

# A Survey on Video Summarization Techniques

Suvarna Patil<sup>1</sup>, Dhanashree Phalke<sup>2</sup>

<sup>1</sup>Student, <sup>2</sup>Professor, Computer Engineering, DYPCOE, Akurdi, Pune)

**Abstract:** This paper provides a survey on video summarization techniques. The video summarization techniques are mostly applied or used in surveillance data. The challenging task of surveillance video is to watch full video because it generates the data in huge amount. By apply summarization technique; it becomes more shorten then original and easy to watch. Video summarization has been proposed to improve faster browsing of large video collection and more efficient contents indexing and access.

**In this survey,** offer a qualitative and elucidatory survey on recent development in the field of Video Summarization. This paper reports on different Video Summarization method. Performance of some selected methods as compared.

**Keywords:** Video Summarization, Video Skimming, Key frames.

## I. INTRODUCTION

Now days, video surveillance technology has become in every sphere of our life. But automated video surveillance generates huge quantities of data, which ultimately does rely upon manual inspection at some stage. 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.

Improvement in video capturing devices and increasing popularity of social media, there are huge volumes of videos being captured and uploaded every second. Creating highlights of sports matches and many more are the fields where video summarization plays an important role. For example the sports highlights from cricket video using techniques of character reorganization. A short video summary is synthesized that includes the frames for the significant events termed as Highlights. The process of sports highlights generation is automated resulting in a summary for the viewer that reduces the time and space requirements [2][8].

Due to the increasing volume of video content on the Web, and the human effort in use to process it, new technologies need to be researched in order to develop well-organized indexing and search techniques to manage effectively and efficiently the huge amount of video data. As the name implies, video summarization is a mechanism to produce a short summary of a video to offer to the user a artificial and helpful visual abstract of video sequence, it can either be a images (key frames) or moving images (video skims).

Video summarization can be represented into two modes: A static video summary (story-board) and a dynamic video skimming. In one hand, static video summary represents a video sequence in a static imagery form (one or more selected representative frames from the original video, or a synthesized image generated from the selected key frames).

### A. Static Video Summarization

This is also called a key frame based video summarization techniques or still image abstract or storyboard. There are some criteria that come across for key frame based techniques, which are as follows: 1. Redundancy: frames with minor difference are selected as key frame. 2. When there are various changes in content it is difficult to make clustering.

The key frame based summarization can be classified in three different ways

- 1) Classification based on sampling: It selects key frame uniformly or randomly without considering the video content.
- 2) Classification based on scene segmentation: It extracts key frames using scene detection; it includes all semantic links in the video.
- 3) Classification based on shot segmentation: It extracts first image and last image as a shot key frame[7][17].

### B. Dynamic video Summarization

The idea of dynamic summarization called as video skimming is a short video composed of informative scenes from original video presented to the user to receive in video format that is it condenses the original video into shorter form while preserving the important content of a video in short time. It also preserves the motion information[9].

This work presents an updated survey on recent development and focus on the future research and development trends in video summarization field. First discuss the review of literature in Section II. In Section III, discuss the different methods for video summarization techniques. In Section IV, compare the selected methods. Finally, concluding remark will be given in Section V.

## II. LITERATURE SURVEY

For better understanding of the difference between the various approaches for the static video summarization and dynamic video summarization which can be found in the literature are discussed next.

Sandra Eliza Fontes de Avila et al [3] proposed VSUMM, a methodology for the creation of static video summaries. It is simple and effective approach for automatic video summarization. The method is based on the extraction of color features from video frames and unsupervised classification and also added new methodology for evaluation the video summarized called as comparison of user summarized i. e. CUS. In this method, the summaries are made by users and compared with other approaches.

S. Zhang, Y. Zhu et al.[4] 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.

Sinnu Susan Thomas, Sumana Gupta et al. [5] proposed a new framework for video summarization i. e. perceptual video summarization. Author introduces for the first time features of Human visual system (HSV) and allow for the emphasis perceptually significant event while concurrently removing perceptual redundancy from the summaries. Author proposes to create an image like panorama registration based on some superior criterion for the choice of the reference frame from the video.

Yifang Yin, Roshan Thapliya et al [6] proposed method for automatic video summary generation with personal adaptation. 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 semantic tree which is more effective for personalized adaptation. In the proposed design of video summarization, it joins the personal interest and visual attention.

Z. Lu and K. Grauman et al.[10] proposed a video summarization approach that discovers the story of an egocentric video. This work for the long video which selects a short chain of video sub-shots depicting the essential events. Author adapts a text analysis technique that connects a new article to visual domain. They also show that how to establish the influence of one visual event on another given their respective objects. Author introduces a novel temporal segmentation method design for egocentric video. They perform a large scale video on the proposed approach which has better sense of story.

Y. J. Lee and K. Grauman et al.[12] proposed a video summarization approach for egocentric or wearable camera data. The proposed method produces a compact storyboard summary of the camera's wearer's day. The proposed method works as first, train a regression from labeled training videos that scores any region to an important person or objects. Author proposes two ways to adjust the compacted of the summary based on either target importance criteria of target summary length the main contribution of this approach is driven by predicted important people and objects.

Rav-Acha, Y. Pritch, and S. Peleg et al. [13] introduced dynamic video synopsis where most of the activities in the video is combined by simultaneous viewing several actions even when that are originally occurred at different time. Author presented two approaches. First approach uses low level graph optimization where each pixel in the synopsis in the video is a node in the graph. In this approach directly obtain the synopsis but the complexity is high. In the second approach directly detect moving object and perform optimization on the detected object.

Y. Hadi, F. Essannouni, and R. O. H. Thami et al. [15] proposed a video summarization algorithm by multiple extraction of key frames in each shot which is based on the K-medoid clustering algorithm for finding the best representative frame for each video shot. In this work the distance between frames is estimated using fast full search block matching algorithm based on the frequency domain. In addition, this approach generates different key frames even in the presence of large motion.

M. Gygli, H. Grabner, H. Riemenschneider, and L. V. Gool et al. [16] proposed a novel approach and a new benchmark for video summarization.

They proposed the novel temporal super frame segmentation for user video and methods to generate informative summary from them. Author proposes a new method to calculate the interestingness of superframe and selecting summary from using 0/1 knapsack optimization. By evaluation proposed method it is generally shown that to create good automatic summaries

## III. VIDEO SUMMARIZATION APPROACHES

### A. Uniform Sampling

It is one of the most primitive algorithms for the video summarization. In this method, basically select every  $K^{\text{th}}$  frame in the summary, where  $K$  is mentioned by the summary length. Uniform sampling often used as baseline score and it is simple in execution. This approach is based on the super frames.

### B. VSUMM Approach

This method used to create the static video summary. It is simple and effective approach for automatic video summarization. The method is based on the extraction of color features from video frames and unsupervised classification and also added new methodology for evaluation the video summary called as comparison of user summaries.

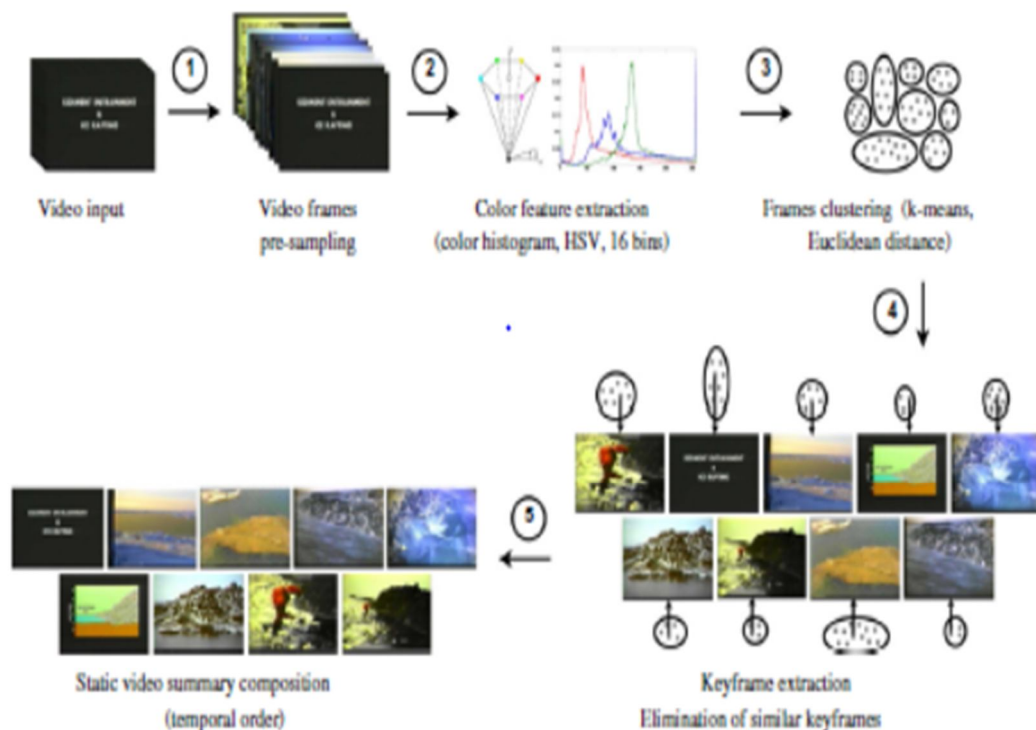


Figure: 1VSUMM approach[3]

In this approach, to produce static video summaries first, the input video is split into frames. In second step, color features are extracted to form a color histogram in HSV color spaces. In addition, the meaningless frames are removed. In step 3, the frames are combined by K-mean clustering algorithm. In step 4, One frame per cluster is selected I.e. key frame. In step 5 to make the key frames which is arranged in temporal order.

### C. Live Light Approach

Live light means Online Video high Lighting (Live Light). In this work, author proposes online video highlighting for generating short video summary of an unedited and unstructured video. Live light enables viewer to understand the video without watching the entire sequence. The live-light scans through the video stream split into a collection of video segments

After processing of first few segments, it starts to make its own dictionary which will be kept updated for new video segment, the current version of dictionary sparsely reconstructed using group sparse coding. If the small reconstruction error occurs that means content is already in current dictionary and this segment excluded from summary and the algorithm mores to next segment.

On other hand, if the new video segment cannot be sparsely reconstructed that means high resolution error is occurred which mean that the unknown content from previous video data and update the dictionary according to video data [11].

### D. KVS Approach

KVS means kernel video summarization approach. In this work firstly, perform an automatic kernel based temporal segmentation which is based on state of the art video frames. This feature automatically selects the number of segments. Then, it is equipped with sum classifier for importance scoring for each segment. Finally to make the video summary of the segments with the highest predicted importance score of frames [14].

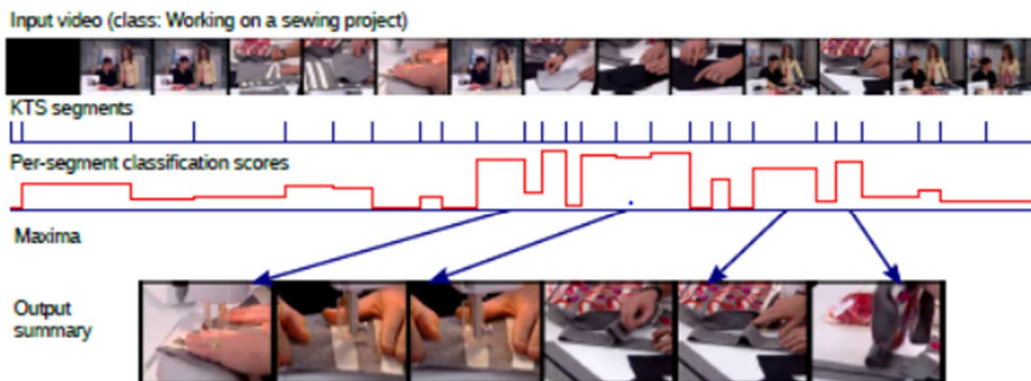


Figure: 2 KVS approach [14]

### E. Clustering Approach

Video summarization algorithm by multiple extraction of key frames in each shot, which is based on the K-medoid clustering algorithm for finding the best representative frame for each video shot. In this work, the distance between frames is estimated using fast full search block matching algorithm based on the frequency domain. In addition, this approach generates different key frames even in the presence of large motion [15].

## IV. RESULT ANALYSIS

Generally, the metrics such as accuracy, precision, recall and F1-measure is used to evaluate performance of classification. Accuracy metric is widely used in machine learning fields, which indicates the overall performance of classification. The precision is the percentage of documents that are correctly classified as positive out of all the documents that are classified as positive, and the recall is the percentage of documents that are correctly classified as positive out of all the documents that are actually positive.

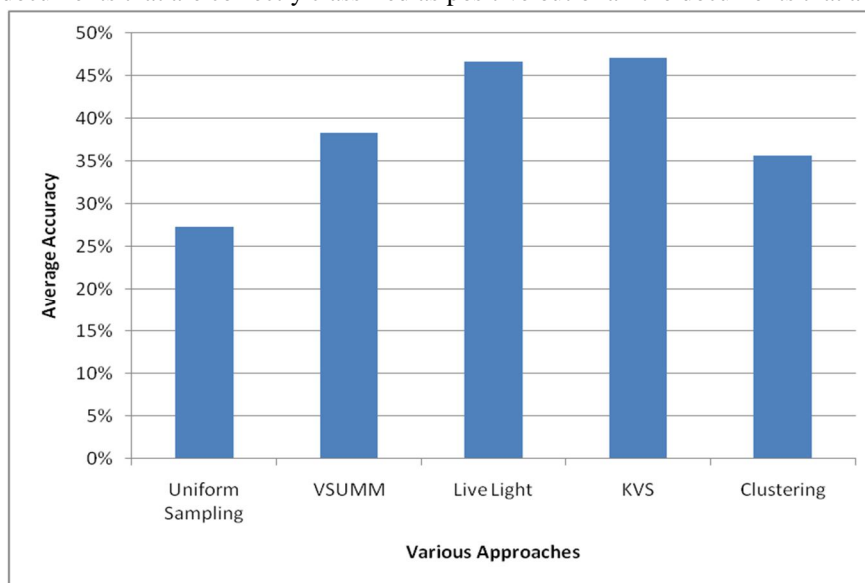


Figure: 3 Chart showing the weighted average accuracy of Existing various video summarization approaches [1]

## V. SUMMARY

This paper is a survey of the current research on various approaches in video summarization. In this work, carried out a review of the research in two main form of the video summarization techniques i.e. Static Video Summarization also known as Key-frame-based method and Dynamic Video Summarization also known as Video Skimming. The average accuracy of various approaches as mentioned in the chart.

## VI. ACKNOWLEDGMENT

The authors would like to thank the publishers and researchers for making their resources available and also thank the college authority for providing the required infrastructure and support. Finally, would like to extend our heartfelt gratitude to friends and family members.

## REFERENCES

- [1] Xuelong Li, Fellow, IEEE, Bin Zhao, and Xiaoqiang Lu, Senior Member, IEEE "A General Framework for Edited Video and Raw Video Summarization", IEEE.
- [2] Ana Garcia del Molino "Summarization of Egocentric Videos: A Comprehensive Survey", IEEE transactions on Video Technology 2016.
- [3] Sandra Eliza Fontes de Avila, Ana Paula Brando Lopes, Antonio da Luz Jr. "VSUMM : A mechanism design to produce static video summaries and novel evaluation method", ELSEVIER 2010.
- [4] S. Zhang, Y. Zhu, and A. K. Roy-Chowdhury "Context-aware surveillance video summarization.", IEEE transactions on Video Technology 2015.
- [5] Sinnu Susan Thomas, Sumana Gupta "Perceptual video summarization - a new framework of video summarization", IEEE transactions on image processing 2017.
- [6] Yifang Yin, Roshan Thapliya "Encoded Semantic Tree for Automatic User Proling Applied To personalized Video Summarization", IEEE transactions on Video Technology 2016.
- [7] P. Mundur, Y. Rao, and Y. Yesha "Keyframe-based video summarization using Delaunay clustering.", IEEE 2006.
- [8] Muhammad Ehsan Anjum, Syed Farooq Ali, Malik Tahir Hassan, Muhammad Adnan "Video Summarization Sports Highlights Generation", IEEE conference 2014.
- [9] Y. Zhang, G. Wang, B. Seo, and R. Zimmermann "Multi-video summary and skim generation of sensor-rich videos in geo-space," ACM-2012.
- [10] Z. Lu and K. Grauman "Story-driven summarization for egocentric video", IEEE Conference CVPR 2013.
- [11] B. Zhao and E. P. Xing, "Quasi real-time summarization for consumer videos," in IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2513–2520.
- [12] Y. J. Lee and K. Grauman "Predicting important objects for egocentric video summarization", International Journal of Computer Vision, 2015.
- [13] A. Rav-Acha, Y. Pritch, and S. Peleg "Making a long video short: Dynamic video synopsis", IEEE Computer Society Conference CVPR 2006.32
- [14] D. Potapov, M. Douze, Z. Harchaoui, and C. Schmid, "Category-specific video summarization," in European Conference on Computer Vision, 2014, pp. 540–555.
- [15] Y. Hadi, F. Essannouni, and R. O. H. Thami "Video summarization by k-medoid clustering", Research Gate 2006
- [16] M. Gygli, H. Grabner, H. Riemenschneider, and L. V. Gool "Creating summaries from user videos", Springer 2014.
- [17] Padmavathi Mundur, Yong Rao, Yelena Yesha "Keyframe-based video summarization using delaunay clustering", Springer 2006