A Study of Privacy Protection of Photos Taken by a Wide-angle Surveillance Camera

Koki Nakai[1], Minoru Kuribayashi, Nobuo Funabiki
Okayama University, Okayama, Japan
E-mail: [1] pc6p5qeu@s.okayama-u.ac.jp

***Abstract*—In this paper, we propose a privacy protection system that detects human faces from images captured by a wide-angle camera, which is assumed to be a surveillance camera, and encrypts the face detection area. In the proposed face detection method, a classifier is created by AdaBoost learning based on Haar-like features, and the face region is detected from the image captured by the wide-angle camera. By creating the training data based on the face images captured by the camera, faces without can be detected compromising the detection accuracy, even for surveillance cameras. We use block-scrambling encryption to protect the privacy of the detected face area. During face detection, minimizing the probability of missing a face and allowing a certain number of false positives are necessary from the privacy protection viewpoint. In the case of false positives in previous encryption methods, the color space of the background cannot be preserved, resulting in visual degradation. Therefore, in the proposed encryption method, visual degradation is suppressed by improving the processing of the color components. Through simulations, we evaluate the effectiveness of the proposed method in terms of detection accuracy and processing speed for face detection, as well as color component change and compression efficiency for encryption.**

I. Introduction

With the increase in crime, the deployment of surveillance cameras and the demand for video recording devices for security has been increasing. Network cameras are the mainstream of surveillance technology because of the shift to Internet of Things (IoT) devices, and the photos and videos obtained from remote locations can be collected at a centralized server. However, these photos and videos may contain sensitive data, and, hence, there is a risk of leakage from the server by hacking, which may infringe on personal privacy. Therefore, the data without should be managed compromising on usability.

One potential issue is the leakage of faces in photos and videos. To prevent leakage, faces should be detected in the first step. From the privacy protection viewpoint, the probability of missing a face must be reduced to the extent possible, while false detection of areas other than faces can be allowed. The next step is to scramble/encrypt the detected face regions. A mosaic operation conceals the characteristics of a face. However, it was reported in [1] that a sophisticated deep learning technology can restore the original face from a mosaic image. Thus, a simple mosaic operation is no longer sufficient to protect privacy. Therefore, encrypting the face is necessary so that ordinary users can see the encrypted face image, and only the users with a secret key can view the original image.

The face detection algorithm in OpenCV[1] can recognize frontal faces in a target image with high accuracy. However, as the faces in surveillance cameras are captured from various angles, profile faces are difficult to detect owing to the reduced number of features. In addition, many surveillance cameras employ wide-angle lenses to eliminate blind spots. Although wide-angle cameras can capture a wide range of images, there is a problem that the accuracy of face detection is degraded owing to distortions in the images.

Block-scrambled encryption was proposed in [2], [3]. This method encrypts an image by scrambling each block to avoid reducing the compression efficiency of the encrypted image by considering block-based compression algorithms, such as Joint Photographic Experts Group (JPEG) compression. However, the characteristics of the color space completely change from the non-encrypted background. If many non-face regions are detected by a face detection algorithm, the direct use of this method causes significant degradation of its visual quality.

In this study, we investigate a face detection algorithm based on the Haar-like feature [4], which can detect face regions with high accuracy from images captured using a wide-angle lens by preparing a training dataset. In the proposed method, we use a surveillance camera to take pictures with a wide-angle lens and create a database of faces from various angles, including the head. Using the database as training data, we train the face detection algorithm under the constraint such that the error rate of missing faces is reduced to the extent possible. Then, we consider the effect of distortion caused by wide-angle cameras.

For the encryption of the detected regions, we propose a block-based encryption that allows for some false positives from the privacy protection viewpoint. In the proposed method, even in the case of false positives, the background color space features are preserved, and encryption is performed without degrading the compression performance. Through simulations, we evaluate the effectiveness of the proposed method in terms of detection accuracy and processing time for face detection, as well as the color component change of the encrypted image and compression efficiency for encryption.

[1] https://opencv.org/

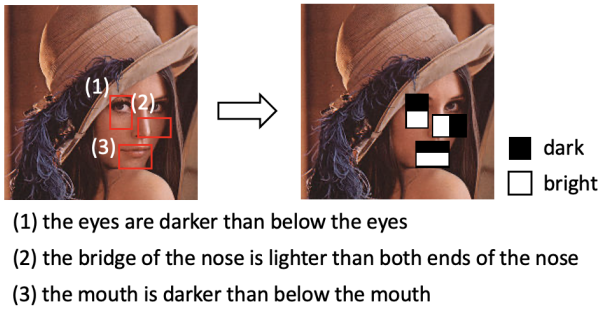


Fig. 1. Example of Haar-like features.

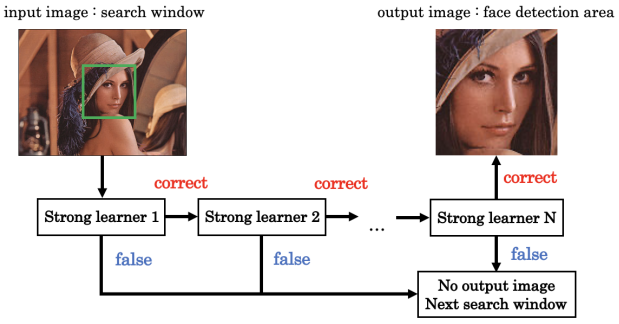


Fig. 2. Cascade classifier.

II. Face Detection

A. Haar Feature-Based Cascade Classifier

1) Haar-Like Feature: The Haar-like feature [4] was calculated by focusing on the difference between light and dark in a face. As shown in Fig. 1, for example, the eyes are darker than the area underneath, the bridge of the nose is lighter than both ends of the nose, and the mouth is darker than the chin.

2) Cascade Classifier Using AdaBoost Learning: A large number of combinations of Haar-like features exists, and creating a classifier by modeling effective features in advance is necessary.

Adaptive boosting (hereinafter, AdaBoost) learning [5] is a type of boosting algorithm model in which multiple weak learners are combined in series to form a strong learner with higher performance. Boosting is a type of ensemble learning, in which weights are assigned to the training data according to the output of each weak learner, and the learning is preformed sequentially while considering previous learning results. Weak learners are created based on Haar-like features.

A cascade classifier consists of several strong learners with different detection accuracies created by AdaBoost learning. Fig. 2 shows the mechanism of the cascade classifier. Each of the strong learners, in turn, determines whether the input image is a face. A face is determined to be a face only when all strong learners identify it as a face from the window-scanned image. In the cascade classifier, the lower the discriminant criterion for the previous strong learner, the higher the false-positive rate. The lower the discriminant criterion, the fewer features are needed, and thus the computation time is reduced. In the case of face detection, because the probability that most of the images in the search window are not faces is high, the cascade classifier can quickly remove the input images that are not identified as faces by the strong learner in front of it; thus, the entire process runs quickly.

B. MTCNN

The multi-task cascaded convolutional neural network (MTCNN) [6] is a deep learning model that uses three convolutional neural networks (i.e., P-net, R-net, and O-net) to detect faces in an image.

P-net: In a convolutional neural network (CNN) called P-net, the detection window and its bounding box regression vector are obtained in a similar manner to the deep dense face detector [7]. The estimated bounding box regression vector was then used for calibration, and non-maximum suppression (NMS) was used to integrate the overlapping detections.

R-net: All detections are fed into a CNN called R-net, which further eliminates many false detections and calibrates them with bounding box regression and NMS detection merge.

O-net: A CNN called O-net aims to detect faces in more detail, similar to R-net. Specifically, it outputs the positions of the five landmarks on the face.

The architecture of a CNN-based face detector was designed by [8]. However, the performance was limited for the following reasons. (1) Some filters lack weight diversity and, thus, may not be able to provide discriminative descriptions. (2) To improve the discriminability of face detection, the number of filters is 5×5, which is computationally expensive and difficult to identify in real time. However, because face detection is a task with only binary classification compared with other multi-class object detection and classification, discriminability is considered high even with a small number of filters. Therefore, the number of filters is reduced from 5×5 to 3×3 to minimize computational complexity. Meanwhile, by increasing the depth, we can increase the discriminability and still obtain the performance to identify at runtime.

For face detection assuming a surveillance camera, the cascade classifier is superior to MTCNN in terms of computational cost and runtime performance. However, MTCNN is better in terms of accuracy. In this study, we study a cascade classifier that is as accurate as MTCNN and superior in runtime.

C. Wide-Angle Camera

Wide-angle lenses are used in surveillance and security cameras because images should be captured over a wide area to eliminate blind spots. Although a wide-angle lens makes it possible to capture images over a wide area, it also causes lens distortion. There are two types of distortion caused by lenses,

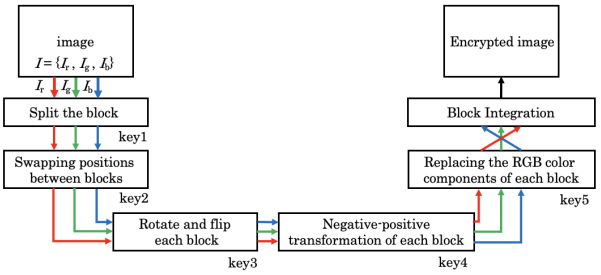
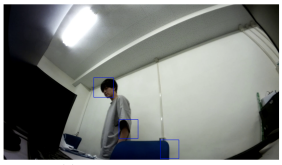
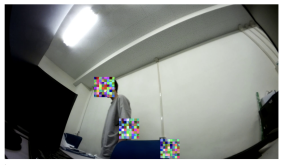


Fig. 3. Flowchart of a block-scrambling process.



(a) Face detection image



(b) Image with encrypted face detection area

Fig. 4. Problem of encrypted images using conventional methods.

namely, barrel-shaped distortion, a phenomenon in which a straight object appears to be bent, and volume distortion (wide-angle distortion), which affects the shape of subjects located at the periphery of the image in a wide-angle lens.

The distortion caused by these lenses deforms the contour of the face of a person located at the periphery of the image during face detection, which affects the detection accuracy.

III. Privacy Protection

In this section, we present an overview of the block-scrambling encryption method in [2], [3], which is a privacy-protecting method, and discuss the effects of JPEG compression and the problems associated with this method.

A. Block Scrambling

The flow of the block-scrambling process is shown in Fig. 3. In block-scrambling encryption, an image is divided into blocks of arbitrary size, and the following processes are applied to produce an encrypted image: interchanging the positions between blocks, rotating and flipping each block, negative–positive flipping, and RGB color component switching. Each process is based on pseudo-random numbers generated by a secret key, and the same key is required to decrypt the image.

Split the block
Split the original image $I = \{I_r, I_g, I_b\}$ into arbitrary block sizes, where I_r, I_g, and I_b represent each RGB component of I.

Swapping positions between blocks
The positions of the blocks divided into RGB components are switched according to the pseudo-random number generated using a secret key.

Rotate and flip each block
According to the generated pseudo-random number, the image is rotated by $(0°, 90°, 180°, 270°)$ for each RGB component and then flipped (horizontally or vertically).

Negative–positive transformation of each block
The pixel value of each RGB component is inverted according to the pseudo-random number generated by the secret key. The pseudo-random number consists of 0 and 1, where 0 indicates no negative–positive inversion and 1 implies negative–positive inversion.

Replacing the RGB color components of each block
The RGB components are switched according to the pseudo-random number generated by using the secret key.

Other encryption methods include RSA, Goldwasser-Micali's[9] method, Paillier's method[10][11], ElGamal's method[12], Okamoto-Uchiyama's method[13]. Unlike these number-theoretic transform based encryption methods, block-scrambling encryption methods are capable of compressing image data. Number-theoretic transform based encryption methods cannot be expected to compress image data. And secure decryption is possible. number-theoretic transform based encryption methods does not tolerate errors in the encrypted data, so if an error occurs on the transmission path, it will be difficult to recover the encrypted data.

B. Compression

JPEG is a compression coding scheme for still images used in computers.

In JPEG compression, the image is divided into blocks of 8×8 pixels. When the input image is a color image, the color difference components (Cb, Cr) are thinned by at most 1/2. Therefore, a block size of at least 16×16 pixels is required to create an 8×8 pixel block with only the color difference component when 1/2 thinning is applied. In the block-scrambling process, the discrete cosine transformation coefficients of the AC component can be kept in the same state as before the block by processing in blocks of $16k \times 16l$ pixels (where k and l are positive integers). This allows us to compress the image without significant degradation in image quality.

C. Problem

In the block-scrambling process of the previous study, the scrambling process locally scrambles the original image, as shown in Fig. 4, causing the color space to change drastically, and the degradation becomes visually noticeable only in that region. To harmonize the scrambled image with the rest of the detection area, the change in color space can be suppressed by using a common secret key for RGB components and introducing a threshold for negative–positive transformation.

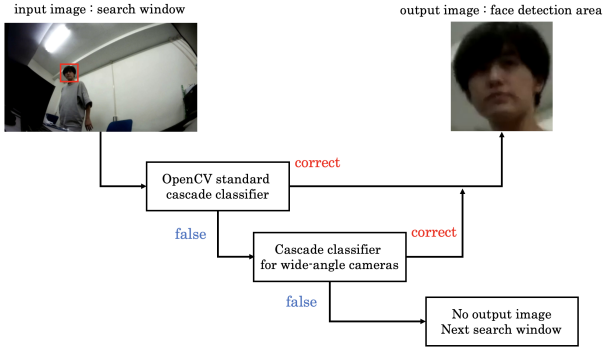


Fig. 5. Hybrid cascade classifier.

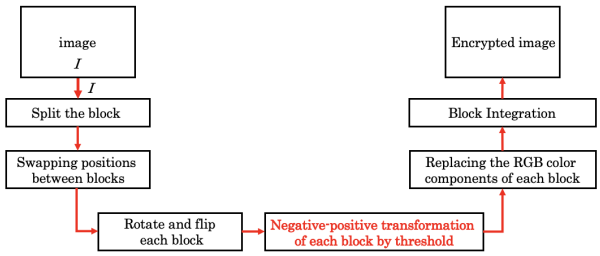


Fig. 6. Modified block-scrambling method.

IV. Proposed Privacy Protection Method

We propose a privacy-preserving system that detects human faces in images captured by a wide-angle camera, assumed to be a surveillance camera, and encrypts the face detection area. In the proposed method, a cascade classifier is created using face data captured by a wide-angle camera as training data and hybridized with the OpenCV standard cascade classifier to perform face detection. Then, the face detection region is privacy-protected using block-scrambling encryption.

A. Retraining of the Cascade Classifier

Wide-angle cameras, used as surveillance cameras, suffer from distortion caused by the lens. Therefore, the detection accuracy of face detection using the OpenCV standard cascade classifier is degraded. Therefore, we improve the detection accuracy by retraining the cascade classifier using the face data captured by the wide-angle camera as training data.

1) Collection of Training Data: In face detection, a correct image (i.e., a face image) and an incorrect image (i.e., not a face image) should be obtained. For face images, we take pictures while changing the angle of the face to correspond to various angles and use the faces of three subjects as the training data.

2) Data Augmentation: To create a highly accurate classifier, a large amount of training data are required. However, in this study, the number of models was quite small to achieve sufficient training. Therefore, we extended the training data by adding various linear transformations to the original data to increase the amount of training data.

3) Creating a Cascade Classifier for Wide-Angle Cameras: Using face images captured by a wide-angle camera as training data, we created a cascade classifier from the training data with data expansion.

B. Hybrid Method

In the proposed method, we hybridized the OpenCV standard cascade classifier with a cascade classifier trained for wide-angle cameras. The overall framework is shown in Fig. 5. The detection accuracy of the OpenCV standard cascade classifier is low owing to the distortion caused by the wide-angle camera. Therefore, as shown in Fig. 5, by using a classifier for wide-angle cameras to compensate for images that cannot be identified by the OpenCV standard cascade classifier, we can expect to achieve high accuracy in both reproduction and fit rates.

C. Modified Block Scrambling Method

In the proposed method, block-scrambling encryption is performed by modifying some of the processes in the conventional method. An outline of the encryption flow is presented in Fig. 6.

The block-scrambling encryption process is modified using a common encryption key for RGB components and negative–positive inversion with a threshold value to suppress changes in color components. In the case of negative–positive inversion, a threshold T is introduced to prevent negative–positive inversion for pixels with extremely low or high average pixel values in each block. Negative–positive inversion by the threshold can be determined using the following steps:

1) Calculate the block average pixel value.
2) Set the threshold T for pixel values. However, let $0 < T < 128$.
3) If the average pixel value of the block x is greater than T and less than $255-T$, the pixel value is inverted. The inversion of the pixel value is given by

$$x' = \begin{cases} 255 - x & T < x < 255 - T \\ x & \text{otherwise} \end{cases} \quad (1)$$

V. Experimental Results

A. Measurements

To confirm the effectiveness of the proposed method, we evaluated it in terms of detection accuracy and processing time for face detection, color component change for encryption, and compression efficiency.

1) Face Detection: In face detection, we classify the face of a person and the non-face of a person. Because it is a binary classification, there are four patterns of correct/false judgments: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Each case is explained in the following:

TP: Number of face areas correctly identified as faces.
TN: Number of non-facial areas correctly judged as non-facial.

TABLE I
ACCURACY OF FACE DETECTION.

Type of face detection	TP	TN	FP	FN	Recall (%)	Precision (%)
OpenCV standard cascade classifier	136	0	45	64	68.0	73.9
MTCNN [6]	190	0	15	10	95.0	92.7
Hybrid cascade classifier	193	0	69	7	96.5	73.7

TABLE II

Face detection type	Processing time (s)
OpenCV standard cascade classifier	0.0859
MTCNN [6]	4.8812
Hybrid cascade classifier	0.1238

FP: Number of non-facial areas mistakenly judged as faces.

FN: Number of face areas incorrectly judged as non-facial.

Recall and precision were used as evaluations of detection accuracy. The following is a summary of each evaluation:

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

2) Preservation of Color Space: One of the advantages of the proposed method is the suppression of the changes in the color components in the block-scrambled area. In the proposed method, we make the policy for face detection such that the FN is managed to be as small as possible while allowing the increase in FP. By revising the block-scrambling operation, we control the visual distortions in the non-face area that are mistakenly detected.

To compare the conventional and proposed methods using the original image in terms of the changes in color components for the encrypted image, we evaluated the changes in terms of mean squared error (MSE). Rather than the RGB color space, the changes in the Cb and Cr components were measured between the original and encrypted images in this experiment. The MSE can then be formulated as

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(f_i - y_i)^2, \tag{4}$$

where N is the number of pixels in the image, f_i is the color component of the original image, and y_i is that of the encrypted image.

B. Accuracy of Face Detection

We used a wide-angle camera to obtain 8,000 correct images (facial images) and 10,000 incorrect images (non-facial images/background), and then we created a cascade classifier using AdaBoost. To evaluate the accuracy of the detection, 200 images were randomly selected from 1,400 images not featured in the training data as test images.

Table I lists the results of comparing the face detection accuracy of the cascade classifier of the OpenCV standard only, the cascade classifier of the proposed method hybrid cascade classifier (OpenCV standard+wide-angle camera support), and MTCNN [6], and the processing times are listed in Table II.

The results show that the hybrid cascade classifier of the proposed method can improve the recall without reducing precision compared with the OpenCV standard cascade classifier for surveillance cameras. However, the precision was lower than that of MTCNN, resulting in many false positives. The advantage of the proposed method is its shorter processing time, because it assumes that the detector is installed in IoT devices, such as surveillance cameras.

(a) Conventional method (b) Proposed method

Fig. 7. Comparison of the file size when varying the threshold T.

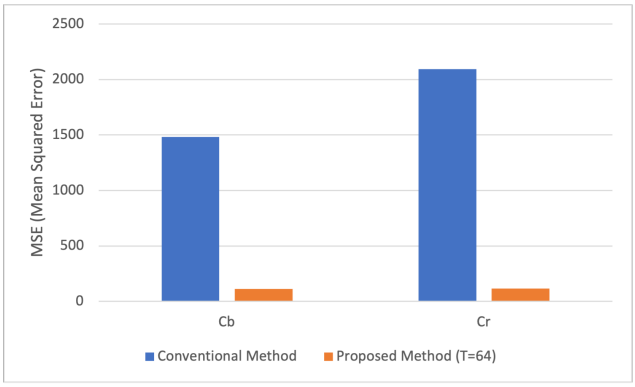


Fig. 8. Comparison of the changes of color components (Cb and Cr).

C. Evaluation of the Privacy Protection Method

We evaluated the color component change and compression performance of the block-scrambling method. For the color component change, we evaluated the color component of the image by comparing it with the encrypted image in the previous study and by changing the threshold of the negative–positive transformation in the proposed block-scrambling method. Then, we calculated the data size of the original image, encrypted image, and uncompressed image and evaluated the compression efficiency. For the simulation, 10 test images (128×128 pixels) were used, and the block

TABLE III
FILE SIZE COMPARISON (BYTES).

Uncompressed image	Original image	Encrypted image					
		$T = 0$	$T = 4$	$T = 8$	$T = 16$	$T = 32$	$T = 64$
49,150	20,096	24,918	25,012	25,080	25,034	25,076	25,009

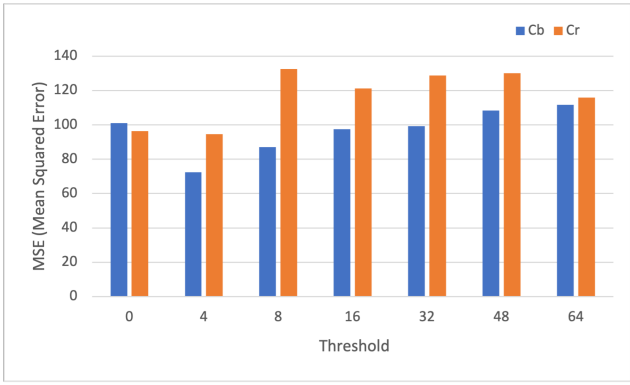


Fig. 9. Comparison of the changes of color components when varying the threshold T.

division size was 16×16 pixels.

First, we checked the change in color components in the encrypted image. Fig. 7 shows the comparison of the encrypted image between the conventional method and the proposed method. The encrypted image of the proposed method shown in Fig. 7 is when the threshold $T = 64$. Fig. 8 shows the results of the comparison between the conventional method and the proposed method ($T = 64$) for color component changes using MSE. The comparison results of the threshold of the proposed method for $T = \{0, 4, 8, 16, 32, 64\}$ are shown in Fig. 9.

A comparison of the file sizes of the original image, the encrypted image ($T = \{4, 8, 16, 32, 64\}$), and the uncompressed image is given in Table III. The file sizes of the original and encrypted images were calculated as the average of 10 face detection images.

The results show that the proposed method can suppress the change in color components compared with the conventional method, thus reducing the visual degradation when false positives occur in face detection. In addition, the encrypted image of the proposed method can reduce the file size and increase the compression efficiency compared with encryption based on number theory transformation (AES, DES).

VI. CONCLUSIONS

In this paper, we proposed a system for face detection and privacy protection in surveillance cameras. For the face detection policy, we have designed a system to minimize missed detection from the privacy protection viewpoint. However, the detection accuracy of the OpenCV standard cascade classifier is low for surveillance cameras. Our approach creates a hybrid of the OpenCV standard cascade classifier with a cascade classifier for wide-angle cameras used in surveillance cameras to achieve high detection accuracy.

The experimental results show that the hybrid cascade classifier has a higher recall than the OpenCV standard cascade classifier and reduces the number of missed faces. In terms of processing time, the computational cost is lower than that of MTCNN, which enables real-time detection. In the future, it will be necessary to further improve recall and precision.

In terms of privacy protection, the use of conventional block-scrambling methods causes visual degradation of the background when false positives occur in face detection. Therefore, our approach can be applied to face detection systems by partially modifying the conventional encryption process to reduce the change in color components.

Owing to the suppression of the color component change, the proposed method sacrifices confidentiality slightly. Future work will include the investigation of a new encryption process other than the change of color components and the setting of an appropriate threshold.

ACKNOWLEDGMENTS

This research was supported by JSPS KAKENHI Grant Number 19K22846, JST SICORP Grant Number JPMJSC20C3, and JST CREST Grant Number JPMJCR20D3, Japan. We would like to thank Editage (www.editage.com) for English language editing.

REFERENCES

[1] S. Menon, A. Damian, S. Hu, N. Ravi, and C. Rudin, "PULSE: Self-supervised photo upsampling via latent space exploration of generative models," in *Proc. CVPR'20*, 2020, pp. 2434–2442.

[2] K. Kurihara, M. Kikuchi, S. Imaizumi, S. Shiota, and H. Kiya, "An encryption-then-compression system for JPEG/motion JPEG standard," *IEICE Trans. Fundamentals*, vol. E98.A, no. 11, pp. 2238–2245, 2015.

[3] S. Imaizumi and H. Kiya, "A block-permutation-based encryption scheme with independent processing of RGB components," *IEICE Trans. Information and Systems*, vol. E101.D, no. 12, pp. 3150–3157, 2018.

[4] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, pp. 137–154, 2004.

[5] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. ICML'96*. 1996, pp. 148–156, Morgan Kaufmann Publishers Inc.

[6] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[7] S. S. Farfade, M. J. Saberian, and L.-J. Li, "Multi-view face detection using deep convolutional neural networks," in *Proc. ICMR'15*, 2015, pp. 643–650.

[8] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," *Proc. CVPR'15*, pp. 5325–5334, 2015.

[9] S. Goldwasser and S. Micali, "Probabilistic encryption and how to play mental poker keeping secret all partial information," *ACM Symposium on Theory of Computing*, pp. 365–377, 1982.

[10] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," *EUROCRYPT'99, LNCS*, vol. 1592, pp. 223–238, 1999.

[11] Chao-Yung Hsu Yi-Chong Zeng, Hon-Yue Chou Yi-Fei Luo, and Hong-Yuan Mark Liao, "Object detection in encryption-based surveillance system," *Proceedings of the Second APSIPA Annual Summit and Conference*, pp. 86–94, 2010.

[12] T. ElGamal, "A public-key cryptosystem and a signature scheme based on discrete logarithms," *IEEE Trans. on Information Theory*, vol. IT-31, no. 4, pp. 469–472, 1985.

[13] T. Okamoto and S. Uchiyama, "A new public-key cryptosystem as secure as factoring," *EUROCRYPT'98, LNCS*, vol. 1403, pp. 308–318, 1998.