IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105238 |

Video Summarization Using Convolutional Neural Network

Sameer Sarode, Aneesh Patil, Pooja Satao, Sneha Bachhav, Prof. Pragati Pandit

UG Students, Dept. of IT Engineering, K. K. Wagh Institute of Engineering Education and Research, Nashik, India

Assistant Professor, Dept. of IT Engineering, K. K. Wagh Institute of Engineering Education and Research, Nashik, India

ABSTRACT: In today's digital world videos play a significant role in information sharing and surveillance systems for security purposes. Many surveillance videos get added to the network due to the increase in the need for security. Video recorded from these surveillance systems is large, requires a huge amount of time for monitoring, and has large storage space. The goal is to find important and informative content in the video where motion is detected. In this project initially, input video is provided by the user, which is further loaded using the OpenCV video capture object.A looping statement is used to read each frame of the input video and is converted to grayscale. The absolute difference between the current frame and the previous is calculated, and the threshold has applied the difference to reduce noise and artifacts pixels with intensity above the threshold are set to the maximum value, and pixels with low intensity than the threshold are set to zero. If motion gets detected, the frame is passed to CNN. A Convolutional Neural Network generates a feature vector for a frame passed by extracting relevant features like texture, shape, color, etc. The Feature vector is used to identify frames with significant motion. Finally, frames are added to list the and are combined to generate a summarized video. Hence this deep learning methodology will be useful to extract the important part of the video and provide the summarized necessary content to the content-seeker.

KEYWORDS: Video Summarization, Keyframe selection, Extraction, Deep Learning, Convolutional Neural Network, Clustering.

I. INTRODUCTION

Today, system comprised of Surveillance cameras has become very useful and important in every field, mostly in the security industry. For doing such a inspection or observation, cameras are placed at various places for example Banks, ATMs, Public transport, Airports etc. Surveillance system basically comprises of such cameras which are placed at public and private premises and are capable to capture videos that can be stored and sent over communication network. These cameras can then be connected to digital computer to display the videos which helps monitoring ongoing activity and extracting useful information from the video. The estimated number of cameras currently in operation are more than 20 crores.

They all together produced more than 10 million GB of data per week. Surveillance videos are sometimes large in size so extracting information from it becomes very tedious task as huge video contains many ideal and repeated scenes which are completely useless and it is not possible for humans to continuously sit in front of the computer and monitor activities from the such a large video as time management is the most important factor for humans. Also, hardware required to store such a video should be in the large size.

Video summarization is the best solution for all the problems faced in surveillance system. Video summarization is basically condensation of huge video into the smaller one which so that we can manage our time and storage efficiently and reduce human work to monitor the large video. Video is nothing but a collection of serial static frames or images so, summarizing a video basically includes deletion of less important frames. Video summarization process uses deep learning algorithm to summarize the video into short clip by extracting important elements such as objects by detecting important objects and moving objects from original video. Video summarization has become very important to create the precise summary of given video. Summarization of the video includes extraction of important scenes, meaningful occurrence and particular objects which gives the core information of the whole video.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105238 |

II. RELATED WORK

According to Mayu Otani, Esa Rahtu, Janne Heikkil, and Naokazu Yokoya. In this paper Author proposed to learn semantic deep features for video summarization. For a deep feature learning they use CNN with two sub-network for videos and descriptions. In this approach the input video is represented by deep features from the segments then segments corresponding to clusters centers are extracted to generate the video summary. But the limitation in this approach is the uniform length videos are extracted. But it is difficult to extract the features from short videos.[1] According to Mrigank Rochan, Yang Wang. In this paper they introduced fully convolutional sequence networks (FCSN) for video summarization. The proposed models are inspired by fully convolutional networks in semantic segmentation. In computer vision, video summarization and semantic segmentation are often studied as two separate problems. It has been shown that these two seemingly unrelated problems have an underlying connection. Hence have adapted popular semantic segmentation networks for video summarization. The models achieve very competitive performance in comparison with other supervised and unsupervised state-of-the-art approaches that mainly use LSTMs believing that fully convolutional models provide a promising alternative to LSTM-based approaches for video summarization. Finally, the proposed method is not limited to FCSN variants.[2]

According to Tejal Chavan, Vruchika Patil, Priyanka Rokade, Surekha Dholay. Surveillance cameras at different areas record videos which are usually connected to the computers and later get monitored by humans in the control room. So, such a surveillance videos got from the cameras are the important sources for information and knowledge extraction but videos which are large may lead to some problems like missing some important part of the video, Taking long time to monitor the whole video, wastage of storage and human power. Hence, Video summarization is the best solution to this problem so that monitoring becomes 3 quick by removing redundant information.[3] Some other key-frame extraction techniques are introduced by various researchers include a method which uses ranking based model to generate the four-thumbnail summaries for web videos which are compact in nature. The model learns from pair-wised labeling which helps in the selection of a compact set of four frames and perfectly fits a compact UI to show the videos on web pages and phones.

III. PROBLEM STATEMENT

Internet videos are quite diverse and play an important role in information sharing as well as in surveillance system for monitoring and security purpose. Many numbers of surveillance cameras get added to the networks as need is increasing day by day. Video recorded from these surveillance cameras are large in size which require huge amount of time for monitoring and large storage space. Cameras are placed at various places for example Banks, ATMs, Public transport, Airports etc. Videos are sometimes large in size so extracting information from it becomes very tedicious task as huge video contains many ideal and repeated scenes which are useless and difficult to monitor activities from such a large video. To overcome this problem video summarization is a solution which consist of an algorithm that takes a video as an input and extract set of important frames to represent the entire video content in an effective manner.

IV. METHODOLOGY

A. Convolutional Neural Network:

A Convolutional Neural Network (CNN, or ConvNet) is a class of artificial neural network (ANN), most commonly applied to analyse visual imagery. CNNs are also known as Shift Invariant or Space Invariant Artificial Neural Networks (SIANN), based on the shared-weight architecture of the convolution kernels or filters that slide along input features and provide translation-equivariant responses known as feature maps. Counter-intuitively, most convolutional neural networks are not invariant to translation, due to the down-sampling operation they apply to the input. They have applications in image and video recognition, recommender systems, image classification, image segmentation, natural language processing 5 A CNN is a deep learning algorithm which takes an input image assigns importance to various objects in image, and is able to differentiate between images.The architecture of CNN is similar to that of connectivity pattern of human in brain.It contains five Layer,Convolutional Layer,pooling Layer,Fully connected Layer and output layer.

B. OpenCV:

Open Source Computer Vision Library is an open-source computer vision and machine learning software library designed to help developers to create an application that can understand visual data. It was originally developed by Intel in 1999 and now is maintained by OpenCV team.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105238 |

OpenCV provides various algorithm to perform image video processing , object detection and recognition.It has become popular choice for computer vision research and application due to ease of use, cross-platform and large community of contributors.It is used for various purpose in project like to read the input video, to convert frame to grayscale using 'cv2.cvtcolor()', to apply thresholding to different image using 'cv2.threshold()', to write output video using 'cv2.VideoWriter()'.

V. SYSTEM ARCHITECTURE

Convolutional Neural Network is a deep learning algorithm which is used for video summarization. Python is the language we have chosen for building the system. The user uploads input video further the algorithm extracts frames from that video. A first frame is chosen and get compared with next frame by computing the pixel distance between frames. For object movement detection image processing tool like OpenCV2 is used.

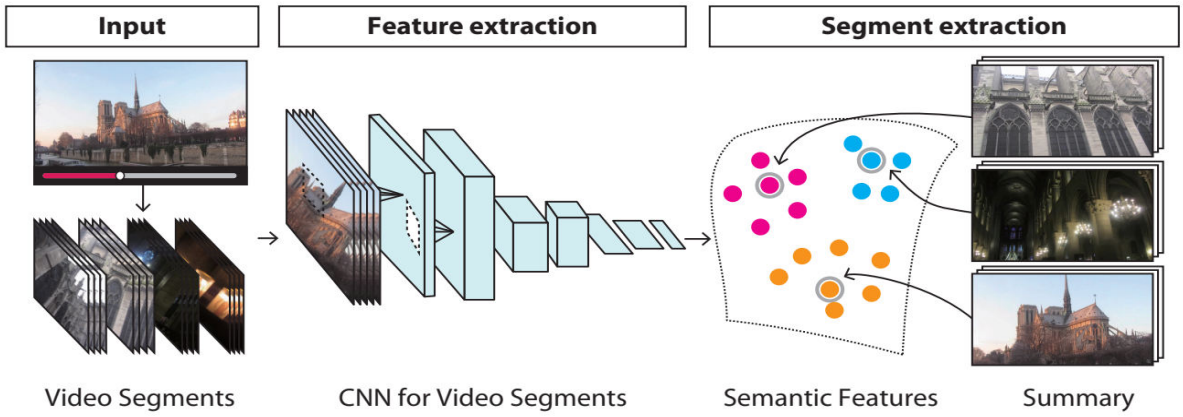

Fig. 1: CNN Model

Key-frames are selected by comparing the object distance with threshold. As shown in Fig. 1. The selected key-frames are clustered to generate final summarized video containing only useful information or a content.

VI. WORKING MODULE

The working module or graphical user interface (GUI) of a project is the front-end that provides a visual representation of the project features and allows users to interact with it. A GUI is designed to be user-friendly and intuitive.The design and functionality of the GUI depends on the project requirements and the needs of user.A well-designed GUI can enhance the user experience.

Here are some snapshots of our Graphical User Interface(GUI).

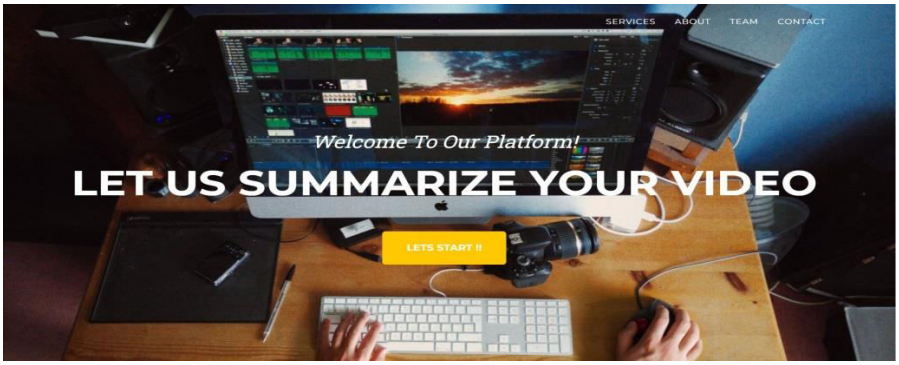

Fig 2. Home Page

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105238 |

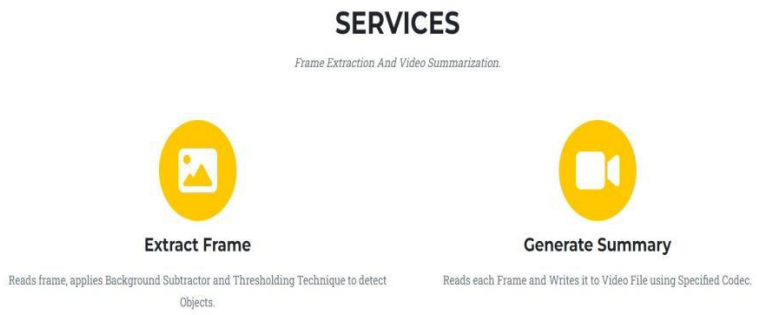

Fig 3. Service Page

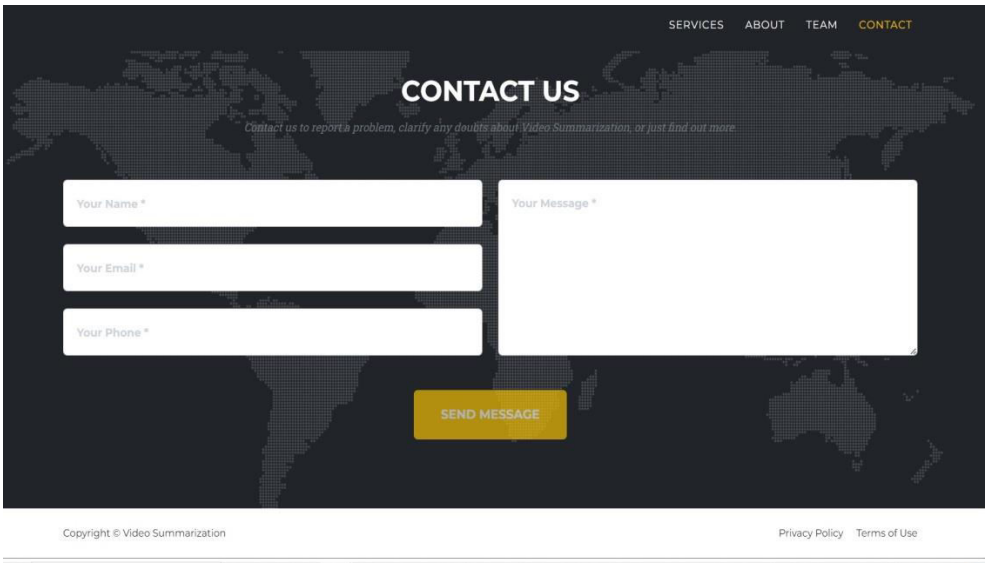

Fig 4. Contact Page

VII. RESULTS

We got the following result as shown in Table I for different input videos got from the surveillance video system of the ATM. We used our algorithm to extract the key frames from the video to get the summarized video and also checked the reduction in size, frames and length of the summarized video with respect to the original video. This result shows how efficiently the system works on different input videos.

Following Fig.5 and Fig.6 are the pictures of frames got after detecting moving object and performing distance measure for content completeness of the video i.e. Object should be in the centre of frame.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105238 |

Video ID	Input size	Input length	Output size	Output length
Video 1	46.70 MB	01 m 01 s	16.09 MB	24.02 s
Video 2	08.88 MB	01 m 44 s	03.55 MB	23.00 s

Table 1. Results

Fig.5 shows the folder containing the resultant frames got after applying frame extraction algorithm on original input video along with their frame number. A folder contains those frames in which motion is detected.

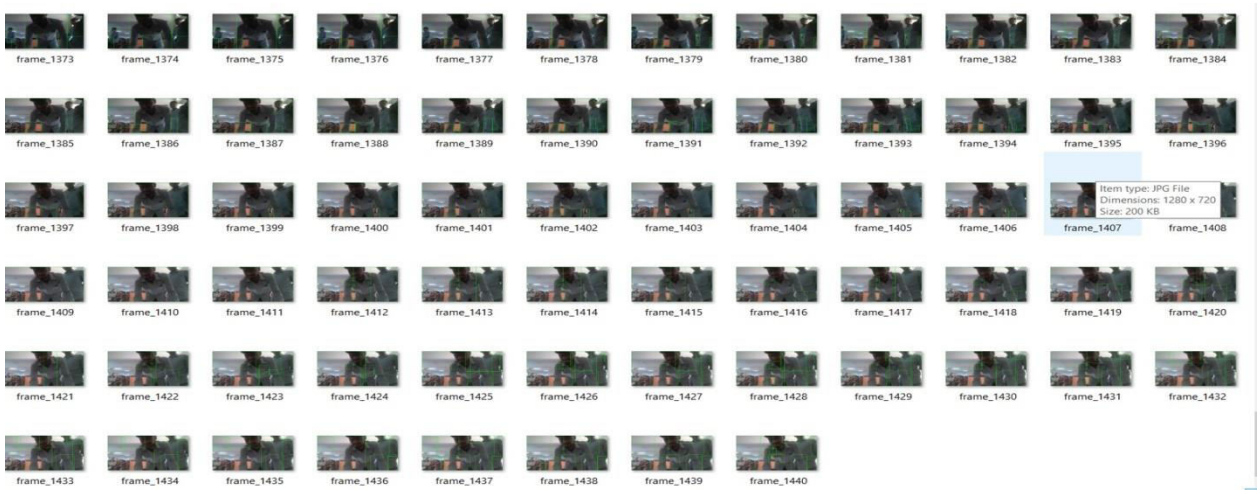

Fig. 5 Folder containing motion detected frames

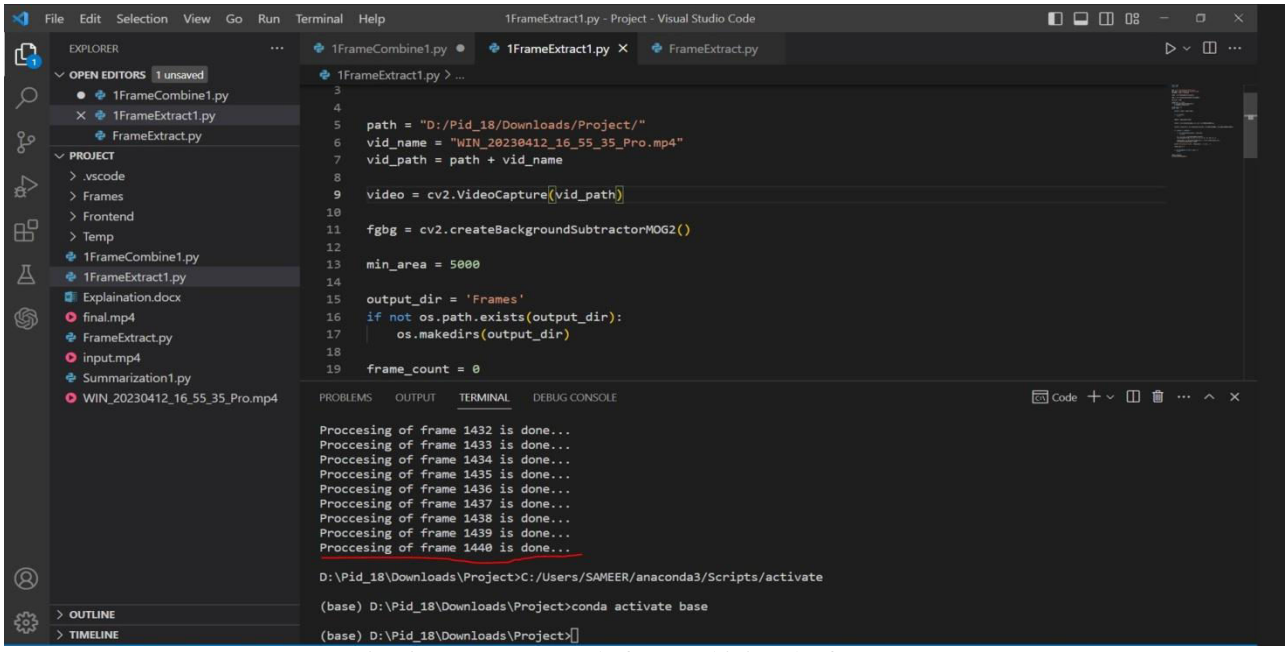

Fig. 6Frames generated after combining the frames

Fig.6 shows the frames as a key frame got after applying combine algorithm which are used to generate summarized video.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105238 |

VIII. CONCLUSION

The proposed system focuses on providing a user an easier and better monitoring of the surveillance video captured from the cameras placed in public and private premises for the security purposes by summarizing the video. Summary of the video contains the important scenes or events present in the original video which gives short and brief idea about the video. Our project removes idle scenes from the video and chooses only those scenes where multiple objects interact with each other i.e. removing useless content and produced the efficient video as output. The system also provides evaluation metrics to measure the performance of summarized video.

REFERENCES

1. Mayu Otani, Esa Rahtu, Janne Heikkil, and Naokazu Yokoya, "Video Summarization using Deep Semantic Features." *In IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR).* (2016) 174-180
2. Mrigank Rochan, Yang Wang, "Video Summarization Using Fully Convolutional Sequence Networks." *In 2018 provided by the Computer Vision Foundation, content of paper is similar to the content of officially published ECCV*, (2018)
3. Tejal Chavan, Vruchika Patil, Priyanka Rokade, Surekha Dholay, "Superintendence Video Summarization." *In International Conference on Emerging Trends in Information Technology and Engineering (ETITE)*, (2020)
4. Gong, Y., Liu, "Video summarization using singular value decomposition." *In Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR)*, (2000) 174180
5. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Transactions on Pattern Analysis and Machine Intelligence (2017)
6. De Avila, S.E.F., Lopes, A.P.B., da Luz, A., de Albuquerque Ara´ujo, A.: Vsumm: A mechanism designed to produce static video summaries and a novel evaluation method. Pattern Recognition Letters 32(1), 56–68 (2011)
7. Eigen, D., Fergus, R.: Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: IEEE International Conference on Computer Vision (2015)

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details