Video Summarization Using Deep Learning

[1]Manasi J. Sutar, [2]Shrutika D. Patil, [3]Sakshi S. Bhutada, [4]Suraj M. Dhale, [5]Mr. V. N. Malavade
[1,2,3,4]UG Student, [5]Asst. Prof. , Sanjay Ghodawat Group of Institution Atigre, Kolhapur & India.
[1]manasisutar14@gmail.com, [2]patilshrutika0202@gmail.com, [3]khushissbhutada123@gmail.com,
[4]surajdhale17@gmail.com, [5]malavade.vn@sginstitute.in

Abstract: **Video summarization is a technique which can extract keyframes and remove duplicate frames from the original video to give a short video which represents the whole video. Video summarization can help us to browse a large number of videos in a short time interval. There are two fundamental ways to generate summaries are static and dynamic. In this paper, we are proposing a methodology for dynamic video summarization.**

Keywords —Aggregation Method, Color Layout Description, Dynamic Video Summation, Keyframes, Static Video Summarization, Video Summarization

I. INTRODUCTION

Now a day's video content on web increasing, due to improvement in digital technology. Social media websites such as YouTube, Facebook are sharing lots of videos daily. If we search for a certain topic we get lots of suggestion regarding that, browsing it may waste our time so video summarization can help us by providing a short and informative summary of that video. The aim of video summarization is to produce a summary of video which is interesting to the user and also include all main content of the video.

Video summarization is a mechanism that facilitates organized storage, fast browsing and retrieval of a large collection. The summary can be generated either by choosing keyframes which best represent the video or through video skimming. There are two different kinds of video summarization techniques. One is static video summarization and other is dynamic video summarization [1], [2]. Static summarization involves a set of key frames from the original video and there is no time restriction and sequence issue. On another side, it selects most consequent, small dynamic portions of video and audio which represent the whole video to generate the video summary. The main objective of video summarization is to produce a summary by taking minimum frames as possible but not missing any important information. Basic steps to start video summarization are shot boundary detection and keyframe extraction [1]. Shot boundary detection divides the video into shots.

II. RELATED WORK

A wide number of contributions have been found in the area of video summarization. In [3] essential individuals and protests in the video got from a wearable camera is distinguished utilizing district-based relapse for creating a storyboard outline. The authors in [4] developed a machine learning system that learns to predict video transitions based on statistical information derived from two successive frames. Paper [5] proposes a system for personalized video summarization that produces customized video summaries by adapting to the user's interest. The proposed system uses a high-level feature extraction to reduce the manual video annotation. The creators in [6] proposed an SVM based structure learning for exchanging information among video and literary highlights for a synopsis. More recently, supervised methods which directly leverage human-edited summary examples to learn how to summarize videos have attracted much attention [7]. In [8], they proposed another system — perceptual video outline where synopsis tuned in to the properties of human recognition assume a key job in deciding the edges to be chosen for the rundown. Conventional unsupervised video summarization methods generate summaries by leveraging handcrafted criteria based on low-level visual or motion cues [9], [10]. To retrieve information from audio or visual content it requires extraction of high-level semantic information from low level audio or visual data. 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. Video abstract can also use as an end product to be shared, digest and enjoyed by the user. Only essential information of a view sequence improves storage, bandwidth and viewing time. Video abstraction is the process has three phases: Video information analysis, meaningful clip selection and output synthesis. RGB histogram is used to provide distribution information of colors for given video frame.

III. COMPARISON

Parameter	Color Layout Descriptor (CLD)	Aggregation Method
Features	Pixel difference, histogram difference, template matching edge change	RGB color space, color histogram, moment of inertia
Input	Frames are converted into 64 coefficient and then used for computation	Whole frame size is used directly for computation
Threshold Value	No need to predefine	According to input videos, it must be defined
Efficiency	If total number of frames are increases, it will not affect efficiency	If total number of frames are increases, then efficiency decreases
Execution Time	Less	More

Table 1. Comparison between static video summarization techniques

IV. PROPOSED WORK

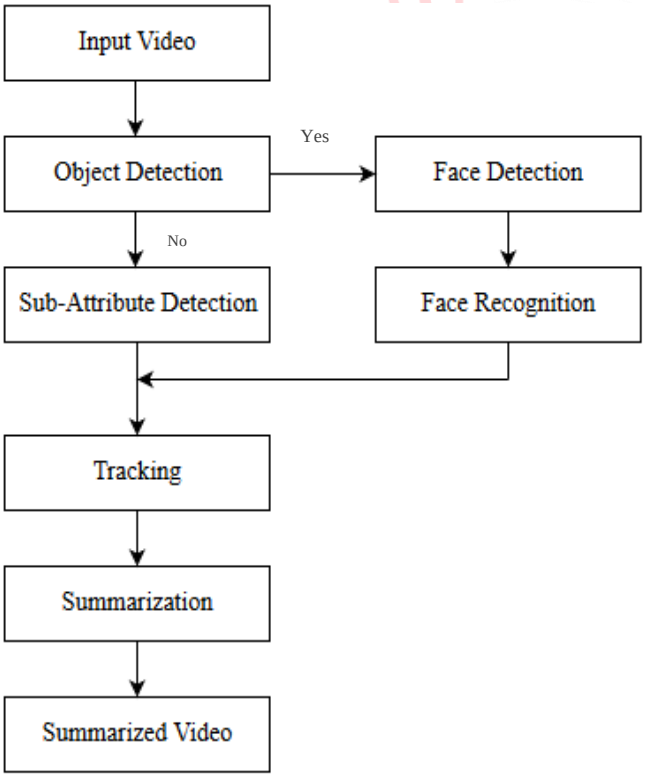

Fig 1. Flowchart for video summarization

In this paper, the architecture of proposed video summarization technique is shown in fig.1. The input video is taken by the user. That input video contains number of frames. That video is given to the module which works on it and gives the summarized video. Now our task is to face and object detection in every frame. Object detection, face recognition is done using the OpenCV. OpenCV is the trained module used for the computer visualization. Object detection and face detection are work on the frames, so that the frames are extracted from the input video. By using VGG16 module the key frames are extracted. The VGG16 is the pre-trained module used for extracting the frames from the input video.

The second section of this project is, after extracting frames, object detection and face recognition. After this, on the basis of color, cluster, data of frames, the necessary key frames are extracted and that key frames are converted into the video format which is very interesting to the user.

V. CONCLUSION

In this work, we presented an effective video summarization framework for video by finding correlation between different frames. This is done by object detection. Face detection and sub attribute detection in every frame. This method allows to user to generate short relevant video from a long video.

REFERENCES

[1] Tinumol Sbastian, Jiby J Puthiyidam, A Survey on Video Summarization Techniques, International Journal of Computer Applications (0975-8887).

[2] Manasa Srinivas, M. M. Manohara Pai, Radhika M. Pai, An Improved Algorithm for Video Summarization – A Rank Based Approach, Procedia Computer Science 89 (2016) 812-819.

[3] Y. J. Lee, J. Ghosh and K. Grauman, "Discovering important people and objects for egocentric video summarization," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1346-1353, June 2012.

[4] Wei REN, Yuesheng ZHU, "A Video Summarization Approach based on Machine Learning," International Conference on Intelligent Information Hiding and Multimedia Signal Processing

[5] Rajkumar Kannan, Gheorghita Ghinea, Sridhar Swaminathan, Suresh Kannaiyan, "Improving Video Summarization based on User Preferences"

[6] L. Li, K. Zhou, G. R. Xue, H. Zha, and Y. Yu, "Video summarization via transferrable structured learning," In Proceedings of the 20th international conference on World wide web, pp. 287-296, March 2011.

[7] B. Gong, W.-L. Chao, K. Grauman, and F. Sha, "Diverse sequential subset selection for supervised video summarization," in Advances in Neural Information Processing Systems, 2014, pp. 2069–2077.

[8] Sinnu Susan Thomas, Sumana Gupta, Venkatesh K. Subramanian, "Smart Surveillance Based on Video Summarization"

[9] Y.-F. Ma, X.-S. Hua, L. Lu, and H.-J. Zhang, "A generic framework of user attention model and its application in video summarization," IEEE transactions on multimedia, vol. 7, no. 5, pp. 907–919, 2005.

[10] J. You, G. Liu, L. Sun, and H. Li, "A multiple visual models based perceptive analysis framework for multilevel video summarization," IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 3, pp. 273–285, 2007.