International Journal of Current Engineering and Technology
©2021 INPRESSCO®, All Rights Reserved

E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
Available at http://inpressco.com/category/ijcet

*Research Article*

# Video Summarization using Keyframe Extraction

**Ajay Mushan and Prof. Pujashree Vidap**

Department of Computer Engineering, PICT, Pune

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

### *Abstract*

*Video summarization plays important role in too many fields, such as video indexing, video browsing, video compression, video analyzing and so on. One of the fundamental units in the video structure analysis is the key frame extraction, Key frame provide a meaningful frames from the video. Key frame consists meaningful frame from the videos which helps for video summarization. In this proposed model, we present an approach which is based on Convolutional Neural Network, keyframe extraction from videos and static video summarization. First, the video should be converted to frames. Then we perform redundancy elimination techniques to reduce the redundancy from frames. Then extract the keyframes from video by using Convolutional Neural Network (CNN) model. From extracted keyframe we form a video summarization. The results obtained from this proposed model is compare with the VSMM model. The proposed model performed on VSUMM dataset. The results are compared with the VSUMM model duration and video static summary. In Convolutional Neural Network (CNN) model has trained by using 50 salad datasets to summarize the videos related to cooking video.*

***Keywords:*** *Video Summarization, Keyframes, Convolutional Neural Network*

## Introduction

One the most effective and efficient way for capturing and storing digital media in today's world is video. A simple and effective key frames extraction method is represented, to generate static video summaries. Key-frames, also called representative frames, are defined as the most informative frames that capture the major elements in a video in terms of content. Key-frames can generate summaries of the videos to provide browsing capabilities to users [12]. Key framing is used to summarize essential video content in short time. The rapid development of digital video capture and editing technology has led to enormous increase in video data, creating the need for effective techniques for video retrieval and analysis. Video summary is a temporally condensed representation of a video. The purpose of the video summary evolves mainly due to viewing time constraints[13]. It helps us to assess the relevance or value of information within a shorter period of time while decision making process[13]. It also plays prime role where the resources like storage, communication bandwidth and power are limited. It finds main applications in security, military, and entertainment. A MPEG (Moving Picture Expert Group) video is segregated into many still images. Such still images are called frames. The unit of a video is frame. For effective multimedia data management video summarization techniques are attracted a lot. There is a requirement of quality measurement for summarizing a video. The primary step of key frame extraction is shot segmentation which basically deals with detecting the transition between successive slots. Video summarization has become a very essential procedure that provides the faster browsing of a large video. Basically, the techniques that are used for video summarization are static video summarization (video summary) and dynamic video summarization (video skimming). If we want to reduce a lengthy, archived video summary static video summarization (video summary) is a fast and powerful application. There are different approaches are applied for different requirements like sports, nature, etc. Video summarization methods has recently attracted researchers and is now an emerging research field[9]. Most of the methods make use of low-level features.

The visual features are the commonly used low- level features. The low-level features can only capture global characteristics of the frames. Color feature, texture, mutual information, motion information and fuzzy color histogram[15]. The high-level feature vectors are extracted using CNN. After that by using the high-level feature vectors to summarize the video. High level features can SIFT, interest point from image, edge detection, etc. Video summarization is one of the most important topics, which potentially enabling faster browsing of large video collections and also more efficient content indexing and access. Essentially, this research area consists of automatically generating a short summary of a video, which can either be a static summary or a dynamic summary[11]. Static video summaries are composed of a set of keyframes extracted from the original video, while dynamic video summaries are composed of a set of shots and are produced taking into account the similarity or domain-specific relationships among all video shots[11].

## Literature Survey

Hana Gharbi, Sahbi Bahroun, et. al, In their paper, they presented a new novel approach for video summarization from keyframes, they are calculated interest point on local description and repeatability matrix to reduce the redundancy then they apply PCA and HCA to extract the keyframes. They try to give the visual summary based on most representative object in the video database[1]. Hana Gharbi, Sahbi Bahroun, et. al, In their paper, They presented a new approach for key frame extraction based on local image description "interest points", repeatabily matrix and modularity and introducing Graph Modularity Clustering method. The results prove that local description can be good alternative in keyframe extraction field[2]. Chaohui Lv, Yiyang Huang, et. al, In this paper, They proposed an improved clustering method based on video feature, and it proves to be effective in keyframe extraction. In their approach the remove the blurry frames by motion blur detection and apply the clustering algorithm to extract the keyframes. It can effectively get the main content information in personal videos and extremely saves process time for video research[3]. Muhammad Asim, Noor Almaadeed, et. al, In their paper, a simple and effective keyframes extraction method was represented, to generate static video summaries. Their method is based on comparing color features extracted from patches of frames, to detect groups of video frames with a similar content. The proposed method can effectively detect sharp and gradual cuts, in a video sequence. Experiments were conducted on a benchmark video dataset for summarization to validate the performance of the proposed approach[4]. Dipti Jadhav and Udhav Bhosle, In their paper, They proposes an efficient static video summarization method using SURF features of the frames. The SURF features is used to generate a video summary that can capture salient frames and actions in the video. Their approach is experiment on the open video dataset V37 and V24 The proposed algorithms can be used across various general of videos[5]. S. S. Thomas, S. Gupta, et. al, In their paper, they proposed a model which is used to smart surveillance. Here the is some predefined vector that is predefine feature vectors are the and surveillance frame are match with these vector if match is found it will take it as keyframe and video retrieval. If not match then it discard the keyframe. It only take similarity frames and make video summarization[6]. H. H. Phadke and H. Mallika, et. al, In their paper they proposed model, in which they extract the text from videos like TV news, and a frame containing text take it as keyframe. The segmentation performed on the frame and extract the text and take it as keyframe. After extraction of keyframe localization is done[7]. Sanjoy Ghatak, In their paper, They discussed on various key frame extraction methods from video. The comparison between frames and calculating to extract the keyframe or not based on threshold value. If calculated value if greater than threshold simply discard and if value is less than threshold take it as keyframe[8].

## Proposed Methodology

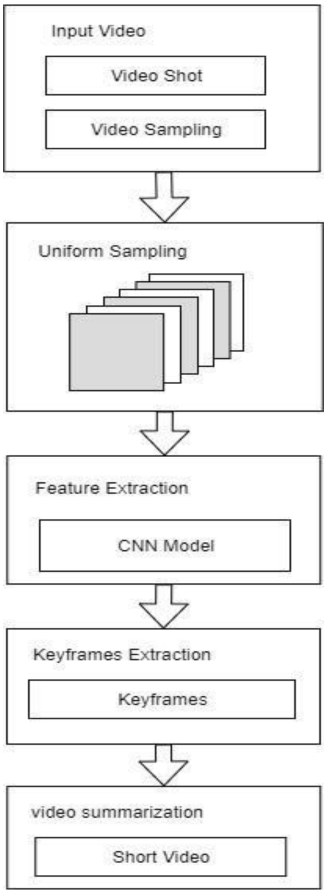

Fig 1 : Overview of the proposed method

In this proposed system, Firstly, we take input as a video, then we extract the frame from videos. After that we perform the Uniform sampling on videos to reduce the redundancy cPGCON 2020 (Post Graduate Conference for Computer Engineering) among frames. After reduction of frames, we extract the features from frame by using CNN. And extracted features compare with model and extract the keyframes. After extraction of keyframe we make summarize video of frames. In this proposed methodology, we are using CNN as shown if Fig. 1, so we extract the high level features from the video which extract the meaningful frames for video summarization.

### *A. Algorithms*

The steps to be followed are:

Step 1 : Take a video as input.
In this we take a video as input.The video is converted into a video shots then the videos shorts will convert into sequence of frames.
Step 2 : Convert a frames into Uniform Sampling. We convert a frames into uniform sampling by taking frame per second. At particular time we take a frame from a sequence of frames. In this way we reduce the redundancy frames from a sequence of frame.
Step 3 :Feature Extraction.
We extract the features from frame by using convolution neural networks.
Step 4 : Feature extraction by CNN.
We extract the features from the frames.
Step 5 : Calculate similarity measure between frame. If threshold is minimum take it as keyframe and of threshold is greater take it as non keyframe.
Step 6 : Now make a video summary of selected keyframes.

The proposed model is used to generate a static video summary of video. The results obtained from this proposed model is compare with VSUMM model[16]. The results obtained in duration of summarized video, precision ,recall and f-measure. The results obtained from the proposed model is comparing with the static summarization video in VSUMM dataset. The comparative analysis is done.

## Dataset

In this dataset, 50 salads[10], it contains videos and accelerometer data. This dataset captured by 25 people preparing 2 mixed salads each and contains over 4h of annotated accelerometer and RGB-D video data[10]. Annotated activities correspond to steps in the recipe and include activity class, activity phase (pre-/ core-/ post), and the ingredient acted upon[10].

The dataset description :

1. RGB video data 640x480 pixels at 30 Hz.
2. Depth maps 640x480 pixels at 30 Hz
3. 3-axis accelerometer data at 50 Hz of devices attached to a knife, a mixing spoon, a small spoon, a peeler, a glass, an oil bottle, and a pepper dispenser.
4. Synchronization parameters for temporal alignment of video and accelerometer data.
5. Annotations as temporal intervals of pre- core- and post- phases of activities corresponding to steps in a recipe[10].

## Experimental Results

For performance evaluation, we will use accuracy and comparation ratio. In accuracy, the overall performance of a system will be tested and accuracy is compared with the existing approach and do the comparative analysis. Keyframe extraction result should not contain many key frames in order to avoid redundancy, That's why we should evaluate the compactness of the summary[1].The compression ratio is computed by dividing the number of key frames in the summary by the length of video sequence[1]. The performance of the automatically generated video summaries is measured by three objective measures[], namely, called Recall, Precision and F-measure [14], defined below:

$$Recall = \frac{\text{matching keyframes}}{Nus} \qquad (1)$$

$$Precision = \frac{\text{matching keyframes}}{Nas} \qquad (2)$$

$$F_{measure} = \frac{2 * Precision * Recall}{Precision + Recall} \qquad (3)$$

Where Nus and Nas represents keyframes selected by users and automatic summaries.

## Conclusions

We discussed on various key frame extraction methods for video summarization and found that most of the methods use motion as one of the important features and few use visual features like global features and local features. In this proposed system, we are using CNN which extract the high-level feature, CNN provides better results as compared to other algorithms. In the future, we can use the different datasets to make comparative analysis and comparing results with other approaches from key frame extraction and try to build system, which can gives better summarization of video.

## References

[1]. Hana Gharbi, Sahbi Bahroun and Ezzeddine Zagrouba, "A Novel Key Frame Extraction Approach for Video Summarization", VISAPP 2016 - International Conference on Computer Vision Theory and Applications.

[2]. Hana Gharbi, Sahbi Bahroun and Ezzeddine Zagrouba, “Key Frames Extraction Using Graph Modularity Clustering For Effective Video Summarization", 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[3]. Chaohui Lv, Yiyang Huang, “Effective Keyframe Extraction From Personal Video By Using Nearest Neighbor Clustering", 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2018).

[4]. Muhammad Asim, Noor Almaadeed and Azeddine Beghdadi, “A Key Frame Based Video Summarization using Color Features", 2018 Colour and Visual Computing Symposium (CVCS).

[5]. Dipti Jadhav and Udhav Bhosle, "SURF based Video Summarization and its Optimization", International Conference on Communication and Signal Processing, April 6-8, 2017, India.

[6]. S. S. Thomas, S. Gupta and V. K. Subramanian, "Smart surveillance based on video summarization", 2017 IEEE Region 10 Symposium (TENSYMP), Cochin, 2017.

[7]. H. H. Phadke and H. Mallika, "Key frame extraction, localization and segmentation of caption text in news videos," 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT), Bangalore, 2017.

[8]. Sanjoy Ghatak, "Key-Frame Extraction Using Threshold Technique", International Journal of Engineering Applied Sciences and Technology, Issue 8, ISSN No. 2455-2143, Pages 51-56, 2016 Vol.

[9]. Ch, Sujatha & Uma, "A Study on Keyframe Extraction Methods for Video Summary", Proceedings - 2011 International Conference on Computational Intelligence and Communication Systems, CICN 2011. 10.1109/CICN.2011.15.

[10]. Sebastian Stein and Stephen J. McKenna," Combining Embedded Accelerometers with Computer Vision for Recognizing Food

[11]. Preparation Activities", the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2013).

[12]. Sandra E. F. de Avila, Ana P. B. Lopes, Antonio da Luz Jr., Arnaldo de A. Araújo, "VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method" ,Pattern Recognition Letters, Volume 32, Issue 1, January 2011, pages 56– 68.

[13]. Shaoshuai Lei, Gang Xie, and Gaowei Yan, "A Novel Key-Frame Extraction Approach for Both Video Summary and Video Index" Hindawi Publishing Corporation Scientific World Journal Volume 2014.

[14]. Deepali Bhawarthi, Prof. Shriniwas Gadage, "Enriching Feature Extraction Using A-priory Algorithm for Cricket Video" , International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 3, MayJun 2012, pp.976-979.

[15]. J. Almeida, N. J. Leite, and R. D. S. Torres, "VISON: VIdeo Summarization for ONline applications," *Pattern*

[16]. *Recognit. Lett.*, vol. 33, no. 4, pp. 397–409, 2012.

[17]. Mundur, Padmavathi & Rao, Yong & Yesha, Yelena. (2006). Keyframe-based video summarization using Delaunay clustering. Int. J. on Digital Libraries. 6. 219-232. 10.1007/s00799-0050129-9.

[18]. Sandra Eliza Fontes de Avila and Ana Paula Brandão Lopes and Antonio da Luz Jr. and Arnaldo de Albuquerque Araújo , "VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method" , Pattern Recognition Letters, volume 32, 2011.