EdgeBox: Live Edge Video Analytics for Near Real-Time Event Detection

Bing Luo
Computer Science
Wayne State University
Detroit, USA
Bing@wayne.edu

Sheng Tan
Computer Science
Wayne State University
Detroit, USA
tan.sheng@wayne.edu

Zhifeng Yu
Edgemind Solutions LLC.
Detroit, USA
Zhifeng.F.Yu@gmail.com

Weisong Shi
Computer Science
Wayne State University
Detroit, USA
Weisong@wayne.edu

***Abstract*—With growing popularity of surveillance cameras in public and workplace safety there is increasing demand for automatic event detection. An edge computing based security solution, EdgeBox, is proposed to detect safety threat events in a near real time fashion and notify related parties when possible. We choose an armed bank robbery scenario as our case study to show how the EdgeBox solution can improve security awareness and prompt responses to safety incidents.**

***Index Terms*—public safety, edge computing, object detection, video analysis, real-time**

I. Background

The ever growing use of video surveillance system in public safety and home security along with the explosive emergence of edge computing is surging demand for automatic video analysis. While motion detection is a common feature built in with many surveillance cameras it is practically little useful due to its rudimentary design. A few high-end cameras recently introduced in the market were designed with specific purposes for tracing objects, detecting range intruding etc. They are very expensive and not expandable for other safety needs. After all, without manual monitoring, even the crime is captured by the surveillance cameras, the video footage is only used for investigation later after the incident is over.

Intrigued by the US Homeland Security Departments campaign If you see something say something, we propose a security edge computing solution (EdgeBox) that enables surveillance system to detect safety threat event near real-time, alert timely for proper responses.

With real-time video analysis, the EdgeBox solution enables surveillance cameras to say something (to alert) when they see something happens (security threat) without human operations. The solution can be deployed with some off line trained basic security threat recognition models and continue learning and adapting itself to the actual environment settings.

Our main contribution is to exploit one of the greatest potentials of surveillance systems: automatically recognize potential safety threats in the first place. With an edge computing based design, EdgeBox can notify related parties any safety threat in a real-time fashion, and only by that, it becomes possible to prevent or reduce actual security incidents and damages.

Another contribution is that the EdgeBox solution is designed to be plug and playable with any existing IP based surveillance systems without requiring an upgrade to high-end expensive cameras.

II. Related Work

A real-time surveillance system for event detection involves edge computing and live video analytics.

Weisong Shi et al. analyzed the issues in IoT and cloud computing and proposed edge computing. They also discussed challenges and visions about edge computing based on several case studies, such as smart home and cloud offloading [1].

Ganesh Ananthanarayanan et al. discussed the current situation of video analysis and identified 4 drivers for building an edge-based live large-scale video analytic system, which are latency, bandwidth, provisioning and the cost of video processing [2]. VLPR is an application of video image recognition technology in Vehicle License Plate Recognition [3]. It can extract and identify license plates of moving vehicles. Scott Gardner et al. introduced an intelligent video threat detection system which can detect guns with deep learning method [4].

III. Platform Design

A. Architecture

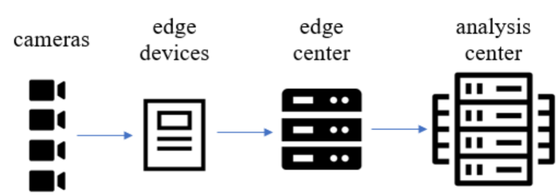

Fig. 1. The Overall Architecture of EdgeBox.

The overall architecture of the EdgeBox solution is shown in Figure 1. An edge device detects and manages a small group of cameras deployed on the same local area network. The edge device could be a desktop, NAS system, etc. It usually covers a relatively small area. Several edge devices are connected to an edge center, which is mainly responsible for real-time video analysis and abnormal detection. It may cover a relatively large area e.g. a building, a factory, etc. All edge centers are connected to the analysis center to perform complex activity detection that may take advantage of both deep learning and computer vision. It also provides a centralized management interface (console interface or web services) to end users and can communicate with cloud services if necessary. With that management interface, the security personnel can replay the video capturing related to a safety threat , remotely monitor the situation real time. It also helps minimize false alarms and unnecessary actions.

The actual deployment architecture may vary to address different levels of complexity and capabilities of surveillance systems. It can be a subset of the components in Figure 1 ranged from only one edge device for all the functionality to distributed multiple analysis centers to with high availability. So the solution can scale up and down based on video analytics demands.

B. Process

The first step is to discover and connect to existing IP cameras. EdgeBox can be plugged to play on the same local area network where the surveillance system runs on and streams each video source for real-time analysis. To do so, we will have an adapter to support standards like ONVIF and mainstream camera manufactory interfaces. Then, the system analyzes the working condition and effectiveness of each camera by examining whether a camera is connected, frozen, etc. This can be done on edge devices because such analysis is not resource hungry. Furthermore, this process prevents useless videos from being analyzed further. After that, the system uses computer vision based algorithms to perform image-based real-time effectiveness analytics including noise, blurring, etc. Motion detection and simple object detection are also done in this step. For example, it can detect escalators in a mall and monitor whether they work in the right direction. This process can be executed on edge center.

Image-based computer vision algorithms are usually good for some specific use cases. Emerging deep learning based AI analysis are more robust in general but requires more computing power. This process can detect complex and abnormal activities, like fighting, robbing, tabbing, etc. Once a potential security threat is detected, EdgeBox will alert predefined parties in multiple forms such as SMS, email etc.

IV. Case Study

A. Scenario Analysis

Consider an armed bank robbery. In a bank cameras are usually installed above and behind the cashiers. When a robbery happens, victims are usually can do nothing. Although the cameras are recording while the incident occurring, security forces won't be notified until someone call them after the incident. With activity detection, however, EdgeBox will recognize any potential safety threat via real time video analytics. It is able to detect any assaulting weapons such as firearms or knives and suspicious behavior such as abnormally forceful motion etc. Once a safety threat is detected, an alert at predefined levels based on threat assessment will be sent to relevant departments. As a result, the robbery can be either prevented or the damage can be controlled with proper response from the security force a timely manner.

Fig. 2. The Binary Image of Hand.

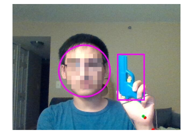

Fig. 3. Firearm Detection.

B. Solution Description

To validate the platform architecture, we designed and implemented a prototype which focuses on firearm detection. The entire software system is deployed on a PC workstation with a web camera connected. It uses a computer vision algorithm to detect pistols. The algorithm uses HSV color space to segment the skin color areas and exclude the non-gesture areas according to the geometric features of gestures. Then the object held in hand is detected in combination with the number of fingertips and the visual characteristics of the target. HSV (Hue-Saturation-Value) [5] is a color model which describes color by tone, saturation, and brightness. The range of skin color in HSV color space is $Hue \in [30, 45]$, $Saturation \in [35, 200]$, $Value \in [20, 255]$. Then a binary image can be obtained by transformation based on this color range. After eliminating some small areas, the part of hand can be marked in figure 2.

Next step is to recognize objects in the area. Seven characteristic values of Hu invariant matrix [6] are obtained with OpenCV edge detection. Those values are used to calculate the aggregation centers by a fuzzy clustering algorithm. Finally, the result is compared with those in the training library based on the Hu invariant matrix and the clustering center. Once the similarity reaches the preset standard, the system marks it as a firearm shown in figure 3 and sends an alert.

V. Conclusion

In this paper, we discussed the requirements of automatic event detection based on real time video analytics and proposed a general conceptual architecture/process of EdgeBox. We then conducted a case study in a specific scenario. Our next step is to integrate different modules to validate the architecture and finalize the system architecture design and hardware specifications.

References

[1] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016.

[2] G. Ananthanarayanan, P. Bahl, P. Bodik, K. Chintalapudi, M. Philipose, L. Ravindranath, and S. Sinha, "Real-time video analytics time video analytics time video analytics–the killer appfor edge computing for edge computing."

[3] S. Du, M. Ibrahim, M. Shehata, and W. Badawy, "Automatic license plate recognition (alpr): A state-of-the-art review," *IEEE Transactions on circuits and systems for video technology*, vol. 23, no. 2, pp. 311–325, 2013.

[4] R. F. P. H. Scott Gardner, Parviz Palangpour, "Gundetect."

[5] V. Oliveira and A. Conci, "Skin detection using hsv color space," in *H. Pedrini, & J. Marques de Carvalho, Workshops of Sibgrapi*. Citeseer, 2009, pp. 1–2.

[6] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE transactions on information theory*, vol. 8, no. 2, pp. 179–187, 1962.