniques are used by the uses based on feature for example if user wants color features, then it's good to pick color-based video summarization technique [6].

B. **Event-based video summarization** [2][3][9]: These types of techniques are mostly used in surveillance video or sports video for event detection and summary generation. It helps to detect the intrusion/malicious behavior, Sports events [3] (Fours/Sixes/Wickets/goals) any abnormal events like robbery scenes, Mobile snatching, and terrorism. Song et al. [9] proposed surveillance video summarization approach based on event detection. First, the trajectories of motion objects are computed and then event detection is determined.

C. **Motion and Color Based video Summarization** [2][8]: Motion-based video summarization techniques is mostly used when the camera is involved. Nada Jasim Habeeb et al.[8] presented a system that extract motion objects from large video data.

D. **Object-Based video summarization** [2][7]: The object-based techniques are helpful in the detection of object from the video, like person, car, cat, etc. like author of [7] used object-based video summarization detect shot boundaries and extract video objects by a 3D graph-based algorithm. Once the objects are obtained, the shape of the objects needs to be represented. The key objects are extracted in a global manner by K-means clustering of shapes.

II. **Literature Survey**

Muhammad Ehsan Anjum et al.[3] proposed techniques to generate sports highlights from cricket video using techniques of optical character recognition. Two cricket videos of one day matches as South Africa vs. Australia and Pakistan vs. South Africa are used for experimentation. The score bar is extracted from the frames then the character recognition techniques are used to extract information for events like sixes, fours and wickets. Brute Force method and Template Matching techniques are used to detect the score from the frame. A short video summary is synthesized that includes the frames for the aforementioned significant events termed as Highlights.

**Advantages:** The Given approach is simple & can be used practically for highlighting the fours and sixes.

**Disadvantages:** 1. there are other important events other than four, Six or wickets which cannot be highlighted are due to low quality video and less extensive library for OCR.

2. Locating a score bar is tedious task.

M.Sridevi et al.[4] proposed a system that generate a video summary, by modeling a two stream architecture containing deep convolutional neural network in each stream for extracting both spatial
and temporal information of a video. Two dimensional Convolutional Neural Network (2D CNN) is used to exploit spatial information whereas a three dimensional Convolutional Neural Network (3D CNN) is used to exploit temporal information to generate highlight scores for segments of the video.

**Advantages:** 1. The Two Stream DCNN gives higher precision rate to capture both spatial and temporal information from a video to generate a well-designed summary.

2. The given method is outperformed than another technique like SVM Based Model, Rule Based Model, Spatial DCNN and Temporal DCNN.

Xuelong Li et al.[5] presented A general summarization framework for edited videos and raw videos. ADL dataset containing 20 long videos and SumMe dataset for short raw videos are used for the experimentation. Four models i.e., importance, representativeness, diversity and storyness are designed to capture the properties of summarized video. A score function is built with the weighted combination to balance the influence of these four property models. The efficiency of the proposed framework is verified on three datasets, including edited videos, short raw videos and long raw videos.

S. Zhang, Y. Zhu et al.[10] proposed the context-aware video summarization (CAVS) framework which is able to find the most informative video portions, from video sequences is given. The sparse coding with generalized sparse group lasso is used to learn a dictionary of video features and a dictionary of spatiotemporal feature correlation graphs. Sparsely gives the most informative features from the video.

**Advantages:** 1. Context-Aware Video Summarization (CAVS), a framework incorporates the event correlations to generate a short video summarizing the most informative parts of a long video sequence.

Yifang Yin, Roshan Thapliya et al. [11] proposed method for automatic video summary generation with personal adaption. The author introduces a novel hierarchical dictionary name semantic tree (SeTree). SeTree is a hierarchy which captures the conceptual relationships between the visual scenes in the codebook. The author proposed the automatic content-based feature encoding approach with a semantic tree which is more effective for personalized adaption. In the proposed design of video summarization, it joins the personal interest and visual attention.

Pushkar Shukla1Hemant Sadana et al.[12] proposed a framework that can automatically generate Cricket highlights considering event-based and excitement-based features. Wickets, boundaries, sixes, and milestones, are extracted using Event-driven features while the remaining important events are identified with the aid of excitement features. The OCR technique is used for detection of Event-driven strategies. Audio-based classifiers and replays are used to detect excitement-based strategy. A CNN+SVM methodology was adopted to classify key frames as starting frames.

**Advantages:** 1. The system provides better quality of cricket highlights compared to previous approaches.

2. The system is very robust in terms of detecting important events.

Juhi Naik et al.[13] proposed a tool for summarizing a video using Convolutional Neural Nets (CNNs). CNNs combined with Long Short-Term Memory (LSTM) modules to find much better video data. The dataset having 50 videos having duration from 1 to 10 minutes are used for the experimentation.

### III. Proposed Work

We are trying to implement a system i.e “Sound Approach for AI Generated Video Highlights” that scrutinizes the salient events from the large video databases without losing important aspects. The highlight generation reduces the time required for watching the entire video.
Figure: Proposed Highlight Generation Framework

We propose a system that initially extracts the frames from video input. Image preprocessing is applied on the extracted frame to remove noise and other disturbances from the image. CNN will be used to train the video analytics engine for recognizing important frames in the video. Figure 1 shows the overall flow of the proposed system that we are going to implement.

IV. Conclusion

It is quite obvious that one cannot watch and search important events visually from video lengths. An automated mechanism helps viewers to pay attention only for the events those are worth watching. Lot of work already has been done in terms of shortening of video to save both space & time. This paper presented different approaches used by the reviewer to generate highlights and summarize the videos of sports or any surveillance. Every view has some of the advantages and disadvantages mentioned above.

Here an attempt is made here to address ever increasing gap between the volumes of actual data generated and the volume that can be reasonably inspected manually with higher accuracy rate.

V. References


