EAI Endorsed Transactions
on Security and Safety
Research Article EAI.EU

Minor Privacy Protection by Real-time Children Identification and Face Scrambling at the Edge

Alem Fitwi, Meng Yuan, Seyed Yahya Nikouei, Yu Chen*

Dept. of Electrical and Computer Engineering, Binghamton University, Binghamton, NY 13902, USA

Abstract

The collection of personal information about individuals, including the minor members of a family, by closed-circuit television (CCTV) cameras creates a lot of privacy concerns. Revealing children's identifications or activities may compromise their well-being. In this paper, we propose a novel Minor Privacy protection solution using Real-time video processing at the Edge (MiPRE). It is refined to be feasible and accurate to identify minors and apply appropriate privacy-preserving measures accordingly. State of the art deep learning architectures are modified and repurposed to maximize the accuracy of MiPRE. A pipeline extracts face from the input frames and identify minors. Then, a lightweight algorithm scrambles the faces of the minors to anonymize them. Over 20,000 labeled sample points collected from open sources are used for classification. The quantitative experimental results show the superiority of MiPRE with an accuracy of 92.1% with near-real-time performance.

Received on 01 May 2020; accepted on 12 May 2020; published on 14 May 2020

Keywords: Child Detection, Minor Privacy Protection, Smart Surveillance, Video Feature Extraction, Decentralization.

doi:10.4108/eai.13-7-2018.164560

1. Introduction

With the increasingly ubiquitous deployment of smart surveillance cameras throughout urban areas where majority of the population live, privacy issues are coming into focus [1–3]. Privacy often defines the boundaries to limit access to an individual's private information and body. Today, we live in an information society where vast quantities of data about us are gathered and analyzed through automated processes and cameras. A lot of private attributes and personal information about individuals are collected by closed-circuit television (CCTV) cameras and streamed to remote cloud servers and viewing stations with no privacy protection mechanism enforced [4].

Surveillance cameras are deployed in important locations to ensure public safety and physical security on top of providing concrete evidence for forensic analysis [5, 6]. The users may vary from the public safety authorities, law enforcement agents to a house owner. If video streams are transmitted by the edge cameras unprotected through the communication network, they may be subject to attacks. As a consequence, these large amount of data collected by the cameras could be intercepted and abused by adversaries. An example of a privacy breach is when a man in the middle manages to view the raw frames on transit [7, 8]. This has caused the public to be more concerned and to demand for change in the way video surveillance works [9, 10].

Specifically, the practice of mass-surveillance can have a profound effect on the understanding of minors about privacy in their later lives [11]. Usually children learn through experience; hence, they should grow up in a privacy-aware environment in order to learn what privacy is and how it works. Besides, many argue that the right experience of privacy is very important to a child's future success and good decision-making in setting correct safety measures and privacy boundaries. Hence, today's pervasive surveillance systems must have the means to protect minor's privacy.

Privacy protection is one of the active research areas in the rise of Internet of Things (IoT) [12], where a huge number of sensors and low powered processors are being connected to the network with none or minimal security measures. One of the more important aspects

*Corresponding author. Email: ychen@binghamton.edu

of this research is to protect the identity of the people in case the data is compromised. Hence, any effort to address the privacy problems in a surveillance system must have techniques for both identifying private attributes on images and for protecting them [13, 14].

Private attributes like face are detected through the use of machine learning or deep learning networks [13]. Following the detection process, these private attributes of individuals are scrambled using apropos cryptographic schemes. These schemes ensure that video streams are not accessed by means of interception attacks and abused by unauthorized people while being transmitted from the cameras to the fog/cloud servers and viewing centers. From the various privacy-preserving requirements, minor children's identity and face protection is more essential to every family to protect the minors from attackers or abusers [15–17].

In this paper, we propose a novel Minor Privacy protection solution using Real-time video processing at the Edge (MiPRE). The MiPRE basically comprises a face-object detector built based on Deep Neural Network (DNN) and a lightweight scrambling algorithm. The face detection method employs the Multitask Convolutional Neutral Network (MTCNN) [18] to detect and align the faces. The face recognition process is realized using the FaceNet model [19], designed by Google. It is employed to extract the important features of minors' faces. The lighter scrambling scheme is designed based on a chaotic map, which is highly sensitive to initial conditions. Then, the face-object detector model is loaded to edge cameras to check every frame so as to detect faces and classify them as adults or minors. Following every successful detection and classification of faces, the scrambling algorithm is called into action to securely denature the children faces before the raw video is transmitted over the communication network to the consumers or storage sites.

The rest of this paper is organized as follows. In section 2, we discuss several methods for children detection as well as the historical efforts in the detection and recognition of human faces. Section 3 presents the system architecture of our MiPRE scheme and its function blocks in detail, including the multi-step pipeline face detection, children recognition, and face scrambling for privacy-protection using a lightweight scheme. Section 4 reports the model training process and the performance of the MiPRE scheme. Finally, Section 5 concludes this paper.

2. Related Work

2.1. Age Recognition based on Face Recognition

With the development of machine learning, computers are becoming more widely used in many vision processing tasks, which reduce manual workload and guarantees high recognition rate [20, 21]. In addition, the community is witnessing the migration of powerful machine learning algorithms to the IoT environments by developing lightweight solutions [6, 22, 23] in recent years.

In the field of face recognition, researches mainly focus on two aspects, authentication [24, 25] and recognition [26, 27]. Then, irrespective of whether it is recognition or authentication, a well-known top-down approach that comprises three major stages is pursued. Firstly, the face object is detected and boxed. Secondly, the important features are extracted from the face. At last, the comparison of features is performed [28] for the purpose of recognition and authentication. Human age recognition is a well-studied sub-area [29] in the process of face recognition. Classifiers are trained to detect the age of the subject or to predict the facial appearance in certain age-group.

Face recognition-based age recognition is the process of extracting age-related facial features to create an age classification model [30, 31]. Then, use this model to evaluate the age range of a given person to categorise this person into different age groups. However, the ability to build an age-recognition model through face recognition is limited due to the fact that human aging and changes are not exclusively attributed to time. Aging is a complex process that could also be affected by health and people's living conditions.

Although the research on face recognition started earlier, there are only few studies on the establishment of age classification models. While some previous works have proposed methods of age classification, the accuracy of their models is far from satisfactory. Today's top-performing techniques of face recognition are based on Multi-task convolutional neural networks. Both Facebook's DeepFace [32] and Google's FaceNet [19] architectures have the highest accuracy. DeepFace uses 6 convolutional (conv.) layers followed by two fully connected layers (FC) that are used to detect and map a face in 3-D space and to map 67 fiducial points on the face. The Facenet is based on an approach of detecting faces that belong to the same person using illumination and Pose invariance architecture. Hence, the MTCNN and FaceNet architectures are employed in our model so as to achieve performance results comparable to the performance of the state-of-the-art techniques.

2.2. Privacy Features Protection Mechanism

These days many people are very concerned about the invasion of their privacy by the widespread use of CCTV cameras [33, 34]. Huge amount of sensitive information is collected by the CCTV cameras and transmitted over an insecure communication channel to cloud servers for video analytics. Consequently, they could be intercepted by adversaries that undermine the privacy of individuals. Besides, they could be misused

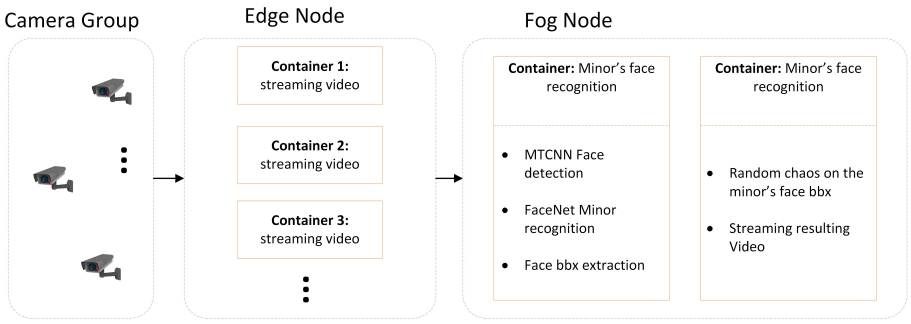

Figure 1. MiPRE system architecture.

by the people in control of the surveillance system. Hence, it is necessary to have a secure video-privacy-preserving measure implemented at the edge. The image/video frame privacy protection usually includes two tasks. Firstly, the detection of private attributes, and secondly enforcing measures to protect these sensitive attributes through de-identification. Machine learning or deep learning techniques are employed for the detection of the sensitive contents of frames or images. Then, the process of de-identification is done using scrambling schemes.

However, the preexisting stream ciphers like Rivest Cipher 4 (RC4) and block ciphers like advanced encryption standard (AES) are not convenient for the encryption and decryption of information-rich images due to their high computational resource requirements and their inability to handle the strong correlations amongst adjacent pixels of images. On the contrary, chaotic scrambling schemes have high degree of sensitivity and randomness. Then, these characteristics make them the more feasible options for video enciphering. However, existing chaotic schemes are slow and they need to be redesigned so as to fit into the resource constrained environment of edge computing [35]. Hence, we focused on designing lightweight scrambling schemes based on computationally simple but robust chaotic schemes. In this paper, we investigated the characteristics of the Peter de Jong map [36] and proposed a secure lightweight minor's face scrambling technique based on it.

3. MiPRE: Minor Privacy Protection at the Edge

3.1. System Overview

Figure 1 presents the architecture of our MiPRE system. It consists of three major function blocks: (1) face detection using a multi-step pipeline model, (2) face recognition based on the extracted features to identify children faces, and (3) face scrambling to protect children's privacy. Each module is implemented in a docker container which promises scalability and faster updates in parts of the system using micro-services architecture [37, 38]. The design rationales and technical details are presented in the following subsections.

3.2. Face Detection

While there are many face detection methods like Dlib, OpenCV, and OpenFace, we adopted MTCNN approach for two main reasons. Firstly, it achieves a high detection accuracy of frontal and lateral faces. Secondly, the FaceNet model was built with an MTCNN interface for face objects detection, which allows us to focus on our target. Basically, the MTCNN is a deep learning model for face detection based on a multi-task cascaded Convolutional Neural Network (CNN). It exploits the inherent correlation between detection and alignment to boost up its performance. In particular, to predict face and landmark locations in a coarse-to-fine manner, the framework used in this paper leverages a cascaded architecture with three stages of carefully designed deep conv. networks [18, 32].

Given an input image, an image pyramid is built by rescaling the image into different scales through a bi-linear interpolation. This step insures scale invariation. Figure 2 shows an example of three cascaded-stages of the MTCNN following the scaling step.

- *P-Net*: This is a full convolutional neural network (FCN). The feature map obtained at each position by a forward propagation is a 32-D feature vector used to determine whether a 12×12 grid area contains a face or not. If it contains human face, a bounding box is regressed and the corresponding area in the original image is grabbed. Next, the bounding box with the highest score is retained by a non-maximum suppression (NMS) step and all other bounding boxes with excessively large overlapping area are discarded.

- *R-Net*: It is a simple CNN stage. Similar to the last stage of the O-Net, the 24×24 box is up-scaled to a 48×48 box in order to have the highest confidence of bounding box detection and facial

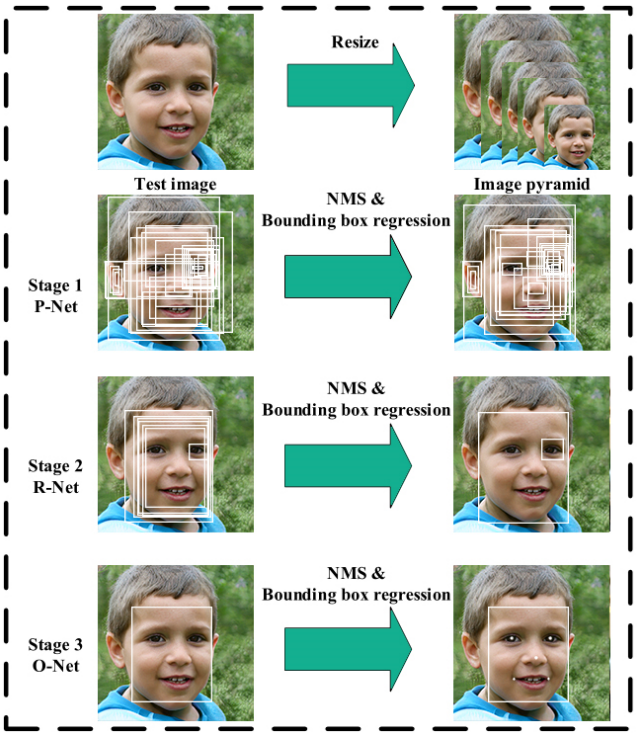

Figure 2. MTCNN: Built to have cascaded architecture to ensure best performance in human face detection and bounding box regression.

landmark extraction when input to the R-Net stage.

- *O-Net*: This is employed to achieve higher accuracy. In this stage, the 12×12 input box produced by P-Net is first up-scaled to 24×24 box using a bilinearly interpolated method. Then, it is input to the O-Net to determine whether a human face exists. If a human face is detected, the regression of bounding box is performed followed by the NMS step.

Figure 3 presents the architecture of the layers employed in each stage of the cascaded MTCNN model. Each step uses different sizes of Conv. filters and different number of layers to produce the same class of results. The outputs have three categories. The face classification score is presented as the first set of outputs using two neurons: one for the presence of a face and the other as its score. Another part of the output is the bounding box regression where the upper left and lower right of the bounding box are represented by four neurons (dx_1, dy_1, dx_2, *and* dy_2). Facial landmark localization regresses the position of five points on left eye, right eye, nose, left mouth corner, and right mouth corner. So, it is a 10-D variable that needs ten neurons for representation.

During the training phase, all the three networks use the landmark positions as supervised signals to

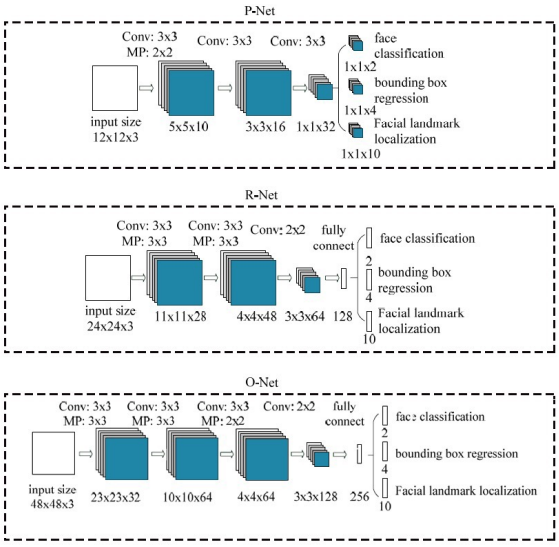

Figure 3. MTCNN: Stage architecture of the model used for face detection and landmark extraction.

guide the learning of the network. In the prediction phase, however, the P-Net and R-Net conduct only face detection; they do not output landmark positions because they are inaccurate in these phases. The landmark position is only outputted in the O-Net. Bounding box and landmarks coordination outputs are normalized relative to the input image.

As mentioned above, there are three tasks that MTCNN archives. They are face classification, bounding box regression and facial landmark localization. Thus, the loss function of the algorithm also has three parts, briefly described in what ensues. Readers interested in getting more details are referred to [32]. The cross-entropy loss function shown in Eq. (1) is employed for face classification.

$$L_i^{det} = -(y_i^{det} log(p_i) + (1 - y_i^{det})(1 - log(p_i))) \quad (1)$$

where y_i^{det} is the ground truth for the i_{th} object and p_i is the network output for the face detection.

Next is the bounding box regression loss where the euclidean distance loss function is employed as stated in Eq. (2).

$$L_i^{bbx} = \|(\hat{y}_i^{bbx} - y_i^{bbx}\|_2^2 \quad (2)$$

Lastly, the same regression loss function is employed for each of the L landmark of each sample i as depicted in Eq. (3).

$$L_i^{landmark} = \|(\hat{y}_i^{landmark} - y_i^{landmark}\|_2^2 \quad (3)$$

3.3. Children Faces Recognition

There are several ways to compare the similarity of two images. The euclidean distance metric is one of the

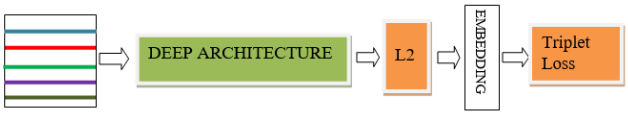

Figure 4. Model structure: This network consists of a batch input and output layer and a deep CNN followed by L2 normalization, which results in the face embedding. This is followed by the triplet loss during training.

most used metric because of the ease in implementation and in-expensive computation. Given a feature map where the features are extracted from the face, this metric is going to show the similarity in the features between the feature set and a known set. This idea is the prime focus of this section. The faces extracted by the face detection step are going to be fed to the FaceNet. Then, the resulting feature map is compared with the datasets that comprise known positive and negative images of children's faces. At last, a similarity threshold is computed and picked to give a final label to the face.

FaceNet is a universal system that can be used for face authentication, recognition and clustering. It learns to map images to an Euclidean space through CNNs. A spatial distance is directly related to the similarity of pictures. Different images of the same person have a small spatial distance; but images of different people have a larger spatial distance. Once the mapping is done, the face recognition task becomes simple [39].

The preexisting DNN-based face recognition models use an FC classification layer. The middle layer in the FC layers, after the Conv. layers, or the last Conv. layer is a vector map of the face image. The FC classifier layer is then placed on top of this vector map. The disadvantages of such methods are indirectness and inefficiency. In contrast, FaceNet directly uses the loss function of triplets-based Large Margin Nearest Neighbor (LMNN) to train the neural network, and the network directly outputs a 128-D vector space. The triplets we selected contain two matching face thumbnails and one non-matching face thumbnail. The goal of the loss function is to distinguish positive and negative classes by distance boundaries. The model structure is shown in Fig. 4.

The purpose of the model is to embed the 2-D face image X into the Euclidean space with D dimensions where $f(X) \in R^d$. In this vector space, the anchor image of a face $x_i^a(anchor)$ is close to other images with the same facial expressions $(x_i^p(positive))$ and far from faces with different characteristics $(x_i^n(negative))$. As illustrated by Fig. 5, the training process migrates the network's behavior from the left side to the right side.

To reach this goal, a triplets loss function is calculated from the triplet of three pictures. The triplet is

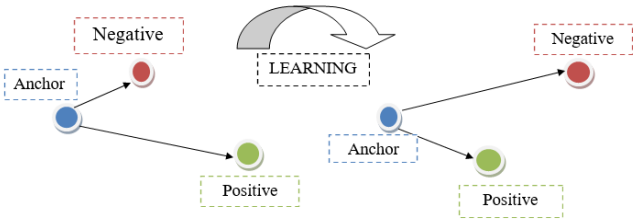

Figure 5. Training goal of the FaceNet network.

composed of Anchor (A), Negative (N), and Positive (P) images. Any image can be used as a base point (A). Then, images which have the same facial characteristics as the base are considered as its (P) and those images that do not share the same characters are considered as its (N). Triplets Loss minimizes the distance between an anchor and a positive, both of which have the same identity, but it maximizes the distance between the anchor and a negative image. Mathematically, the loss function can be formulated as stated in Eq. 4:

$$L = \sum_{i}^{N} [\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \epsilon] \quad (4)$$

where ϵ is the safe boundary between the positive and negative images.

Theoretically speaking, the best images for training purposes are those with the highest distance between the (A) and (N) but lowest distance between the (A) and (P). However, in practice this approach creates a local minimum and a global solution is not going to be reached. A remedy to this problem is to select all positive image-pairs in a mini-batch, which can make the training process more stable. The selection of the (N), on the other hand, is made according to the condition set in Eq. (5) to train the network.

$$\|f(x_i^a) - f(x_i^p)\|_2^2 < \|f(x_i^a) - f(x_i^n)\|_2^2 \quad (5)$$

Training: The network was trained using 10,000 images of children and 10,000 images of adult faces collected from an open source [40]. The children ages are variable between 6 to 14 years old (dictated by the datasets). For the positive selection, we selected 100 face images for each of the children and adult categories. The images are fed through the MTCNN to have a Bounding Box around the face. The face rectangle is then fed to the FaceNet for euclidean distance calculation. Figure 6 presents the data flow of our model for minor's face detection.

In Fig. 6, the positive and negative dataset against which a test image is compared is prepared beforehand. This feature set is called embedding dataset that contains the feature maps of the aforementioned 200 images for comparison. Total inference time, thus, is divided into two parts: face detection time and feature

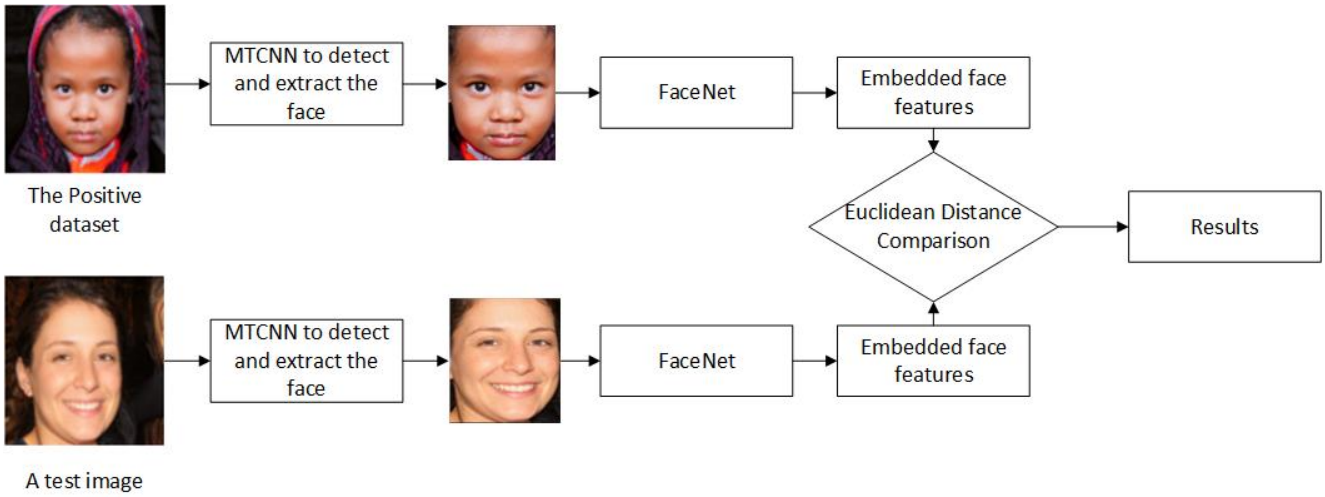

Figure 6. Dataflow in our model shows how the MTCNN is used to detect and crop the faces from each input image and then use the FaceNet to calculate the euclidean distance between the anchor face and positive and negative images to detect children's faces.

comparison using the FaceNet network. More details are presented in Section 4.

3.4. Lightweight Children-Face Protection Scheme

Due to their high sensitivity to slight changes, chaotic methods are more suitable for image privacy protection [36, 41, 42]. But existing chaotic methods do not fit into a resource constrained devices like edge cameras. In this paper, we developed a lightweight minor's privacy protection scheme based on the Peter De Jong Map [36]. The Peter de Jong map is a type of 2D recursive system stated in Eq. (6). The choice of parameter values and initial conditions will generate a different attractor. It has four parameters and two initial conditions. However, the Peter de Jong map cannot be used for chaotic encryption as it is. It does not meet the security requirements. For instance, Fig. 7 shows a non-uniform distribution of the pixels of chaos generated by Eq. (6), signifying that it is insecure to be used as it is for scrambling purpose.

$$\begin{aligned} x_{n+1} &= sin(a * y_n) - cos(b * x_n) \\ y_{n+1} &= sin(c * x_n) - cos(d * y_n) \end{aligned} \tag{6}$$

After an extensive experimental study, in this paper we proposed a one-way scrambling of children faces in video frames using an improved version of De Jong map and vectorized pixel-array multiplication. Our proposed scheme is stated in Eq. (7). We added four more parameters and identified the secure range for every parameter value. It generates a random and uniform output that is secure enough to be used for cryptographic purpose. In other words, it generates a random chaotic sequence that passes all standard security tests like entropy, sensitivity, correlation, and statistical analyses. Besides, it has a key space far

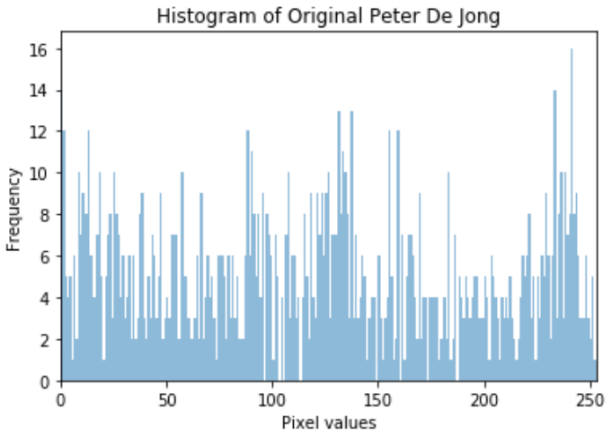

Figure 7. Distribution of the De Jong Chaos: it is not uniform revealing its weakness against histogram analysis attacks.

greater than the lower secure key space boundary (128 $bits$). The key contains eight double-precision floating value elements giving a key length of 64 $bits \times 8 = 512$ $bits$ and a key space of 2^{512}.

$$\begin{aligned} key &= [k_0, k_1, k_2, k_3, k_4, k_5, k_6, k_7] \\ x_0, y_0 &= k_0, k_1 \\ x_1 &= sin(k_2 * y_0) + k_3 * cos(k_4 * x_0) \\ y_1 &= sin(k_5 * x_0) + k_6 * cos(k_7 * y_0) \\ x_0, y_0 &= x_1, y_1 \end{aligned} \tag{7}$$

Algorithm 1 shows the scrambling algorithm. A key comprising eight elements is first generated followed by the generation of a random chaos used to scramble the minors' faces on video frames detected by the object-detection method. Equation (8) illustrates the secure range of the eight elements of the key employed for the generation of a chaos equal to the size of the region of

interest (ROI). Every element of the key is a double-precision floating-point value randomly selected from its respective range defined in Eq. (8). The ranges were defined after an extensive security analysis and they produce secure chaotic outcomes when inserted into Eq. (7). The random chaos generated is of the same size as the ROI, which is a minor's face in this case, and a vectorized point-wise multiplication operation is performed, which results in secure cipher.

Algorithm 1 Privacy-protection Scheme

```
ROI ← Object-dtector(cam.video())
procedure GENERATEKEY
    K ← secRand([k0, k1, ..., k9])
    return ← K
key ← generateKey()
procedure GENERATEKEY(key, W, H)
    chaos ← Eq.7
    return ← chaos
chaos ← generateKey(key)
procedure DENATUREFACES(chaos, ROIxy)
    fdenatured ← np.multiply(ROI, chaos)
    return ← f denatured
```

$$\begin{aligned} key &= [k_0, k_1, k_2, k_3, k_4, k_5, k_6, k_7] \\ k_0, k_1 &= random(1,4),\ random(1,4) \\ k_2, k_5 &= random(-1,1),\ random(-1,1) \\ k_3, k_6 &= random(-1,-0.99),\ random(-1,-0.99) \\ k_4, k_7 &= random(-3,-3),\ random(-3,-3) \end{aligned} \tag{8}$$

Figure 8 demonstrates a sample histogram analysis. Figure 8(a) is a clear input image and its ciphered version is shown by Fig. 8(b). Figure 8(c) shows that the frequency distribution of the input image in Fig. 8(a) is not uniform. In contrast, the frequency distribution of scrambled image in Fig. 8(b) has become uniform as portrayed in Fig. 8(d). This validates that the privacy-protection scheme is secure against any frequency analysis attack. Besides, it is computationally efficient and secure in that it passes all other security tests described in sub-parts of subsection 4.3.

4. Experimental Results

4.1. Experimental Setup

The multi-level MiPRE architecture is tested in an environment comprising a fog server and an edge device. A relatively powerful machine, that has an AMD Ryzen 7 2700X processor with 8 cores and 3.7 GHz clock, is considered as the edge server; whereas a Raspberry PI is employed as an edge module. The server has dual 8GB memory modules and a windows 10 Pro edition operating system is installed on it. During

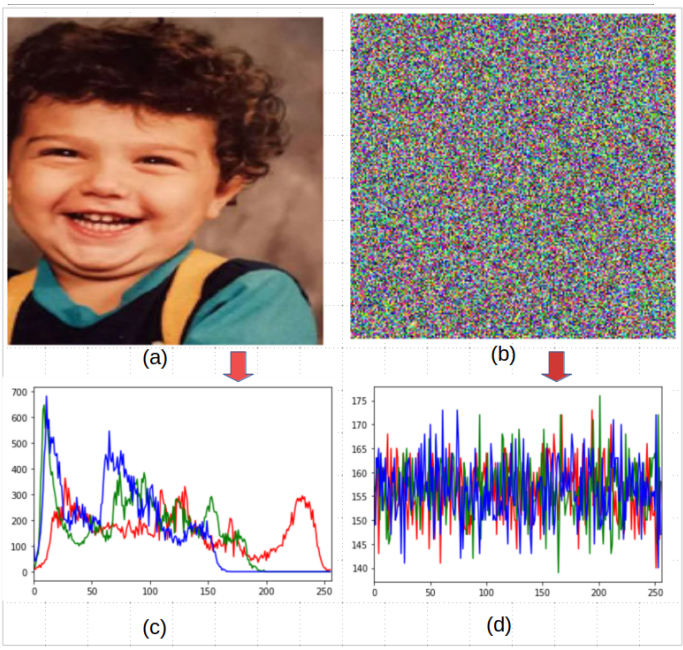

Figure 8. Histogram: (a) plain image, (b) cipher of plain image (a), (c) histogram of plain image (a), and (d) randomized histogram of cipher image (b).

inference, we observed an average CPU utilization of 18%, which is acceptable considering the need to connect several edge nodes to an edge server. On the other hand, the memory utilisation of the process is higher at about 10GB on average. Given the loading of several stages of CNN models, the higher memory need is not surprising. But it should be considered when deploying this model.

4.2. Accuracy of Face Recognition

We compared the accuracy of our MiPRE model with the state of the art models for age recognition based on facial components, as depicted in table 1. The approach reported in [43] tries to divide faces into multiple components and uses their changes as the features to predict the age of the subject face. Levi et al. [44] uses a CNN to classify primary objects in an image between gender and age. Meanwhile, a rule based method has been also proposed that divides the image into sections and implement privacy measures based on rule sets [45]. The last method we compared our scheme with is [46], which tries to extract facial features and accurately detect the age of each face. As shown by Table 1, our MiPRE scheme achieves a better performance in terms of accuracy than the other efforts.

Table 2 shows the ratio at which the multi staged model we proposed in MiPRE works. The model achieves a miss detection rate (MDR) of 7.9% in the testing set, which means that 158 out of 2000 images

Model	Accuracy
Otto et al. [43]	81.27 %
Levi et al. [44]	84.7 %
Teixeira et al. [45]	91.14 %
Du et al. [46]	79.24 %
MiPRE	92.1 %

Table 1. Accuracy of face classification based on the age. Our multi staged model has achieved a higher average accuracy.

Miss Detection	MDR	Detection	DR
158	0.079	1842	0.921

Table 2. Miss classification rate based on the 2000 images that are used for testing.

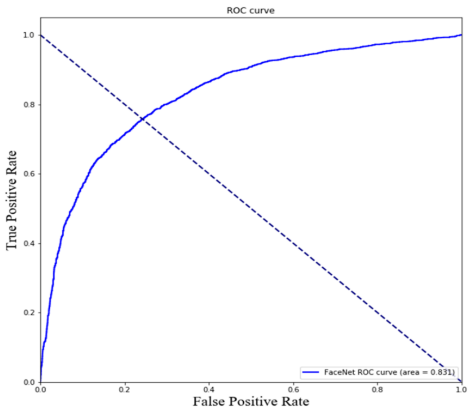

Figure 9. ROC curve.

are classified to the wrong category. The detection rate (DR) is 92.1%.

Figure 9 is the ROC curve that shows detection rate, shown as True Positive Rate versus the False Positive rate. This curve gives some intuitive insight to the best possible threshold to be set for the child detection. A bigger area under this curve implies that the system performs better with higher true positive and lower false positive rate. During implementation, for example, if a true positive rate of 0.7 is needed, then a 0.2 false positive rate is expected.

4.3. Lightweight Face Scrambling Scheme

Functional Test. The privacy-conserving technique was integrated with the face-detection scheme. Figure 10(a) shows two minors' faces accurately detected and bounded. Figure 10(b) depicts the enciphered faces of the two minors following the detection process. The experimental result also verifies that our proposed minor's privacy protection scheme is lightweight. It accomplishes the operation of scrambling a minor's face with a size of $150 \times 150 \times 4$ pixels in less than 45 ms.

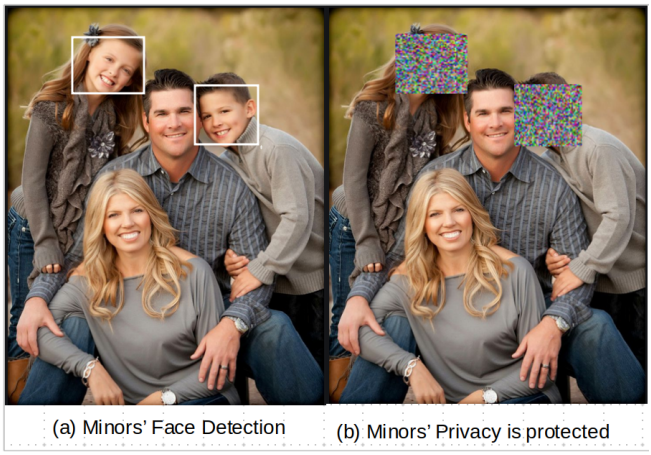

Figure 10. Detected and Scrambled Minors' faces

Finally, Fig. 11 and Fig. 12 show some of the positive samples along with the processing time of each individual image underneath. In general, face detection takes about 240 ms, but the recognition process is much faster at about 37 ms. Moreover, the label that the machine assigns and the confidence in the score are reported for each sample. The higher the confidence, the better the label is. But a confidence of 50% is a random guess between the outputs.

Comparative Security and Performance Analysis. On top of the histogram and functional analysis, we have also compared our scrambling scheme with preexisting popular stream and block ciphers like RC4 and AES, respectively. AES is the most secure and most widely used block cipher whereas RC4 is a simple stream cipher. One of the performance metrics employed is the encryption time. The other security parameters considered are statistics, key space (K_s), key sensitivity(K_σ) measured in terms of Number of Pixels Change Rate (NPCR) and Unified Average Changing Intensity (UACI), information entropy ($Info_e$), correlations in horizontal ($corr_h$), vertical ($corr_v$), and diagonal ($corr_d$) directions. The K_s tells us whether a scheme is resistant against brute force attacks. The K_σ measures how much the cipher varies when the original key is slightly changed, like by one bit. The $Info_e$ measures how random the pixels are distributed. For an image (I) comprising 8-bit pixel values, the ideal entropy value is $H(I) = 8$. Hence, a scheme that produces an entropy value closer to 8 is said to resistant to entropy analysis attack. Lastly, the correlation parameter calculates the correlation among the adjacent pixels of the cipher image in horizontal, vertical, and diagonal directions, which is expected to be very close to zero.

In terms of encryption time, Table 3 shows that our scheme is faster than both RC4 and AES. Besides, the ideal mean and standard deviation (STD) of a uniformly distributed pixels of an image are 127.5 and 73.901. Table 3 also shows that our scheme has mean and STD

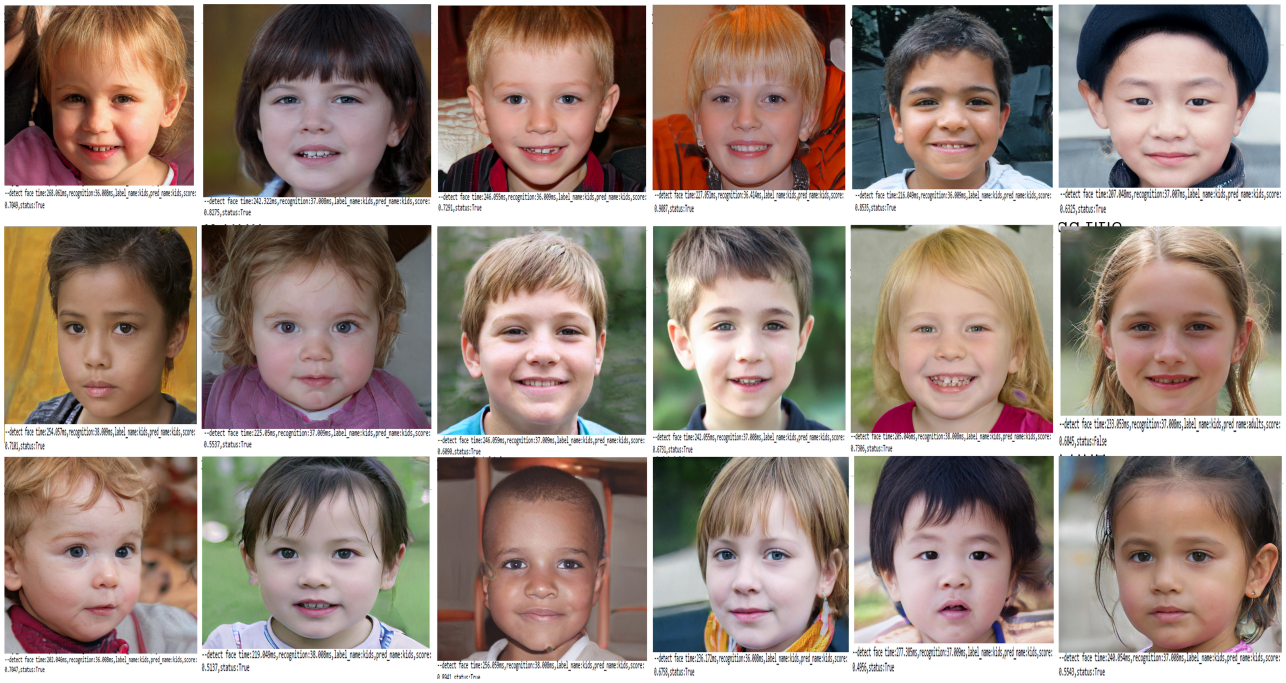

Figure 11. Sample examples of the proposed children's face detection model.

Figure 12. Sample examples of the proposed adult's face detection.

values closer to the ideal value signifying uniformity in the distribution of the cipher pixels. Generally, the traditional encryption algorithms like RC4 stream cipher and block cipher AES are not suitable for image encryption due to image's intrinsic properties such as bulky data capacity, strong redundancy and strong correlations among adjacent pixels. On the other hand, chaotic schemes have many properties, including uniform randomness, unpredictability, aperiodicity, sensitive dependence on initial conditions. These properties have helped chaotic systems become popular in image encryption.

Table 3. Comparative Security and Performance Analysis

Parameter	Our Scheme	AES	RC4
Time(ms)	44.73	94.228	63.169
K_s	2^{512}	2^{256}	2^{2048}
K_σ			
UACI	48.031%	25.079%	33.227%
NPCR	99.616%	74.730%	99.590%
$Info_e$	7.909 bits	6.784 bits	7.991 bits
$Corr_h$	8.899e-5	8.93e-4	0.011
$Corr_v$	5.52e-4	0.0012	0.021
$Corr_d$	0.00812	0.0132	0.0724
$Mean$	126.501	159.666	159.188
STD	77.554	84.499	84.521

5. Concluding Remarks

As the number of CCTV cameras increases, families have grown more concerned about the privacy of their members and data. Minimizing appearance of the minors in unauthorized videos is one of the responsibilities of parents. Hence, scrambling face, the most powerful human identifying attribute, can effectively anonymize individuals. In this work, a novel lightweight minor privacy protection scheme named MiPRE is proposed. The MiPRE scheme is designed by leveraging multi-stage DNN based face recognition approaches and a lighweight chaotic scrambling algorithm. It detects and scrambles children's faces in video frames to ensure de-identification of the minors at the edge of the network, just before the video is streamed over the Internet to distant monitors. The MiPRE scheme is tested on a platform consisting of a smart camera and an edge server, and the experimental results verified that the MiPRE scheme meets the design goal. It achieved a high accuracy in children face recognition, 93%, and is able to finish the secure face-scrambling operation in 45 ms.

Our on-going efforts mainly focus on identifying additional attributes that have significant impacts on children privacy, which allows us to extend the coverage of the MiPRE scheme. To consider more privacy-attributes other than faces, we will continue investigating lightweight machine learning algorithms to fit the next version of the MiPRE scheme in the edge environments.

References

[1] Cavallaro, A. (2007) Privacy in video surveillance [in the spotlight]. *IEEE Signal Processing Magazine* **2**(24): 168–166.

[2] Dufaux, F. (2011) Video scrambling for privacy protection in video surveillance: recent results and validation framework. In *Mobile Multimedia/Image Processing, Security, and Applications 2011* (International Society for Optics and Photonics), **8063**: 806302.

[3] Nikouei, S.Y., Chen, Y., Aved, A. and Blasch, E. (2020) I-vise: Interactive video surveillance as an edge service using unsupervised feature queries. *arXiv preprint arXiv:2003.04169* .

[4] Taylor, E. (2010) I spy with my little eye: the use of cctv in schools and the impact on privacy. *The Sociological Review* **58**(3): 381–405.

[5] Chen, N., Chen, Y., Blasch, E., Ling, H., You, Y. and Ye, X. (2017) Enabling smart urban surveillance at the edge. In *2017 IEEE International Conference on Smart Cloud (SmartCloud)* (IEEE): 109–119.

[6] Xu, R., Nikouei, S.Y., Chen, Y., Song, S., Polunchenko, A., Deng, C. and Faughnan, T. (2018) Real-time human object tracking for smart surveillance at the edge. In *the IEEE International Conference on Communications (ICC), Selected Areas in Communications Symposium Smart Cities Track* (IEEE).

[7] Kumar, V. and Svensson, J. (2015) *Promoting social change and democracy through information technology* (IGI Global).

[8] Newton, E.M., Sweeney, L. and Malin, B. (2005) Preserving privacy by de-identifying face images. *IEEE transactions on Knowledge and Data Engineering* **17**(2): 232–243.

[9] Fitwi, A., Chen, Y. and Zhou, N. (2019) An agent-administrator-based security mechanism for distributed sensors and drones for smart grid monitoring. In *Signal Processing, Sensor/Information Fusion, and Target Recognition XXVIII* (International Society for Optics and Photonics), **11018**: 110180L.

[10] Lyon, D. (2010) Surveillance, power and everyday life. In *Emerging digital spaces in contemporary society* (Springer), 107–120.

[11] Waters, S. (2018) The effects of mass surveillance on journalists' relations with confidential sources: A constant comparative study. *Digital Journalism* **6**(10): 1294–1313.

[12] Yang, Y., Wu, L., Yin, G., Li, L. and Zhao, H. (2017) A survey on security and privacy issues in internet-of-things. *IEEE Internet of Things Journal* **4**(5): 1250–1258.

[13] Fitwi, A., Chen, Y. and Zhu, S. (2019) No peeking through my windows: Conserving privacy in personal drones. *arXiv preprint arXiv:1908.09935* .

[14] Nikouei, S.Y., Chen, Y., Aved, A. and Blasch, E. (2020) I-vise: Interactive video surveillance as an edge service using unsupervised feature queries. *arXiv preprint arXiv:2003.04169* .

[15] Berson, I.R. and Berson, M.J. (2006) Children and their digital dossiers: Lessons in privacy rights in the digital age. *International Journal of Social Education* **21**(1): 135–147.

[16] Lwin, M.O., Stanaland, A.J. and Miyazaki, A.D. (2008) Protecting children's privacy online: How parental mediation strategies affect website safeguard effectiveness. *Journal of Retailing* **84**(2): 205–217.
[17] Shmueli, B. and Blecher-Prigat, A. (2010) Privacy for children. *Colum. Hum. Rts. L. Rev.* **42**: 759.
[18] Zhang, K., Zhang, Z., Li, Z. and Qiao, Y. (2016) Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters* **23**(10): 1499–1503.
[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*: 815–823.
[20] Amos, B., Ludwiczuk, B., Satyanarayanan, M. *et al.* (2016) Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science* **6**: 2.
[21] Cox, D. and Pinto, N. (2011) Beyond simple features: A large-scale feature search approach to unconstrained face recognition. In *Face and Gesture 2011* (IEEE): 8–15.
[22] Nikouei, S.Y., Chen, Y., Song, S., Xu, R., Choi, B.Y. and Faughnan, T. (2018) Real-time human detection as an edge service enabled by a lightweight cnn. In *the IEEE International Conference on Edge Computing* (IEEE).
[23] Nikouei, S.Y., Chen, Y., Song, S., Xu, R., Choi, B.Y. and Faughnan, T. (2018) Smart surveillance as an edge network service: From harr-cascade, svm to a lightweight cnn. In *2018 ieee 4th international conference on collaboration and internet computing (cic)* (IEEE): 256–265.
[24] Fathy, M.E., Patel, V.M. and Chellappa, R. (2015) Face-based active authentication on mobile devices. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE): 1687–1691.
[25] Vazquez-Fernandez, E. and Gonzalez-Jimenez, D. (2016) Face recognition for authentication on mobile devices. *Image and Vision Computing* **55**: 31–33.
[26] Sharif, M., Bhagavatula, S., Bauer, L. and Reiter, M.K. (2016) Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. In *Proceedings of the 2016 acm sigsac conference on computer and communications security*: 1528–1540.
[27] Yaman, M.A., Subasi, A. and Rattay, F. (2018) Comparison of random subspace and voting ensemble machine learning methods for face recognition. *Symmetry* **10**(11): 651.
[28] Uiboupin, T., Rasti, P., Anbarjafari, G. and Demirel, H. (2016) Facial image super resolution using sparse representation for improving face recognition in surveillance monitoring. In *2016 24th Signal Processing and Communication Application Conference (SIU)* (IEEE): 437–440.
[29] Li, Y., Wang, G., Nie, L., Wang, Q. and Tan, W. (2018) Distance metric optimization driven convolutional neural network for age invariant face recognition. *Pattern Recognition* **75**: 51–62.
[30] Bhattacharya, S. and Gupta, M. (2019) A survey on: Facial emotion recognition invariant to pose, illumination and age. In *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)* (IEEE): 1–6.
[31] Zhou, H. and Lam, K.M. (2018) Age-invariant face recognition based on identity inference from appearance age. *Pattern recognition* **76**: 191–202.
[32] Taigman, Y., Yang, M., Ranzato, M. and Wolf, L. (2014) Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*: 1701–1708.
[33] Streiffer, C., Srivastava, A., Orlikowski, V., Velasco, Y., Martin, V., Raval, N., Machanavajjhala, A. *et al.* (2017) eprivateeye: To the edge and beyond! In *Proceedings of the Second ACM/IEEE Symposium on Edge Computing* (ACM): 18.
[34] Yu, J., Zhang, B., Kuang, Z., Lin, D. and Fan, J. (2017) iprivacy: image privacy protection by identifying sensitive objects via deep multi-task learning. *IEEE Transactions on Information Forensics and Security* **12**(5): 1005–1016.
[35] Feigenbaum, M.J. (1979) The onset spectrum of turbulence. *Physics Letters A* **74**(6): 375–378.
[36] Gleick, J. (2011) *Chaos: Making a new science* (Open Road Media).
[37] Nagothu, D., Xu, R., Nikouei, S.Y. and Chen, Y. (2018) A microservice-enabled architecture for smart surveillance using blockchain technology. In *2018 IEEE International Smart Cities Conference (ISC2)* (IEEE): 1–4.
[38] Nikouei, S.Y., Xu, R., Chen, Y., Aved, A. and Blasch, E. (2019) Decentralized smart surveillance through microservices platform. In *Sensors and Systems for Space Applications XII* (International Society for Optics and Photonics), **11017**: 110170K.
[39] Jose, E., Greeshma, M., TP, M.H. and Supriya, M. (2019) Face recognition based surveillance system using facenet and mtcnn on jetson tx2. In *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)* (IEEE): 608–613.
[40] seeprettyface (2018) http ://www.seeprettyface.com/mydataset.html, last accessed on may 05, 2020 .
[41] Fitwi, A.H., Nagothu, D., Chen, Y. and Blasch, E. (2019) A distributed agent-based framework for a constellation of drones in a military operation. In *2019 Winter Simulation Conference (WSC)* (IEEE): 2548–2559.
[42] Fitwi, A.H. and Nouh, S. (2011) Performance analysis of chaotic encryption using a shared image as a key. *Zede Journal* **28**: 17–29.
[43] Otto, C., Han, H. and Jain, A. (2012) How does aging affect facial components? In *European Conference on Computer Vision* (Springer): 189–198.
[44] Levi, G. and Hassner, T. (2015) Age and gender classification using convolutional neural networks. In *Proceedings of the iEEE conference on computer vision and pattern recognition workshops*: 34–42.
[45] Teixeira, L., Maffra, F., Lelu, K., Al-Obaidi, A. and Badii, A. (2014) A rule-based methodology and assessment for context-aware privacy. In *2014 IEEE 6th International Conference on Awareness Science and Technology (iCAST)* (IEEE): 1–6.
[46] Du, J.X., Zhai, C.M. and Ye, Y.Q. (2011) Face aging simulation based on nmf algorithm with sparseness constraints. In *International Conference on Intelligent*

Computing (Springer): 516–522.