A Privacy-Preserving Framework for Surveillance Systems

Kok-Seng, Wong
VinUniversity, Hanoi, Vietnam
wong.ks@vinuni.edu.vn

Nguyen Anh, Tu
Nazarbayev University, Nur-Sultan, Kazakhstan

Anuar, Maratkhan
Nazarbayev University, Nur-Sultan, Kazakhstan

M. Fatih, Demirci
Nazarbayev University, Nur-Sultan, Kazakhstan

ABSTRACT

The ability to visually track people present in the scene is essential for any surveillance system. However, the widespread deployment and increased advancement of video surveillance systems have raised awareness of privacy to the public, i.e., human identity in the videos. The existing indoor surveillance systems allow people to be watched remotely and recorded continuously but do not prevent any party from viewing activities and collecting personal visual information of people in the videos. Because of this problem, we propose a privacy-preserving framework to provide each user (e.g., parents) with a personalized video where the user see only selected target subjects (e.g., child, teacher, and intruder) while other faces are dynamically masked. The primary services in our framework consist of a video streaming service and a personalized service. The video streaming service is responsible for detecting, segmenting, recognizing, and masking face images of the human subjects in the video. Notably, it classifies human subjects into insider and outsider classes and then applies the de-identification (i.e., masking) to those in the insider class, including the target subjects. Subsequently, the personalized service receives the visual information (i.e., masked and unmasked faces) from the streaming service and processes it at the user's mobile device. The output is then a personalized video for each user. For security reasons, we require the surveillance videos stored in the cloud in an encrypted form. To ensure an individual remains anonymous in a group, we propose a dynamic masking approach to mask the human subjects in the video. Our framework can deliver both reliable visual privacy protection and video utility. For instance, users can have confidence that their target subjects are anonymized in other views. To utilize the personalized video, users can use analytics software installed on their mobile devices to analyze the activities of their target subjects.

CCS CONCEPTS

• **Security and privacy**; • **Human and societal aspects of security and privacy**; • **Privacy protection**;

ICCNS 2020, November 27–29, 2020, Tokyo, Japan

ACM ISBN 978-1-4503-8903-7/20/11…$15.00
https://doi.org/10.1145/3442520.3442524

KEYWORDS

Privacy-preserving framework, Surveillance systems, Visual privacy protection, Human de-identification

ACM Reference Format:
Kok-Seng, Wong, Nguyen Anh, Tu, Anuar, Maratkhan, and M. Fatih, Demirci. 2020. A Privacy-Preserving Framework for Surveillance Systems. In *2020 the 10th International Conference on Communication and Network Security (ICCNS 2020), November 27–29, 2020, Tokyo, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3442520.3442524

1 INTRODUCTION

The environment in which a computer vision (CV) algorithm operates has driven recent progress in video surveillance. The effectiveness of video surveillance technology is continuously improving as a safety and management tool for risk monitoring, managing staff, and crime monitoring. One of the main goals of CV is to extract useful information from visual data to support applications in agriculture, healthcare, and other industries. For example, vision-based surveillance systems employed various CV tasks, including face recognition, and human re-identification, to identify a person from the captured visual data [11]. Ideally, one would like surveillance to be carried out in a way that no personal information can be revealed from the videos. However, this is not possible when people are being watched remotely and recorded continuously. Furthermore, the CCTV operator, researchers, policymakers, and others can access surveillance videos.

The widespread deployment and increased advancement of video surveillance systems have raised awareness of privacy to the public. Recently, a primary school in Hangzhou China is using facial recognition to monitor the behavior and attentiveness of their students [14]. The system received many critics from the parents, students, and others who have concerns about what information is being collected or shared. The main reason is that the facial attributes not only provides details about the identity of a subject, but also reveals other person-related information such as gender, race, and age [5]. Such information is useful to profile a person when they are combined with external resources such as images and videos on social media. Specifically, in most person identification systems, facial features are commonly used to identify a human subject because the face is an accurate point to determine the identity compared to other visual information such as body and gait [3]. The detected face, together with human identity, can be used to reveal additional person-related information such as gender, race, age, and emotion automatically by using classification algorithms. Without the permission of users, the outputs of these vision tasks can be exploited for various analysis purposes, which again threaten the privacy of

individuals [22]. Hence, when human is the dominant objects of interest in CV applications, there is a need to protect the relationship between collection, dissemination, and storage of visual data.

To further explain our problem statement, let us consider a video surveillance system used at a nursery place. In common practice, parents are monitoring kids at nursery via CCTV stream that is linked to their mobile devices. This convenience also allows other parents to observe the behavior of other children in the CCTV footage, and hence, increases the risk of identity theft because the intruders can learn the lifestyle of a target victim. In this use case, we can see several types of visual privacy invasion based on different user's behavior. For example, parents or staff may want to collect personal visual information from others. In another case, a party may upload a funny CCTV footage in social media and causes a severe cyberbullying problem that may affect every part of a victim's lives and causing deep emotional issues. With the rise of malware designed to attack mobile device features, the attacker can access the victim's device to carry out malicious surveillance tasks such as audio and screen recording. A possible solution to protect children's identity is to mask their faces and show different views for different users based on access permission. For instance, when two users are viewing the same stream video, they only can see their child while other kids are masked. Subsequently, each child remains anonymous in the video such that no one can tell if a child is present or absent in a scene (except the parents). We can adopt an appropriate access control mechanism to minimize the chances of privacy invasion and to make surveillance systems widely acceptable.

1.1 Contribution and Plan of this Paper

This paper proposes a privacy-preserving framework for indoor surveillance systems. Our framework outputs a personalized surveillance video for each user. The personalized video only allows the users to see their target subjects (e.g., child, teacher, and intruder) while other human subjects are dynamically masked. We summarize our contributions as follows:

- Our framework provides visual privacy protection to the human subjects in the video and to prevent any party from collecting personal visual information of others.
- To ensure an individual remains anonymous, we propose to use a dynamic masking approach that masks the human subjects in the video with an updated average face when someone is leaving or newly appear in the scene.
- Our framework not only increases the confidence of the users concerning the privacy of their target subjects but also allows the user to utilize the personalized video for analytics purposes.

The remainder of this paper is structured as follows: In Section 2, we discuss related work, and in Section 3, we present the technical preliminaries of this work. Section 4 presents our proposed framework, followed by analysis and discussion in Section 5. The conclusion and future directions are in Section 6.

2 RELATED WORK

Protecting the visual privacy of human subjects in surveillance video across several different domains has become increasingly crucial as cameras become ubiquitous. In CV solutions, a common practice to protect the visual privacy of individuals is to anonymize people's faces in datasets composed of images or videos. Although the faces cannot be recognized, humans can easily analyze the rest of the image or video. In this subsection, we briefly review work closely related to our framework design for visual privacy protection.

In most of the existing works in computer vision, the focus is on how to prevent a machine from identifying an individual. For example, Saheb et al. [5] proposed an adversarial perturbation based algorithm to anonymize selected attributes of the facial image. The proposed algorithm causes the attribute prediction algorithm to output incorrect classification results. Although their work can preserve identity information and visual content from the machine, it does not surpass human observers to identify people or actions in the scene.

The majority of visual privacy protection schemes currently used in practice rely on ad-hoc methods such as mosaicking, blurring, and pixelation. Early work proposed by Schiff et al. [18] capable of obscuring the faces of human subjects present in the scene if they wear a visual marker such as a hat or vest. The video surveillance system will locate the face of the target subject and obscure it with an ellipse. Jana et al. [10] proposed an approach that performs sketching transform on a human face image at different privacy levels. However, the increase in privacy level will cause information loss and affects the video utility, e.g., the transformed outputs cannot be used to track back and search for the person and objects related to an incident. The contradiction of using ad-hoc methods is between high privacy protection and low video utility. Heavy pixelation or blurring can cause vision disturbance, which violates the purpose of the surveillance system. Also, they cannot leverage the utility of visual data, although they can alleviate some of the critical privacy concerns.

Face replacement (or face-swapping) is another way to protect the visual privacy of human subjects in the surveillance system. In the work of Bitouk et al. [4], they introduce a system that replaces faces by selecting a similar face from a database of real face images. However, the construction of such a database for practical applications is ethically and legally problematic. Newton et al. [15] propose the k-same face de-idenfication algorithm to guarantees that each de-identified face image could be representative of k faces. The k-same algorithm can remove privacy-sensitive information, but the resulting images suffered from mismatches in image alignment.

In more recent work on face de-identification by Hukkel et al. [9] the DeepPrivacy algorithm is introduced, which automatically anonymize faces in images while keeping the distribution of original data. Visual cryptography is another way to protect visual privacy in the surveillance system. Du and Li [8] proposed a privacy-preserving scheme for data security in video surveillance. In their work, they obscure human objects using motion blur, and then each blurred foreground object is encrypted into N shares by visual cryptography. Some works in the literature intentionally captured or processed video to be in special low-quality conditions [6, 17]. This kind of anonymization technique only allows for the recognition of some target events or activities in the video.

Unlike the existing works, we consider a different setting in our framework design. Instead of de-identify all human subjects in the

video, we allow the users to personalize stream video such that only he or she can see the selected target subjects in the video. Hence, the users can observe the identities and activities of the targeted subjects in the scene while protecting others, i.e., without knowing others' identities. Also, we propose to mask the target subject's faces (in other user's view) dynamically by using an average face computed from the face images in the database.

3 TECHNICAL PRELIMINARIES

Let $\mathcal{S}$ denote a group of n human subjects in a classroom C such that $\mathcal{S} = \{s_1 \cup s_2 \cup \ldots \cup s_n\}$. We consider different scenarios for a nursery surveillance system based on the present or absent of a target subject $s_t = \{child,\ teacher,\ intruder\}$ in C where $s_t \in \mathcal{S}$. Particularly, we use the following cases in our framework design:

1. s_t is present in C.
2. s_t is absent in C.
3. Two (or more) groups of students in .
4. C is empty.

For human subjects other than s_t, we replace their faces in the video with an average face computed from a database $\mathcal{D}$ where $\mathcal{S} \in \mathcal{D}$. In the presence of two or more groups, two average faces will be computed separately based on the face images from each respective group's database, i.e., $\mathcal{D}_1$ and $\mathcal{D}_2$. Subsequently, user i only see s_t^i in the video while other human subjects are masked with the average face. We refer intruders as those who have not registered with the client or existing users who have been deleted from the user database.

3.1 Security Requirements

To ensure visual data security, we define the following security requirements for our proposed framework:

1. Device (hardware) security: The user should be able to run the mobile app securely. We assume that the user's devices are equipped with secure hardware elements that prevent physical tampering (i.e., they are tamper-proof). Hence, they can store information and perform the required computations in a safe environment.
2. Device (software) security: The mobile app installed must be genuine, and its communication with the streaming server is secure.
3. Mutual authentication: We require mutual authentication between the user (mobile app) and the service provider (streaming server) before the video streaming begins.
4. Session key agreement: The session key will be generated between the mobile app and the streaming server after mutual authentication has been successful. This agreement is established to provide secure communications for message exchange.
5. Link security: The transmission of information between the mobile app and the streaming server must be protected.

Before the video streaming begins, a mutual authentication process is activated between the streaming app and the service provider. This process is essential to ensure that the same device used to install the streaming app provided by the service provider during the enrollment process and vice versa. We can assume that this is a continuous process that runs in the background between video streaming service and the streaming app. It will be automatically activated to establish a connection once the user unlocks the streaming app during the live broadcast.

3.2 Secure Cloud Storage

In our framework design, we store the surveillance videos in the cloud. To prevent the leakage of streaming videos, we require the service provider to store the videos in an encrypted form. In video encryption, the video content is first compressed, and then the compressed bitstream is entirely encrypted using a standard cipher such as DES, AES, IDEA, etc. [16]. Techniques such as partial encryption, selective encryption, soft encryption, and perceptual encryption are commonly used in video encryption. Recently, Elliptic Curve Cryptography (ECC)-based solution has been proposed for real-time video streaming applications [20]. The authors in [23] proposed an encryption method based on the ZUC algorithm for the mobile video surveillance system. Since video encryption is not the focus of this work, we refer authors to a survey paper in [12].

In our scenario, the service provider uses the client's public key to encrypt the original stream videos (i.e., unmasked video) while the users encrypt the personalized videos using their public keys.

4 PROPOSED FRAMEWORK

The idea behind the design of our framework is to provide each user with a personalized video where the user sees only selected target subjects while other faces are dynamically masked. Our solution consists of two processes: enrollment and personalized video streaming. In the rest of this section, we first introduce the system components used in our framework, then describe the steps in both enrollment and personalized video streaming processes followed by the details of average face construction.

4.1 System Components

As shown in Figure 1, there are three major players in our framework: a client, a service provider, and a group of users.

For instance, a nursery (client) subscribes to the video streaming service from a security surveillance company (service provider) to broadcast videos from security cameras installed in the classroom to parents (users). More specifically, the system components used in our framework design are as follows:

1. Client: A subscriber who has an active account with the service provider.
2. Service provider: An organization that provides video streaming service to its subscribers, e.g., a security surveillance company broadcasts videos from security cameras.
3. User: A registered user of the client who wants to use the streaming services.
4. Video streaming service: This service is used to detect, segment, recognize, and mask the face images of the human subjects in the video.
5. Cloud storage: Cloud storage is used to store subscriber information and surveillance videos.
6. Streaming app: A mobile application created and maintained by the service provider. The personalized service running in this app outputs a personalized video for each user.

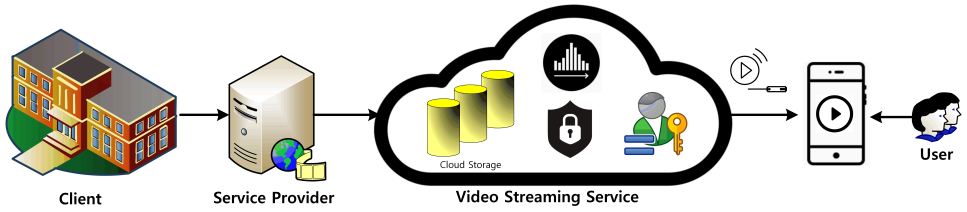

Figure 1: Overview of the Proposed Framework

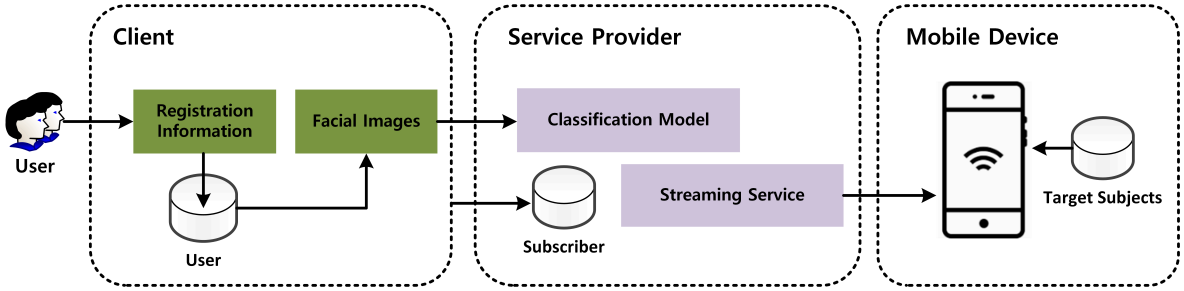

Figure 2: Enrollment Process

7. Streaming app interface: The streaming app interface that allows the users to watch the streaming videos, upload the personalized videos to the cloud, and access the personalized videos stored in the cloud.

4.2 Enrollment Process

The enrollment process is a mandatory operation that is performed by all players. The overview of the enrollment process is shown in Figure 2, and the details are as follows:

Step 1: The client subscribes to a surveillance system provided by the service provider. All subscriber information, such as type of organization, location, contact number, etc., will be collected by the service provider and stored in the cloud storage, i.e., subscriber database. The service provider then provides all necessary hardware (e.g., CCTV camera), software (e.g., front-end system), installation facilities, and training to the client.

Step 2: To enroll users into the system, the client first collects registration information such as name, social number, address, contact number, and facial image from its users. The collected information will be stored on the client-side, i.e., user database.

Step 3: Next, the client retrieves facial images from the user database to generate a training dataset for the service provider. Note that the face images in the training dataset should not contain other personal information of the users.

Step 4: The service provider uses the training dataset to build a classification model. This classification model aims to classify a given face image as an insider or outsider. We refer to an insider as a registered human subject in the user database, whereas outsider is a non-registered user.

Step 5: The service provider develops a mobile streaming app for the clients (to be used by registered users). This app can be customized according to the needs and requirements of the client.

Step 6: The client broadcasts a download link of the streaming app to its registered users. Also, it sends a target subjects database that consists of face images of the user, his or her child, the child's class teachers, and nursery staff to each registered user. Each user installs the streaming app and stores the target subjects database into a mobile device (e.g., smartphone or tablet). The learned classification model will be integrated into the streaming app when the user updates the app.

Step 7: To complete the enrollment process, each user activates the mobile app by verifying its identity with the client. We require the user to pass two verification phases before using the video streaming service. Accurately, the user first enters the account credential into the mobile app (i.e., user authentication with the client) and then uses in-app face verification (i.e., identification and matching with the face image stored in the target database) to access the stream videos.

4.3 Personalized Video Streaming Process

To ensure the surveillance videos do not violate the visual privacy of any individual, we provide a personalized video for each user. We illustrate the overall video streaming process in Figure 3. The details of each operation are as follows:

Step 1: The client performs authentication using the front-end system provided by the service provider. Once the authentication is successful, the communication between the CCTV camera and the video streaming service will be established.

Step 2: At this step, the service provider performs two operations on the incoming video captured by the CCTV camera. First, it encrypts the video using the client's public key. The encrypted video will be stored in the stream video database. Next, the service provider detects face images from the incoming video and then

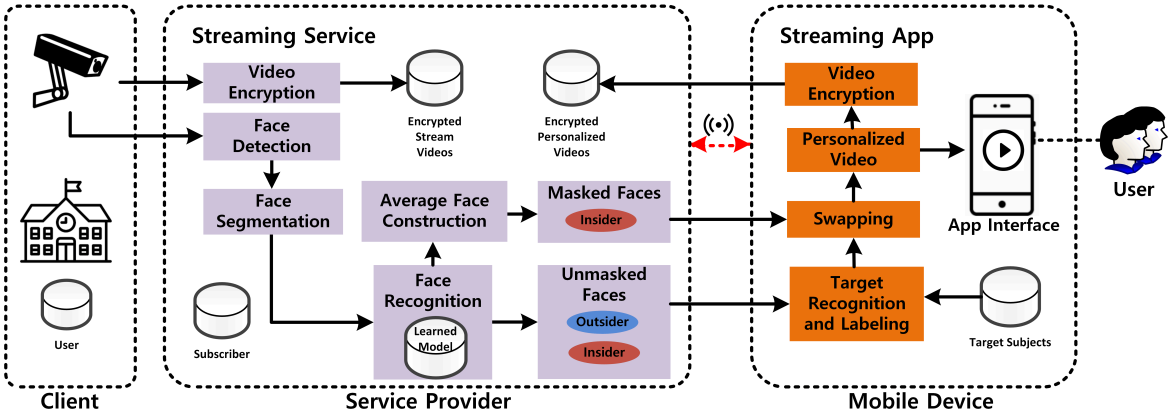

Figure 3: Video Streaming Process. The red dotted line represents mutual authentication between the streaming server and the mobile app.

segments the detected images and uses them as the inputs for the classification task.

Step 3: The learned model classifies the segmented images into insider and outsider classes. Based on the classified results, the service provider sends the unmasked face images to the streaming app for target subjects recognition and labeling. Also, the recognized insider face images will be sent for face averaging construction. We provide the details of face averaging construction in Section 4.5.

Step 4: To mask the insider face images, the service provider replaces each insider face with an average face computed from Step 3. The masked images will be sent to the personalized service for image swapping purposes.

Step 5: Assuming that the mutual authentication between the service provider and the streaming app is established. The streaming app first performs matching between unmasked insider class images with those stored in the target subjects database. Next, the streaming app labels all matched images according to the name or role of the target subject. For instance, the user's children will be labeled with their names while the teacher and staff will be labeled using their roles, such as "teacher" or "staff." For unmatched face images, the streaming app labels them as "unknown."

Step 6: After the target subjects are labeled, the streaming app swaps the face images of masked insiders in Step 4 with those matched target subjects in Step 5. In other words, this swapping step unmasked all target subjects in the video. Note that the images of masked and unmasked insiders are arranged and placed in the same position.

Step 7: Upon the completion of the swapping operation, the user watches the personalized video (except the target subjects, other insiders are masked) from the streaming app interface.

Step 8: To securely store the personalized videos at the cloud, the users first encrypt the videos by using their public keys and then upload the encrypted videos into the cloud storage via the streaming app interface. Alternatively, the service provider can help the users to encrypt the personalized videos to reduce the computation overhead at the mobile device. Furthermore, video encryption can be performed in both online or offline mode.

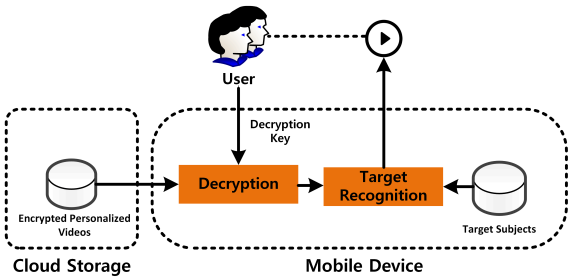

Figure 4: Retrieving Personalized Videos from the Cloud.

4.4 Accessing Personalized Surveillance Video

To retrieve encrypted personalized videos in the cloud storage, the user needs to perform decryption operation, as shown in Figure 4, and the details are as follows:

Step 1: The user submits credentials such as password or fingerprint on streaming app to access the videos list from the cloud storage.

Step 2: Once the authentication is successful, the user downloads the chosen video into the mobile device. Next, the user decrypts the video by using his or her private key.

Step 3: After the decryption, the streaming app performs target recognition to check if the unmasked face images in the video are matched with those in the target subjects database.

Step 4: Assuming that the target recognition is successful, the user can use the personalized video on the mobile device for activities analytics or replay purposes.

4.5 Average Face Construction for Preserving Privacy

Our approach to preserving privacy consists of several steps. Namely, we utilize face detection, average face construction, face

Figure 5: The Face Detection Results for Two Sample Video Frames Taken from the YouTube-8M Dataset [2].

Figure 6: Face Averaging and Face Recognition Results for Two Different Users.

recognition, and face masking for our privacy-preserving application. As noted before, the faces that the current user is not authorized to see, e.g., other students in the class, are masked with an average face.

The goal of a face detector is to identify whether an image contains a human face and, if so, find its location. The face detection method in the proposed framework is based on the Viola-Jones algorithm [21], which uses Haar-like features and utilizes an ensemble approach, where features are put through a sequence of weak classifiers, and their detection results are combined. In each row of Figure 5, the right image presents face detection results for the image shown on the left.

Once the faces are detected, we proceed with face recognition and average face construction systems from the detected faces. Face recognition aims at identifying or verifying a person from an input image. In general, face recognition systems work by comparing selected facial features from the input image with faces within a database. In the proposed paper, we use a recently introduced popular and powerful face recognition framework, FaceNet [19], which directly learns a mapping from face images to a compact Euclidean space where distances correspond to a measure of face similarity.

To generate the average face, we first align the faces so that their scale (the size of the face detection box) and the position of their eyes match with each other. We then obtain the average face image by averaging the pixel values of the aligned images. Note that in case the number of detected faces is low in an image, the average looks very similar to them, resulting in privacy leaks. To deal with this problem, we include the pre-computed average face image from the database to construct a representative average face for the group. After the average face is computed, we then use it to replace each face that a current user is not allowed to see. The images in Figure 6 present how the result of face averaging and face recognition look as it is seen from two user's smartphones for the input images given in Figure 5. Since the two users are allowed to see their kid along with the teacher, the other faces are replaced with the average face to preserve their identities.

5 ANALYSIS AND DISCUSSION

5.1 Security Analysis

In our framework design, the original stream videos and personalized videos are encrypted using public keys from the client and the user, respectively. Hence, only the owner of the videos has the decryption key to access the encrypted videos in the cloud storage. An attacker who obtains the key via malicious activities such as theft, extortion, or dumpster diving cannot view the content of the videos because our framework requires a matching operation to verify if the target subjects of the user are present in the video. Therefore, we can ensure that the storage of the surveillance videos in our framework is secure.

During the enrollment process, the client generates a distinct target subject database for each registered user. Then this database is used to personalize the surveillance video. To ensure the security of the framework, the users cannot update the database locally. Otherwise, it is possible for a malicious user to include other human subjects into the database and hence, bypass the masking operation. When the users want to add or remove a target subject in the database, an update request should be submitted to the client. Also, the client needs to update the database when the staff is leaving, or new staff is joining the nursery.

Although we assume that the hardware used by the users are secure, it is worth to mention that the surveillance system can be compromised if the attackers tamper the surveillance cameras. For instance, the attackers can mount a man-in-the-middle (MITM) attack to force the surveillance recorder to replay fake footage instead of showing the real live footage from the camera [7]. The attacker can attack the surveillance system via other internet-connected Internet of things (IoT) devices.

5.2 Visual Privacy and Video Utility Analysis

Unlike other solutions, we do not anonymize all human subjects in the surveillance video. Instead, we allow the target subjects to stay unmasked in the video while others are masked with an average face. To protect the visual privacy of the human subjects, we provide personalized video for each user, i.e., show different views for different users. Nevertheless, in the case where all target subjects

of the users are absent, they will have the same view. Consequently, the users can observe the activities of the target subjects in the scene but cannot identify the identities and collecting unwanted personal data of others.

In our framework, we construct a representative average face image for each group of human subjects by using the pre-computed average face and those that appear in the scene. When a human subject is leaving or newly appear in the scene, we will update the average face image dynamically. This kind of dynamic masking approach can prevent anyone from learning if an individual's face image is used to construct the average face when two or more groups present in the video, different average face images will be constructed to anonymize the human subjects.

How to make the right trade-off between privacy protection and video utility has become an essential question in vision-based applications. Our framework achieves both reliable privacy protection and high utility. For instance, the users can have confidence that their target subjects are anonymized in other views while the video analytics software can be installed on the mobile device to analyze the activities of their target subjects and notifies them of the finding (e.g., encouraging the children to do exercises if they slept too much). Also, a particular algorithm can be used by the service provider to perform a security-related function, such as determine if suspicious behavior is occurring when the intruders are in the scene.

5.3 Discussions

In face reconstruction, the output maintains facial features of the original 2D face image, which may breach the identity of an individual because the generated face can be easily recognized with human observation. Because of this, it is essential to suppress some face attributes while preserving other attributes of the image. Hence, in this work, we propose to utilize attributes from the faces to generate an average face image.

On the other hand, the effectiveness of the proposed privacy-preserving framework critically depends on the accuracy of the face detection approach. In case a partially occluded face is not detected by the system, then its identity can be determined by non-authorized users. According to our experimental results, we have obtained 85.63% face detection accuracy measured by intersection over union (IOU) for 1092 videos. In other words, the intersection of the bounding boxes computed by our face detection approach and the ground truth over their union is 85.63%. We plan to further improve this ratio by including face tracking and person detection.

As noted before, we bring the detected faces to the same size and align them by their eyes before face averaging. However, other facial features should also be considered for generating more natural face averages. Specifically, one needs to take into account the facial feature correspondences to create the average. Moreover, the face averaging process can also be improved by incorporating some other techniques. Precisely, the detected faces can be moved to 3D based on some previous works such as [1], the average face can be constructed in 3D and projected back to 2D. We expect that the average face will be more natural and will better protect the identity since the depth information will be taken into consideration during the average face construction process. Our preliminary result on 3D face averaging can be seen in Figure 7.

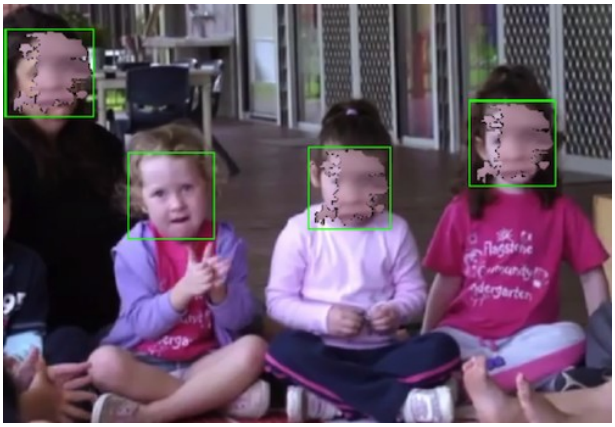

Figure 7: Preliminary Result of Face Averaging in 3D

6 CONCLUSION AND FUTURE DIRECTIONS

In this work, we have proposed a solution for visual privacy protection in the indoor surveillance systems. We utilized face images of human subjects in an insider group for average face construction. Our framework not only increases the confidence of users to use video surveillance systems but also raises public awareness about technologies that place privacy at risk. We protect the visual privacy in the personalized video and allow the user to utilize it for analytics purposes.

In our future work, we will consider combining both face reconstruction and face interpolation [13] to generate an artificial face. We plan to apply our framework to other scenarios such as live streaming on social media (e.g., Facebook) and visual privacy protection for online video platforms (e.g., YouTube and Instagram).

REFERENCES

[1] Aaron, S. J., Adrian, B., Vasileios A., and Georgios, T. 2017. Large pose 3D face reconstruction from a single image via direct volumetric CNN regression. In Proceedings of the IEEE International Conference on Computer Vision (Venice, Italy, October 22–29, 2017). ICCV'17. IEEE, 1031–1039. DOI=http://doi.acm.org/10.1109/ICCV.2017.117.

[2] Abu-El-Haija, S., Kothari, N., Lee, J., Natsev, P., Toderici, G., Varadarajan, B. and Vijayanarasimhan, S., 2016. Youtube-8m: A large-scale video classification benchmark. arXiv:1609.08675. Retrieved from https://arxiv.org/abs/1609.08675.

[3] Alice, J. O., Jonathon, P., Samuel, W., Dana, A. R., Julianne, A., Robert, B., and Joseph, D. 2011. Recognizing people from dynamic and static faces and bodies: Dissecting identity with a fusion approach. Vision research, 51, 1 (2011), 74–83. DOI=https://doi.org/10.1016/j.visres.2010.09.035.

[4] Bitouk, D., Kumar, N., Dhillon, S., Belhumeur, P. and Nayar, S.K., 2008. Face swapping: automatically replacing faces in photographs. ACM Trans. Graph. 27, 3 (August 2008), 1–8. DOI=https://doi.org/10.1145/1360612.1360638.

[5] Chhabra, S., Singh, R., Vatsa, M. and Gupta, G., 2018. Anonymizing k-facial attributes via adversarial perturbations. In Proceedings of the 27th International Joint Conference on Artificial Intelligence (Stockholm, Sweden, July 13-19). IJCAI'18. IJCAI, California, 656-662. DOI=https://doi.org/10.24963/ijcai.2018/91.

[6] Dai, J., Saghafi, B., Wu, J., Konrad, J. and Ishwar, P., 2015, September. Towards privacy-preserving recognition of human activities. In Proceedings of the IEEE International Conference on Image Processing (Quebec City, Canada, September 27-30). ICIP'15. IEEE Computer Society, 4238-4242. DOI=https://doi.org/10.1109/ICIP.2015.7351605.

[7] Danny, B. 2019. Researchers Hack Camera in Fake Video Attack. (August 2019). Retrieved May 3, 2020 from https://nakedsecurity.sophos.com/2019/08/01/researchers-hack-camera-in-fake-video-attack/.

[8] Du, L. and Li, Y., 2014, October. Privacy preserving for human object in video surveillance via visual cryptography. In Proceedings IEEE International Conference on Security, Pattern Analysis, and Cybernetics (Wuhan, China, October 18-19). SPAC'14. IEEE Computer Society, 80-85. DOI=https://doi.org/10.1109/SPAC.2014.6982661.

[9] Hukkelås, H., Mester, R. and Lindseth, F., 2019. Deepprivacy: A generative adversarial network for face anonymization. In Proceedings of the International Symposium on Visual Computing (Lake Tahoe, NV, USA, October 07-09). ISVC'19. Springer, Cham, 565-578. DOI=https://doi.org/10.1007/978-3-030-33720-9_44.

[10] Jana, S., Narayanan, A. and Shmatikov, V., 2013, May. A scanner darkly: Protecting user privacy from perceptual applications. In Proceedings of the 34th IEEE symposium on security and privacy (San Fancisco, CA, May 19-22). S&P'13. IEEE Computer Society, 349-363. DOI=https://doi.org/10.1109/SP.2013.31.

[11] Jourabloo, A., Yin, X. and Liu, X. 2015. Attribute preserved face de-identification. In Proceedings of the International conference on biometrics (Phuket, Thailand, May 19-22, 2015). ICB'15. IEEE Computer Society, 278-285. DOI=https://doi.org/10.1109/ICB.2015.7139096.

[12] Liu, F. and Koenig, H., 2010. A survey of video encryption algorithms. Comput. Security. 29, 1 (Feb. 2010), 3-15. DOI= https://doi.org/10.1016/j.cose.2009.06.004.

[13] Lu, Y., Tai, Y.W. and Tang, C.K., 2018. Attribute-guided face generation using conditional cyclegan. In Proceedings of the 15th European Conference on Computer Vision (Munich Germany, September 08-14). ECCV'18. Springer, Berlin, Heidelberg, 282-297. DOI=https://doi.org/10.1007/978-3-030-01258-8_18.

[14] Neil, C. 2018. Chinese school uses facial recognition to monitor student attention in class. (May 2018) Retrieved June 4, 2020 from https://www.telegraph.co.uk/news/2018/05/17/chinese-school-uses-facial-recognition-monitor-student-attention/.

[15] Newton, E.M., Sweeney, L. and Malin, B., 2005. Preserving privacy by de-identifying face images. IEEE Trans. Knowl. Data Eng. 17, 2 (Feb. 2005), 232-243. DOI=https://doi.org/10.1109/TKDE.2005.32.

[16] Parameshachari, B. D., and Soyjaudah, K. M. S. 2012. Analysis and comparison of fully layered image encryption techniques and partial image encryption techniques. In Proceedings of the International Conference on Information Processing (Bangalore, India, August 10-12). ICIP'12. Springer, Berlin, Heidelberg, 599–604. DOI= https://doi.org/10.1007/978-3-642-31686-9_70.

[17] Ryoo, M.S., Rothrock, B., Fleming, C. and Yang, H.J., 2017. Privacy-preserving human activity recognition from extreme low resolution. In Proceedings of the 31st AAAI Conference on Artificial Intelligence (San Francisco, California, February 04-09). AAAI'17. ACM, New York, NY, 4255–4262. DOI=https://doi.org/10.5555/3298023.3298185.

[18] Schiff, J., Meingast, M., Mulligan, D.K., Sastry, S. and Goldberg, K., 2009. Respectful cameras: Detecting visual markers in real-time to address privacy concerns. In Protecting Privacy in Video Surveillance. 65-89, Springer, London. DOI=https://doi.org/10.1007/978-1-84882-301-3_5.

[19] Schroff, F., Kalenichenko, D. and Philbin, J., 2015. FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Boston, MA, USA, June 07-12). CVPR'15. IEEE Computer Society, 815–823. DOI=https://doi.org/10.1109/CVPR.2015.7298682.

[20] Sen, N., Dantu, R., Vempati, J. and Thompson, M., 2018, June. Performance Analysis of Elliptic Curves for Real-Time Video Encryption. In Proccedings of the National Cyber Summit (Huntsville, AL, USA, June 05-07). NCS'18. IEEE Computer Society, 67-71. DOI=https://doi.org/10.1109/NCS.2018.00015.

[21] Viola, P. and Jones, M.J., 2004. Robust real-time face detection. Int. J. Comput. Vision, 57, 2 (May. 2014) 137-154. DOI=https://doi.org/10.1023/B:VISI.0000013087.49260.fb.

[22] Wong, K.S., Maratkhan, A., Tu, N.A. and Demirci, M.F., 2019. Towards Self-Enforcing Privacy Protection for Surveillance System. (June 2019). Retrieved May 20, 2020 from https://cvcops19.cispa.saarland/#.

[23] Zeng, H. and Ji, L., 2018. An encryption method for mobile video surveillance system based on ZUC algorithm. In Proceedings of the 8th International Congress of Information and Communication Technology (Nanning, China, January 11–13). ICICT'18. Elsevier, 282-288. DOI=https://doi.org/10.1016/j.procs.2018.04.215.