International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-9 Issue-2, July 2020

Video Summarization using Keyframe Extraction Methods

Ajay Mushan, P. S. Vidap

Abstract: ***Video summarization plays an 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 keyframe extraction, Keyframe provides meaningful frames from the video. The keyframe consists of the meaningful frame from the videos which help for video summarization. In this proposed model, we presented an approach that 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 the Convolutional Neural Network(CNN) model. From the extracted keyframe, we form a video summarization.***

Keywords: ***Video Summarization, Keyframes, Convolutional Neural Network, key frame extraction, interest point.***

I. INTRODUCTION

In today's world, One of the most effective and efficient ways of capturing and storing digital media is video[4]. Key-frames are called representative frames[12]. Keyframes contents the most informative frames, which may capture the main information in a video in terms of content. Keyframing is used to summarize essential video content in a short time. Keyframes can generate summaries of the videos to provide browsing capabilities to users[12]. There are two techniques which are used for video summarization are static video summarization that is video summary and dynamic video summarization that is video skimming. If we want to reduce a length of the video which is archived by video summary static video summarization[14] that is video summary, is a fast and powerful application[15]. There are different approaches are applied for different requirements like sports, nature, etc. Video summarization methods have recently attracted most of the researchers. The high-level feature vectors that is deep features are extracted using CNN. After that by using the high-level feature vectors to summarize the video[14]. High-level features can SIFT, interest point from an image, edge detection, etc.

As shown in following figure 1, We are passing the input video which is segmented into shots using the shot change detection techniques by identified the shot, the keyframes can be extracted from the candidate frames which is used to represent each shot[1]. At the end, all the keyframes can be clubbed together to create a video summary. Video summary Which will represent the video as a whole[1].

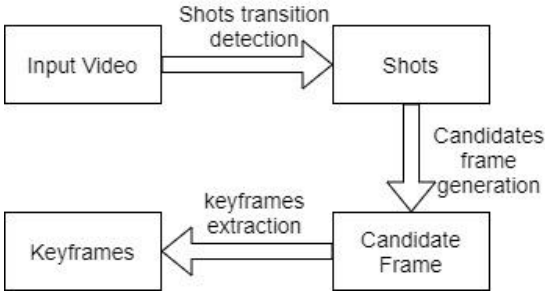

Figure 1. Key frames extraction steps[1].

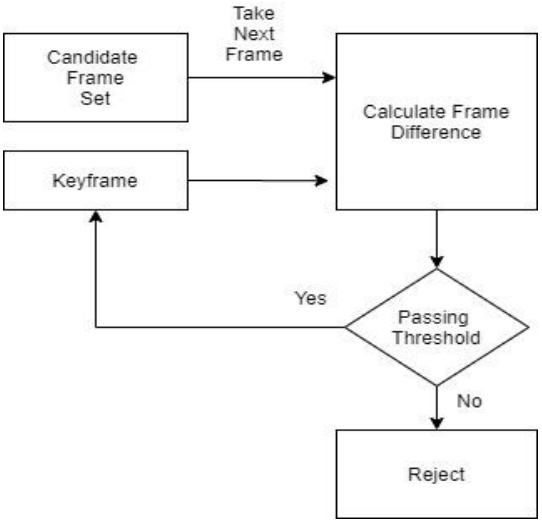

Figure 2. Selection of Keyframes.

In above figure 2, The frames are sequential, initial frames are compare with next frames if frame threshold value is grater than we take it as keyframe. Otherwise, we reject the frame.

II. LITERATURE SURVEY

H. Gharbi, S. Bahroun and Ezzeddine Zagrouba, 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 a visual summary based on the most representative object in the video database[1].

Manuscript received on May 25, 2020.
Revised Manuscript received on June 29, 2020.
Manuscript published on July 30, 2020.
***** Correspondence Author
Ajay Mushan, Student, Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India. E-mail: ajaymushan1519@gmail.com
Prof. Pujashree Vidap, Associate Professor, Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India.

Hana Gharbi, Sahbi Bahroun, et. al, In their paper, They were presented a new approach for keyframe extraction method which is based on local image description that is "interest points", repeatability matrix for redundancy remove between the frames and modularity and introducing Graph Modularity Clustering method.

The obtained results proved that local descriptions can be a good alternative for the keyframe extraction field[2].

Chaohui Lv, Yiyang Huang, et. al, In this approach, They proposed an improved clustering method which is based upon local features. It proved that the in keyframe extraction were effective and efficient. In their approach, the removes the unwanted frames by motion blur detection[3] and apply the clustering algorithm to extract the keyframes. It can effectively give the valuable information from the openly available videos and extremely saves processes for video anylsis time required for video research[3].

Dipti Jadhav and Udhav Bhosle, In their paper,Authors proposed an static video summarization method by using SURF features[4] of the frames. SURF is used to calculate interest point in the frame. Their approach was an experimented on Open Video Dataset(OVD) V37 and V24. The proposed algorithms can be used various common videos which available[5].

S. S. Thomas, S. Gupta, et. al, In their paper, they proposed a model that is used to smart surveillance. There was some predefined vector that predefined feature vectors are the and surveillance frame are a match with feature vector if a match was found it will take it as keyframe and video retrieval. If not match then it discards the keyframe. It only takes similarity frames and makes video summarization[6].

H. H. Phadke and H. Mallika, et. al, In their paper they proposed a model, in which they extract the text from videos like TV news, and a frame containing text take it as a keyframe. The segmentation performed on the frame and extract the text and take it as a keyframe. After extraction of a keyframe, localization is done[7].

Sanjoy Ghatak, In their paper, They discussed on various keyframe extraction methods from a video. The comparison between frames and calculating to extract the keyframe or not based on a threshold value. If calculated value if greater than the threshold simply discard and if a value is less than threshold take it as keyframe[8].

III. PROPOSED METHODOLOGY

The Proposed methodology steps as follows:

Step 1: Take a video as input.
In this we take a video as input.

Step 2: Convert a video into frames.
We convert a video into frames

Step 3: Redundancy Elimination.
We eliminate the same frames by using similarity measure techniques and reduce the redundancy among frames.

Step 4: Feature extraction by CNN.

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.

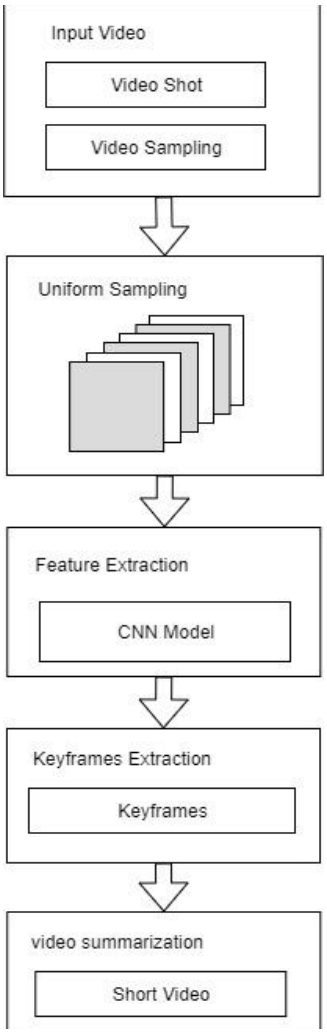

Figure 3. Proposed methodology

IV. DATASET

In this dataset, 50 salads[9], it contains videos and accelerometer data. This dataset created by 25 people. The dataset contains over 4h of annotated accelerometer and RGB-D video data[9].

V. EXPERIMENTALRESULTS

For performance evaluation, we will use the 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. The keyframe extraction result may contain many keyframe which is remove by keyframe analysis , That is a respond to evaluate the redundancy of number of frames of the summary[1].

The performance of the automatically generated video summaries is calculated by following objective measures, namely, called Recall, Precision and F-measure[14], which is defined as below:

$Recall$ = matching keyframes / Nus (1)

$Precisio$ = matching keyframes / Nas (2)

$Fmeasure = 2*Precision*Recall \,/\, Precision+Recal$(3)

Where Nus and Nas represent keyframes choose by users and automatic generated summaries.

VI. CONCLUSION

We discussed various types of keyframe extraction methods for video summarization, and by analysing method we observe that most of the methods use motion as a important features and very less methods 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 a comparative analysis and comparing results with other approaches from keyframe extraction and try to build a system, which can give better summarization of video.

REFERENCES

1. H. Gharbi, S. Bahroun and Ezzeddine Zagrouba, "A Novel Key Frame Extraction Approach for Video Summarization", VISAPP 2016 -International Conference on Computer Vision Theory and Applications.
2. H. Gharbi, S. Bahroun and Ezzeddine Zagrouba, "Key Frames Extraction Using Graph Modularity Clustering for Effective Video Summarization", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2017.
3. Chaohui Lv, Yiyang Huang, "Effective Keyframe Extraction From Personal Video By Using Nearest Neighbor Clustering", 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", Colour and Visual Computing Symposium (CVCS) 2018.
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 Preparation Activities", the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2013).
11. 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.
12. 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.
13. J. Almeida, N. J. Leite, and R. D. S. Torres, "VISON: Video Summarization for ONline applications," Pattern Recognit. Lett., vol. 33, no. 4, pp. 397–409, 2012.
14. Mundur, Padmavathi & Rao, Yong & Yesha, Yelena. (2006). Keyframe-based video summarization using Delaunay clustering. Int. J.on Digital Libraries. 6. 219-232. 10.1007/s00799-005-0129-9.
15. Shivangi Pandey, Prashant Dwivedy, Sunil Meena, Anjali Potnis. "A survey on key frame extraction methods of a MPEG video", 2017 International Conference on Computing, Communication and Automation (ICCCA), 2017.

AUTHORS PROFILE

Ajay Mushan, Completed engineering from Walchand Institute of Technology, under the Solapur University. Currently pursuing Masters in Computer Engineering from Pune
Institute of Computer Technology under Savitribai Phule Pune University, Pune.

Prof. P. S. Vidap, Associate professor at Pune Institute of Computer Technology, Her research areas are Compiler Design, Pattern Recognition ,Computer Vision.