Video Synopsis

Sushma Amudalapally[1], Komal Rana[1], Priti Vichare[1], Akshata Sarode[1], Jyoti Kolap[2]

[1]Student, Department of Electronics & Telecommunication Engineering, Atharva College of Engineering, Mumbai, Maharashtra, India
[2]Professor, Department of Electronics & Telecommunication Engineering, Atharva College of Engineering, Mumbai, Maharashtra, India

Publication Info

Article history:
Received : 15 February 2020
Accepted : 00 May 2020

Keywords:
Dynamic video synopsis, Video stream, Video surveillance, Video summarization.

Corresponding author:
Sushma Amudalapally
e-mail: *sushma.amudalapally@gmail.com*

Abstract

In the era of technological advancements, the demand for digital access of real-time data has increased tremendously. The accelerating pace of human intellect has also driven individuals to find out ways of accessing the digitalized data in a smarter way, with high accuracy and reduced time duration. One such ideology accustomed in the Image Processing domain includes the retrieval of video frames by using techniques of Video synopsis goals to produce a compact video representation while retaining the various activities of the first video. We depict dynamic video synopsis, where maximum video frames are condensed by way of simultaneously showing various activities. Previous procedures for video abstraction addressed in large part the temporal redundancy by using deciding on representative keyframes or time intervals. In dynamic video synopsis, the method is to shift the activity of interest into a substantially shorter period, at some point of which the interest is a way denser.

1. INTRODUCTION

Due to the limited time constraints, compact storage, and communication systems, there is an outgrowing need for video summarization in numerous applications.

For example, in a campus security application, it accelerates investigation, i.e., it increases campus security and effectively reviews video to understand threats and investigate incidents, attain situational awareness, and derive operational intelligence.

Keyframes are selected from the original video and then included in the video summary based on these applications. Papers aim to improve the efficiency in the browsing of surveillance video, and make it faster and more accurate. We provide with the technique of video synopsis with clustering of previous activities, and showing together only similar activities of the same cluster.

The main aim is to achieve minimum summary length and maximum information coverage in the framework. We propose a strategy by selecting different test recordings that contain video information, the proposed technique breaks them into two processes and applies video rundown by casting off a certain frame in the video as per multi-level calculations that reduces irrelevant edges in resultant video.

The video input in Fig.1 shows a person walking, and after a certain period of inactiveness displays a moving car. A condensed video synopsis can be produced by presenting the car and the person simultaneously.

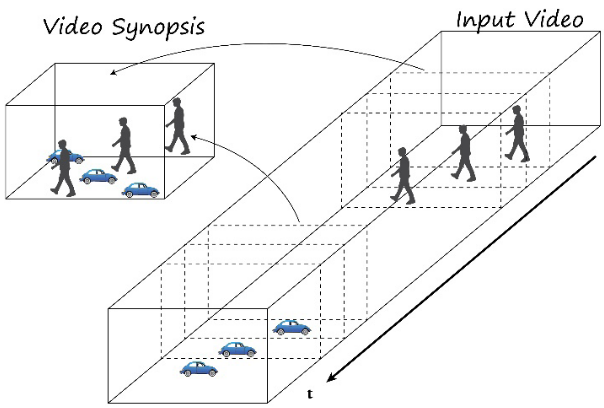

Fig. 1: A compact video synopsis

Fig. 2: Video fragmented into Key Frames

2. LITERATURE SURVEY

The scheme for the project was coined after acknowledging the issues faced by the people in the working environment and gathering information from the various sources. Prof. Shmuel Peleg from the Hebrew University of Jerusalem held the technical foundations to the startup company BriefCam Ltd.,[1], which creates short summaries of long surveillance video, able to summarize hours of surveillance video in minutes. In the former works, researchers have done related literature where they have proposed various techniques of video synopsis.[2],[3] A video clip describes visible activities on time, and reducing the time axis permits observant an outline of such a clip in a more condensed time. Dynamic and Static are the different techniques that can be used for converting long video into a short clip. Video synopsis is the technique of condensing a video based on activities taking place, and the purpose is to display maximum activities as possible simultaneously in minimum time. The videos based on activities consolidation was introduced by Rav-Acha et al. (2006) [1] under the title of video synopsis, an innovative procedure that diverts recognized activities in the time domain to show them concurrently in a less period, as represented.

3. APPROACHES FOR VIDEO SYNOPSIS

3.1. Static Video Summarization

Static video summarization is also known as a reference frame (R-Frame). Static video summarization consists of 3 types of summarization. They are:

- Shot segment summarization
- Scheme segment summarization
- Video-based summarization.

The video takes the keyframe uniformly, or it can take keyframes randomly.[4] Scene segmentation is done by scene detection, included in the scenes are all parts that have an exposition link in the original video or at the same time. In scene segmentation, the adopted keyframes are extracted as shown in Fig. 2. The techniques may differ is terms of features like the color histogram, luminance, as well as clustering algorithm as k-means, hierarchical.

3.2. Dynamic video summarization

Dynamic video summarization referred to generate a short video from its original video. A technique used to get a dynamic video application is known as the skimming semantic analysis.[5] In the segmentation of videos into scenarios with semantically meaning where shorter videos are acquired, the video relying on 2-dimensional histogram entropy of pixel of an image is adapted.

The common techniques under dynamic video summarization are the motion model and Singular Value Decomposition (SVD). Finally, the clustering technique is used where clustering together the same shot frames, followed by extraction of the same frame per cluster as a key frame.

3.2.1. Singular Value Decomposition (SVD)

Initially, at the input, the feature frame matrix A is created, and then we perform SVD on it. In the refined featured space, initially, the static frame cluster is found, defined as a content unit, and later on, use this content value as the threshold to cluster the rest to the frames. Because of all this, the optimal video is generated.[6]

3.2.2. Key Frame Extraction

The output frame hence obtained, is the replica of the original frame with maximum accuracy. It can obtain maximum results by providing the best summary if the original video where most of the common techniques used in the method of keyframe extraction are based on Shot activity and finally based on motion analysis.

4. METHODOLOGY

4.1. Single Shot MultiBox Detector

The object detection plays a major role in video synopsis. Multiple objects with varying scales/dimensions are present at various locations in videos. Thus the detection of an object becomes an important factor in achieving the relevant information[7]. The presence of an object can be detected after its conversion into frames, and the image hence obtained shows the objects of interest that play a vital role in video synopsis. The Single Shot Multi-Box Detector (SSD) is employed within the project that uses a single convolutional neutral network to discover an object in a picture. SSD presents the primary deep network-based object detector that doesn't resample pixels or features for bounding box hypotheses and is as correct as approaches that do. Faster R-CNN[8], YOLO do perform the task of detecting the object, but the accuracy of SSD is far better than these methods[9]. The heart of SSD is prophesying section counts and box offsets for a predefined set of error

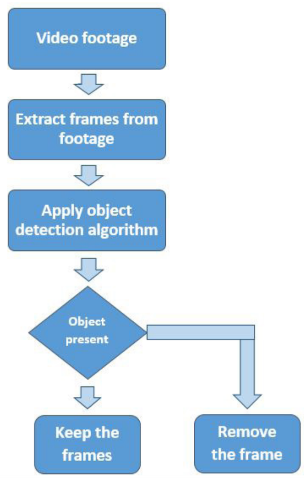

Fig. 3: Video Synopsis

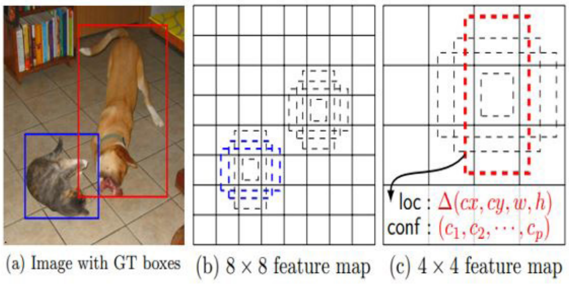

Fig. 4: Applying SSD using COCO Dataset

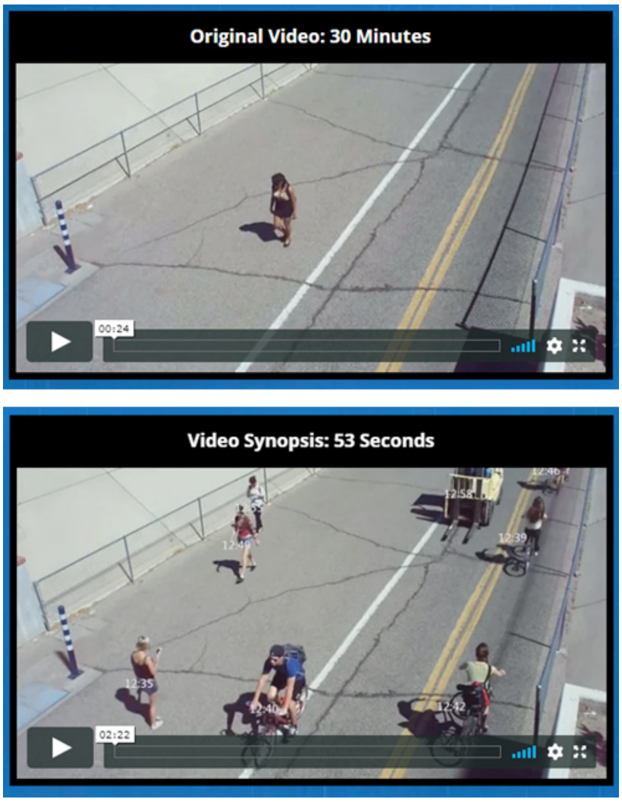

Fig. 5: Outcome of Video Synopsis

bounding boxes using small convolutional filters enforced for featured maps[10]. The accuracy of the model is analyzed by varying input objects using the COCO dataset. [11]

4.2. COCO Dataset

The term COCO stands for Common Objects in Context. In this dataset, the objects are predefined with tags of various categories and super categories, which can be used in image segmentation, image classification, and object detection. It helps us classify the objects concerning the category of 2D icons views that it belongs to.

In Fig. 4, the cat and dog of super categories from the animal category are detected using the coco dataset with predefined tags. Further, using the 8x8 or 4x4 feature maps, the accuracy of the object detected can be calculated.

5. OUTPUT

With respect to Fig.3, Once all the objects are detected from their respected class and categories, the probabilities of the object recognition are determined. After this, the mapping of the objects takes place in a specific video frame for the object of interest. While doing so, the presence of such objects of interest in each frame is observed; and the frames with these objects are retained, casting off the other unwanted frames.

In Fig. 5, it can be observed that footage undergoing the technique of video synopsis has a lesser number of video frames hence producing a compact version of the original video.

6. CONCLUSION

Video summarization has recently attracted vital analysis. Accordingly, techniques and mechanisms are instructed. The paper presents the effectiveness of the investigation by specializing in the reduction of power consumption in mobile applications and extra price within the method of downloading an exact video signal. The paper splits the first video into video content victimization using a projected algorithmic program, the video is being divided into its contained frames, and every frame split consistent with a time issue, then take a median for the set of block frames. Finally, merge the output from previous steps to construct the newly processed video.

7. REFERENCES

[1] Rav-Acha, A., Pritch, Y., & Peleg, S. (2006, June). Making a long video short: Dynamic video synopsis. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (Vol. 1, pp. 435-441). IEEE.

[2] Pritch, Y., Rav-Acha, A., Gutman, A., & Peleg, S. (2007, October). Webcam synopsis: Peeking around the world. In 2007 IEEE 11th International Conference on Computer Vision (pp. 1-8). IEEE

[3] Pritch, Y., Rav-Acha, A., & Peleg, S. (2008). Nonchronological video synopsis and indexing. IEEE transactions on pattern analysis and machine intelligence, 30(11), 1971-1984.

[4] Khan, S., & Pawar, S. (2015). Video summarization: survey on event detection and summarization in soccer videos. International Journal of Advanced Computer Science and Applications, 6(11).

[5] Dilawari, A., & Khan, M. U. G. (2019). ASoVS: Abstractive summarization of video sequences. IEEE Access, 7, 29253-29263.

[6] Gong, Y., & Liu, X. (2000, June). Video summarization using singular value decomposition. In Proceedings IEEE

Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662) (Vol. 2, pp. 174-180). IEEE.

[7] Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search for object recognition. International journal of computer vision, 104(2), 154-171.

[8] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).

[9] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778)

[10] .Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2013). Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229.

[11] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).