Privacy-Preserving Surveillance for Smart Cities

Ijaz Ahmad
Dept. of Computer Engineering
Chosun University
Gwangju, Korea
ahmadijaz@chosun.kr

Eunkyung Kim
Dept. of Artificial Intelligence Software
Hanbat National University
Daejeon, Korea
ekim@hanbat.ac.kr

Suk-seung Hwang
Dept. of Electrical Engineering
Chosun University
Gwangju, Korea
hwangss@chosun.ac.kr

Seokjoo Shin
Dept. of Computer Engineering
Chosun University
Gwangju, Korea
sjshin@chosun.ac.kr (Corresponding author)

***Abstract*— This work presents privacy-preserving surveillance system for smart cities based on perceptual encryption algorithm (PE). The encryption is block-based and provides necessary level of security while preserving intrinsic properties of an image necessary for compression. Unlike existing PE methods, the proposed method retains color information, thus can enable processing in encryption domain. The analysis shows that our method achieves same compression performance as existing methods while providing better security. In addition, we have performed face recognition on the encrypted images and demonstrated that the proposed method delivers the same recognition accuracy as that of the plain images.**

Keywords—perceptual encryption; privacy-preserving Machine Learning; image compression; face recognition

I. INTRODUCTION

Video surveillance is one of the main building blocks of smart cities, which provides safer communities and efficient city operations. However, the convenience comes at a cost of relinquishing personal data and privacy. Given the large volume of data, cloud-based storage is emerging as a cost effective and efficient solution. In addition, the data is often outsourced to third-party computational resource providers for performing several computer vision tasks such as action and activity recognition, people counting, age and gender estimation, fire and smoke detection and vehicle detection [1]. The data collected by the surveillance system often consists of privacy sensitive data that can be exploited to recognize an individual. Therefore, it is important to keep the citizens data secure while providing them with the facilities. When transmitting data over unprotected public channels, the traditional encryption algorithms can be used for the protection of multimedia data. However, when the goal is to enable other requirements like low computational complexity, format compliancy and processing in the encryption domain then the number theory based encryption algorithms are not adequate. On the other hand, for privacy-preserving techniques like federating learning, differential privacy, and homomorphic encryption, there is a privacy and model accuracy tradeoff [2], [3]. Therefore, to solve these problems a new class of encryption techniques are emerging for protecting image data called perceptual encryption algorithms.

This research is supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07048338).

The perceptual encryption (PE) algorithms have simple computational steps that protects the human perceivable information while retaining the intrinsic properties of images to enable several applications. The algorithms are block based and performs four steps: block permutation, block rotation and inversion, and pixel level negative and positive transformation. The main advantage of the methods is that the encrypted images are JPEG compressible and are referred to as encryption-then-compression (EtC) methods. The applications of EtC schemes have been extended to social networking services and cloud-based photo stage [4], [5], image retrieval systems [6] and for protecting medical images [7]. Several studies have improved the encryption as well as compression performance of the EtC schemes. For example, [8] proposed color image based EtC system (Color-EtC) with an additional step to permute the blocks in the color channels for improved encryption efficiency. However, the scheme has a limitation on the block size. The smallest block size that can be used is 16×16 in order to avoid block distortion in the recovered image. Therefore, [4] proposed to represent the input image as grayscale image by combining the color channels along the horizontal or vertical direction. Such representation allows using a smaller block size of 8×8 and can improve the encryption efficiency. However, the grayscale image based EtC does not consider JPEG color subsampling. An alternative grayscale image based EtC is proposed in [5], which can enable color subsampling by using YCbCr color space.

The grayscale EtC methods (GS-EtC) have improved encryption efficiency; however, the lack of color information limits their applications. In this paper, we proposed an efficient PE method that uses different keys for each color channel in rotation-inversion, and negative-positive transformation steps. Thus, improves encryption efficiency of the existing methods without compromising their compression savings. In addition, the presence of color information in the encrypted images makes them suitable for privacy-preserving machine learning (ML) tasks. Color is crucial for face recognition task and improves performance of an algorithm significantly [9]. As an application of the proposed method, we have implemented ML based face recognition for color images in the encryption domain as opposed to [10] which is only applicable to grayscale images. The main advantage of the proposed method is that privacy sensitive images do not require to be exposed to the third-party

cloud owners for storage and/or computation. The proposed PE algorithm can be integrated as a component of the surveillance system used in smart cities.

II. METHODS

A perceptual encryption (PE) algorithm generally consists of the following two steps:

Step 1. Input image representation.

Step 2. Block-based geometric and color transformations. The geometric transformations change position and orientation of a block and color transformation changes pixels values in a block.

A. Conventional Perceptual Encryption Method [4]

When an input image is represented as a true color image (an image with three-color channels) as proposed in [8], then the block-based transformations are performed on block size size no smaller than 16×16 in order to avoid block artifacts in the recovered image. The larger blocks may make the scheme vulnerable to jigsaw puzzle attacks [17]. An alternative approach has been proposed in [4], which allows smaller block size. The method is described as below:

Step 1. Input image representation.

A color image is represented as a pseudo grayscale image by combining its color channels either in horizontal or vertical direction. When chroma subsampling is desirable then the method proposed in [5] can be used. The method first converts an input image into YCbCr color space and then the chroma components are down sampled. Finally, luminance and color components are concatenated to form a grayscale image.

Step 2. Block-based transformations

a) Divide the grayscale image from the previous step into blocks and permute them by using a random secret key K_1.

b) Rotate and invert each block randomly by using a key K_2 where each entry represents a different combination of rotation and inversion.

c) Apply negative-positive transformation to each block by using a uniformly distributed key K_3 as:

$$p'^{(i)} = \begin{cases} p(i), & K_3(i) = 0 \\ 255 - p(i), & K_3(i) = 1 \end{cases} \quad (1)$$

The cipher image obtained in the final step can be compressed by the JPEG image standard in grayscale mode.

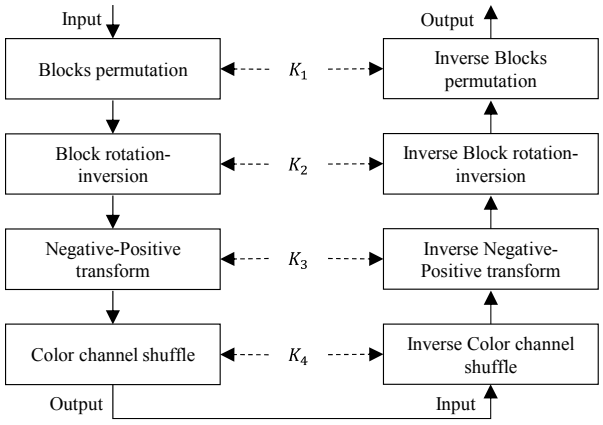

Fig. 1. Proposed block-based perceptual encryption method.

B. Proposed Perceptual Encryption Method

The proposed method is a block-based perceptual encryption (PE) algorithm that makes an image difficult to be recognized visually. Fig. 1. shows a high level illustration of the proposed method encryption and decryption processes and Fig. 2. shows the detail description of each step. The proposed method consists of the following steps:

Step 1. Input image representation.

The proposed method represents an input image as a true color image in order to preserve color and spatial information of the image.

Step 2. Block-based transformations

a) Divide an image with W × H pixels into blocks, each with $B_w \times B_h$ pixels, and permute the divided blocks by using a randomly generated secret key K_1. The same permutation key is used in each color channel, which is important to preserve same spatial information in each color channel.

b) Randomly rotate and invert each block by using a key K_2 where each entry represents a different combination of rotation and inversion. The key $K_2 \in \{K_2^R, K_2^G, K_2^B\}$ for the color channels red (R), green (G), and blue (B), where $K_2^R \neq K_2^G \neq K_2^B$.

c) Randomly apply negative-positive transformation to each block by using a uniformly distributed key K_3 as in (1). The key $K_3 \in \{K_3^R, K_3^G, K_3^B\}$ for the color channels red (R), green (G), and blue (B), where $K_3^R \neq K_3^G \neq K_3^B$.

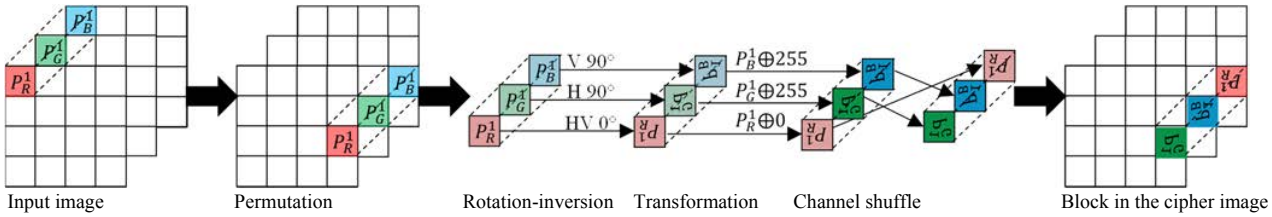

Fig. 2. Illustration of the proposed method encrypting a single block. Flip directions: {Vertical (V), Horizontal (H)}; Rotation degrees: {0°, 90°}; P_C^i: Block i of image with color channels as $C \in$ {Red (R), Green (G), Blue (B)}

d) Shuffle the blocks in the three channels randomly by key K_4 where the key elements represents a permutation of the channels.

The encrypted image obtained in the last step can be compressed by the JPEG standard in the RGB or YCbCr mode. The original image can be recovered by performing the above steps in a reverse order with the same keys.

The conventional color PE method [8] uses the same key for each color channel in the second and third steps while the proposed method uses different keys for each color channel in the same steps, and thus improves the encryption efficiency. On the other hand, the conventional grayscale PE method improves the security efficiency of the algorithm. However, lack of color information and disoriented spatial information limit their applications. Since, the proposed method represents an input image in color and preserves almost the same spatial information in each color component; therefore, overcomes limitations of exiting PE methods.

C. JPEG Compression

The JPEG standard created in 1992 is widely used image compression standard. It is applicable to both color and grayscale images. The baseline JPEG encoding consists of the following steps.

Step 1. The input image is first represented in YCbCr colorspace in order to separate the luminance from the color components of the image. For additional compression savings, the chroma subsampling is applied to the color components depending on the application requirements.

Step 2. The YCbCr image is divided into 8×8 blocks, which then goes through a discrete cosine transform (DCT) function [11]. The transform reduces the data correlation and gives a compact representation to large amount of information as few data samples, which aids to the compression savings in later steps.

Step 3. The DCT coefficients are quantized by using predefined quantization tables of different quality factors.

Step 4. The DC coefficients values in the adjacent blocks are represented as difference values. The AC coefficients are scan in zigzag order to benefit from the larger number of zeros.

Step 5. The final step is to encode the DC and AC coefficients obtained in the last step by using Huffman encoder. The Huffman encoder assigns a variable length code to each symbol depending on its frequency and stores them in a table. The most probable symbols have short codes while the less probable symbols have longer codes.

The same steps in a reverse order can carry out the JPEG decompression with the same quantization tables and Huffman encoding tables used during compression.

D. Machine Learning-based Face Recognition

Support vector machines (SVMs) [12] are a type of supervised machine learning (ML) algorithms for discriminative classification, which finds a line in two dimension or manifold in multiple dimension data to separate classes from each other. In general, there exist several separators to differentiate between the classes, which makes it difficult to choose a best fit. Instead of making a zero-width line as a decision boundary, we can draw a margin of some width on both sides of each line up to the nearest data point. SVMs choose a line that maximizes the margin, as an optimal model. The points that touch the margin are called support vectors. When fitting the model, loss function is computed based on the support vectors and any points beyond the margins do not modify the fit. In addition, for faster computation, instead of fitting the model on the original images, they can be treated as a vector in a high-dimensional space to derive a lower dimensional representation. One example of the methods is called principal component analysis (PCA). The PCA is an unsupervised algorithm that describes a dataset by finding a list of principal components, which are strictly eigenvectors, and are often called eigenfaces [13] when used for face recognition.

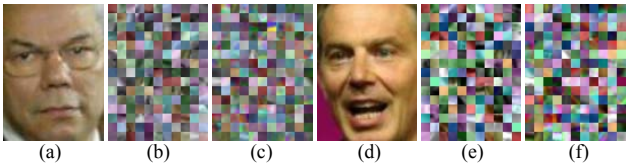

Fig. 3. Example images from the dataset. (a) and (d) are the original images. Their corresponding EtC encrypted images are (b) and (e), and PE encrypted images are (c) and (f).

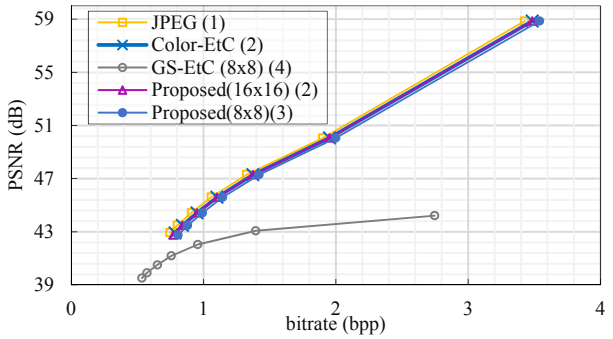

Fig. 4. Rate-distortion curves for different encoding schemes. The performance rank of each method is given in the end of each method name in the legend.

III. Results and Discussion

This section presents our simulation results. For compression and encryption performance analysis, the experiments are carried out on Tecnick dataset [14], which consists of 120 color images of dimension 1200×1200. For the baseline methods, we have implemented Color-EtC schemes and GS-EtC schemes without JPEG chroma subsampling. For accuracy analysis of the face recognition algorithm, Labeled Faces in the Wild (LFW)[15] have been used. Fig. 3. shows example images used as input for the face recognition model.

A. Compression Analysis

For the compression analysis, Fig. 4. plots the rate distortion (RD) curve for the image quality (dB) and bitrate (bpp) savings. The RD curves are for JPEG quality factors $Q_f =$

{70,75,80,85,90,95,100}. The JPEG compression of encrypted images obtained from the proposed and color EtC methods have the same performance as the compression of the original images while the GS-EtC method fails to deliver the same quality images. In order to quantify the difference between the RD curves with respect to JPEG compression, we used Bjontegaard delta (BD) measures [16]. The BD rate gives the percentage rate differences for the equivalent quality and BD quality measures the dB average differences for the equivalent rate. Fig. 5. shows the BD measures for different methods compared to JPEG compressed plain images. For the compression of encrypted images obtained from the proposed method and color EtC scheme only requires 2.7% more bitrate as compared to the plain images compression. However, when block size in the proposed method is reduced to 8×8 then the bitrate difference increases to 5.4%. On the other hand, compression of GS-EtC images requires bitrate up to 127% to achieve the same quality images.

B. *Encryption Analysis*

The keyspace of an encryption algorithm should be large enough to resist brute force attack. The encryption algorithm of the block-based PE system consists of four keys: key for permutation step K_1, key for block rotation and inversion K_2, key for positive and negative transformation of the pixel values K_3 and key for color channel shuffling step K_4. For example, if an image I of dimension W × H pixels is divided into blocks of size $B_{\mathrm{w}} \times B_{\mathrm{h}}$, then the number of blocks n is given as

$$n = \left\lfloor \frac{\mathrm{W} \times \mathrm{H}}{B_{\mathrm{w}} \times B_{\mathrm{h}}} \right\rfloor \tag{2}$$

When either dimension of the image is not divisible by the block size, then padding of the required size should be added. The key space K^C of the conventional algorithm is

$$\begin{aligned} K^C &= K_1^C \cdot K_2^C \cdot K_3^C \cdot K_4^C \\ &= n! \cdot 8^n \cdot 2^n \cdot 3!^n \end{aligned} \tag{3}$$

and the expanded key space K^P of the proposed algorithm is

$$\begin{aligned} K^P &= K_1^P \cdot K_2^P \cdot K_3^P \cdot K_4^P \\ &= n! \cdot 8^{3\times n} \cdot 2^{3\times n} \cdot 3!^n \end{aligned} \tag{4}$$

The conventional PE method uses the same key for each color channel in the second and third steps as $K_i \in \left\{K_i^{C^R}, K_i^{C^G}, K_i^{C^B}\right\}$ where $K_i^{C^R} = K_i^{C^G} = K_i^{C^B}$ for $i \in \{2,3\}$ and R, G, and B are red, green and blue channels, respectively. The proposed method uses different keys for each color channel in the same steps as $K_i \in \left\{K_i^{P^R}, K_i^{P^G}, K_i^{P^B}\right\}$ where $K_i^{P^R} \neq K_i^{P^G} \neq K_i^{P^B}$ for $i \in \{2,3\}$, and thus improves the encryption efficiency. Note that the keys in permutation and channel shuffling steps remain the same size for both methods. The larger keyspace guarantees resistance against brute force attack. Besides brute force attack, the security of the EtC systems have been analyzed intensively in [17] against extended Jigsaw puzzle attacks. For an EtC scheme to resist a jigsaw puzzle attack should have [4]: *large number of blocks, smaller block size and JPEG distortion*.

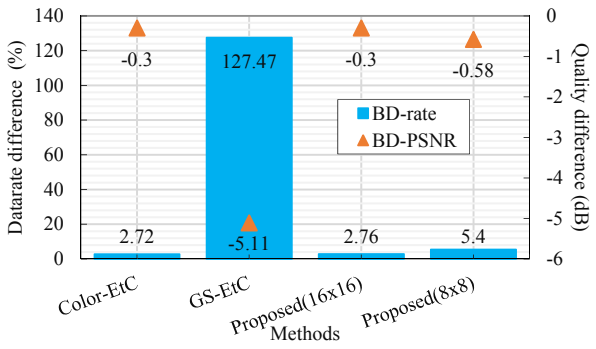

Fig. 5. BD measures for JPEG compression on encrypted images with respect compression of plain images.

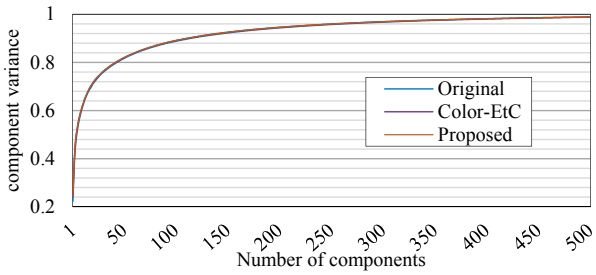

Fig. 6. Number of components required to represent the variance of LFW dataset.

The proposed scheme have larger number of blocks and includes JPEG distortion to resist jigsaw puzzle attacks.

C. *Accuracy Analysis*

For privacy-preserving face recognition task, we have implemented support vector machines (SVMs) as discussed in Section II. C. In the experiment, we have chosen people with at least 70 images from the LFW dataset. As a result, we have 7 different classes and 1,288 images in total. The images were resized to 120×88 pixels in order to avoid padding required to fit the block size (i.e., 8) of the encryption algorithm. For training, 75% of the images were used. In addition, for the dimensionality reduction, we have used randomized principal component analysis (PCA) instead of standard PCA for its faster computation. When using PCA for dimensionality reduction, only the largest principal components that represent the maximal data variance are preserved and the rest are zeroed out. The number of components needed for describing the data can be determined by the cumulative explained variance ratio as a function of the principal components number as shown in Fig. 6. One important thing to note here is that the block-based perceptual encryption has no effect on the PCA. Fig. 9. shows some example principal components (also called eigenfaces) for the plain dataset, EtC encrypted dataset and PE encrypted dataset. The first few eigenfaces show the angle of lighting on the face and the later corresponds to more details of the face. In experiments, principal components $\boldsymbol{N} = (500,250,150,100)$ have been used for face recognition, which accounts for 99%, 96%, 92% and 89% of the variance, respectively. Fig. 7. gives classification accuracy on the test dataset and the training time required for each value of $\boldsymbol{N}$. The training time increases as the number of components increases. The accuracy of SVM for face

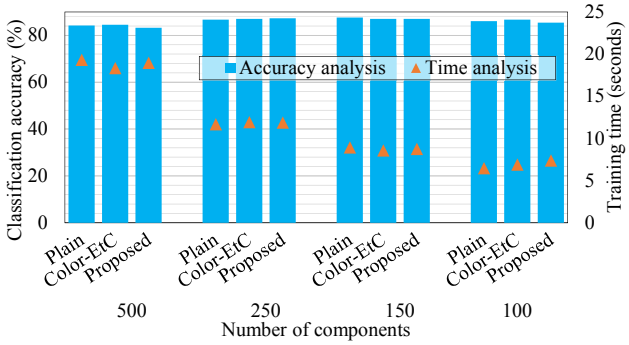

Fig. 7. Face recognition accuracy for different methods by varying the number of eigenfaces.

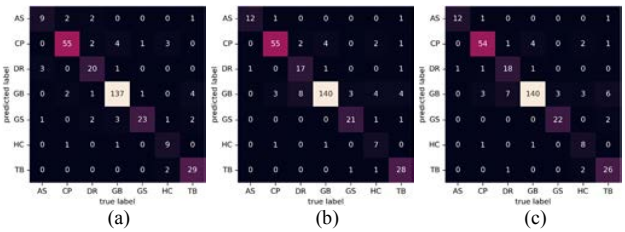

Fig. 8. Confusion matrices of SVM predictions on plain images (a), EtC encrypted images (b) and proposed encrypted images (c). The labels are the name initials where AS: Ariel Sharon; CP: Colin Powel; DR: Donald Rumsfeld; GB: George W. Bush; GS: Gerhard Schroeder; HC: Hugo Chavez; TB: Tony Blair.

recognition on plain and encrypted images remains almost the same with a negligible difference. The best accuracy is achieved with only using 92% variance with acceptable time. To get a better understanding of the trained estimator for $\boldsymbol{N} = 150$ components, Fig. 8. shows confusion matrices, which gives the labels that are likely to be missed by the estimator.

IV. Conclusion

In this work, we proposed block-based perceptual encryption algorithm for secure image data transmission and storage. The encryption is carried out in such a way that the cipher image retains intrinsic properties of the original image. Thereby, can enable computation in the encryption domain. The main advantage of the proposed method is that it retains color information, which makes it suitable for privacy-preserving machine learning (ML). As an application, we have implemented face recognition for privacy-preserving surveillance that can be used in smart cities. The analysis shows that the encryption has no effect on the algorithm accuracy.

References

[1] M. A. Ezzat, M. A. Abd El Ghany, S. Almotairi, and M. A.-M. Salem, "Horizontal Review on Video Surveillance for Smart Cities: Edge Devices, Applications, Datasets, and Future Trends," *Sensors*, vol. 21, no. 9, p. 3222, May 2021, doi: 10.3390/s21093222.

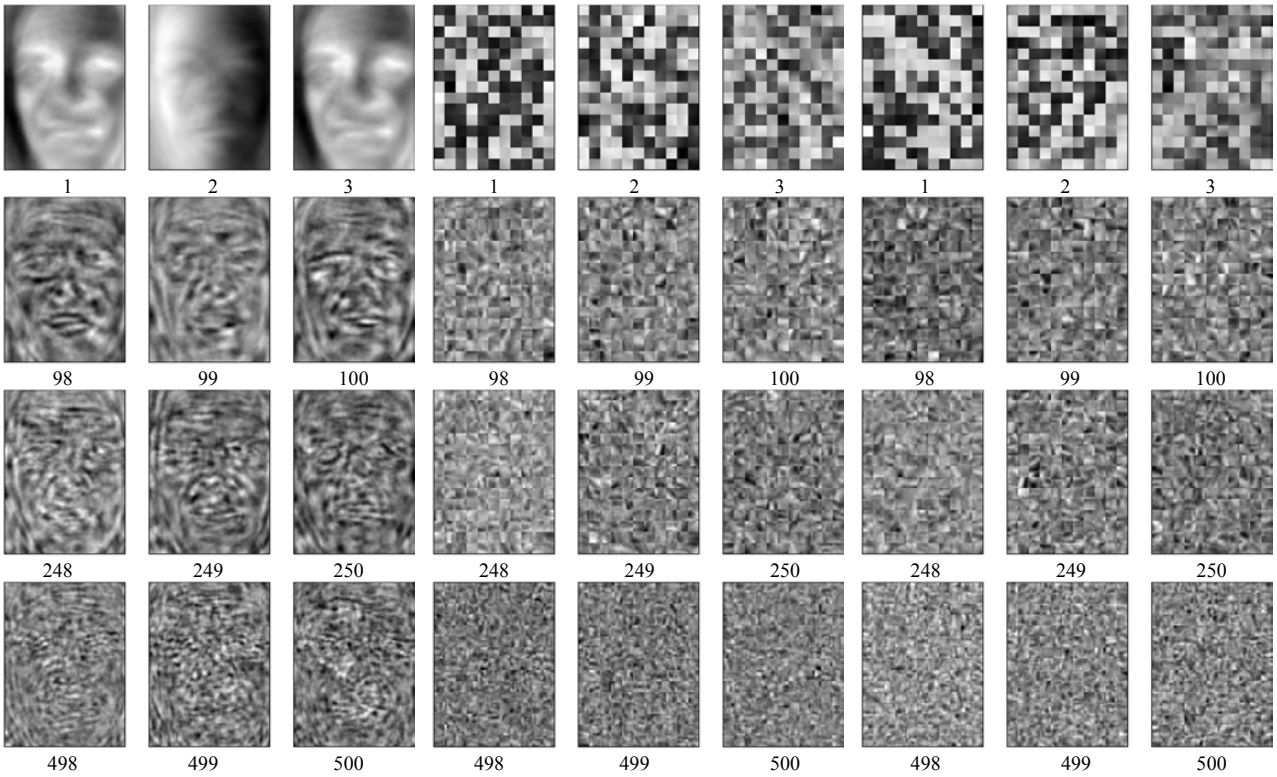

Fig. 9. Eigenfaces obtained from the plain images (column 1 to 3), EtC encrypted images (column 4 to 6), and proposed method encrypted images (column 7 to 9). The number below each figure shows the component in features vector space.

[2] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nat. Mach. Intell.*, vol. 2, no. 6, pp. 305–311, Jun. 2020, doi: 10.1038/s42256-020-0186-1.

[3] W. Kim and J. Seok, "Privacy-preserving collaborative machine learning in biomedical applications," in *2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Jeju Island, Korea, Republic of, Feb. 2022, pp. 179–183. doi: 10.1109/ICAIIC54071.2022.9722703.

[4] T. Chuman, W. Sirichotedumrong, and H. Kiya, "Encryption-Then-Compression Systems Using Grayscale-Based Image Encryption for JPEG Images," *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 6, pp. 1515–1525, Jun. 2019, doi: 10.1109/TIFS.2018.2881677.

[5] W. Sirichotedumrong and H. Kiya, "Grayscale-based block scrambling image encryption using YCbCr color space for encryption-then-compression systems," *APSIPA Trans. Signal Inf. Process.*, vol. 8, 2019, doi: 10.1017/ATSIP.2018.33.

[6] K. Iida and H. Kiya, "Privacy-Preserving Content-Based Image Retrieval Using Compressible Encrypted Images," *IEEE Access*, vol. 8, pp. 200038–200050, 2020, doi: 10.1109/ACCESS.2020.3035563.

[7] I. Ahmad and S. Shin, "Encryption-then-Compression System for Cloud-based Medical Image Services," in *2022 International Conference on Information Networking (ICOIN)*, Jeju-si, Korea, Republic of, Jan. 2022, pp. 30–33. doi: 10.1109/ICOIN53446.2022.9687214.

[8] K. Kurihara, S. Shiota, and H. Kiya, "An encryption-then-compression system for JPEG standard," in *2015 Picture Coding Symposium (PCS)*, Cairns, Australia, May 2015, pp. 119–123. doi: 10.1109/PCS.2015.7170059.

[9] B. Karimi and A. Krzyzak, "A Study on Significance of Color in Face Recognition using Several Eigenface Algorithms," in *2007 Canadian Conference on Electrical and Computer Engineering*, Vancouver, BC, Canada, 2007, pp. 1309–1312. doi: 10.1109/CCECE.2007.333.

[10] A. Kawamura, Y. Kinoshita, T. Nakachi, S. Shiota, and H. Kiya, "A Privacy-Preserving Machine Learning Scheme Using EtC Images," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. E103.A, no. 12, pp. 1571–1578, Dec. 2020, doi: 10.1587/transfun.2020SMP0022.

[11] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete Cosine Transform," *IEEE Trans. Comput.*, vol. C–23, no. 1, pp. 90–93, Jan. 1974, doi: 10.1109/T-C.1974.223784.

[12] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[13] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Maui, HI, USA, 1991, pp. 586–591. doi: 10.1109/CVPR.1991.139758.

[14] N. Asuni and A. Giachetti, "TESTIMAGES: a Large-scale Archive for Testing Visual Devices and Basic Image Processing Algorithms," *Smart Tools Apps Graph. - Eurographics Ital. Chapter Conf.*, p. 8 pages, 2014, doi: 10.2312/STAG.20141242.

[15] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," University of Massachusetts, Amherst, 07–49, Oct. 2007.

[16] Gisle Bjontegaard, "Calculation of average PSNR differences between RD-curves." VCEG-M33 ITU-T Q6/16, Apr. 2001.

[17] T. Chuman, K. Kurihara, and H. Kiya, "Security evaluation for block scrambling-based ETC systems against extended jigsaw puzzle solver attacks," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, Hong Kong, Hong Kong, Jul. 2017, pp. 229–234. doi: 10.1109/ICME.2017.8019487.