SUMMARIZATION OF SURVEILLANCE VIDEOS BASED ON OBJECTS DETECTED

Ms. Resha Govekar Khandeparkar[1], Ms. Vaishali Shirodkar[2]

[1]Dept. of Information Technology, Goa College of Engineering, Goa, India
[2]Assistant Professor, Dept. of Information Technology, Goa College of Engineering, Goa, India

---***---

Abstract- *The surveillance cameras are installed in public and private premises are on rise in recent trends, hence there is an exponential growth in the amount of surveillance videos available.*

These surveillance videos produce a huge amount of data which may be used as information by not only the owner but can be useful from the analysis or investigation point of view to the Police officials, Criminal Investigation Department or to the security personnel.

The storage, retrieval, management and analysis of these video can be a challenging task. Hence there is a need of efficient technique which will produce a short skim of these surveillance videos by retaining the most prominent scenes and discarding the least important information.

In this paper, we propose a method which will produce an efficient video summary of these surveillance videos based on deep Convolutional Neural Network.

The proposed methodology works by identifying the various objects present the video and keeping only those video frames which has prominent objects and discarding the remaining frames; thus, it maintains the interest appeal of the user.

Keywords: Segmentation, Object Detection, Key Frame Extraction, Summarization.

1. INTRODUCTION

Security is become a primary concern in the recent trends and surveillance plays a major role in security. The number of CCTV cameras installed in the public or private premises has been on high rise for over a decade now. These surveillance cameras produce enormous amount of data on daily basis. The storage, retrieval, management and analysis of the data can be a challenging task.

These surveillance video data can be of great use. Besides the owner, these videos are often used by Police officials, security personnel and by Criminal Investigation Department as a part of investigation as these videos can serve as an evidence.

Processing videos can be time consuming, as to extract the relevant or the required useful information the user has to watch the entire long videos. Moreover, a large part of these videos is either non-essential or repetitive. Manually watching these long videos for hours can be time consuming and tiring. Besides, focusing on the same video or video screen for long hours can be erroneous. The chances for a person to miss out any important event in the video are high.

Hence it is very important to develop an efficient technique that allows the analysist to choose the most illuminative part of the video and still keeping the integrity of the video intact.

There are multiple techniques which provide video summarization. But most of these techniques deal with videos such as sport, entertainment and so on. These summarization techniques normally produce a video summary in the form of highlights of sports videos, trailers of entertainment videos etc.

There are various ways in which video summarization can be achieved. Video summary is basically a short summary or highlights of a long video. Video summary of an original video should contain the most important aspects of the video and redundant information must be discarded.

Video summarization are of two types: Static video summarization and Dynamic Video Summarization. In static video summarization key frames or main frames are extracted and it produces a series of Images combined together as a short video. In Dynamic video summarization video skimming is done which produces a short video as an output.

One approach for creating summary of surveillance video is based on motion detection [1]. Here Omar Elharrouss, Noor Al-Maadeed and Somaya Al-Maadeed proposed a a block-based similarity evaluation module after which the most likely moving objects within the motion blocks are determined based on block-based morphological erosion and dilation operations. Their module like an eliminates the unwanted examination of the background region and allows the object extraction to process the blocks containing moving objects.

Khan Muhammad, Tanveer Hussain, Sung Wook Baik, proposed Efficient CNN based summarization of surveillance videos for resource-constrained devices [4]. Their framework was based on three phases, shot segmentation using deep features, computing image memorability and entropy for each frame of the shots, and keyframes selection from each shot for summary generation. Certain images gain more attention of the users. Frames which contain scenes like fight, romance are more prominent then the others [4]. Such frames are captured to form a summary.

2. PROPOSED METHODOLOGY

The primary objective of our proposed system is to be generate a summary of the surveillance video using the CNN model. Video summary be generated based on the importance of the scene in the frames of the videos. Frames with prominent scenes will be selected using the CNN model. Prominent scenes here mean the frames in which an object is detected.

The advantage of the project will be mainly for owner of the private property or to the investigating officer to identify the important parts of the video. It will be easier to identify the humans present in the video and easily trace the actions of any person and get the information of any vehicle from the scene.

2.1 System Architecture

The proposed system will generate a short summary of the input video using CNN (Convolution Neural Network) based on the objects detected the various frames of the video.
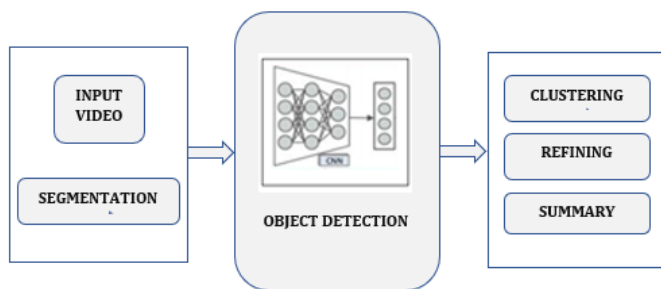

Figure 1. Proposed System Architecture

2.2 Proposed Algorithm

Input: Surveillance Video

Output: Summary of the surveillance video

Step 1: Input the surveillance video.

Step 2: Extract the video frames.

Step 3: Feed the Frames to the CNN.

Step 4: Detect the Objects using pretrained model

Step 5: Select the frames with Objects detected.

Step 7: Find the similar frames.

Step 8: Discard the redundant frames and retain distinct frame.

Step 9: Generate the video summary.

2.3 Video Segmentation

The surveillance videos contain the still cameras. The background of such videos remains constant and they do not have any shots. Video segmentation is a process of converting video into images. The process does it by splitting the entire video into different frames.

2.4 Object Detection

This process here detects the various objects present in the different frames of the video. The various objects such as Human, Animal or dog would be detected using CNN.

2.5 Video Summarization

Video summarization is the final phase of this project where only frames would be selected where in an object is detected. The frames with object present in it will be combined together to form the video summary.

3. EXPERIMENTAL RESULTS

In this section, we demonstrate relevance of our proposed method by providing the experimental results. The approach is implemented using Python programming language. The tests are performed on the input videos. The input data set is a surveillance video taken from a private place. The input video is split into frames. From each frame the objects such as human, animal or vehicle is detected. Only those frames which contains the object are then taken for further processing. If the activities of the person detected needs to be tracked and all the selected frames are considered for video summary. Else clustering of the frames is carried out, where in the redundant frames are discarded and only the distinct one is kept. And the summary is created from these selected frames.

For evaluation and testing purpose we have taken video footages from the surveillance system installed at the entrance of a house. The input video was divided into different frames. Only those frames with person being detected were then considered for video summary. Figure 4 shows the process of video summary being calculated for an input video.

International Research Journal of Engineering and Technology (IRJET)
Volume: 08 Issue: 08 | Aug 2021   www.irjet.net

e-ISSN: 2395-0056
p-ISSN: 2395-0072

Figure 2. Layout of Frames of the input video.

Figure 3. Layout of Output Frames.

Figure 4. Video Summarization Process

Figure 5. Prediction Score of a Trained Prediction Model.

CONCLUSION

The video summary will be generated by detecting the objects present in the surveillance video. The proposed project is based on a fully Convolutional Neural Network that will detect the objects present in the surveillance video. There are two major modules in the proposed methodology. First module performs the segmentation of the input video and the second module detects the objects using a CNN. The main focus was to highlight the important aspect of a Surveillance video, which is to identify the objects present in the video such as Person, Vehicle etc, selecting only those frames and discarding the idle frames. The proposed methodology could generate more accurate video summary.

4. REFERENCES

[1] "Video Summarization based on Motion Detection for Surveillance Systems", Omar Elharrouss, Noor Al-Maadeed, Somaya Al-Maadeed¸ IEEE 2019.

[2] A Deterministic Key-Frame Indexing and Selection for Surveillance Video Summarization T Michael Moses, K Balachandran, IEEE 2019.

[3] A Survey on Video Summarization Techniques Mahesh Kini M, Karthik Pai, 2019 Innovations in Power and Advanced Computing Technologies (i-PACT).

[4] Efficient CNN based summarization of surveillance videos for resource-constrained devices Khan Muhammad, Tanveer Hussain, Sung Wook Baik, Elsevier 2018.

[5] Video Summarization of Surveillance Cameras, Po Kong Lai, Marc Decombas, Kelvin Moutet, IEEE 2016.

[6] Smart Surveillance Based on Video Summarization, Sinnu Susan Thomas, Sumana Gupta, Venkatesh K. Subramanian, 2017 IEEE Region 10 Symposium (TENSYMP).

[6] Selecting Attentive Frames From Visually Coherrent Video Chunks For Surveillance Video Summarization, Wenzhong Wang Qiaoqiao Zhang Bin Luo Jin Tang Rui Ruan Chenglong Li, IEEE 2017.
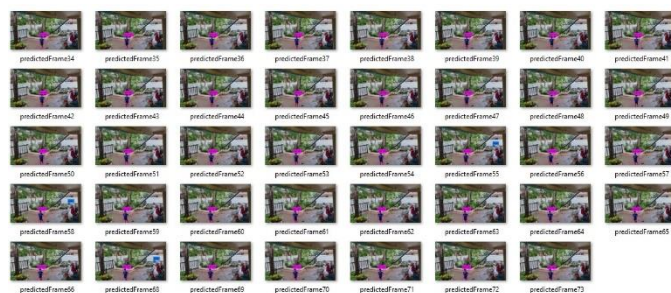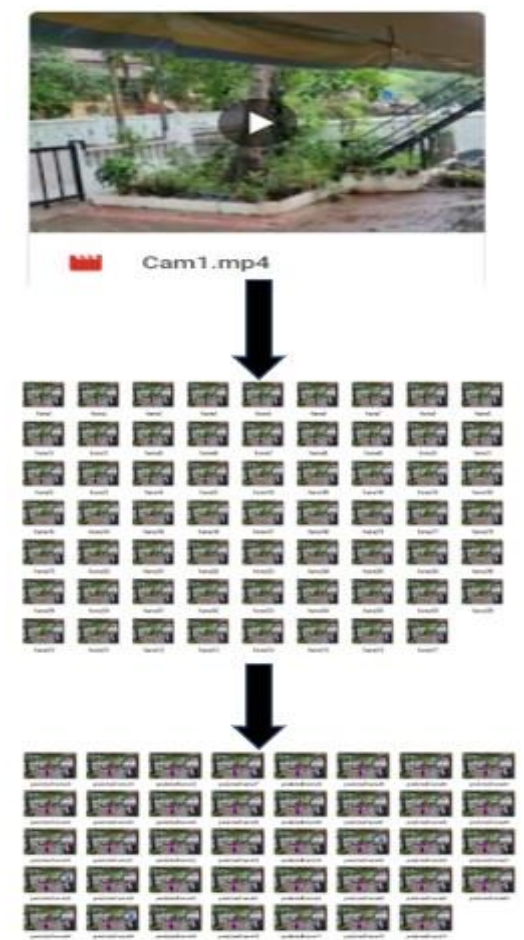
[7] Context-Aware Surveillance Video Summarization Shu

Zhang, Yingying Zhu, and Amit K. Roy-Chowdhury, IEEE Transactions On Image Processing, Vol. 25, No. 11, November 2016