



Techniques and Methods of Summarizing Video Data

Nazarov F^{1*}, Hamidov B.B², Esanov O.O.³, Salimova M.⁴

^{1,2,3,4}Samarkand State University

DOI: <https://doi.org/10.xxxxx/xxxxx>

*Correspondence: **Nazarov F.**

Email:

Received: Dec-18, 2023

Accepted: Jan-12, 2024

Published: Feb-10, 2024



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract: In this research, research was conducted on video data summarization techniques. The relevance, features and methods of summarizing video data are covered in detail. Here, video summarization based on key frames, video summarization based on hierarchical cluster, video summarization using speech transcripts, video summarization based on video features, trajectory-based video summarization methods are mentioned.

Keywords. Keyframes, clustering based, speech based, video features, trajectory based.

INTRODUCTION

At the same time that all spheres of society are rapidly developing, information technology is entering all spheres. Today, due to the large flow of information, the problem of processing them in a short time is also increasing. Widespread use of the Internet and affordable video recording devices have dramatically changed the landscape of video production and consumption. In particular, with the evolution of video streaming services and social media, user-generated videos are more common than ever. The main part of the information circulating on many large social networks and sites corresponds to video information. For example, about 400 hours of video data are uploaded to the YouTube platform every minute. Based on this, it is possible to achieve some economic efficiency based on obtaining conclusions about this video without fully downloading the large volume of video data on the Internet. In this research work, video data inference techniques are studied (Figure 1).

Creating a video summary is the final part of the video summarization process. In the process of video summarization, it is first determined what type the video belongs to. Video data comes in the form of video conferencing, sports broadcasts, video lessons, movies, documentaries, personal videos, and more. Once the type of video is determined, the appropriate video summarization technique is used. For example, a generalization technique based on the trajectory of objects in the video is used to generalize broadcasts of sports games. Information in the form of video lessons is often summarized based on the transcript of the video. In this case, a conclusion is drawn from the speech itself, and a summary of the video is determined based on this speech. The following video data summarization techniques are presented (Figure 1).

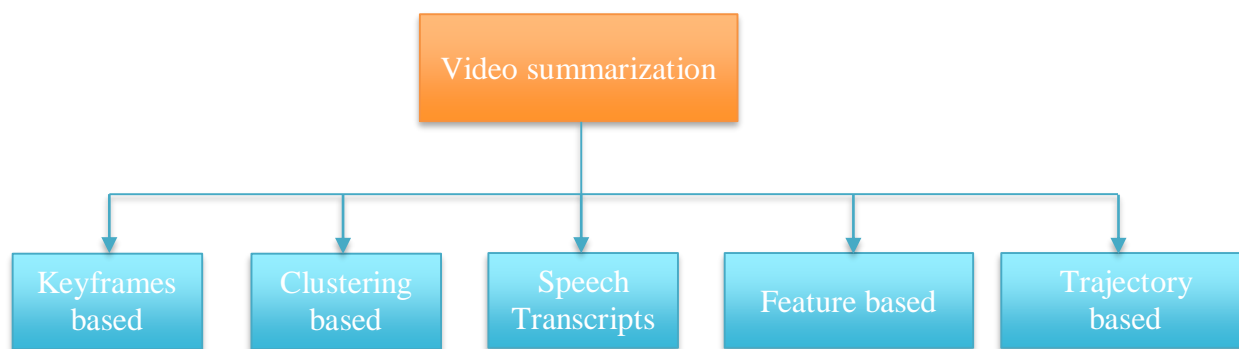


Fig 1. Video summary generation techniques.

Summarize videos based on key frames

Data inference through keyframe extraction focuses on keyframe extraction techniques that can be used for object recognition and information extraction. A keyframe is an image that represents an important event in the video. Keyframe-based video summarization works by extracting frames, so the first step is extracting specific frames from the frame sequence of the original video. The video frames obtained in the second step are clustered, which have redundant content and eliminate overlapping frames. The selection of the main frames is continued in the third step. This process is shown in Figure 2 below.

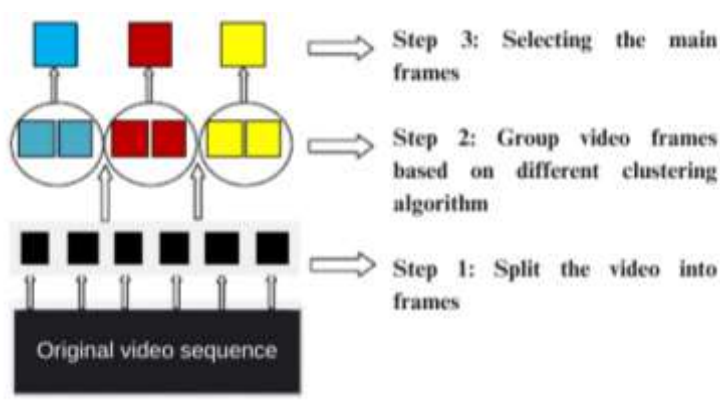


Fig 2. Selecting main frames.

Video inference based on hierarchical clustering

In this, a hierarchical clustering-based keyframe extraction method is proposed to generate video summaries that are closer to the videos that the user is more likely to watch. First, the color and texture features in the video image are extracted and combined. Second, initial clustering parameters are obtained by hierarchical clustering, and K-means clustering is used to optimize the initial results and classify similar frames. Finally, the frame closest to the cluster center is selected as the keyframe.

The interpretation of the experimental results shows that the accuracy of the obtained keyframes improves by an average of 0.71 when using the clustering method. In addition, the average recall rate is 0.76 and the average F-score is 0.73, which is higher than other tested methods and effectively improves the overall effect of video inference. It also produces video keyframes that are closer to human inferences[2].

Video summary using speech transcripts

Video summarization using speech transcripts mainly involves two tasks: first, segmenting the video into small, coherent segments, and second, ranking the resulting segments. Video inference using speech transcripts evaluates segments based on word frequency analysis of speech transcripts. A summary is then generated by selecting the segments with the highest scores and duration ratios and combining them. The comparison is made using a random segment selection scheme based on algorithm and statistical analysis of user learning results[3].

Video summarization based on feature

Digital video includes many features such as color, motion, sound, and more. These methods work well if the user wants to focus on the features of the video. For example, if the user wants to see color features, it is better to choose color-based video summarization methods. Feature-based video summarization methods are classified based on motion, color, dynamic content, gesture, audio-visual, speech transcription, objects, and more.

Video summarization based on trajectory

In trajectory-based video summarization, the moving parts in the video are detected using background subtraction and the binary image is divided into frames. Human detection is performed using a Histogram of Oriented Gradient (HOG) using a Support Vector Machine (SVM) classifier. Then, the movement of people is tracked through successive frames using a Kalman filter, and the trajectory of each person is extracted. Trajectory analysis leads to a meaningful summary that covers only the important parts of the video. You can also specify a region of interest to add to the summary.



Fig 3. Determining the trajectory of objects.

Experimental results show that Trajectory-based video summarization can reduce a long video into a meaningful summary and save a lot of time and cost for storage, indexing, and browsing[4].

Conclusion

In this research work, video summarization techniques were investigated to understand the video data without having to download the entire video data. Since video data comes in different forms, each type requires a customized summation technique to achieve optimal results. In the research work, these methods were analyzed in general. The keyframe inference method is based on the selected feature of the video frames. Hierarchical clustering method divides frames based on features of interest to the user. Trajectory-based inference methods are inferred by selecting one or more objects and observing them. The most commonly used method is to extract the speech from the video and get a conclusion from it. In this work, such video summarization methods were analyzed.

References

Nazarov, F. M., Esanov, O., & Hamidov, B. (сана йил). Static and dynamic methods of summarizing video data. Маҳсулотнинг жамият-илимий конференциясида тезлик, "Modern innovative information and pedagogical technologies in digitalization of education: problems and solutions" дарақчаси, ISBN 978-9943-6558-6-7.

Wang, F., et al. (2020). Hierarchical Clustering-Based Video Summarization. КИТОБНИНГ редакторлари: Kountchev, R., Patnaik, S., Shi, J., Favorskaya, M. Advances in 3D Image and Graphics Representation, Analysis, Computing and Information Technology. Smart Innovation, Systems and Technologies, том 179. Springer, Сингапур. https://doi.org/10.1007/978-981-15-3863-6_4

Taskiran, C., Pizlo, Z., Amir, A., Ponceleon, D., & Delp, E. (2006). Automated Video Program Summarization Using Speech Transcripts. *IEEE Multimedia*, 8, 775-791. <https://doi.org/10.1109/TMM.2006.876282>.

Ajmal, M., Naseer, M., Ahmad, F., & Saleem, A. (2017). Human Motion Trajectory Analysis Based Video Summarization. 550-555. <https://doi.org/10.1109/ICMLA.2017.0-103>.

Nazarov, F. M., O'G'li, Y. S. S., & O'G'li, E. B. S. (сана йил). Algorithms To Increase Data Reliability In Video Transcription. 16th IEEE International Conference on Application of Information and Communication Technologies, AICT 2022 - Proceedings. <https://doi.org/10.1109/AICT55583.2022.10013558>.