Video Summarization based on Keyframe Selection using Connectivity Centroid Clustering

MSc Research Project
Data Analytics

Arghadeep Chowdhury
Student ID: X20189940

School of Computing
National College of Ireland

Supervisor: Dr. Christian Horn

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Arghadeep Chowdhury
Student ID:	X20189940
Programme:	MSc Data Analytics **Year:** 2021-2022
Module:	Research Project
Supervisor:	Dr. Christian Horn
Submission Due Date:	15th August 2022
Project Title:	Video Summarization based on keyframe selection using connectivity centroid clustering
Word Count:	**21 Page Count 5360**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Arghadeep Chowdhury

Date: 15th August 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Video Summarization based on keyframe selection using connectivity centroid clustering

Arghadeep Chowdhury
X20189940

Abstract

Video summarization helps in obtaining important parts of a video by retaining and summarizing its vital information. This process is vital in video-sharing websites such as YouTube, and OTT platforms like Netflix and Prime Video. Summarizing a video, movie or show helps the user to get an overview of what the video contains. This study develops a video summarization technique based on the unsupervised learning method of connectivity clustering to identify keyframes from a video. The keyframes obtained from the video are used to create a video summary. The algorithm makes use of the singular value decomposition technique to obtain the similarity between the frames to cluster them hierarchically. The developed model is tested with the recently developed VSSUM algorithm for evaluation and the results obtained conclude that the developed algorithm is comparable to the VSSUM.

1 Introduction

Over the period a strong growth in video stimulating applications and websites can be identified to occur due to the extensive use of the internet and technology. Based on the estimations in the recent period, it has been observed that YouTube presently receives more than 500 hours of video content each minute and the situation is nearly similar in different other video posting web services, such as Dailymotion, different social media platforms, like Facebook, Twitter and Instagram and others, and news and media agencies online archives as well. These platforms are also associated with hosting massive amounts of video content as well. In this regard, it has been observed that with the implication of automated video summarization technologies these video streaming platforms could effectively create a concise description that supports highlighting the most relevant and crucial aspects of the full-length video as well. As a result of this effective video summarization the overall exploration and navigation of big video collections have become seems too easy for the viewers of the same and this, in turn, contributed to enhancing the audience engagement and consumption of the content as well.

In this aspect, Rochan *et al.* (2018) highlighted in their studies that video summarization provided a proper means of identifying the key contents of a particularly large video and through which the audiences could identify the interesting factors from the same. Hence an effective implication of video summarization tools or technology could deliver the content creators create an initial enthusiasm among the audiences in an effective and efficient manner before uploading or screening an entire big video (Zhang *et al.*, 2018).

1.1 Motivation

In this present big data era, it has been observed that finding and watching the desirable videos will become highly difficult to perform in the absence of proper implication of video summarization tools or techniques. Hence, in this aspect, it is required to be stated that video summarization is a real-time deal or demand for the users of the same. Generally, it has been observed that with the proper implication of deep learning techniques or processes data analysis could be performed effectively in a video summarization system. Over time performing a supervised video summarizing process has acquired a lot of attention among the users due to its higher benefits in the video summarization system. This is because it utilizes videos for input as a mean of training data through which data could be chosen with the proper implication of a supervised learning approach to determine how humans could summarize videos effectively. Thus, undertaking the proposed research approach in the video summarization process is expected to deliver higher value in the video editing and unprocessed video summarization system.

1.2 Objectives

In any research process developing a suitable objective could be identified as a key matter of concern as based on the research objective a researcher can reach its ultimate research aim or goal. For this research paper, the objective could be identified as follows-

- To improve the level of accuracy through which summarization of comprehensive video could be performed effectively.

- To determine the best techniques to be implemented for better summarization of comprehensive videos and associated strengths and limitations.

- To assess the ways for decreasing the time of computation of the present best techniques in the video summarization process.

- To identify the best techniques of validation through which results could be validated for the proposed technique of video summarization.

Consideration of these research objectives could deliver a higher value for the users of video summarizing tools and techniques over time. Apart from that, the implication of key shots-based video summarization techniques could be assessed over the period in an appropriate and effective manner that could deliver higher value.

1.3 Supportive of Research Questions

Forming appropriate research questions could be identified as a key aspect in any research paper as it supports identifying the key aspects that the entire research must focus on. In this regard, the following research questions could be developed for this research paper-
- How do improve the accuracy to summarize a comprehensive video?
- Which are the best techniques or methods to summarize comprehensive videos and what are the key limitations and strengths?
- How can we decrease the computation time of the present best techniques?
- Which are the best techniques of validation through which the results could be validated for our proposed technique?

First of all, this research paper just focused on identifying the motivation and research objectives based on which the entire research process could be performed in an efficient and effective way. Following that, depending on the research objective suitable research questions have also been formed.

Secondly, it has targeted to perform a detailed and suitable literature review depending on key literature works that were performed previously by different researchers. The key aim of performing this literature review is to gain proper insight into the focused area of research.

Thirdly it is considered to develop a proper methodology based on which the focus research has been performed and, in this process, it has followed the KDD process.

Following that, an architecture of model and interpretation of model have also been performed where a detailed comparison of these models along with the models identified from recent research papers have also been performed.

Based on that the results have been discussed and formulated effective conclusion to identify whether future work could be possible in the focused research area.

2 Literature Review

In any research process, a literature review is a key aspect to consider as with the help of a literature review researchers can choose relevant previous works and theories which are applicable to the chosen research area. With the proper application of literature review, an effective theoretical background can be formed based on which the existing literature gap could be identified in an effective and efficient manner. This section of the research paper is going to focus on numerous studies and research papers which were performed in this particular research area and identify the key gaps that are required to be dealt with in this paper. In this context, it is required to be highlighted that this paper has particularly focused on performing a literature review based on the different methods or types of video summarization techniques followed by the users. This is because a better understanding of the present situation over different aspects of the chosen topic would be effective to identify the literature gap and proceed accordingly to attain the research aim and objectives.

2.1 Concepts and approaches of video summarization

Video summarization is basically a process of formulating a short summary of the key content of a longer video by choosing and presenting the most informative or crucial materials for the targeted audience of potential users. In other terms, it aims in formulating a short synopsis that summarizes the key or main content of the video by selecting the most informative and important parts of that entire video (Apostolidis *et al.*, 2021). It supports the targeted audiences to acquire or develop insight about the content that the video is going to present and creates an urge among them to watch the entire video as well. Over the decade different approaches have been formed and the present situation of the art is shown by different methods that depend on modern deep neural network architectures (Basavarajaiah and Sharma, 2019). With the growth in technology and the use of different video hosting platforms, social networks and online repositories of media which uploads a large content of data video summarization is identified to be highly applicable for those organizations. Considering the plethora of video content on different web platforms the effective summarization of video facilitates the viewers to browse and navigate large video collections and this, in turn, supports in enhancing the engagement of viewers and the level of consumption of content.

With the growing changes or advancements in technology, different approaches have been formed for the video summarization process. In this aspect, it has been identified that video summarization is mainly categorized into two key types or segments, which are- single video summarization (SVS) and multi-video summarization (MVS). In this aspect, Cai *et al.* (2018) highlighted that in the case of a single video summarization process the key problem is linked with the subjective understanding of the users relating to the content of a particular video. Different data-driven approaches, like deep neural networks, could be effective to a certain extent to deal with the identified ambiguity that is inherent in this task however it becomes highly expensive to gather the temporal annotations for the large-scale dataset of a certain video. Even Sen and Raman (2019) have highlighted in their study that SVS is effective to segregate or digest a single long video. On the other hand, envious is identified to support summarizing a large number of short videos which are acquired from a wave video query. Due to this nature, MVS is also considered a query-based video summarization process. Hence the approach of processing videos significantly differs in the case of MVS compared to SVS as it needs to deal with a different type of query-based videos, and this may utilize different query information as well. In this regard, Wu *et al.* (2020) have highlighted in their study that MVS is an effective and efficient tool available for users through which they can browse multiple videos. Apart from that, a multi-view video summarization aspect is also identified to keep in focus, and it is mainly used in to develop surveillance settings to comprise the films which are gathered by a different number of cameras (Hussain *et al.*, 2021). Apart from that from the perspective of the learning model supervised and unsupervised video summarizing methodologies can also be identified to be present.

2.2 Supervised and unsupervised video summarization

In the process of video, summarization supervised and unsupervised these two aspects can also be identified to present. Generally, the unsupervised techniques are generally made or formed to confirm that the formed video summary complies with certain criteria which are representativeness, consciousness and informativeness. Due to consideration of all these aspects unsupervised techniques have been witnessed to maintain a higher dominance in the video summarization process. Apart from that, it has also been identified that there are certain selection criteria, like coverage relevance frequency and attention of users, based on which summaries are needed to be formed. Depending on these different criteria different number of techniques have also been formed over the period to meet the purpose of the users. Among all those different approaches clustering-based approach has been observed to be utilized by the users in an extensive manner. According to Hussain *et al.* (2019), the proof of clustering based is implacable for organizing visually comparable shots or frames into certain groups. Among those groups, the centres of the group so as a representative factor of a video and hence that is considered a key frame or shot. Even in this regard, it has also been observed that dictionary learning is one of the most prominent approaches to be implemented in the case of the summarization of unsupervised video (Ma *et al.*, 2020). The basis vectors of the dictionary model are undertaken as the key shots or frames as those are the best approaches to recreate the visual content of a particular original video.

2.3 Literature Gap

Video summarization has emerged as a crucial aspect to consider with the growth of different media platforms that promote video content to upload and operate. In this process, through the evaluation of the findings and discussions of the previous studies (such as journals, articles and others) a significant change in the process of video summarization could be identified to take place over the period. However, based on the previous literary work it can be clearly identified that most of the studies have focused on identifying different methods of video summarization along with identifying the processes employed by the users for selecting the key shots. In order to mitigate the identified literature gap, this study will particularly focus on assessing video summarization based on the selection of keyframes by the utilization of the connectivity clustering technique. This focus of this study would be effective to deliver higher flexibility in the research process and also is expected to add a new dimension to the focused area of research as well based on which the users of video summarization could identify alternative ways to deliver higher value.

2.4 Summary

The previous parts of this section have specifically focused on analyzing the different studies performed related to the video summarization processes. Based on the findings of different studies it could be identified that over the period different methods have been utilized by the researchers to determine the effective video summarization processes. However, in this process lack of research in terms of Attention-based LSTM techniques could be identified which is the key background for this research paper. In this aspect, it is expected that through the completion of this research paper a better outlook or insight could be acquired over the chosen research area and provide a better means through which proper video summarization could be performed as well.

2.5 Comparison of literature in the domain of video summarization:

Author(s)	Model	Performance
Wu, Zhong, and Liu (2020)	Dynamic Graph CNN	F1-score(mean) = 65%
Cai et al (2018)	Weakly supervised model with Variational Encoder-Decoder	mAP score (mean) = 0.76 on web videos and 0.50 on TVSum dataset
He et al (2019)	Attentive Conditional Generative Adversarial Networks	F1-score = 46% on SumMe dataset and 58.5% on TVSum dataset
Elfeki & Borji (2019)	Multilayer Perceptron with CNN based spatial encoder and GRU based temporal encoder	Better at detecting actionness in the video compared to other models
Hussain et al (2019)	Deep Bi-directional LSTM	High F1-score of 90% on YouTube videos
Ji et al (2018)	Attention based Encoder-Decoder networks	F1-score of 44.4% for SumMe and 61% for TVSum
Liang et al (2021)	Convolutional Attentive Adversarial Network	F1-score of 50.81% for SumMe and 59.58% for TVSum
Zhang, Grauman & Sha (2018)	Retrospective Encoders	F1-score of 44.9% for SumMe, and 63.9% for TVSum

Above table lists the performances of the past models developed for video summarization. From the table it can be seen that most of the methodologies incorporated in these studies used supervised learning for the identification of keyframes. The data for supervised classification has to be made such that the key frames in the video are needed to be labeled. This is not easy in case of videos. Videos can contain a large number of frames in them and labelling them all is a very difficult task. Hence implementing an unsupervised algorithm for video summarization always beneficial.

3 Research Methodology

This section discusses the methodology implemented in the study. This section focuses on the steps that are performed to achieve the desired results. This study involves, identifying the most important frames in a video that are combined to construct a video summary. The process of achieving this is depicted in the figure below.

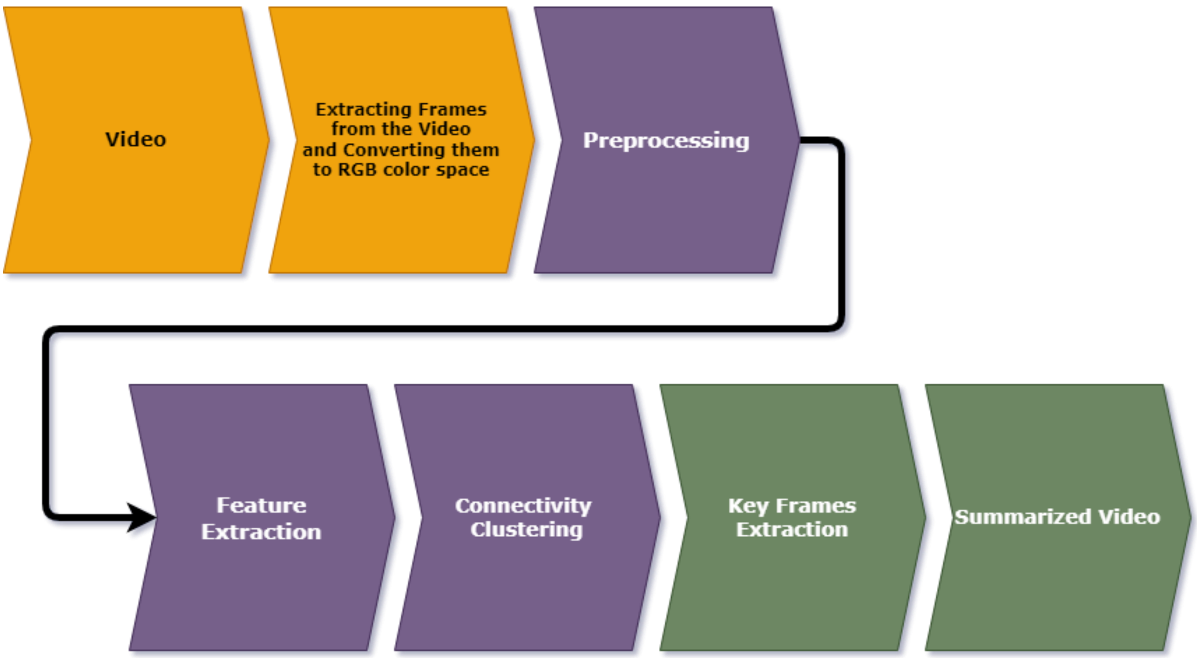

Figure 1: Methodology

As seen in the figure, the methodology follows the steps given below.

3.1 Reading the video file:

The video that has to be summarized is taken as the input for the algorithm. A video is essentially a collection of images called frames that are scrolled at a speed calculated in frames per second (fps). The video that has been utilized in the study is v21.mpg.

3.2 Extraction of Frames and Colorspace Conversion:

Next step in the methodology is to extract the frames from the video. Individual frames that make up a video are extracted in this step. The extracted frames in this step act as images. The CV2 python library extract images that are present in the video in Blue Green Red (BGR) colorspace. However, they have to be converted to RGB colorspace for summarizing the video at the last. These images are then processed further in subsequent steps of the methodology.

3.3 Pre-processing:

This is an important step in processing the video. In this step, 8 separate regions of the video are accessed. In other words, the image is sliced into 8 sections. This helps to process smaller regions of the image to get features relevant to them. This is because it can happen that the background of the image might not change in video and most of the important part is present in the foreground. Splitting the image hence is essential for getting most relevant information from the image.

3.4 Feature Extraction:

Once the image is split into four parts, features are extracted for each of the parts. Feature extraction in the study is done in two steps viz.

3.4.1 Histogram Calculations

Histogram is a method of calculating the representing each of the colors in a colorspace. In an RGB image, histogram measures the number of pixels corresponding to each Red, green and blue color. Figure 2 below depicts the histograms for 2 different frames in the video. Note that the number of pixels given on the y-axis are different for both of the frames. This essentially means that a histogram can be a good choice as a feature.

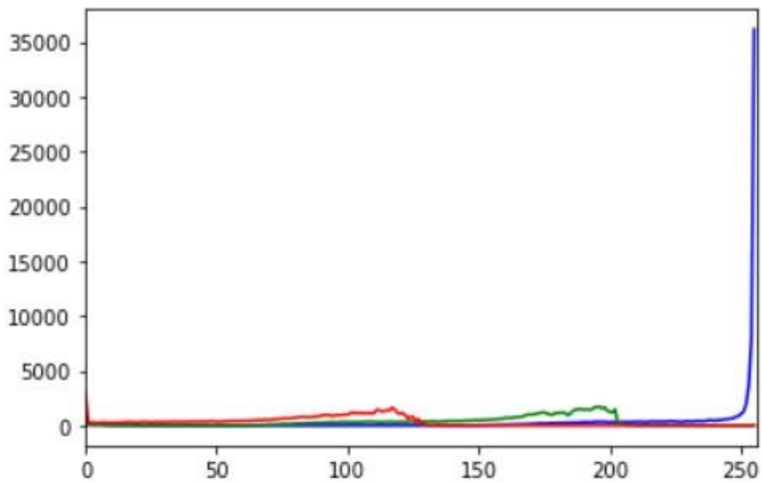

Figure 2: Histograms of 2 frames from different clusters in the video

Using histograms, a matrix of values containing histograms for each frame is created which is further operated upon to extract features.

3.4.2 b. Singular Value Decomposition of the Histogram matrix

Singular value decomposition (SVD) is a method in linear algebra of factorization of a matrix. It basically generates three matrices that convey important geometrical information about the linear transformation in the matrix from which they are obtained. It is essential for the matrix to be a square matrix for performing the SVD. When a matrix such as the matrix produced using the histograms of the frames is not square, it has to be converted to a square matrix using a sparse matrix (a matrix with a large number of zeros). SVD simplifies the Eigen decomposition of a matrix with an orthonormal eigenbasis to any matrix. Figure 3 below illustrates the SVD of a 2x2 matrix.

The orthonormal projections of the singular matrices produced after SVD are considered as the features for the analysis. These projections are obtained for each of the frames present in the video.

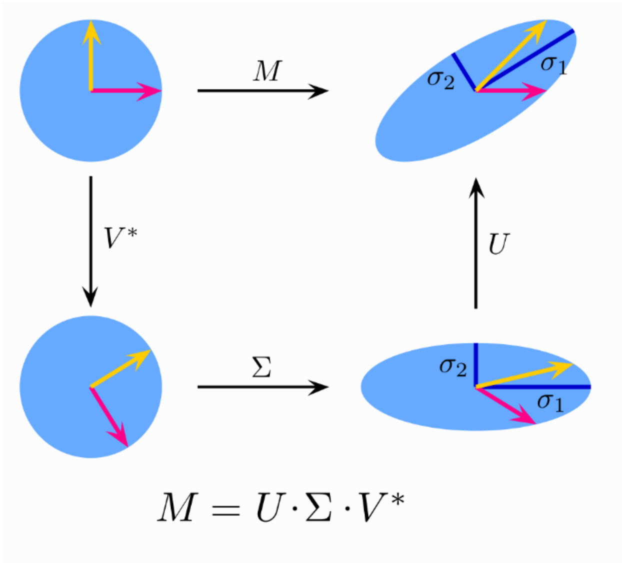

Figure 3: SVD of a matrix

3.5 Connectivity Clustering:

Connectivity clustering also known as hierarchical clustering groups the data based on the dissimilarity between them. The number of clusters that are to be created is based on the nature of the task. As the task at hand requires 2 clusters viz. a cluster with similar frames and a cluster with dissimilar frames. Connectivity clustering can be achieved using various methods such as

a. Complete linkage clustering: Involves clustering based on the dissimilarity between each of the elements of the two clusters by considering the largest value of dissimilarity.

b. Single linkage clustering: Involves clustering based on the dissimilarity between each of the elements of the two clusters by considering the smallest value of dissimilarity.

c. Mean linkage clustering: Considers the average of the dissimilarities between the elements of the clusters.

d. Centroid linkage clustering: Finds the distance between the centroids of the clusters.

e. Ward's method: It involves minimizing the total variance inside the cluster.

This study makes use of the centroid linkage clustering approach for connectivity based clustering. In the study, the centroid linkage method is implemented by calculating the dissimilarity between the projections of the unitary matrix obtained from the SVD of the histogram matrix. Clustering hence groups similar projections. Based on this the most important frames of the video are identified.

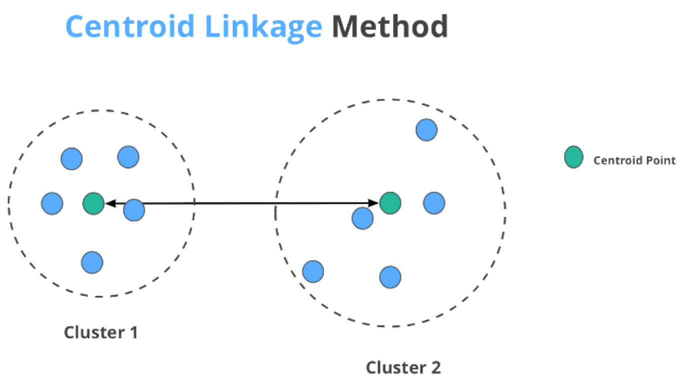

Figure 4: Centroid Linkage Connectivity-based Clustering

In this way, connectivity clustering is done in the study. The output of this clustering helps in identifying the frames that are similar or dissimilar. Similar frames are then used for the summarization of the original video.

4 Design Specification

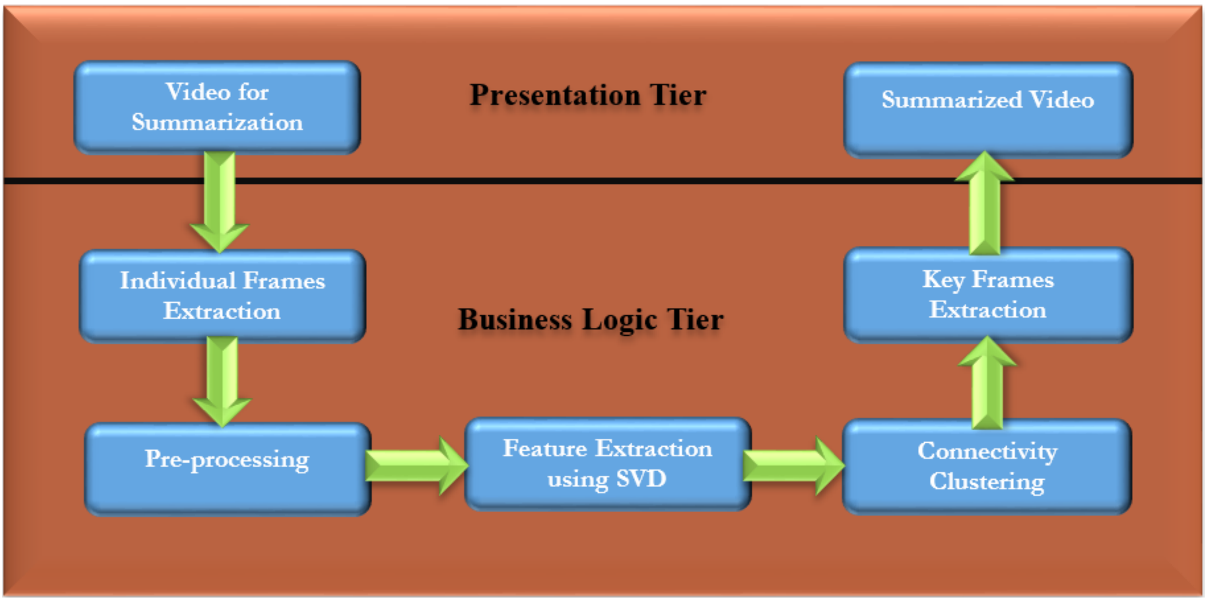

Figure 5: Architecture of the Study

Figure 5 above depicts the final architecture that has been utilized in the study. From the architecture, two tiers can be seen viz. presentation tier and the business logic tier. The presentation layer consists of the input and the output of the study. The business contains the method through which the important part of the processing is done. In the first step the video which needs to be summarized must be loaded into the system. Post doing that the video needs to be broken down into frames in the form of image for frame extraction. Once we get the images, they need to be pre-processed after which we will use SVD for the feature selection which is one key step of the design. We do connectivity clustering after that to get hold of the keyframes by comparing the images for RGB. When the clustering process is done, we are going to use the keyframes to generate our final summarized video.

5 Implementation

The implementation of the system is performed using the Python programming language in the Jupyter notebook computing framework. Various libraries of Python have been utilized in the implementation of the system.

Extraction of constituent frames: CV2 library of Python supports the operation to extract the frames. The extracted frames are stored in a directory on the local drive. CV2 by default extracts the frames embedded in a video in BGR format. For more precise operations, the color space is converted from the BGR to RGB color space. This is done using the cvtColor() function of the CV2 library.

Getting the histograms of the frames: After the conversion of the color space, histograms are calculated for the frames to identify the number of pixels related to each of the colors. These histograms act as a feature set for clustering. Histograms of the features are calculated using the calcHist() function of the CV2 library.

Feature Extraction and Connectivity-based Clustering:

The connectivity clustering in the study is done on the singular value decomposition (SVD) of the feature vector generated using the histograms of the video frames. This study calculates 96 singular values for the processing. SVD in the study is performed using the svds() function in the Scipy library of Python. Function svds() calculates the partial singular value decomposition of the sparse matrix of the feature vector.

2. The output of the svds() is a unitary matrix. Using this matrix, the projections of it are obtained through the multiplication of its transpose and a diagonal matrix. This is done using the transpose() function of Python.

3. These projections are then used along with the cluster centroids to calculate dissimilarity between the features. This helps in clustering the features together.

4. Similar features are clustered in one cluster and dissimilar features are clustered in other.

6 Evaluation and Results

This section discusses the results obtained from the study. Figure 6 below shows the keyframes obtained from the video that will be used for summarizing the video. From the figure, it can be observed that the contents in the video are relatively similar.

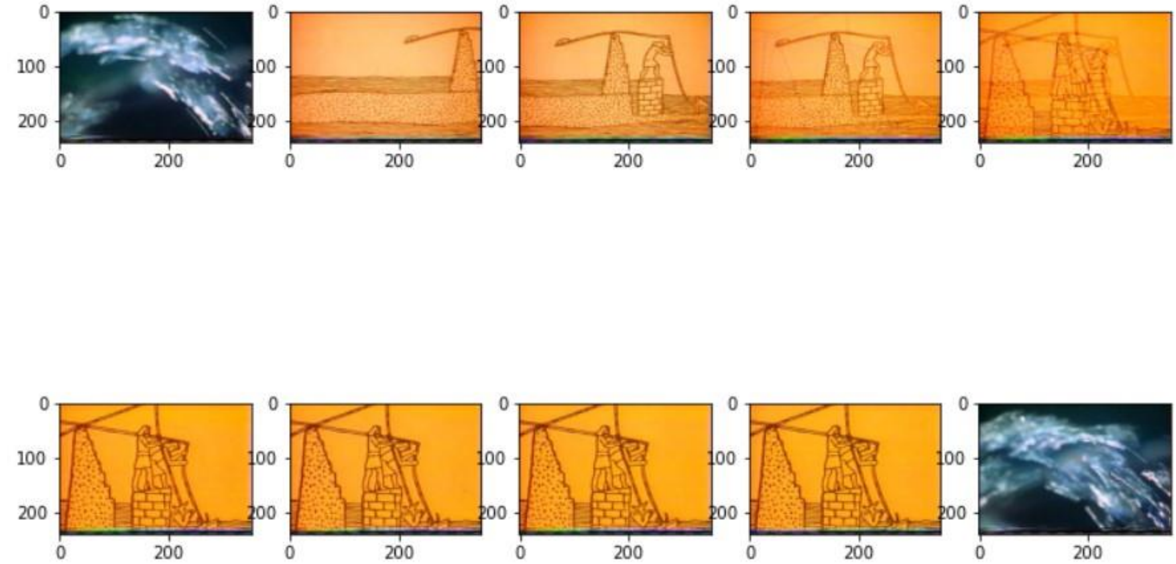

Figure 6: Extracted keyframes from the video

The frames generated are combined to make a video of 10 fps using the fourcc codec. This is done using the VideoWriter() function available in the CV2 library.

The video summarization technique is compared with the VSUMM algorithm. Results of the same are depicted in figure 7 below.

```
For summary of v21
They are different
Number of keyframes is more
The similarity scores of the keyframes is 2.028155672577972
-------------------------------------------------------
For summary of v22
They are different
Number of keyframes is more
The similarity scores of the keyframes is 0.9070976222626665
-------------------------------------------------------
For summary of v23
They are different
Number of keyframes is more
The similarity scores of the keyframes is 3.4671601671595447
-------------------------------------------------------
For summary of v24
They are different
Number of keyframes is more
The similarity scores of the keyframes is 3.7229090180168822
-------------------------------------------------------
For summary of v25
They are different
Number of keyframes is more
The similarity scores of the keyframes is 2.7132763323764975
-------------------------------------------------------
```

Figure 7: Comparison of the model with the VSUMM algorithm

The figure above depicts the comparison of the keyframes generated through the method used in the study and the VSUMM algorithm. The keyframes extracted from both algorithms are compared based on structural similarities. Higher similarity scores ensure that the algorithm developed in the study is working at par with VSUMM for videos, v21, v23, v24, and v25 except for that of v22.

7 Conclusion and Future Work

Video summarization is an important process to obtain critical information from a long video. It finds its uses in various applications across multiple domains. One of the best examples of this is video surveillance. Obtaining important information from a long video such as a day's recording can be obtained from video summarization. Another use-case of the technique is its use in sports. Video summarization is beneficial in obtaining highlights of a sports event which is always a lengthy event.

The algorithm that has been developed in the study helps to obtain a summarized video. The algorithm works using the connectivity clustering technique to obtain the most relevant keyframes. The algorithm after comparison with a previously developed VSUMM model ensures its reliability in video summarization. Hence, in conclusion, a video summarization model can be developed using the connectivity clustering technique used in the study.

Future work of the study involves using semi-supervised learning techniques to automatically summarize a video using the learning obtained from the clustering method along with the advantages of supervised learning. The accuracy of the video summarization can be further improved by using sound and captions in the video. This will help in video summarization of various important programs or events such as lectures, seminars, board meetings, and video sharing websites like YouTube, etc.

References

Apostolidis, E., Adamantidou, E., Metsai, A.I., Mezaris, V., and Patras, I., 2021. Video summarization using deep neural networks: A survey. Proceedings of the IEEE, 109(11), pp.1838-1863.

Apostolidis, K., Apostolidis, E. and Mezaris, V., 2018, February. A motion-driven approach for fine-grained temporal segmentation of user-generated videos. In *International Conference on Multimedia Modeling* (pp. 29-41). Springer, Cham.

Basavarajaiah, M. and Sharma, P., 2019. Survey of compressed domain video summarization techniques. *ACM Computing Surveys (CSUR)*, *52*(6), pp.1-29.

Cai, S., Zuo, W., Davis, L.S. and Zhang, L., 2018. Weakly-supervised video summarization using variational encoder-decoder and web prior. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 184-200).

Elfeki, M. and Borji, A., 2019, January. Video summarization via actions ranking. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 754-763). IEEE.

Haq, H.B.U., Asif, M. and Ahmad, M.B., 2020. Video summarization techniques: a review. *International Journal of Scientific & Technology Research*, *9*, pp.146-153.

He, X., Hua, Y., Song, T., Zhang, Z., Xue, Z., Ma, R., Robertson, N. and Guan, H., 2019, October. Unsupervised video summarization with attentive conditional generative adversarial networks. In *Proceedings of the 27th ACM International Conference on multimedia* (pp. 2296-2304).

Hussain, T., Muhammad, K., Ding, W., Lloret, J., Baik, S.W. and de Albuquerque, V.H.C., 2021. A comprehensive survey of multi-view video summarization. *Pattern Recognition*, *109*, p.107567.

Hussain, T., Muhammad, K., Ullah, A., Cao, Z., Baik, S.W. and de Albuquerque, V.H.C., 2019. Cloud-assisted multiview video summarization using CNN and bidirectional LSTM. *IEEE Transactions on Industrial Informatics*, *16*(1), pp.77-86.

Ji, Z., Xiong, K., Pang, Y. and Li, X., 2019. Video summarization with attention-based encoder-decoder networks. *IEEE Transactions on Circuits and Systems for Video Technology*, *30*(6), pp.1709-1717.

Li, Y., Wang, L., Yang, T. and Gong, B., 2018. How local is the local diversity? reinforcing sequential determinantal point processes with dynamic ground sets for supervised video summarization. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 151-167).

Liang, G., Lv, Y., Li, S., Zhang, S. and Zhang, Y., 2022. Video summarization with a convolutional attentive adversarial network. *Pattern Recognition*, *131*, p.108840.

Lin, J., Zhong, S.H. and Fares, A., 2022. Deep hierarchical LSTM networks with attention to video summarization. *Computers & Electrical Engineering*, *97*, p.107618.

Liu, Y.T., Li, Y.J. and Wang, Y.C.F., 2020. Transforming multi-concept attention into video summarization. In *Proceedings of the Asian Conference on Computer Vision*.

Ma, M., Mei, S., Wan, S., Hou, J., Wang, Z. and Feng, D.D., 2020. Video summarization via block sparse dictionary selection. *Neurocomputing*, *378*, pp.197-209.

Ozkose, Y.E., Celikkale, B., Erdem, E. and Erdem, A., 2019, November. Diverse neural photo album summarization. In *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)* (pp. 1-6). IEEE.

Rochan, M., Ye, L. and Wang, Y., 2018. Video summarization using fully convolutional sequence networks. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 347-363).

Sen, D. and Raman, B., 2019. Video skimming: Taxonomy and comprehensive survey. *arXiv preprint arXiv:1909.12948*.

Sharghi, A., Borji, A., Li, C., Yang, T. and Gong, B., 2018. Improving sequential determinantal point processes for supervised video summarization. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 517-533).

Trevnenski, G., 2021. Evaluation of the SUM-GAN-AAE method for Video Summarization.
Workie, A., Sharma, R. and Chung, Y., 2020. Digital Video Summarization Techniques: A Survey. *International Journal of Engineering Research and*, V9(01).

Wu, J., Zhong, S.H. and Liu, Y., 2020. Dynamic graph convolutional network for multi-video summarization. *Pattern Recognition*, *107*, p.107382.

Wu, T., 2019, February. CNN and RNN-based deep learning methods for digital signal demodulation. In *Proceedings of the 2019 International Conference on Image, Video and Signal Processing* (pp. 122-127).

Zhang, K., Grauman, K. and Sha, F., 2018. Retrospective encoders for video summarization. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 383-399).

Zhao, B., Li, H., Lu, X. and Li, X., 2021. Reconstructive sequence-graph network for video summarization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(5), pp.2793-2801.