다시점 영상에 대한 이상 물체 탐지 기반 영상 시놉시스 프레임워크

팔라시 잉글[1,2], 유진용[1,2], 김영갑[1,2,†]
[1]세종대학교 정보보호학과
[2]세종대학교 지능형드론융합전공
palash@sju.ac.kr, instrol30@gmail.com, alwaysgabi@sejong.ac.kr

Abnormal Object Detection-based Video Synopsis Framework in Multiview Video

Palash Yuvraj Ingle*, Jin-Yong Yu*, Young-Gab Kim*,†
*Dept. of Computer and Information Security, and Convergence Engineering for Intelligent Drone, Sejong University

Abstract

There has been an increase in video surveillance for public safety and security, which increases the video data, leading to analysis, and storage issues. Furthermore, most surveillance videos contain an empty frame of hours of video footage; thus, extracting useful information is crucial. The prominent framework used in surveillance for efficient storage and analysis is video synopsis. However, the existing video synopsis procedure is not applicable for creating an abnormal object-based synopsis. Therefore, we proposed a lightweight synopsis methodology that initially detects and extracts abnormal foreground objects and their respective backgrounds, which is stitched to construct a synopsis.

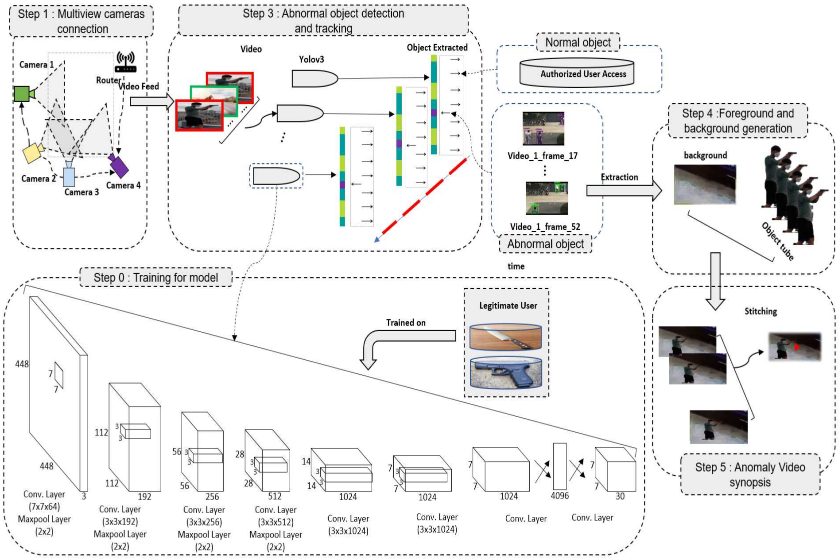
1. Introduction

With the innovations in connectivity and technological advancement surveillance has increased to record daily life activities this has significantly generated an enormous amount of video data. Most of the content in video footage that is obtained is empty frames and it takes a lot of time to analyze and extract useful information from it. Intelligent surveillance can prevent acts like a crowded fight bomb blast with earlier detection. The United Nations Office on Drugs and Crime (UNODC) states that the object used is a gun for committing a crime [1]. The most widely accepted and utilized video condensation method is the synopsis which deals with shifting the foreground object in time and domain space, thus creating a shorter video. As surveillance video contains various objects, analyzing enormous video footage to classify such an abnormal object from the rest of the objects. However, traditional synopsis methods construct a synopsis for fixed field of view in a single camera; the obtained synopsis is rather crowded and suffers from a collision. In addition, the existing synopsis approach does not provide a reliable solution for extracting and creating a shorter anomaly object video from multiple cameras.

Therefore, this paper proposed a lightweight synopsis framework for categorizing and extracting abnormal objects in multiple camera, thus creating a condensed video for analysis and storage. The paper's main contribution is summarized below: 1) We proposed a lightweight synopsis framework that synchronously extracts the anomaly object tube from multiple cameras for constructing a shorter video. 2) We evaluated our study with the state-of-the-art synopsis methodology.

The remainder of the paper is organized as

† Corresponding author



(Figure 1) Process view of the AVS framework

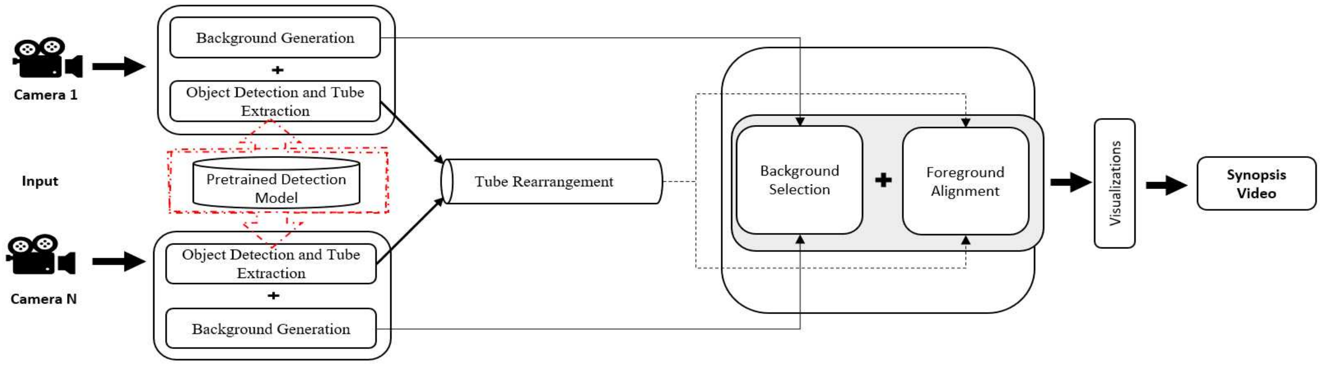
follow. In section 2, we briefly review the related work, and the proposed framework is described in section 3. Finally, section 4 provides the conclusion of this study.

2. Related Work

Initially, in 2006, based on 3d markov random field, a low-level optimization framework in which they detect the activity while tracking and extracting them in sequence, after which they perform alignment of the foreground with the background using simulated annealing, thus stitching them to construct a video synopsis [3]. However, their study can only build a synopsis for a limited period; therefore, to over this problem, based on a user query simultaneously [4], multiple object tubes were extracted to construct a single long synopsis; major problem faced by this study was a collision [5]. Using a set theory approach [6] and shifting the object in the group achieved a maximized visual content of the synopsis. A non-chronological [7] method was used to construct a synopsis of different time zone, and it wasn't evident for content visualization. Similarly, a background modeling and foreground extraction method were implemented to generate a summary of interest [8]. A multiple kernel [9] similarity selection criteria for generating tube was used for critical observation, thus excluding the content redundancy. Using a CNN for detecting and extracting an object was incorporated by the swarm algorithm; however, it suffers from high computational complexity [10]. Most of the existing approaches suffer from challenges like high collision, dense synopsis, computational complexity, and background synthesis while creating a video synopsis.

3. Abnormal video synopsis

Abnormal video synopsis (AVS) can detect and extract only abnormal objects in time and domain space, thus constructing an abnormal content synopsis. Fig. 1 illustrates an process view of the proposed framework for classifying and extracting the abnormal object tubes in real-time from multiple cameras to construct an abnormal content video synopsis. AVS consists of the following steps: In the first step, we trained and a CNN (Yolov3) model for detecting an abnormal object (i.e., a person carrying a gun or knife is unauthorized). We synchronously aligned the multiple cameras; we later performed the stitching based on this alignment. In the third step, based on the object's detection and classification, we provide access to the services. If the detected object is an unauthorized user, later on, the best selected annotated unauthorized

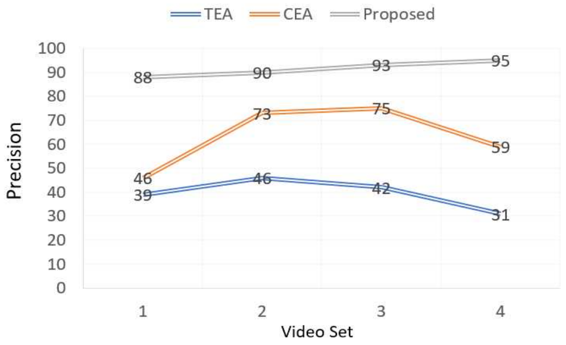


(Figure 2) System view of the video synopsis framework

object is tracked and extracted using a key timestamp. Then, for each video sequence, a respective background is extracted using dynamic programming, and the foreground object tube is queued concerning the alignment. Finally, we stitched the anomaly foreground object tube with the background template, thus creating a shorter video for analysis. A system view of the AVS framework is depicted in Fig. 2. A background is generated using a gaussian mixture model, and the foreground extraction of the object is done based on the pre-trained object detection algorithm. For example, let abnormal objects present in the video be A_{ob}, and normal objects present in the video be N_{ob}. At this stage, the N_{ob} object tube is excluded. Then, only the A_{ob} tube is stitched with the respective background. Finally, the optimization is done using simulated annealing to avoid a collision, such as the obtained synopsis video only contains the abnormal object.

4. Experiment Results

We evaluated the proposed AVS framework using an AMD Ryzen 5 3500X 6-core processor with 16 GB RAM. For testing the synopsis framework, we used a Logitech camera. The object detection model was trained on the ImageNet [11] dataset using TensorFlow, which has achieved a Mean Average precision of 96.54% for detecting abnormal objects. Additionally, we tested the framework with the existing techniques, such as s Time Equidistant Algorithm (TEA) [12] and the Constant Equidistance Algorithm (CEA) [13] on a customized dataset where the videos were obtained from IMFD [14]. Our observations show both this algorithm suffers from distortion while constructing the synopsis. There is a significant difference between the original video and the obtained synopsis video. A characteristic difference between the current study and the proposed studies is depicted in Fig. 3. The proposed algorithm outperforms the existing methodology for constructing an abnormal video synopsis; thus, the precision obtained for creating its synopsis ranges from 88% to 95% for the video data set.



(Figure 3) Precision chart of synopsis technique

5. Conclusion

Abnormal video synopsis framework is a lightweight framework designed for resources constrained devices. As we only create a synopsis for abnormal objects, the resultant synopsis is crucially in information and short in size. In the proposed methodology, we create a synopsis based on the classification of the object, which can be extended based on user requirements. The proposed framework does not suffer from noise, distortion, or collisions as we performed the stitching into avoid overlapping. The best selection of the abnormal objects helps to create smooth synopsis for better visualization.

ACKNOWLEDGMENT

This work was supported in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korean Government (MSIT) under Grant 2019-0-00231, and in part by the Development of artificial Intelligence-Based Video Security Technology and Systems for Public Infrastructure Safety.

Reference

[1] K. Office on Drugs and Crime (UNODC). Global study on homicide 2019. Data: UNODC Homicide Statistics 2019, 2019

[2] Baskurt, K.B. and Samet. R., “Video synopsis: a survey.”, Computer Vision and Image Understanding, 181, pp.26-38, 2019

[3] A. Rav-Acha, Y. Pritch and S. Peleg, "Making a Long Video Short: Dynamic Video Synopsis,", 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, 2006, pp. 435-441

[4] Y. Pritch, A. Rav-Acha, A. Gutman and S. Peleg, "Webcam Synopsis: Peeking Around the World,", 2007 IEEE 11th International Conference on Computer Vision, Rio de Janeiro, 2007, pp. 1-8

[5] Nie, Y., Xiao, C., Sun, H. and Li, P., “Compact video synopsis via global spatiotemporal optimization.”, IEEE transactions on visualization and computer graphics, 19(10), pp.1664-1676, 2012

[6] Xu, Min, Stan Z. Li, Bin Li, Xiao-Tong Yuan, and Shi-Ming Xiang, "A set theoretical method for video synopsis.", 2008 In Proceedings of the 1st ACM international conference on Multimedia information retrieval, New York, 2008, pp. 366-370

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

[8] Sun, H., Cao, L., Xie, Y. and Zhao, M., "The method of video synopsis based on maximum motion power.", IEEE In 2011 Third Chinese Conference on Intelligent Visual Surveillance, Beijing, 2011, pp. 37-40

[9] Zhu, X., Liu, J., Wang, J. and Lu, H., “Key observation selection for effective video synopsis.”, IEEE In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, 2012, pp. 2528-2531

[10] Moussa, M.M. and Shoitan, R., “Object-based video synopsis approach using particle swarm optimization.”, Signal, Image and Video Processing, pp.1-8, 2020

[11] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L,. “Imagenet: A large-scale hierarchical image database.”, In 2009 IEEE conference on computer vision and pattern recognition, Miami, 2009, pp. 248-255

[12] Andreas Girgensohn and John Boreczky. “Time-constrained keyframe selection tech┐nique.” , Multimedia Tools and Applications, 11(3), pp. 347 - 358, 2000

[13] Costas Panagiotakis, Anastasios Doulamis, and Georgios Tziritas, “Equivalent key frames selection based on iso-content principles.”, IEEE Transactions on circuits and systems for video technology, 19(3):447 - 451, 2009

[14] Jordy Gosselt, Joris Van Hoof, Bas┐tiaan Gent, and Jean-Paul Fox, “Violent Frames: analyzing internet movie database reviewers' text descriptions of media violence and gen┐der differences from 39 years of us action, thriller, crime, and adventure movies.”, Inter┐national journal of communication, 9(21), 2015