Automatic Video Summarization

[1]Preeti Vishnu Chandgude, [2]Gayatri Kailas Solanke,
[3]Ankita Sudhakar Avhad, [4]Priyanka Ashok Sanap

Department of Computer Engineering,Matoshri College ofEngineering and Research
Centre, Eklahare,Nashik.

Abstract
The primary goal of this system is to propose an improved method for summarising videos in order to present important and concise information to end users. In video, description and keywords are important factors in deciding which video to watch. The proposed method's main idea is to generate video descriptions automatically. Our approach is critical in reducing the time spent searching for the right video. It aims to save users' time spent watching unwanted videos by utilising timestamps. One of our approach's primary goals is actual keyword extraction. Extracted keywords aid in the discovery of videos containing significant video keywords.Frames, emotions, and speech are used to summarise the video. First, the video content is displayed in the frame, followed by a summary of the video content. Second, emotion and how it changes over time are combined with the outputted frame summarization. Third, audio transcription into text takes place, producing an abstractive summary of the audio track. Finally, using natural language processing techniques, the fusion of all summarizations (audio, video, emotion) occurs.

***Key Words*: text summarization, video summarization, Latent Semantic Analysis**

Published in IJIRMPS (E-ISSN: 2349-7300), Volume 11, Issue 2, March-April 2023

INTRODUCTION

Our time has come to see the fundamental significance of advanced innovation in our day-by-day lives. It allows us to open a colossal assortment of data. We have infinite media content ranging from pictures to large-scale videos. Many videos are transferred to YouTube, Dailymotion, Vimeo, and other video-sharing sites each moment. These large volumes of digital data require significant manpower and resources to retrieve and process important and relevant information. Various sources of digital media such as documentaries, sports matches, and educational videos can be found on the internet. Processing large-scale media can become tedious, time-consuming, and heavy on hardware. Therefore, summarization strategies are incredibly expected to consume the ever-developing measure of information accessible on the web. In essence, summarization is intended to assist us with consuming important data faster.

PROBLEM DEFINATION

We have infinite media content ranging from pictures to large-scale videos. Many videos are transferred to YouTube, Dailymotion, Vimeo, and other video-sharing sites each moment. These large volumes of digital data require significant manpower and resources to retrieve and process important and relevant information. Various sources of digital media such as documentaries, sports matches, and educational videos can be found on the internet. Processing large-scale media can become tedious, time consuming, and heavy on hardware. Therefore, summarization strategies are incred1 1 Automatic Video summarization Automatic Video Summarization ibly expected to consume the ever-developing measure of information accessible on the web.

ADVANTAGES OF SYSTEM

1. Time Reducing system
2. Reducing Manpower.
3. Stay connected

LITERATURE SURVEY:

1. “M. Zhu, ”Video Captioning in Compressed Video,” 2021 6th International Conference on Image, Vision and Computing (ICIVC), 2021, pp. 336-341, doi: 10.1109/ICIVC52351.2021.9526927 In this paper, Authors propose a video captioning method which operates directly on the stored compressed videos. To learn a discriminative visual representation for video captioning, author design a residuals-assisted encoder (RAE), which spots regions of interest in I-frames under the assistance of the residuals frames. First, then obtain the spatial attention weights by extracting features of residuals as the saliency value of each location in I-frame and design a spatial attention module to refine the attention weights and further propose a temporal gate module to determine how much the attended features contribute to the caption generation, which enables the model to resist the disturbance of some noisy signals in the compressed videos. Finally, Long Short-Term Memory is utilized to decode the visual representations into descriptions. We evaluate our method on MSRVTT dataset and demonstrate the effectiveness of our approach.

2. ”C. Jung, Su Young Lee and J. Kim, ”Robust detection of key captions for sports video understanding,” 2008 15th IEEE International Conference on Image Processing, 2008, pp. 2520-2523 In this paper, we provide a robust detection method of key captions in sports videos. We also provide a dual binarization method easily segmenting texts with different color polarities (i.e. dark and bright texts) from the background in the key captions. From this text information, we create the efficient sports navigation system. By conducting experiments on a large database, the system performance is demonstrated to be accurate and robust.

3. “S. -H. Han, B. -W. Go and H. -J. Choi, ”Multiple Videos Captioning Model for Video Storytelling,” 2019 IEEE International Conference on Big Data and Smart Computing (BigComp), 2019, pp. 1-4. In this paper, Author propose a novel video captioning model that utilizes context information of correlated clips. Unlike the ordinary “one clip - one caption” 4 Automatic Video summarization algorithms, author concatenate multiple neighboring clips as a chunk and train the network in “one chunk - multiple caption” manner. then train and evaluate our algorithm using M-VAD dataset and report the performance of caption and keyword generation. The model is a foundation model for generating a video story using several captions. Therefore, 4 Automatic Video Summarization in this paper.

4. ”J. Vaishnavi and V. Narmatha, ”Video Captioning based on Image Captioning as Subsidiary Content,” 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), 2022, pp. 1-6. The proposed model constructed with the option of generating captions with high diversity. Image captions are taken as subsidiary content to enlarge the diversity for captioning the videos. Attention mechanism is utilized for the generation process. Generator and three different discriminators are utilized to contribute an appropriate caption which enriches the captioning process. ActivityNet caption dataset is used to demonstrate the proposed model. Microsoft coco image dataset is considered as subsidiary content for captioning. The benchmark metrics BLEU and METEOR are used to estimate the performance of the proposed model.

SYSTEM ARCHITECTURE

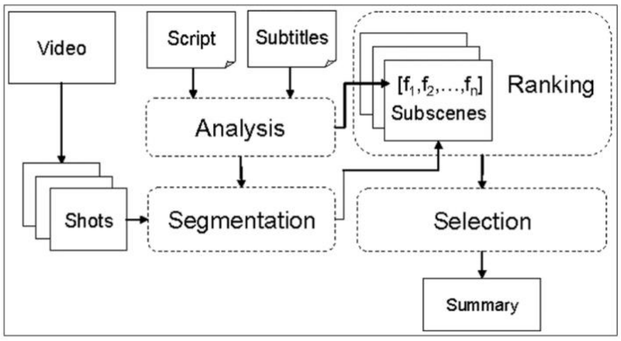
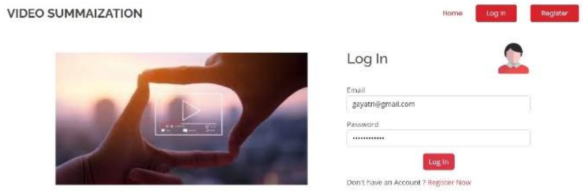


Fig -1: System Architecture Diagram

ALGORITHMS

- Hashing & Mapping: A cryptographic hash function (CHF)is a mathematical algorithm that maps data of an arbitrary size (often called the "message") to a bit array of a fixed size (the "hash value", "hash", or "message digest").
- It is a one-way function, that is, a function for which it is practically infeasible to invert or reverse the computation. Ideally, the only way to find a message that produces a given hash is to attempt a brute-force search of possible inputs to see if they produce a match, or use a rainbow table of matched hashes. Cryptographic hash functions are a basic tool of modern cryptography.
- SVM : In machine learning, support vector machines (SVMs, also support vector networks[1]) are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis. Developed at AT&T Bell Laboratories by Vladimir Vapnik with colleagues (Boser et al., 1992, Guyon et al., 1993, Cortes and Vapnik, 1995,[1] Vapnik et al., 1997[citation needed]) SVMs are one of the most robust prediction methods, being based on statistical learning frameworks or VC theory proposed by Vapnik (1982, 1995) and Chervonenkis (1974). Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). SVM maps training examples to points in space so as to maximise the width of the gap between the two categories. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

RESULTS



VIDEO SUMMAIZATION

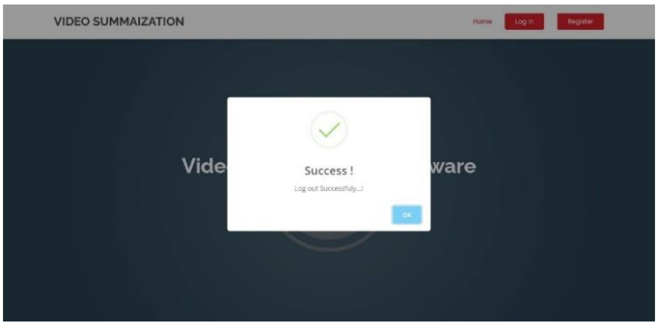
Log In

Email: gayatri@gmail.com

Password: ••••••••••••

Log In

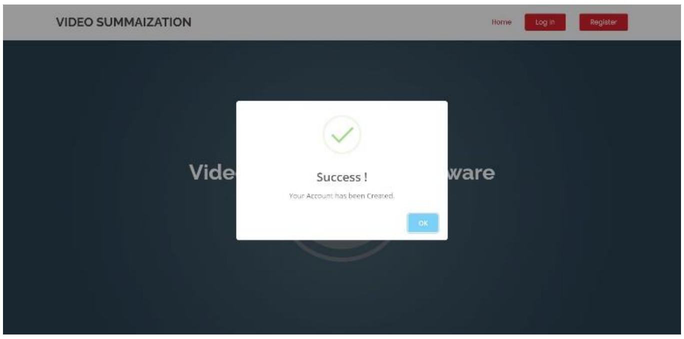
Don't have an Account ? Register Now



VIDEO SUMMAIZATION

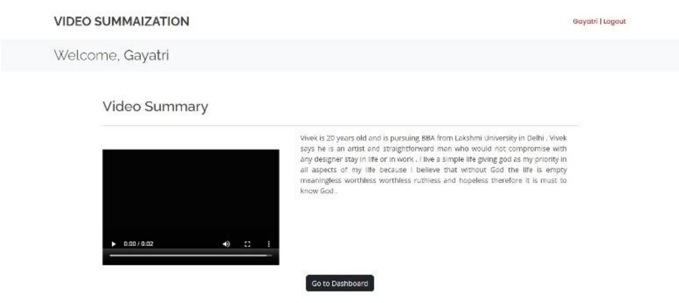
Success !

Log out Successfuly...!

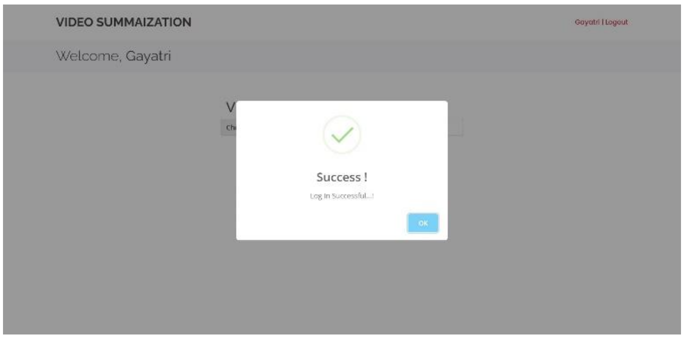


Success !

Your Account has been Created.

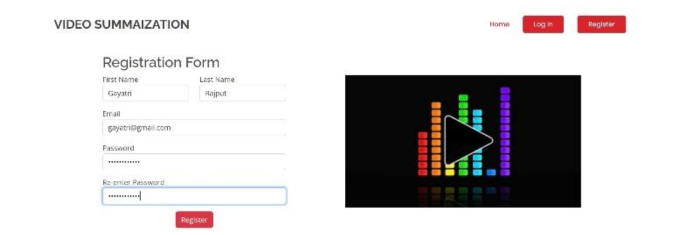


Vivek is 20 years old and is pursuing BBA from Lakshmi University in Delhi . Vivek says he is an artist and straightforward man who would not compromise with any designer stay in life or in work . I live a simple life giving god as my priority in all aspects of my life because I believe that without God the life is empty meaningless worthless worthless ruthless and hopeless therefore it is must to know God .



Success !

Log In Successful...!



Registration Form

First Name: Gayatri

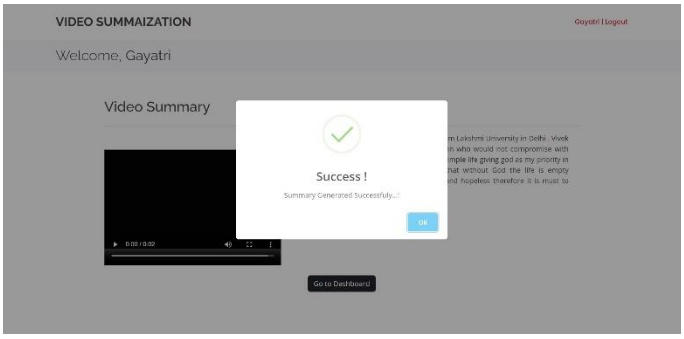
Last Name: Rajput

Email: gayatri@gmail.com

Password: ••••••••••••

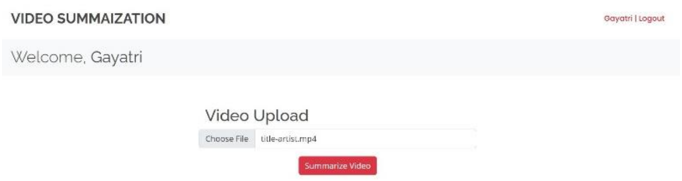
Re-enter Password: ••••••••••••

Register


VIDEO SUMMAIZATION
Gayatri | Logout
Welcome, Gayatri
Video Summary
m Lakshmi University in Delhi . Vivek
n who would not compromise with
imple life giving god as my priority in
hat without God the life is empty
nd hopeless therefore it is must to
0:00 / 0:02
OK
Go to Dashboard

Success !

Summary Generated Successfuly...!



VIDEO SUMMAIZATION

Welcome, Gayatri

Video Upload

Choose File title-artist.mp4

CONCLUSION

The proposed system uses text summarization as the primary method of summarizing videos. Automatic text summarization is the problem in the field of data science of creating a short and accurate summary from a drawn-out report. Automatic text summarization can be utilized in an assortment of uses. The increase in popularity of video content on the internet requires an efficient way of representing or managing the video. This can be done by representing the videos based on their summary.

REFERENCES

1. J. S. Park, M. Rohrbach, T. Darrell and A. Rohrbach, ”Adversarial Inference for Multi-Sentence Video Description,” 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 6591- 6601.
2. S. H. Abdulhussain, S. A. R. Al-Haddad, M. I. Saripan, B. M. Mahmmod and A. Hussien, ”Fast Temporal Video Segmentation Based on Krawtchouk-Tchebichef Moments,” in IEEE Access, vol. 8, pp. 72347-72359, 2020.
3. S. H. Abdulhussain et al., ”A Fast Feature Extraction Algorithm for Image and Video Processing,” 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 2019, pp. 1-8. . Pan, Z. Xu, Y. Yang, F. Wu and Y. Zhuang, ”Hierarchical Recurrent Neural Encoder for Video Representation with Application to Captioning,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 1029-1038.
4. Y. Yang et al., ”Video Captioning by Adversarial LSTM,” in IEEE Transactions on Image Processing, vol. 27, no. 11, pp. 5600-5611, Nov. 2018.
5. R. Shetty, M. Rohrbach, L. A. Hendricks, M. Fritz and B. Schiele, ”Speaking the Same Language: Matching Machine to Human Captions by Adversarial Training,” 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 4155-4164 .
6. Yu, J. Wang, Z. Huang, Y. Yang and W. Xu, ”Video Paragraph Captioning Using Hierarchical Recurrent Neural Networks,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 4584-4593.
7. B. Dai, S. Fidler, R. Urtasun and D. Lin, ”Towards Diverse and Natural Image Descriptions via a Conditional GAN,” 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 2989-2998.
8. X. Liang, Z. Hu, H. Zhang, C. Gan and E. P. Xing, ”Recurrent Topic-Transition GAN for Visual Paragraph Generation,” 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 3382-3391