

Implementation of Video Summarization using Deep Learning Techniques

Mr. Sushant Savita Madhukar Gandhi¹, Dr. Mukesh Shrimali² and Dr. Pradip Mane³

¹PhD Scholar, Pacific University

²Directors, Udaipur Pacific University, Udaipur

³Associate Professor, VPPCOE&VA, Mumbai

Email: Gandhisushant67@gmail.com, mukesh_shrimali@yahoo.com, pradipmane510@gmail.com

Abstract -A very large number of recording video around the clock producing huge volumes. Process these huge volumes of videos demands much more resources like time, man power and hardware storage etc. Video summarization plays vital role in this context, it helps in efficient storage, quick browsing and retrieval of large collection of video data without losing important aspects. Summary of video data nicely expands video browsing and saves a lot of time. Summarization done on the basis of ranking of sentences and that is done using words and their ranks. The current video summarization applications are lack of properly trained model which can create its own word dictionary and rank the words or sentences accordingly which can be trained to get near to appropriate results. Video summarization aims at producing compact version of a full-length video while preserving the significant content of the original video.

I. INTRODUCTION

Now day's digital video recording technology is growing at rapid speed, due to advanced technologies it becomes very easy to record huge volume of videos. From March 2020 because of Covid-19 there is all educational lectures, sessions, corporate meetings conducted through online video communication application. During these sessions or meetings huge video recording data is producing. Most of this information is in form of long-term videos. Summary of accurate video sequences for users and allow them too easily and quickly look at large batches of video data has quickly become one of the most important topics in content-based video processing.

There is need of automated methods to create video summaries, using video summarization we can create small volumes videos, it will helpful to minimize the time period to watch the video, we can get only important aspects from the long-length video. There are two important ways to summarize the video in the books: - A stable video summary (video summary) and a powerful video summary (video summary).

Summarization on produced type of video data (like technical lectures) is to find important segments that contain rich information. Users could obtain the important messages by reading summaries rather than full documents.. The former put emphasis on the detection of key frames, while the later focused on the extraction of important concepts. First identifies all caption words, and then adopts a technique to find the important segments. [2].

Text summarization has been extensively used in various fields like science, medicine, law, engineering, etc. Researchers have focused on generating summaries of doctor's prescriptions, and that has been proved very useful to patients. Similarly, long news articles have been summarized and this way readers can gain a lot of information on several topics within a short span of time. [8].

II. LITERATURE REVIEW

The surveyed about various papers related some kind of video summarization work. In review, understand many research papers which are specifically about video summarization. From literature survey classify the existing system as below:-

A. Manual Summarization

One of the widely used technique to summarize the video manual efforts. Through this technique watch the full length video and understand which points are important and edit the full length video manually to summarize or final product of summarization video. [3][4]

B. Key frame based Video Summarization

A static video summary is a collection of representative frames (key frames) extracted from the original video sequence. The latter aims to extract video segments i.e. a collection of frames, which represent the most interesting and meaningful content of a video. Among these, the key frame based video summarization methods have been studied extensively because of their simplicity and usefulness. [1]

C. Video skim based Video Summarization

Video skimming, also known as dynamic video summarization, generates a temporally abridged version of a given video. Also called as chalkboard story or sequence summary. In this original video divided into various small parts and summarize them as short video.

III. PROBLEM DEFINITION

For large scale video manual summarization is not feasible, also man power required to summarize the video continuously. Another option for video summarization using application which converts video to text and summarizes text but it will does not convert back text to video which is major issue. Similar there is key frame technique, the drawback is do not convert summarized video in video format.

There is no accurate video summarization application developed yet which can summarize the all different type of videos as resp. to area of interest or subject.

IV. PROPOSED METHODOLOGY

Proposed system videos create a system which will use the speech in the video summarization.

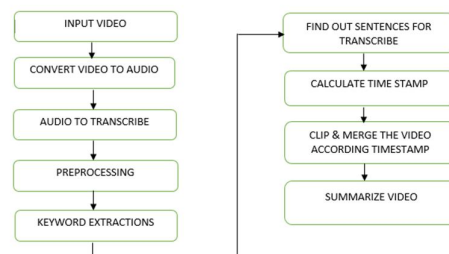


Figure 1. Proposed Methodology

As per proposed system, first have to extract the audio from the input video and generate the speech transcript. Based on this transcript of the video will be use later to processed for summarization of the video, this will be ideal for generate video skims for many different purposes, and then we have to use text summarization techniques for summarization of the generated transcript.

- Steps for the actual method by which video will be summarized on the basis of input & output:
- First step collect video as input.
- Convert video to audio using AWS tools.
- Transcript will be text which is equivalent to video.

- Transcript will be generated using automated speech recognition system.
- First extract the audio from the given video and use this to get a text transcript along with a timestamp.
- Segments of the text based will be formed on paused detection. For pause detection, the time difference between the ending of i th word and the starting of $(i+1)$ th word is used.
- This approach will enable us to get the most important segment of the video by using text summarization technique.
- This will be ideal for generating video skims for information rich video programs, such as documentaries, educational video and presentations using statistical analysis based on speech transcripts.
- These segments will be scored and summarized using statistical and text summarization techniques.
- The segments with highest scores will be including in summary.
- After receiving the selected segments, clipped the video based on the timestamps of the segments.
- All the shortlisted or selected segments will be concatenated and the summarized video will be formed containing the most important segments of the video.

We can find a large amount of digital data for different purposes on the internet and relatively it's very hard to summarize this data manually. Automatic Text Summarization (ATS) is the subsequent big one that could simply summarize the source data and give us a short version that could preserve the content and the overall meaning. While the concept of ATS is started long back in 1950's, this field is still struggling to give the best and efficient summaries. ATS proceeds towards 2 methods, Extractive and Abstractive Summarization. The Extractive and Abstractive methods had a process to improve text summarization technique. Text Summarization is implemented with NLP due to packages and methods in Python. Different approaches are present for summarizing the text and having few algorithms with which we can implement it. Text Rank is what to extractive text summarization and it is an unsupervised learning. Text Rank algorithm also uses undirected graphs, weighted graphs. Keyword extraction, sentence extraction.[14]

V. EVALUATION OF PROPOSED SYSTEM

Frequency of word is using to create summary of long video. This type of videos not contain the close-caption text. First of all the whole video is segmented into small coherent segments and also audio pause boundary detection can be done through temporal analysis of pauses between all words. Segment boundary is declared between these words. Ranking of each segment is done by detecting dominant word pairs using an adaptation method. The rank of those segments increases that contain the first 30 of dominant word pairs. Video skims are generated by selecting those segments that maximize the cumulating score of summary and the duration of summary is provided by user. This technique is useful for those videos that contain gradual pauses between speeches of the speaker, while the audience is silent. It fails in noisy videos and also for those in which there is continuous talking. [13]

NLP is part of Artificial Intelligence reasoning that easily manages analyzing, understanding, finding and producing the dialects that people use in a characteristic manner to make interface with PCs in both composed and spoken settings utilizing common human dialects rather than codes. [14]

The increasing growth of the Internet has made a huge amount of information available. It is difficult for humans to summarize large amounts of text. Thus, there is an immense need for automatic summarization tools in this age of information overload

VI. RESULT & DISCUSSION

Proposed system is using various summarization ratios and tested them on various types of the video. Most of the time the majority of the important segments were included in the video as expected. However summarization ratio below 0.1 might not cover all the important segments of the video. The threshold value will be changed according to the requirement of summarized video. The total size of the summarized video was found to be 181.8secs which is approx. 30 percent of the total video.[1]

According to Shi Ziyang in [6], Summarization could not bring accurate results when the word has a lot of meanings. So there is a need for the particular domain knowledge of the main theme of the document as well. This brings the domain-specific text summarization into the limelight. But the problem arises when the referring is done inaccurately.

The evaluation score indicates a significant improvement from the baseline models that is constant throughout all the test case that we use in the evaluation. The score shows improvement as much three times better compared to

the lowest performing baseline for a specific lecture. The average score evaluation of all the methods also indicates that the proposed model is more suitable for the lecture summarization task compared to the baseline methods.[7]

VII. CONCLUSION

In this work we provided a systematic review of the deep learning- based video summarization. This review allowed to discuss how the summarization technology has evolved over the last years and what the potential is for the future, as well as to raise awareness to the relevant community with respect to promising future directions and open issues.

Video summarization is to process video programs, which usually contain semantic redundancy, and make them more interesting or useful for users. There are various techniques proposed for this task. This method will be useful for information rich video as will be using text summarization techniques on the transcript of the video.

In this, will be using extractive text summarization as it meets the need of solution. In extractive text summarization graph-based page rank algorithm will be used. This will be ideal for generating video-skims for information-rich video programs, such as documentaries, educational videos, and presentations, using statistical analysis based on speech transcripts.

Single video is used into proposed system and create a word dictionary on count of words repeated and rank the words accordingly. Those words will be seen in a sentence and the average rank of each sentence will be calculated on the basis of the text rank formulas. Later, create a model for word dictionaries of each video and rank them dynamically. Here supervised learning will be used to train the model with loads of data, that is, amount of text input file and their corresponding word dictionary, so that the model automatically generates word Dictionaries of each video and video can be summarized easily.

REFERENCES

- [1] S. Radke, P. Mane, et.al. (2021), "Summarization of Video using NLP Techniques." International Conference on Convergence of Smart Technologies (IC2ST-2021), 2021.
- [2] Yu-Chyeh Wu, Yue-Shi Lee and Chia-Hui Chang, "VSUM: summarizing from videos," *IEEE Sixth International Symposium on Multimedia Software Engineering*, 2004, pp. 302-309, doi: 10.1109/MMSE.2004.90.
- [3] Thakkar, Khushboo S., Rajiv V. Dharaskar, and M. B. Chandak. "Graph-based algorithms for text summarization." 2010 3rd International Conference on Emerging Trends in Engineering and Technology. IEEE, 2010.
- [4] Gupta, Vishal, and Gurpreet Singh Lehal. "A survey of text summarization extractive techniques." *Journal of emerging technologies in web intelligence* 2, no. 3 (2010): 258-268.
- [5] M. Taskiran, Z. Pizlo, A. Amir, D. Ponceleon and E. J. Delp, "Automated video program summarization using speech transcripts," in *IEEE Transactions on Multimedia*, vol. 8, no. 4, pp. 775-791, Aug. 2006, doi: 10.1109/TMM.2006.876282.
- [6] Shi Ziyang " The Design and Implementation of Domain-specific Text Summarization ion System based on Co-reference Resolution Algorithm" 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery.
- [7] M. B. Andra and T. Usagawa, "Automatic Lecture Video Content Summarization with Attention-based Recurrent Neural Network," 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIIT), 2019, pp. 54-59, doi: 10.1109/ICAIIIT.2019.8834514.
- [8] Rahul, S. Adhikari and Monika, "NLP based Machine Learning Approaches for Text Summarization," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp. 535-538, doi: 10.1109/ICCMC48092.2020.ICCMC-00099.
- [9] Lai, Po Kong, et al. "Video summarization of surveillance cameras." 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, 2016.
- [10] A. Stefanidis, P. Partsinevelos, P. Agouris, and P. Doucette, "Summarizing video datasets in the spatiotemporal domain," in *Proc. Int. Workshop on Advanced Spatial Data Management (ASDM'2000)*, Greenwich, U.K., Sep. 6–7, 2000.
- [11] Rene Arnulfo Garcia-Herandez and Yulia Ledeneva, "Word Sequence Models for Single Text Summarization", *IEEE*,44-48, 2009.
- [12] J. N. Madhuri and R. Ganesh Kumar, "Extractive Text Summarization Using Sentence Ranking," 2019 Int. Conf. Data Sci. Commun. IconDSC 2019, pp. 1–3, 2019, doi:10.1109/IconDSC.2019.8817040.
- [13] Muhammad Ajmal, Muhammad Husnain Ashraf, Muhammad Shakir, Yasir Abbas, Faiz Ali Shah "Video Summarization: Techniques and Classification", *Computer Vision and Graphics*, 2012, Volume 7594, ISBN : 978-3-642-33563-1
- [14] G. Vijay Kumar , Arvind Yadav "Text Summarizing Using NLP", *Recent Trends in Intensive Computing*, PP-61-69, Dec- 2021.