Towards Self-Enforcing Privacy Protection for Surveillance System

Kok-Seng Wong   Anuar Maratkhan   Nguyen Anh Tu   M. Fatih Demirci

Department of Computer Science, Nazarbayev University, Nur-Sultan, Kazakhstan
{kokseng.wong, anuar.maratkhan, tu.nguyen, muhammed.demirci@nu.edu.kz}

Abstract

The implementation and use of video surveillance (closed circuit television (CCTV)) technologies have raised awareness in privacy concern when people are being watched remotely and recorded continuously. The existing solution cannot prevent any party from viewing activities and collecting unwanted personal data of others. In this paper, we propose a privacy-preserving solution to ensure that only registered user can see the selected target in CCTV monitoring system. Our solution aims to protect individual privacy (i.e., personal identity) and to prevent cyberbullying if CCTV footage is released online. The privacy protected video will be used to detects, tracks, extracts, and classifies objects or persons in video analytics.

1. Introduction

The effectiveness of video surveillance technology is continuously improving as a safety and management tool for risk monitoring, managing staff, and crime monitoring (i.e., crime detection, prevention, and prosecution). Despite a number of advantages associated with this technology, there are some disadvantages to CCTV cameras, which are related to privacy issue. For example, the monitoring at the nurseries is linked to parent's mobile devices to allow them to keep watch over their children. It means that other parents also can view the activities and collecting unwanted personal data of other children and staff. This increases the risk of identity theft because the intruders can learn the lifestyle of a target victim. Hence, it is important to ensure that only registered user can see a selected target in CCTV monitoring system (e.g., parents see only their child and intruders in nursery stream video). Furthermore, the misuse of CCTV footage (e.g., share funny CCTV footage in social media) can cause a severe cyberbullying problem that may affect every part of a victim's lives and causing deep emotional issues.

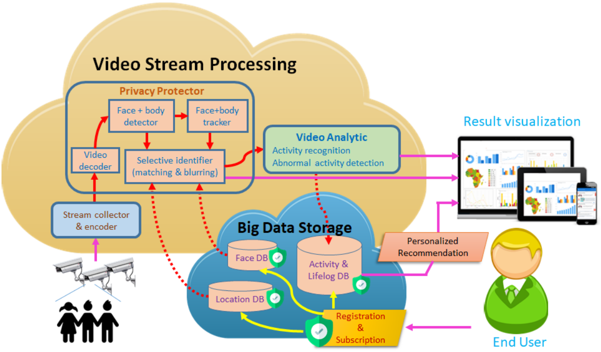

Figure 1. Proposed framework.

In this paper, we propose a privacy-preserving framework to address three challenges: (1) to protect the identity and personal life of individual (e.g., children and staff) in real-time, (2) to perform video analytic on privacy-protected video, and (3) to prevent cyberbullying. The first challenge can be solved by masking the CCTV footage in real-time (i.e., masking of multiple identities including face, objects, and body). In the second challenge, we require video analytics to detects, tracks, extracts, and recognizes objects or person's behaviors from privacy protected video. The third challenge requires each player to help others to preserve their own privacy. Naturally, the parents are not willing to reveal CCTV footage of their children. Hence, we assume that a dishonest player cannot obtain CCTV footage from other players. We adopt the following self-enforcing privacy in our solution:

Definition 1 (Self-enforcing Privacy): All players are aware of the risk of privacy leakage and dishonest players do not gain an advantage from leaking their privacy.

To the best of our knowledge, this paper is the first framework in the literature to support these features.

2. Methodology

In this section, we present our proposed methodology and illustrate its framework in Figure 1. Particularly, the proposed framework describes how we perform video stream processing with privacy protection in the cloud and how a user interact with the system. Our framework is composed of three main components including user registration, privacy protector, and video analytic:

User Registration. Firstly, users need to register our streaming service through the application installed on their personal mobile devices. After the registration, the user will upload their personal data to private cloud storage which is associated with our stream processing cloud service. We categorize the personal data into the facial database (contains face images and user information) and location-based objects database (e.g., school locations and logos).

Privacy Protector. This is the main component in our framework which modifies the content of video frames to protect the privacy. We propose to use a distributed messaging architecture known as Kafka [1] to process large-scale stream data that we collect from multiple cameras. We then extract and encode the information from video frames, and transfer the encoded frames to the privacy-preserving component. After the video frame is decoded, object detection mechanisms are performed to locate bounding boxes around each face, body, and location-based object in the frame. To improve efficiency, we combine an object detector with the tracker to continuously track a person or object across the sequence of frames. Following the object detection, bounding boxes of detected objects are matched with the facial data and the location data stored on the secure cloud storage. These data are streamed to the selective identifier to recognize the identities of interest and objects of interest. In this module, we can either utilize the face recognition approach [4] for each camera view or apply the person re-identification approach [7] to track and retrieve the identity of a specific person across multiple cameras. Finally, the pixels within the bounding boxes of the recognized objects and the unrecognized faces (and its associated body region) are modified (blurred or masked).

Video Analytics. This component incorporates artificial intelligence to CCTV cameras by automatically analyzing video content, sending out alerts and appropriate knowledge to users as well as security personnel, and hence reducing the need for manual monitoring. In particular, the video analytics primarily uses the data processed by privacy protector to describe human behaviors. Our video analytic makes use of Continuous Activity Recognition (CAR) and Abnormal Activity Detection (AAD). Within the former, the sequence of actions is identified by the activity segmentation technique [3] which jointly segments and classifies continuous actions on video frames in the offline manner. The output of CAR corresponds to a sequence of activity such as “sitting”, “walking”, and “running”. The stored CAR outputs in the form of activity data or lifelog will be used to generate a personalized recommendation (e.g. “today, your kid slept too much”). Meanwhile, AAD is to detect the abnormal (e.g., the kid is falling) or suspicious activities (e.g., an intruder breaks into the school) of the subject by using the activity detection technique [5]. The detected activities will be notified immediately to the user in the real-time manner.

3. Case Study

To demonstrate one case study, we apply our framework to some sample kindergarten videos taken from YouTube-8M Dataset, which consists of more than six million YouTube videos [2]. In our case study, two registered users views the same scene from two different smart phones. After performing face detection and face recognition, our system shows kids faces to the authorized smart phones only. In particular, the first smart phone is able to see the third kid from the right, while seeing the other faces blurred. The second smart phone, on the other hand, can see the first kid from the right. The images captured from these smart phones are presented in Figure 2 and Figure 3, respectively.

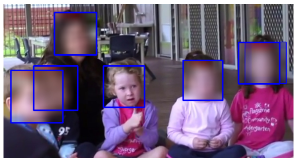

Figure 2. Image as seen from the first smart phone.

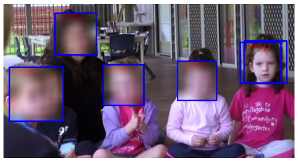

Figure 3. Image as seen from the second smart phone.

The proposed framework can be improved in several ways. Since the accuracy of face detection directly influences the effectiveness of our framework, missing a face causes a big privacy leak. Specifically, when a face is not detected, it will be shown to both authorized and unauthorized smart phones. In addition, if a face is detected properly but the integrated face recognition module yields false negative, i.e., it marks a known face as unknown, this face is also shown to all users, resulting in an other privacy violation. Thus, it is crucial for our system to have both robust and effective face detection and recognition approaches. In our preliminary experiments, we use the Haar Cascade classifier [6] for face detection and FaceNet architecture [4] for face recognition. In addition to these modules, one may also consider applying body detection to better preserve the privacy.

4. Conclusion and Future Plan

Our framework is designed to preserve individual privacy on stream CCTV while the privacy-protected videos still can be used to support intelligent analytics. At the same time, we also prevent cyberbullying if CCTV footage is released online. We demonstrated a case study on how to protect individual privacy based on face recognition.

In the future, to accomplish the objectives of this paper, we shall perform extensive experiments to find a balance between the privacy protection and the utility of video analytic. We also plan to extend our framework to problem such as automatic detection of cyberbullying in stream video. We believe that the future will see a growth in the demand for privacy protection for stream video.

References

[1] Apache kafka. https://kafka.apache.org/. 2

[2] Youtube-8m dataset. https://research.google.com/youtube8m/index.html. Accessed: 2019-4-8. 2

[3] K. Kulkarni, G. Evangelidis, J. Cech, and R. Horaud. Continuous action recognition based on sequence alignment. *Int'l J. Computer Vision*, 112(1):90–114, 2015. 2

[4] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. *CoRR*, abs/1503.03832, 2015. 2

[5] G. Singh, S. Saha, M. Sapienza, P. H. Torr, and F. Cuzzolin. Online real-time multiple spatiotemporal action localisation and prediction. In *ICCV*, pages 3637–3646, 2017. 2

[6] P. Viola and M. Jones. Robust real-time face detection. *Int. J. Comput. Vision*, 57(2):137–154, May 2004. 2

[7] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. Scalable person re-identification: A benchmark. In *ICCV*, pages 1116–1124, 2015. 2