IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 7 September 2022, accepted 26 September 2022, date of publication 3 October 2022, date of current version 12 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3211657

RESEARCH ARTICLE

Privacy Protection in Surveillance Videos Using Block Scrambling-Based Encryption and DCNN-Based Face Detection

KHALID M. HOSNY[1], (Senior Member, IEEE), MOHAMED A. ZAKI[1], HANAA M. HAMZA[1], MOSTAFA M. FOUDA[2], (Senior Member, IEEE), AND NABIL A. LASHIN[1]

[1]Department of Information Technology, Zagazig University, Zagazig 44519, Egypt
[2]Department of Electrical and Computer Engineering, Idaho State University, Pocatello, ID 83209, USA

Corresponding author: Khalid M. Hosny (k_hosny@yahoo.com)

ABSTRACT Surely surveillance cameras are certainly important in all aspects of life. We have become in an era where we need to use surveillance cameras everywhere, homes, schools, banks, hospitals, and companies, even in the general streets, to monitor everything that happens and follow the progress of those places with all safety by surveillance videos. However, the pervasiveness of surveillance cameras has become an issue for people's privacy. This paper proposes a novel method for surveillance video privacy protection using block scrambling-based encryption and DCNN-based object detection. An object detection model based on DCNN You Only Look Once version 3 (YOLOv3) is used to detect the faces of the people. Then, the detected faces are scrambled using the fast block scrambling technique. Finally, the scrambled faces are encrypted using a secret key produced from a chaotic logistic map. The bounding boxes that output from the YOLOv3 are modified to include the entire edges of the detected faces to prevent any leaks of the sensitive regions. The simulation results and security analysis confirmed the proposed method's effectiveness in protecting the surveillance videos' privacy.

INDEX TERMS Video encryption, IoT, chaotic logistic map, surveillance camera, YOLO.

I. INTRODUCTION

Internet of Things (IoT) applications could bring massive value to our lives. With revolutionary computing capabilities, newer wireless networks, and superior sensors, IoT could be the next frontier in the race for its share of the wallet. Imagine an intelligent device such as a traffic camera. The camera can monitor the streets for accidents, weather conditions, traffic congestion, etc. The users can track anything that happened or is happening in their homes on their mobile devices. Because the surveillance videos are streamed over different networks, the people's privacy in the surveillance videos may be violated by an attacker. So the surveillance videos' privacy must be protected. Encryption is one of the most efficient mechanisms to protect surveillance videos [1], [2]. The surveillance video encryption techniques can be used in two manners: 1) entire video encryption [3], [4], [5], [6], and 2) encrypt only the regions of interest (ROIs) that are considered sensitive information [7], [8], [9], [10], [11], [12]. There is no need to encrypt the entire surveillance video, especially if the surveillance video has information from public places and does not have sensitive information in all regions. Figure 1 shows the two mentioned manners.

Additionally, encrypting entire frames of a surveillance video is computationally expensive. So encrypting only ROIs and keeping the non-ROIs unprotected increase the efficiency of the surveillance video protection process. This work focuses on ROI-based protection, and the ROIs represent people's faces. The proposed technique uses an object detector to extract the locations of the ROIs from surveillance video. Then the proposed encryption algorithm encrypts the ROIS using a fast block scrambling technique and keys from a chaotic map. The work's contributions are summarized as follows:

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenhua Guo.

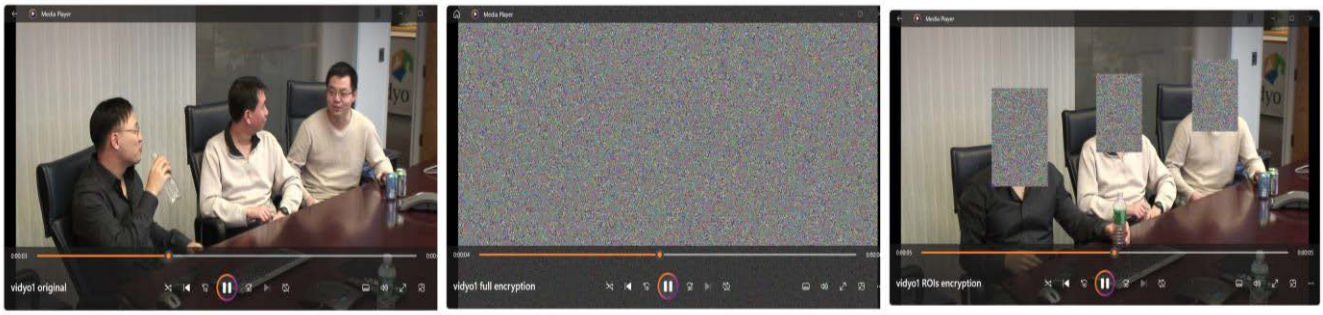


FIGURE 1. Full video encryption and ROIs video encryption.

1. The edges of the detected ROI are protected to prevent any sensitive leaks.
2. A novel splitting technique is used to generate blocks and sub-blocks from ROI.
3. Remove correlation between ROI pixels using a zigzag pattern, rotation, and blocks permutation to generate a scrambled ROI.
4. Encrypting different ROIs with different keys increases the security of encrypted ROIS.
5. The key used in the encryption process is based on the logistic map and the input ROI.

The rest of this paper is organized as follows. Section 2 explains the related work; Section 3 demonstrates the proposed privacy protection method in detail. In section 4, the simulation results and security analysis are presented. In section 5, the work is concluded.

II. REVIEW OF RELATED WORKS

Also, traditional encryption techniques such as 3-DES and AES are used in [13] and [14] to protect sensitive information. These techniques have a high computational cost [13]. Still, they are not the best solution for digital surveillance video privacy protection since the surveillance video frames have a high correlation between neighboring pixels and a large amount of redundancy. Recently, several techniques for privacy protection have been proposed [15], [16], [17], [18]. A secure video surveillance model is proposed in [15]. A secure authentication protocol is implemented to resist replay attacks and man-in-the-middle attacks. Lee and Park [16] have exploited blockchain technology in the network of surveillance systems. The method ensures the high security of cloud-based intelligent surveillance systems. Du et al. [17] proposed a method to quickly find the ROIs in videos, then protect the privacy of videos by encrypting the ROIs. Chu et al. [18] proposed a method for real-time privacy preserving moving object detection in the cloud. The method has some cons: 1) the encryption method is not strong if the chaotic map is not used in their randomness functions, and 2) the contours of the foreground objects are available at the server, violating privacy. Newton et al. [19] proposed a technique to protect privacy by de-identifying faces. The similarity between faces is calculated using a distance metric; then, new faces are generated by averaging components of the image. Korshunov et al. [20] developed a method to obfuscate faces in video surveillance based on well-known warping techniques. Ma et al. [21] proposed a reversible full privacy region protection method for cloud video surveillance. Du et al. [22] developed a privacy protection method in video surveillance that addresses video anonymization, behavior preservation, recoverability, and compressibility problems in one unified system. Rahman et al. [23] presented a scrambling technique based on chaos cryptography to protect ROIs that contain sensitive data in video surveillance. Zhang et al. [24] proposed a lightweight encryption technique based on layered cellular automata for privacy protection in surveillance videos. The ROIs are encrypted and stored on the camera side, where only authenticated users can access the encrypted ROIs. Any user can watch surveillance videos without ROIs.

ROI-based protection techniques protect the regions that have located by a detection algorithm. The security requirement for ROI protection should not depend only on the design of the encryption algorithm but also on the detection algorithm that detects the entire object to prevent any leaks of sensitive regions. The detection algorithm should detect objects accurately and efficiently. Wen et al. [25] use a geometric active contour model to detect the target regions. Kanso et al. [26] presented a technique to locate ROIs in a medical image. The input image is divided into blocks. Then each block is processed to determine whether it is a significant region or not based on a statistical measure. In [27], ROI with irregular shapes is chosen and detected arbitrarily. A Gaussian mixture model and HOG feature extraction are used in [27] and [28] for ROI detection. Many encryption and object detection techniques have been proposed. They have some shortcomings:

- Object detection time is high.
- The object detectors do not detect all sensitive regions.
- The running time of the encryption algorithm is high.
- The evaluation of the proposed work is performed based on test images and does not investigate the test videos.

FIGURE 2. The proposed method's pipeline.

FIGURE 3. Bounding box.

Motivated by these vulnerabilities, a fast and efficient encryption technique is proposed in this paper to protect the ROIs in surveillance videos to improve such drawbacks. The proposed technique uses an object detector to extract the locations of the ROIs. Then the proposed encryption algorithm protects the ROIS. An object detector called YOLOv3 [30] is used to locate multiple faces in a surveillance video. YOLOv3 is an improved version of YOLO [31]. YOLOv3 has advantages in the object detection process regarding speed and accuracy. Bounding boxes are generated from the detection process and represent the location of objects in a surveillance video. The proposed work modifies the bounding boxes to protect the edges of detected faces and prevent any sensitive information leaks. The proposed encryption algorithm consists of four steps. First, splitting the input ROI into blocks and sub-blocks. A series of operations are applied to the blocks and sub-blocks to scramble the ROI in the second step. Third, a secret key is generated using a chaotic logistic map. Finally, the scrambled ROI is encrypted using the generated secret key by applying the XOR operation.

III. THE PROPOSED PRIVACY PROTECTION METHOD

This section describes all steps of the proposed method in detail. YOLOv3 processes the surveillance video to locate the ROIs from each frame. Then the ROIs are fed into the encryption algorithm to encrypt sensitive data. Finally, the encrypted ROIs are placed into their original places in the original video frames. Figure 2 shows the proposed method pipeline.

A. FACE DETECTION PROCESS

Face detection is the first process in the proposed method to determine the locations of the people's faces in surveillance video. YOLOv3 is used to detect the people's faces in each frame. A pre-trained YOLOv3 weights file [32] which is trained on the wider face: a face detection benchmark dataset [33], is used in this work. The bounding boxes that represent the location of each face are modified by scaling them to include the edges of detected faces and prevent any sensitive leaks. Figure 3 shows a comparison between the original bounding boxes and the modified ones.

B. ENCRYPTION PROCESS

Before the detected ROIs are encrypted, they need to be preprocessed. The encryption and decryption algorithm requires the input size to be multiple of the block size number used in the ROI channel splitting step. Therefore the size of ROIs is processed and padded with neighboring pixels if needed. The proposed method for encrypting an ROI has four steps

applied to the three channels (red, green, and blue) of an ROI. Figure 4 shows the steps of the ROI encryption process.

1) ROI CHANNEL SPLITTING STEP

An ROI channel is split into blocks that have the same size. The block sizes the user can use are (i.e. 16, 32, and 64). Then a random number is generated for each block. For each random number of each block, the block is split into sub-blocks or kept as it is.

2) ROI CHANNEL CONFUSION STEP

In this step, the blocks and sub-blocks positions in an ROI channel are reordered randomly by applying more than one operation:

a. Change the arrangements of pixels in each block and sub-blocks by using a zigzag pattern.
b. Rotate the blocks and sub-blocks by ninety degrees.
c. Generate a random sequence with a length equal to the number of blocks in the ROI channel.
d. Generate a confused ROI channel by changing the arrangements of the blocks depending on the random sequence.

3) ROI KEY GENERATION STEP

Each ROI in the video is encrypted using a different key to make the proposed method more robust. The chaotic logistic map is used to generate the keys. The generated key is used in the ROI diffusion step. The key can be calculated using:

$$X_{n+1} = \lambda X_n(1 - X_n) \tag{1}$$

where the initial value X_0 is X_n when $n = 0$, $0 < X_0 < 1$, and $0 < \lambda \leq 4$. When $\lambda \in [3.57, 4]$, the map is chaotic. The initial value X_0 depends on the original ROI. The steps for key generation are:

a. Generate X_0 of the logistic map by:

$$X_0 = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} ROI(i, j, 1)}{M \times N \times 255 \times 3} + 10^{-20} \tag{2}$$

where M and N are the sizes of the input ROI.
b. Generate a sequence called SEQ_{tmp} by calculating the formula (1) $N_0 + MN$ times, where N_0 is a user-defined variable.
c. Generate a new sequence called S by skipping the first N_0 values of S_{temp}.
d. Calculate the key vector K by:

$$K\,(i) = mod\left(floor\left(S\,(i) \times 10^{14}\right), 256\right) \quad i = 1\ to\ MN \tag{3}$$

4) ROI CHANNEL DIFFUSION STEP

In this step, an ROI channel pixel's values are substituted with another value by applying XOR operation using the generated key vector K to generate the final encrypted ROI channel.

The encryption steps have clarified that the encryption process is strong and could protect the privacy of videos against different attacks without affecting the normal use of video, as shown in figure 5. If the encryption quality is reduced as the naked eye has a relatively low resolution, the video privacy seems to be encrypted well to the naked eye. Still, as to attackers, the sensitive data may be exploited as the encryption quality is reduced. For example, the encryption complexity would be reduced if all ROIs were in the same video frame or all video frames were encrypted with the same key instead of one for every ROI. Still, if the attacker could decrypt one ROI, all ROIs in the video would be leaked. Algorithm 1 presents the encryption process in detail.

Algorithm 1 The Proposed Encryption Process for the Detected ROI

Input: Plain ROI $\boldsymbol{P}$, block size $\mathbf{b}$, a parameter of the logistic map $\boldsymbol{\lambda}$, and the iterations number $\boldsymbol{N_0}$.
Output: Encrypted ROI $\boldsymbol{E}$

1) Separate the channels of P, so C_1 contains the red channel, C_2 contains the blue channel and C_3 contains the green channel.
2) Create an empty matrix E with a size equal to P to store the encrypted channels of the ROI.
3) Calculate the initial value of the logistic map by formula (2).
4) Execute the formula (1) $N_0 + MN$ times, then skip N_0 element and store the result in a vector S.
5) For $i = 1$ to MN
6) K (i) = mod (floor (S (i) $\times 10^{14}$) , 256).
7) End for
8) For $j = 1to3$
9) Split C_j into blocks B, with a dimension size b.
10) Generate a random vector R, *with length* $= lenght(B)$, *where* $2 \leq R_i \leq log_2(b)$.$i = 1, 2, .., MN/b^2$
11) For $i = 1$ to length(R)
12) Split B_i into sub-blocks or keep it as it is, depending on R_i.
13) End for
14) Change the pixel's positions of the blocks and sub-blocks by using a zigzag scan.
15) Change the pixel's positions of the blocks and sub-blocks by Rotating 90°.
16) Generate a random sequence R, *with length* (MN/b^2).
17) Change the positions of the blocks using the sequence R to obtain a permuted ROI Y.
18) Generate $Y' = reshape(Y, 1, MN)$.
19) X = Y′ ⊕ K
20) Generate $X'(j) = reshape(X, M, N)$.
21) E(:, :, j) = X′(j).
22) End for

C. DECRYPTION PROCESS

In this process, the original surveillance video frames are retrieved from the encrypted one by inverting the encryption

steps and obtaining the locations of the encrypted ROIS and keys used in the encryption process. The steps of the decryption process are:

1. Generate the scrambled ROI from the encrypted one by applying the XOR operation between the encrypted ROI and its key vector.
2. Reorder the positions of the blocks of the ROI channels using the generated random sequences.
3. Rotate the blocks and sub-blocks of the ROI channels by ninety degrees in the counter direction.
4. Apply an inverse zigzag pattern to the blocks and sub-blocks of the ROI channels to generate the plain ROI channels.

IV. SIMULATION RESULTS AND SECURITY ANALYSIS

This section uses different statistical tests and measurements to analyze the proposed work. Tested videos from [34] are used as surveillance videos. Table 1 shows the properties of the tested videos. MATLAB (R2021a) is used to execute the proposed work on a device that has Intel(R) Core(TM) i7-8750H CPU @ 2.20GHz 2.21 GHz, 16 GB memory, and Windows 11 OS. The used parameters in the proposed algorithm are: $b = 16$, $\lambda = 3.9$, *and* $N_0 = 1500$.

TABLE 1. Test videos properties.

Video Name	Number of Frames	Resolution
Salesman	449	QCIF
Akiyo	300	CIF
Paris	1065	CIF
Crew	600	4CIF
vidyo4	600	720p
vidyo1	600	720p

A. VISUAL ANALYSIS

All detected faces in test videos are encrypted and decrypted to evaluate the proposed work from the visual inspection. All results are shown in figure 5. From the figure, it is clear that the proposed work can protect the privacy of the persons in the videos without any leaks. Also, the decryption process can retrieve the original videos successfully.

B. HISTOGRAM ANALYSIS

A histogram is an important tool for representing the frequency distributions of the intensities in a grayscale image. The histogram is applied to the original ROIs and their corresponding encrypted ROIs in frame number 312 of the vidyo1 test video. Figures 6-8 show the histogram for the original and encrypted ROIs of left, middle, and right persons in the test video. The figures state that the intensities distributions of the encrypted ROIs are uniform compared to the original ROIs. Consequently, the proposed work can mask any patterns in the original ROIs.

TABLE 2. Entropy for original, encrypted, and decrypted ROIs of vidyo1.

Video Object	Channel	Original ROI	Encrypted ROI	Decrypted ROI
Left ROI	R	7.3509	7.9959	7.3509
	G	7.2380	7.9959	7.2380
	B	7.2106	7.9957	7.2106
Middle ROI	R	7.3554	7.9935	7.3554
	G	7.1347	7.9932	7.1347
	B	7.0999	7.9940	7.0999
Right ROI	R	7.5490	7.9927	7.5490
	G	7.6972	7.9925	7.6972
	B	7.6181	7.9932	7.6181

TABLE 3. Correlation coefficient for original and encrypted ROIs.

		Original ROI			Encrypted ROI		
		V	H	D	V	H	D
Left ROI	R	0.9906	0.9909	0.9845	-0.0064	-0.0061	0.0123
	G	0.9890	0.9897	0.9798	-0.0099	-0.0129	-0.0278
	B	0.9887	0.9892	0.9785	-0.0278	-0.0044	-0.0118
Middle ROI	R	0.9895	0.9879	0.9803	0.0165	-0.0178	-0.0406
	G	0.9873	0.9840	0.9738	-0.0138	0.0043	-0.0322
	B	0.9838	0.9806	0.9717	-0.0129	-0.0151	-0.0005
Right ROI	R	0.9918	0.991	0.9828	0.0085	0.0114	0.0117
	G	0.9925	0.9897	0.9832	0.0022	0.0026	-0.0358
	B	0.9917	0.9882	0.9834	-0.0132	0.0126	-0.0134

C. INFORMATION ENTROPY ANALYSIS

The entropy is carried out to measure the unpredictability and randomness of the original ROIs and their corresponding encrypted ROIs of the test videos. The mathematical formula of the entropy is defined by:

$$H(X) = \sum_{i=0}^{255} p(X_i) \, log_2 \frac{1}{p(X_i)} \tag{4}$$

where X_i represents the grey level value of the input ROI channel, and the probability of X_i is $p(X_i)$. The larger the ROI entropy is, the more randomness of the gray values are. The vidyo1 test video is used in this experiment. Table 2 shows the entropy values for frame number 312. This table shows that the entropy values of the encrypted ROIs are near the theoretical value 8. So the ROIs encrypted by the proposed method are against entropy attacks.

D. CORRELATION ANALYSIS

Usually, there are strong correlations between neighboring pixels because they have similar values. Such relationships among pixels must be eliminated in any effective encryption algorithm. The correlation coefficient between adjacent pixels in the horizontal, vertical, and diagonal directions can be calculated to evaluate the correlation strength.

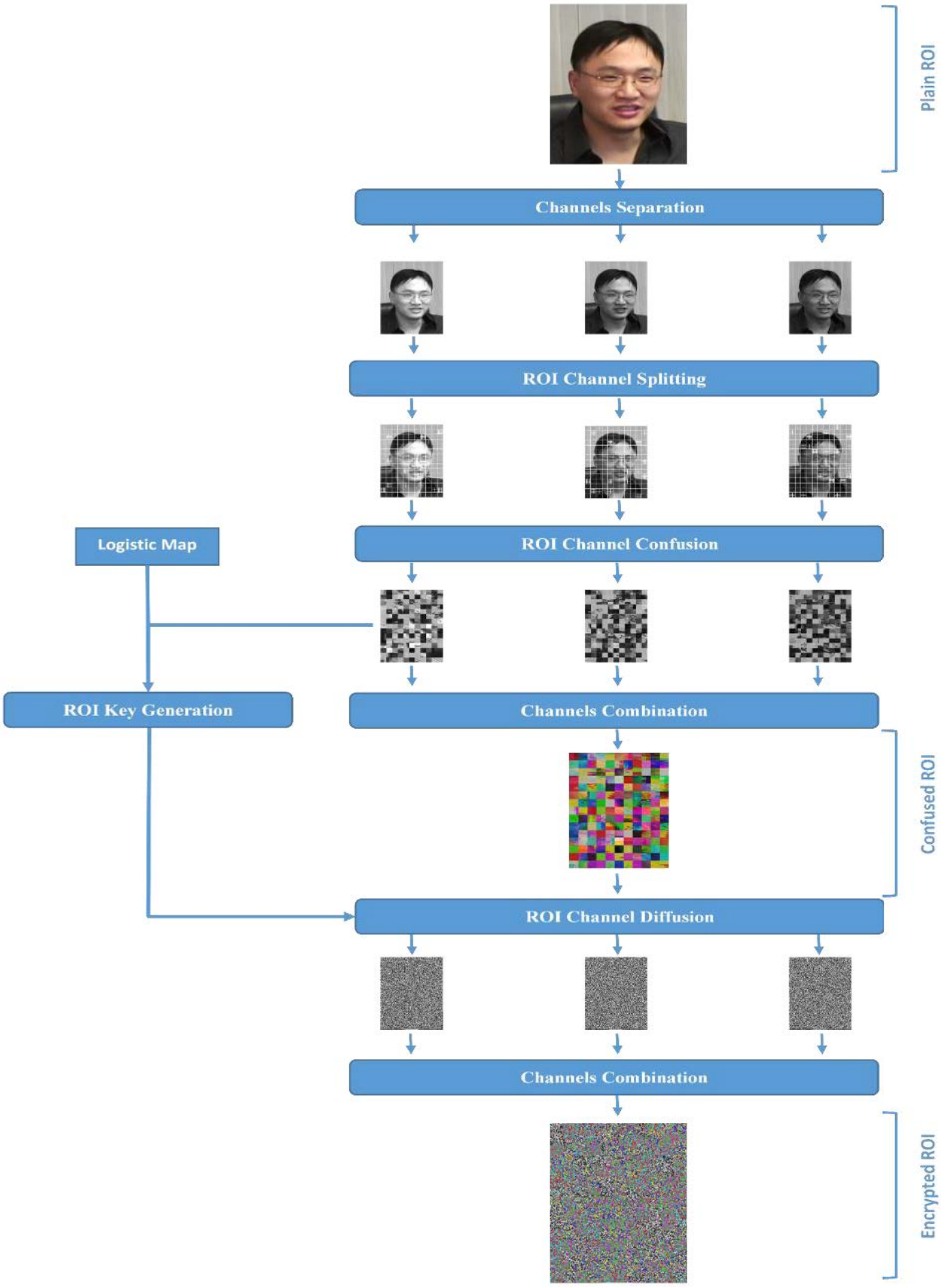


FIGURE 4. ROI encryption diagram.

FIGURE 5. Original, encrypted, and decrypted videos.

Mathematically, it is calculated by:

$$r_{u,v} = \frac{E((u-E(u))(v-E(v)))}{\sqrt{D(u)D(v)}} \tag{5}$$

$$E(u) = \frac{1}{s}\sum_{i=1}^{s} u_i \tag{6}$$

$$D(u) = \frac{1}{s}\sum_{i=1}^{s} (u_i - E(u))^2 \tag{7}$$

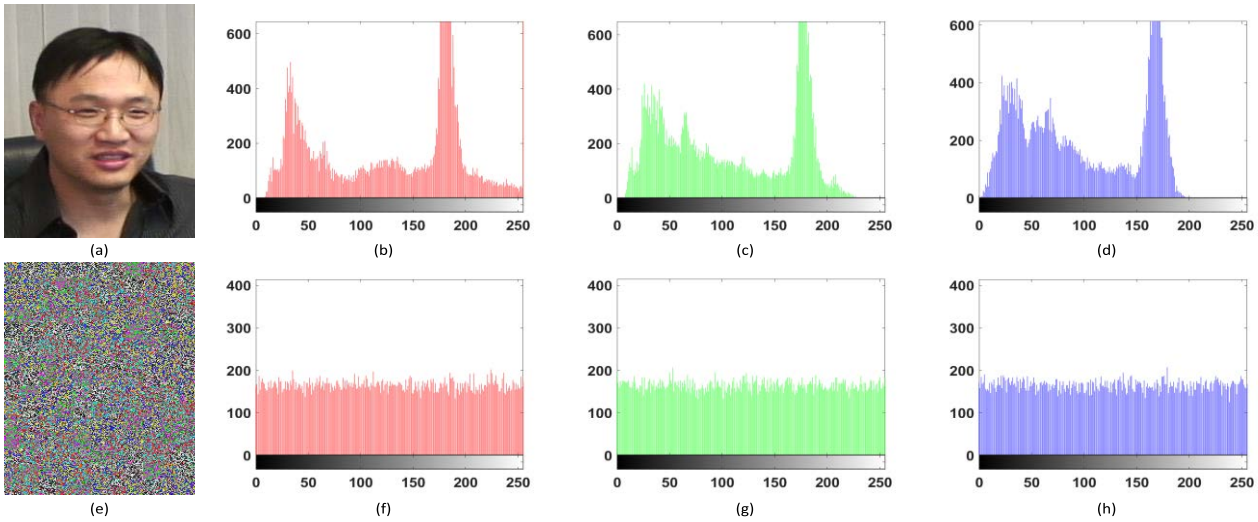


FIGURE 6. Histogram for original and encrypted left ROI of vidyo1.

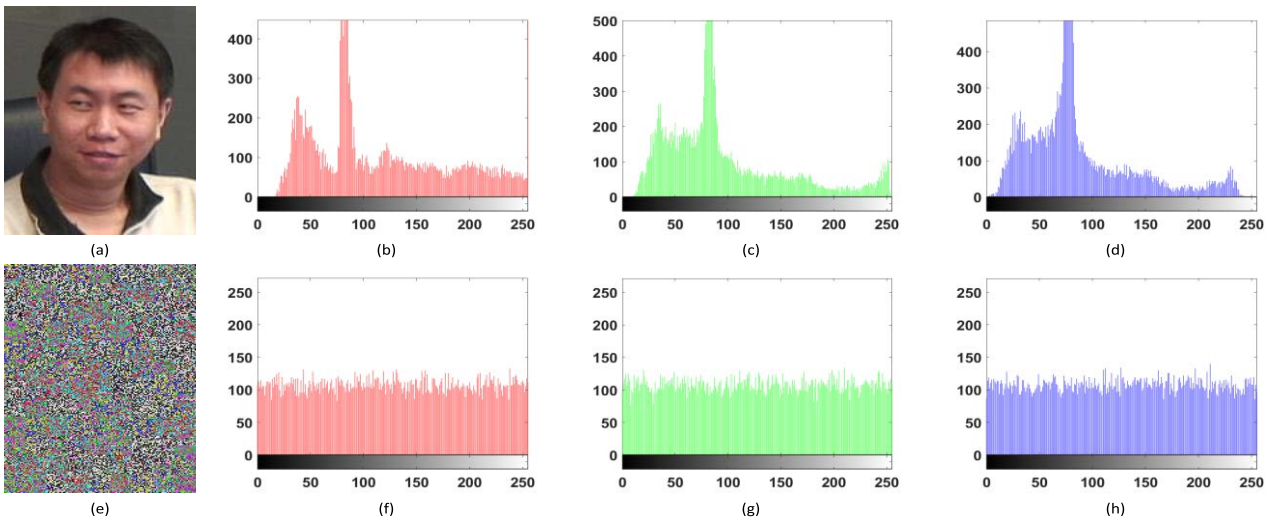


FIGURE 7. Histogram for original and encrypted middle ROI of vidyo1.

where u and v represent two adjacent pixel values, and s is the number of sampled pixel pairs. Five thousand pairs of neighboring pixels in vertical, horizontal, and diagonal directions are sampled from the color channels of the original ROIs and their corresponding encrypted ROIs of the vidyo1 test video. Figures 9-11 show the correlation distribution from the three directions for the left, middle and right ROIs in frame 312. Table 3 shows the results of the correlation coefficients of frame number 312. From these results, the coefficient values are close to one in the original ROIs, while the values are close to zero in the encrypted ROIs. Consequently, the proposed method has removed the correlation between pixels and can resist statistical attacks.

E. DIFFERENTIAL ATTACK ANALYSIS

In a differential attack, the adversary modifies a plain ROI p_1 by changing a one-bit pixel to get a modified plain ROI p_2. Then, p_1 and p_2 are encrypted using the same key to get two encrypted ROIs I_1 and I_2. The adversary then searches for the relationships between the plain and encrypted ROIs. The encryption algorithm must be sensitive to any small change in the original ROI. NPCR (Number of Pixels Changing Rate) and UACI (Unified Average Changing Intensity) are two criteria used to analyze this sensitivity, and they are identified by:

$$NPCR = \frac{1}{WH} \sum_{i=1}^{M} \sum_{j=1}^{N} D\,(i,j) \times 100(\%)$$

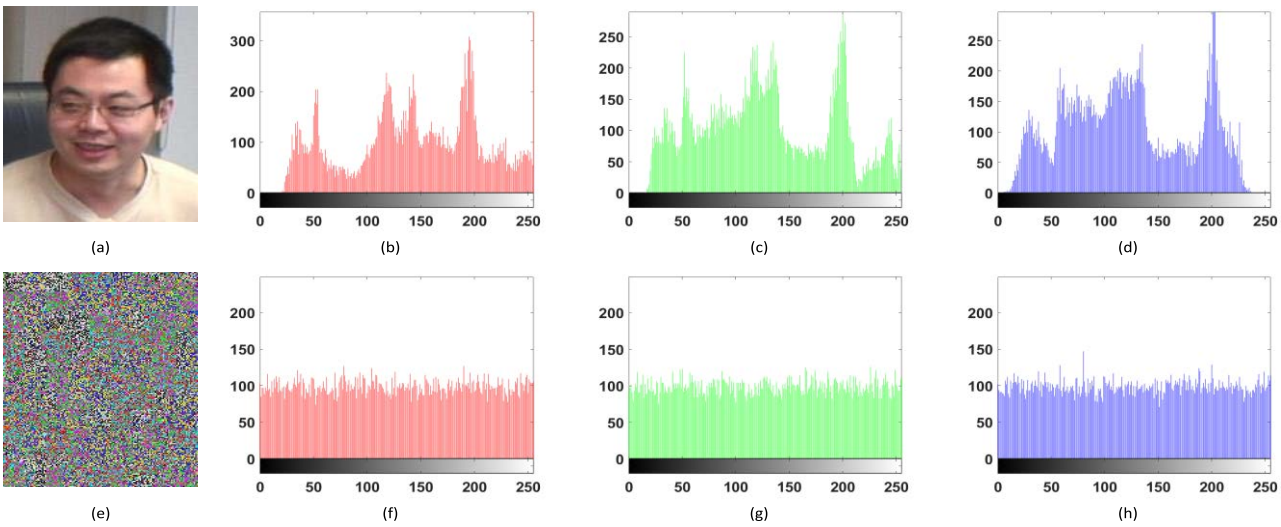


FIGURE 8. Histogram for original and encrypted right ROI of vidyo1.

where

$$D(i,j) = \begin{cases} 0 & if\ I_1(i,j) = I_2(i,j), \\ 1 & if\ I_1(i,j) \neq I_2(i,j), \end{cases} \tag{8}$$

$$UACI = \frac{1}{WH} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{|I_1(i,j) - I_2(i,j)|}{255} \times 100(\%) \tag{9}$$

where I_1 and I_2 represent two encrypted ROIs. The ROI width and height are represented by W and H. The NPCR and UACI theoretical values are 99.6094% and 33.4635%, respectively. Two versions of the vidyo1 test video are used in this experiment. The first version is the plain video, and the second version is the plain video with a one-bit pixel change in each ROI. Both video versions are encrypted, and then the NPCR and UACI values are calculated. Table 4 shows the results of frame number 312. The results are near the expected values, and the developed method can withstand the differential attack.

TABLE 4. PSNR, SSIM, and FSIM results between original and corresponding encrypted ROIs.

Video Object	Channel	NPCR	UACI
Left ROI	R	99.5881	33.4166
	G	99.5691	33.5704
	B	99.6188	33.3275
Middle ROI	R	99.7106	33.6079
	G	99.6781	33.6712
	B	99.5696	33.5516
Right ROI	R	99.6054	33.2439
	G	99.5502	33.5130
	B	99.5739	33.4714

F. PSNR, SSIM, AND FSIM ANALYSIS

In this section, peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and feature similarity (FSIM) values are calculated to assess the quality performance of the encryption and decryption processes.

The PSNR is used to measure the ratio between the maximum value of a pixel to the mean square error (MSE) between the encrypted ROI and the original ROI and can be identified by:

$$PSNR = 10 \times \log_{10}\left(\frac{255 \times 255}{MSE}\right)(db)$$

$$MSE = \frac{1}{MN} \times \sum_{i=1}^{m} \sum_{j=1}^{n} |R_1(i,j) - R_2(i,j)|^2$$

where, R_1 and R_2 are the original and the encrypted ROI. The encryption process is good if the value of PSNR between the original ROI and the corresponding encrypted one is small. The results of PSNR values between the original and encrypted ROIs in frame number 250 for the vidyo1 test video are presented in table 5. From the table, the PSNR values are small, indicating the proposed encryption process is efficient. Also, the PSNR values between the original and decrypted ROIs in frame number 250 for the vidyo1 test video are presented in table 6. The PSNR has infinity values from the table, indicating the proposed decryption process is efficient.

The SSIM index estimates the degree of similarity between two ROIs. SSIM range from -1 to 1, where one indicates that the two ROIs are identical. It is calculated by:

$$SSIM(o,e) = \frac{(2m_o m_e + c_1)(2\sigma_{oe} + c_2)}{(m_o^2 + m_e^2 + c_1)(\sigma_o^2 + \sigma_e^2 + c_2)}$$

where m_o and m_e, respectively represent the average value of original ROI o and encrypted ROI e, σ_o^2 is the variance

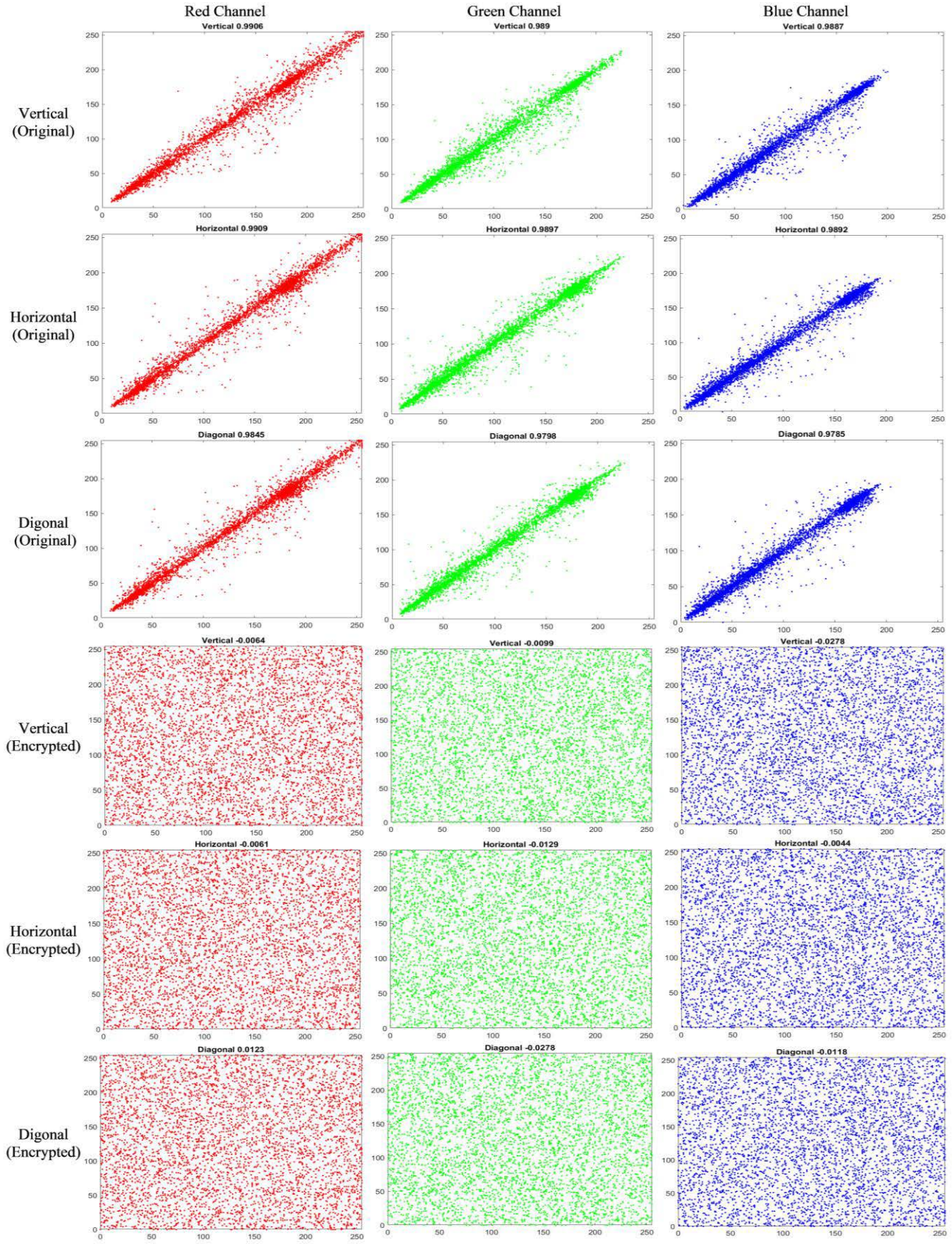


FIGURE 9. Correlation distribution of the original and encrypted left ROI.

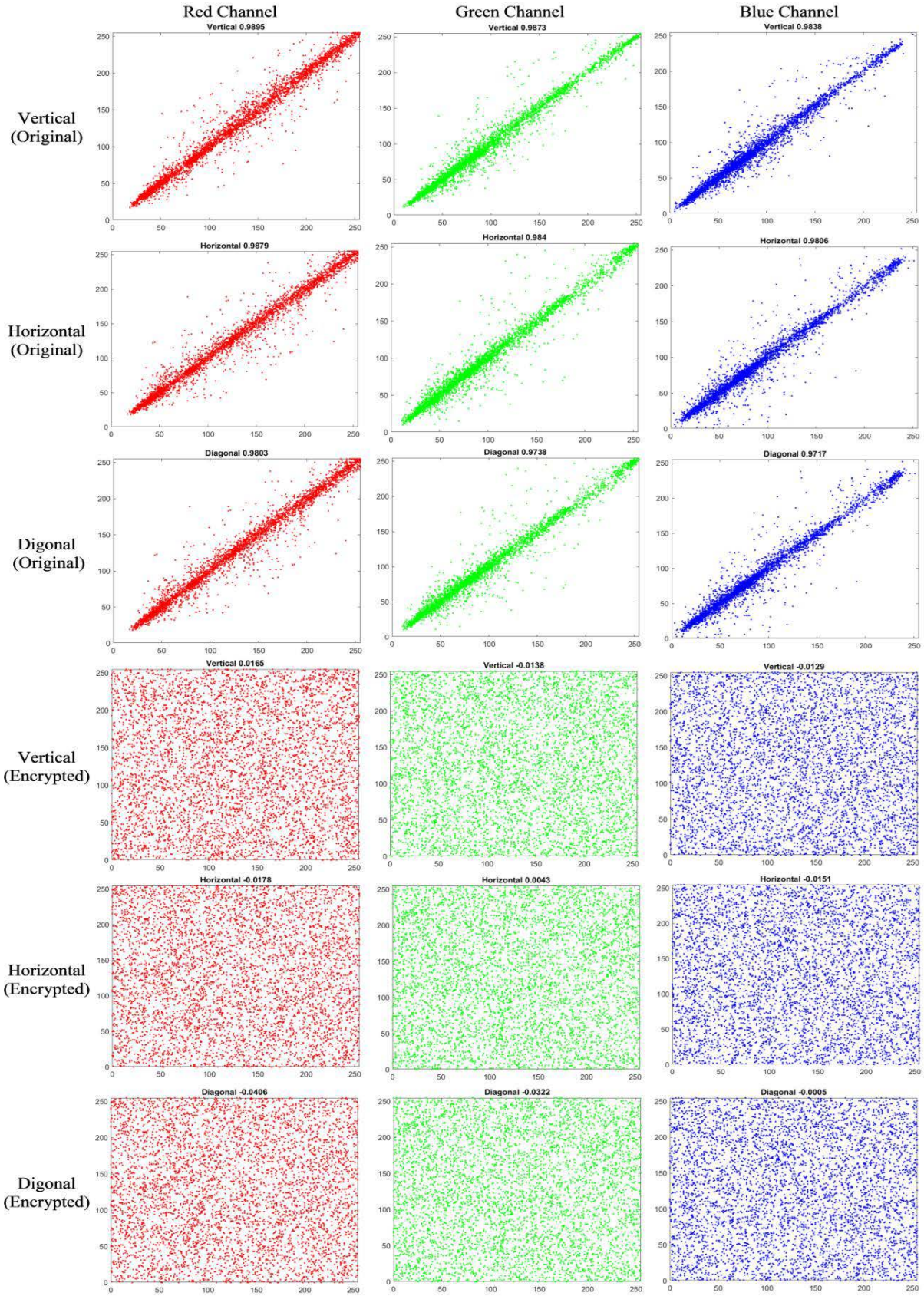


FIGURE 10. Correlation distribution of the original and encrypted middle ROI.

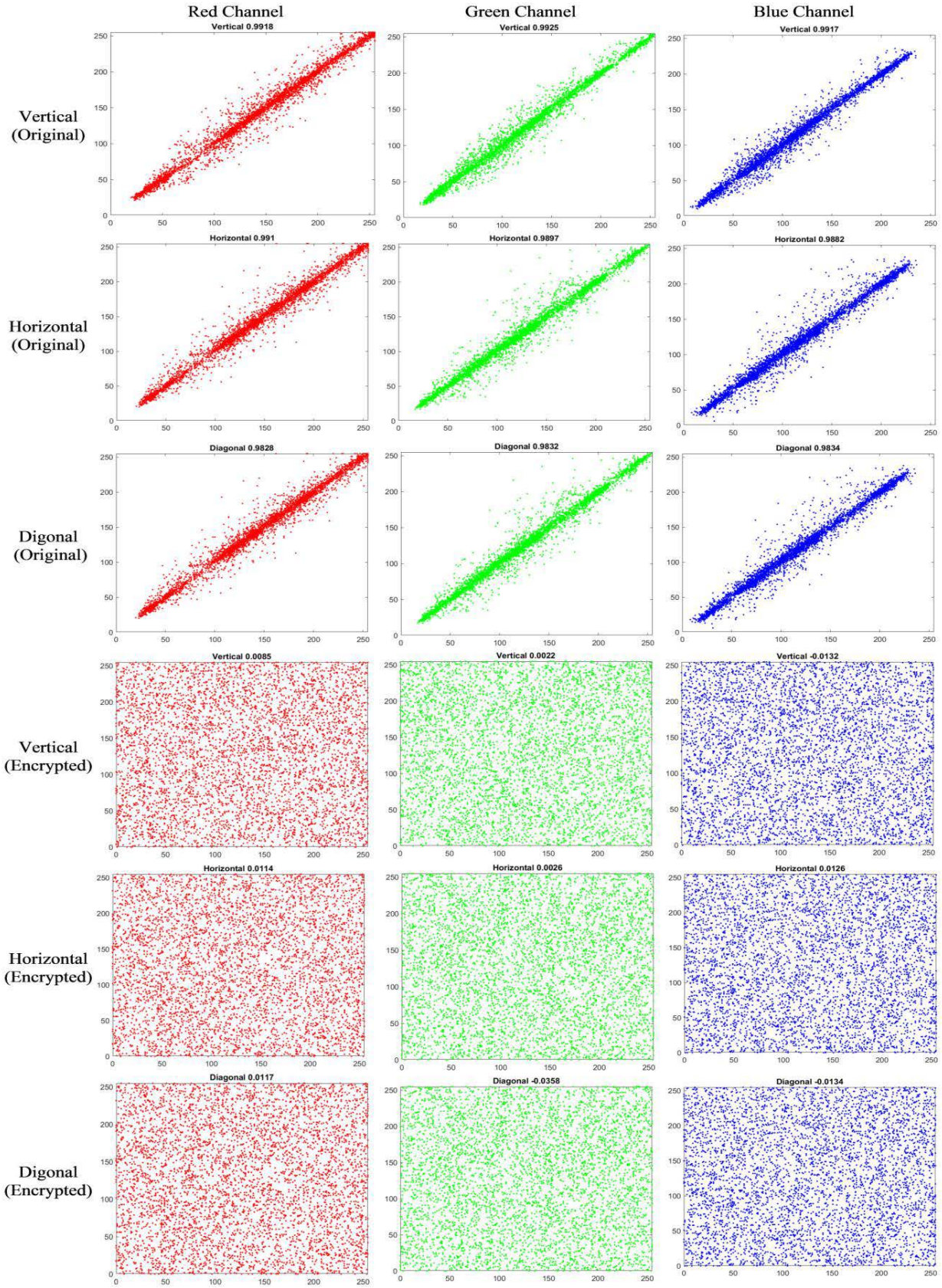


FIGURE 11. Correlation distribution of the original and encrypted right ROI.

TABLE 5. PSNR, SSIM, and FSIM results between original and corresponding encrypted ROIs.

Video Object	Channel	PSNR	SSIM	FSIM
Left ROI	R	8.1823	0.0091	0.2821
	G	8.4497	0.0079	0.2729
	B	8.4085	0.0082	0.2739
Middle ROI	R	8.3276	0.0109	0.2871
	G	8.4579	0.0096	0.2767
	B	8.4139	0.0083	0.2850
Right ROI	R	8.1955	0.0086	0.3059
	G	8.5891	0.0122	0.2933
	B	8.6787	0.0087	0.2921

TABLE 6. PSNR, SSIM, and FSIM results between original and corresponding decrypted ROIs.

Video Object	Channel	PSNR	SSIM	FSIM
Left ROI	R	Inf.	1	1
	G	Inf.	1	1
	B	Inf.	1	1
Middle ROI	R	Inf.	1	1
	G	Inf.	1	1
	B	Inf.	1	1
Right ROI	R	Inf.	1	1
	G	Inf.	1	1
	B	Inf.	1	1

of original ROI, σ_e^2 the variance of encrypted ROI, σ_{or} represents the covariance of o *and* r, and $c1$ and $c2$ are constants. The results of SSIM values between the original and encrypted ROIs in frame number 250 for the vidyo1 test video are presented in table 5. Getting lower SSIM values between the original and encrypted ROIs is recommended to prove the encryption process's efficiency. From the table, the SSIM values are small, indicating the proposed encryption process equality is high. Also, the SSIM values between the original and decrypted ROIs in frame number 250 for the vidyo1 test video are presented in table 6. It is recommended to get higher SSIM values between the original and decrypted ROIs to prove the efficiency of the decryption process. From the table, the SSIM values are equal to 1, indicating the proposed decryption process equality is high.

FSIM calculates the local similarity between the original ROI and the corresponding encrypted one. It is calculated by:

$$FSIM = \frac{\sum_{x\in\Omega} S_L(x).PC_m(x)}{\sum_{x\in\Omega} PC_m(x)}$$

where $S_L(x)$ defines the total anticipated similarity between the two ROIs, Ω denotes the spatial domain for the ROI, while $PC_m(x)$ represents the congruency phase value.

The results of FSIM values between the original and encrypted ROIs in frame number 250 for the vidyo1 test video are presented in table 5. It is recommended to get small FSIM values between the original and encrypted ROIs to prove the efficiency of the encryption process. From the table, the FSIM values are low, indicating the proposed encryption process equality is high. Also, the SSIM values between the original and decrypted ROIs in frame number 250 for the vidyo1 test video are presented in table 6. It is recommended to get higher FSIM values between the original and decrypted ROIS to prove the efficiency of the decryption process. From the table, the FSIM values are equal to 1, indicating the proposed decryption process equality is high.

TABLE 7. EDR values between the original and corresponding encrypted ROIs.

Video Object	EDR
Left ROI	0.9226
Middle ROI	0.9243
Right ROI	0.9143

G. EDGE DETECTION ANALYSIS

The proposed method should protect the information on the edges of the encrypted ROIs. The proposed method uses the edge differential ratio (EDR) metric to estimate the edge distortion. It is calculated as follows:

$$EDR = \frac{\sum_{i,j=1}^{k} \left|P(i,j) - \overline{P}(i,j)\right|}{\sum_{i,j=1}^{k} \left|P(i,j) + \overline{P}(i,j)\right|}$$

where $P(i,j)$ and $\overline{P}(i,j)$ denote the pixel values in the edges within the binary form of the original ROI and corresponding encrypted one, respectively. The value of EDR should be close to 1 to ensure the dissimilarity between the original ROI and the corresponding encrypted one. The EDR values between the original and corresponding encrypted ROIs in frame number 250 for the vidyo1 test video are presented in table 7. The values in the table are close to one; hence, the proposed method guarantees that the original and the corresponding encrypted ROIs are different. Also, the Laplacian of Gaussian edge detection for the original, corresponding encrypted, and decrypted ROIs in frame number 250 for the vidyo1 test video are displayed in figure 12. There is a big difference between the original and encrypted ROIs on the edges in the displayed results. So the proposed method can disappear the main details in the encrypted videos. Also, the edges in decrypted ROIs are similar to those in original ROIs, which ensures the efficiency of the proposed method in the decryption process.

H. KEYSPACE ANALYSIS

The keyspace used in the encryption of ROIs must be large enough to make the proposed work secure against brute-force attacks. The proposed work can resist brute-force attacks when the keyspace $\geq 2^{100}$. In this work, logistic map initial parameter X_0, logistic map parameter λ, and iterations number parameter N_0 are used to generate a secret key.

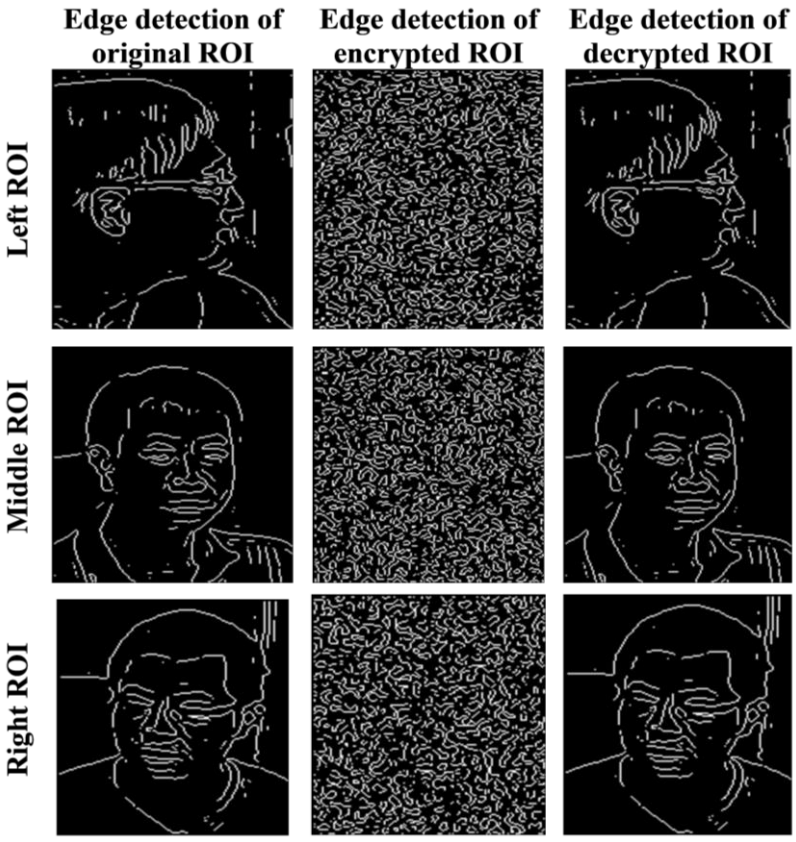


FIGURE 12. Laplacian of gaussian edge detection results for the original, encrypted, and decrypted ROIs for the vidyo1 test video.

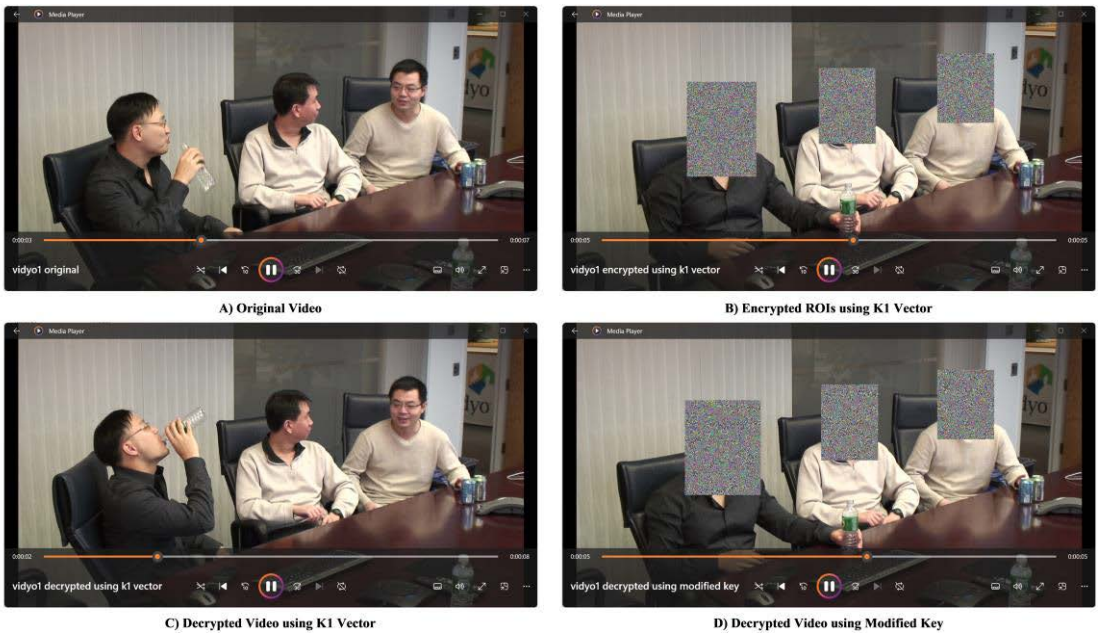


FIGURE 13. Sensitivity of the key.

The precision of b *and* λ is considered to be 10^{16}, and the precision of N_0 is considered to be 10^3. So 10^{35} is the total keyspace.

I. KEY SENSITIVITY ANALYSIS

The attacker may use a key similar to an original key to break encrypted ROIs, so the proposed encryption algorithm must

FIGURE 14. Salt and pepper noise.

FIGURE 15. Gaussian noise.

FIGURE 16. Occlusion noise.

be sensitive to the secret key. To test the key sensitivity, the vidyo1 test video ROIs are encrypted using a key vector K_1. The initial values of the logistic map used to generate K_1 are saved in a vector X_0. The encrypted video is decrypted twice. Once using the key vector K_1 and again with a key vector K_2 where the initial values of the logistic map used to generate K_2 are saved in a vector XX_0 where $XX_0 = X_0 + 10^{-10}$. Figure 13 shows the results of this experiment. This figure shows that the encryption algorithm is sensitive to the secret key.

J. NOISE ATTACKS ANALYSIS

The encrypted video may be affected by noises during video transmission through different communication channels. The security method should be insensitive to such noises to recover the original video successfully. In this experiment, different types of noises are used to prove the efficiency of the proposed method.

Salt & Pepper Noise: This noise results in black and white dots on the affected regions. The salt and pepper noise is added to the encrypted ROIs for different test videos with a variance value of 0.005. Then the noisy ROIs are decrypted. Figure 14 shows the noisy encrypted ROIs and the corresponding decrypted ones for various test videos. It is clear from the figure that the decrypted ROIs are still intelligible, despite the effect of the noise. Consequently, the proposed method is robust to salt and pepper noise.

Gaussian Noise: the limitation of the sensor during the acquisition of the video under low-light conditions may cause such type of noise. The Gaussian noise is added to the encrypted ROIs for different test videos with a variance value of 0.005. Then the noisy ROIs are decrypted. Figure 15 shows the noisy encrypted ROIs and the corresponding decrypted ones for various test videos. It is clear from the figure that the decrypted ROIs are still intelligible, despite the effect of the noise. Consequently, the proposed method is robust to Gaussian noise.

Occlusion Noise: this type of noise may occur during the transmission of the encrypted video, and part of it has been dropped or lost in this experiment. The occlusion noise is performed on the encrypted ROIs for different test videos. Then the noisy ROIs are decrypted. Figure 16 shows the noisy encrypted ROIs and the corresponding decrypted ones for various test videos. It is clear from the figure that the decrypted ROIs are still intelligible, despite the effect of the noise. Consequently, the proposed method is robust to occlusion noise.

K. ENCRYPTION TIME ANALYSIS

The efficient security method should protect the privacy of surveillance videos with low processing time. In this experiment, various test videos are used to estimate the encryption time of the proposed method. The experiment is performed multiple times, and the average values are calculated. Table 8 presents the average encryption time for various test videos. From the table, the time of the encryption process is low, revealing the proposed method's power in protecting the videos, which will be stored on the cloud for future use or sent across in a small time. Additionally, the encryption time may be further reduced to apply the technique in real-time cases if the technique is implemented in parallel.

TABLE 8. Average encryption time.

Video Name	Max ROIs per Frame	Time of Encryption (second)
salesman	1	1.33
akiyo	1	6.35
Paris	2	14.37
crew	12	34.28
vidyo4	1	50.31
vidyo1	3	42.88

V. CONCLUSION

This paper proposed a practical method for surveillance video privacy protection based on block scrambling and face detection using YOLOv3. Multiple faces from a video frame can be detected and encrypted using a key for each detected face to increase security. The proposed method is reversible for displaying the faces to an authorized person. The proposed method performance was evaluated using visual analysis, histogram analysis, information entropy analysis, correlation analysis, differential attack, PSNR, SSIM, FSIM, edge detection analysis, keyspace analysis, key sensitivity, noise attacks analysis, and encryption time analysis. The results proved that the proposed mechanism could successfully detect and protect people's faces without any leaks, and the method could withstand potential attacks.

REFERENCES

[1] K. M. Hosny, S. T. Kamal, and M. M. Darwish, "A novel color image encryption based on fractional shifted Gegenbauer moments and 2D logistic-sine map," *Vis. Comput.*, vol. 38, pp. 1–18, 2022, doi: 10.1007/S00371-021-02382-1.

[2] K. M. Hosny, S. T. Kamal, and M. M. Darwish, "Novel encryption for color images using fractional-order hyperchaotic system," *J. Ambient Intell. Humanized Comput.*, vol. 13, no. 2, pp. 973–988, Feb. 2022, doi: 10.1007/S12652-021-03675-Y.

[3] X. Li, H. Yu, H. Zhang, X. Jin, H. Sun, and J. Liu, "Video encryption based on hyperchaotic system," *Multimedia Tools Appl.*, vol. 79, nos. 33–34, pp. 23995–24011, Sep. 2020, doi: 10.1007/S11042-020-09200-1.

[4] X.-H. Song, H.-Q. Wang, S. E. Venegas-Andraca, and A. A. A. El-Latif, "Quantum video encryption based on qubit-planes controlled-XOR operations and improved logistic map," *Phys. A, Stat. Mech. Appl.*, vol. 537, Jan. 2020, Art. no. 122660, doi: 10.1016/J.PHYSA.2019.122660.

[5] A. Hafsa, M. Fradi, A. Sghaier, J. Malek, and M. Machhout, "Real-time video security system using chaos-improved advanced encryption standard (IAES)," *Multimedia Tools Appl.*, vol. 81, no. 2, pp. 2275–2298, Jan. 2022, doi: 10.1007/s11042-021-11668-4.

[6] A. Alarifi, S. Sankar, T. Altameem, K. C. Jithin, M. Amoon, and W. El-Shafai, "A novel hybrid cryptosystem for secure streaming of high efficiency H.265 compressed videos in IoT multimedia applications," *IEEE Access*, vol. 8, pp. 128548–128573, 2020, doi: 10.1109/ACCESS.2020.3008644.

[7] Q. Xue, H. Zhu, X. Ju, H. Zhu, F. Li, X. Zheng, and B. Zuo, "A video-selection-encryption privacy protection scheme based on machine learning in smart home environment," in *Proc. Int. Conf. Artif. Intell. Commun. Netw.*, vol. 287, 2019, pp. 65–76, doi: 10.1007/978-3-030-22971-9_6.

[8] A. Shifa, M. N. Asghar, M. Fleury, N. Kanwal, M. S. Ansari, B. Lee, M. Herbst, and Y. Qiao, "MuLViS: Multi-level encryption based security system for surveillance videos," *IEEE Access*, vol. 8, pp. 177131–177155, 2020, doi: 10.1109/ACCESS.2020.3024926.

[9] S. Fahmeeda Sultana and D. C. Shubhangi, "Privacy preservation of videos using neutrosophic logic based selective encryption," *ICTACT J. Image Video Process.*, vol. 9, no. 3, pp. 1960–1965, Feb. 2019, doi: 10.21917/IJIVP.2019.0278.

[10] M. A. Taha, W. Hamidouche, N. Sidaty, M. Viitanen, J. Vanne, S. El Assad, and O. Deforges, "Privacy protection in real time HEVC standard using chaotic system," *Cryptography*, vol. 4, no. 2, p. 18, Jun. 2020, doi: 10.3390/CRYPTOGRAPHY4020018.

[11] A. Shifa, M. B. Imtiaz, M. N. Asghar, and M. Fleury, "Skin detection and lightweight encryption for privacy protection in real-time surveillance applications," *Image Vis. Comput.*, vol. 94, Feb. 2020, Art. no. 103859, doi: 10.1016/J.IMAVIS.2019.103859.

[12] M. Farajallah, W. Hamidouche, O. Deforges, and S. E. Assad, "ROI encryption for the HEVC coded video contents," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 3096–3100, doi: 10.1109/ICIP.2015.7351373.

[13] Z. Shahid, M. Chaumont, and W. Puech, "Fast protection of H.264/AVC by selective encryption of CAVLC and CABAC for I and P frames," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 5, pp. 565–576, May 2011, doi: 10.1109/TCSVT.2011.2129090.

[14] Y. Zhao, L. Zhuo, M. Niansheng, J. Zhang, and X. Li, "An object-based unequal encryption method for H.264 compressed surveillance videos," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Aug. 2012, pp. 419–424, doi: 10.1109/ICSPCC.2012.6335618.

[15] H. Li, T. Xiezhang, C. Yang, L. Deng, and P. Yi, "Secure video surveillance framework in smart city," *Sensors*, vol. 21, no. 13, p. 4419, Jun. 2021, doi: 10.3390/S21134419.

[16] D. Lee and N. Park, "Blockchain-based privacy-preserving multimedia intelligent video surveillance using secure Merkle tree," *Multimedia Tools Appl.*, vol. 80, nos. 26–27, pp. 34517–34534, Nov. 2021, doi: 10.1007/S11042-020-08776-y.

[17] H. Du, L. Chen, J. Qian, J. Hou, T. Jung, and X.-Y. Li, "PatronuS: A system for privacy-preserving cloud video surveillance," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 6, pp. 1252–1261, Jun. 2020, doi: 10.1109/JSAC.2020.2986665.

[18] K.-Y. Chu, Y.-H. Kuo, and W. H. Hsu, "Real-time privacy-preserving moving object detection in the cloud," in *Proc. 21st ACM Int. Conf. Multimedia*, Oct. 2013, pp. 597–600, doi: 10.1145/2502081.2502157.

[19] E. M. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 2, pp. 232–243, Feb. 2005, doi: 10.1109/TKDE.2005.32.

[20] P. Korshunov and T. Ebrahimi, "Using warping for privacy protection in video surveillance," in *Proc. 18th Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2013, pp. 1–6, doi: 10.1109/ICDSP.2013.6622791.

[21] X. Ma, L. T. Yang, Y. Xiang, W. K. Zeng, D. Zou, and H. Jin, "Fully reversible privacy region protection for cloud video surveillance," *IEEE Trans. Cloud Comput.*, vol. 5, no. 3, pp. 510–522, Jul. 2017, doi: 10.1109/TCC.2015.2469651.

[22] L. Du, W. Zhang, H. Fu, W. Ren, and X. Zhang, "An efficient privacy protection scheme for data security in video surveillance," *J. Vis. Commun. Image Represent.*, vol. 59, pp. 347–362, Feb. 2019, doi: 10.1016/J.JVCIR.2019.01.027.

[23] S. M. Mizanur Rahman, M. A. Hossain, H. Mouftah, A. El Saddik, and E. Okamoto, "Chaos-cryptography-based privacy preservation technique for video surveillance," *Multimedia Syst.*, vol. 18, no. 2, pp. 145–155, Mar. 2012, doi: 10.1007/S00530-011-0246-9.

[24] X. Zhang, S.-H. Seo, and C. Wang, "A lightweight encryption method for privacy protection in surveillance videos," *IEEE Access*, vol. 6, pp. 18074–18087, 2018, doi: 10.1109/ACCESS.2018.2820724.

[25] W. Wen, Y. Zhang, Z. Fang, and J.-X. Chen, "Infrared target-based selective encryption by chaotic maps," *Opt. Commun.*, vol. 341, pp. 131–139, Apr. 2015, doi: 10.1016/J.OPTCOM.2014.12.026.

[26] A. Kanso and M. Ghebleh, "An efficient and robust image encryption scheme for medical applications," *Commun. Nonlinear Sci. Numer. Simulat.*, vol. 24, nos. 1–3, pp. 98–116, Jul. 2015, doi: 10.1016/J.CNSNS.2014.12.005.

[27] D. Xiao, Q. Fu, T. Xiang, and Y. Zhang, "Chaotic image encryption of regions of interest," *Int. J. Bifurcation Chaos*, vol. 26, no. 11, Oct. 2016, Art. no. 1650193, doi: 10.1142/S0218127416501935.

[28] Y. Liu, J. Zhang, D. Han, P. Wu, Y. Sun, and Y. S. Moon, "A multidimensional chaotic image encryption algorithm based on the region of interest," *Multimedia Tools Appl.*, vol. 79, nos. 25–26, pp. 17669–17705, Jul. 2020, doi: 10.1007/S11042-020-08645-8.

[29] H.-W. Xue, J. Du, S.-L. Li, and W.-J. Ma, "Region of interest encryption for color images based on a hyperchaotic system with three positive Lyapunov exponets," *Opt. Laser Technol.*, vol. 106, pp. 506–516, Oct. 2018, doi: 10.1016/J.OPTLASTEC.2018.04.030.

[30] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[31] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[32] *GitHub Sthanhng/Yoloface: Deep Learning-Based Face Detection Using the YOLOv3 Algorithm*. Accessed: May 21, 2022. [Online]. Available: https://github.com/sthanhng/yoloface

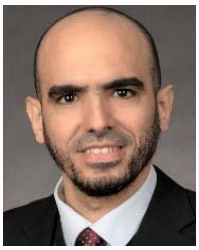
[33] *WIDER FACE: A Face Detection Benchmark*. Accessed: May 21, 2022. [Online]. Available: http://shuoyang1213.me/WIDERFACE/

[34] *Xiph.Org: Derf's Test Media Collection*. Accessed: Jun. 26, 2022. [Online]. Available: https://media.xiph.org/video/derf/

KHALID M. HOSNY (Senior Member, IEEE) was born in Zagazig, Egypt, in 1966. He received the B.Sc., M.Sc., and Ph.D. degrees from Zagazig University, Egypt, in 1988, 1994, and 2000, respectively. He is currently a Professor in information technology at the Faculty of Computers and Informatics, Zagazig University. From 1997 to 1999, he was a Visiting Scholar at the University of Michigan, Ann Arbor, and the University of Cincinnati, Cincinnati, USA. He is a Senior Member of ACM. He has published four edited books and more than 100 papers in international journals. His research interests include image processing, pattern recognition, multimedia, and computer vision. He is an editor and a scientific reviewer for more than 50 international journals. He is among the top 2% of scientists according to Stanford Report 2020 and 2021.

MOHAMED A. ZAKI was born in Abu-Kabir, Egypt, in 1995. He received the B.Sc. degree in information technology from the Faculty of Computers and Informatics, Zagazig University, Egypt, in 2017. He is currently working as a Teaching Assistant with the Information Technology Department, Faculty of Computers and Informatics, Zagazig University.

HANAA M. HAMZA received the B.Sc., M.Sc., and Ph.D. degrees from the Faculty of Computers and Informatics, Zagazig University, Egypt. She is currently a Lecturer at the Faculty of Computers and Informatics, Zagazig University.



MOSTAFA M. FOUDA (Senior Member, IEEE) received the Ph.D. degree in information sciences from Tohoku University, Japan, in 2011. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Idaho State University, ID, USA. He is also an Associate Professor with Benha University, Egypt. He has worked as an Assistant Professor at Tohoku University. He was a Postdoctoral Research Associate with Tennessee Technological University, TN, USA. He has published more than 60 papers in prestigious peer-reviewed journals and conferences. His research interests include cybersecurity, communication networks, wireless mobile communications, smart healthcare, smart grids, AI, blockchain, and the IoT. He received the prestigious first place award during his graduation from the Faculty of Engineering, Benha University, Shoubra, Egypt, in 2002. He has served as the Symposium Chair/the Track Chair of the IEEE VTC2021-Fall Conference. He has also served as the Workshop Chair, the Session Chair, a Technical Program Committee (TPC) Member, and a Designated Reviewer in leading international conferences, such as IEEE GLOBECOM, ICC, PIMRC, ICCVE, IWCMC, and 5G World Forum. He also served as a Guest Editor for some special issues of several top-ranked journals, such as IEEE Wireless Communications and *IEEE Internet of Things Magazine*. He also serves as a Referee for some renowned IEEE journals and magazines, such as IEEE Communications Surveys and Tutorials, IEEE Wireless Communications, IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on Smart Grid, IEEE Access, IEEE Transactions on Network and Service Management, IEEE Transactions on Emerging Topics in Computing, and *IEEE Network*. He is an Editor of IEEE Transactions on Vehicular Technology (TVT) and an Associate Editor of IEEE Access.

NABIL A. LASHIN received the B.Sc. degree in communication and electronics engineering from Zagazig University, Egypt, in 1993, the M.Sc. degree in communication and electronics engineering from Cairo University, in 1999, and the Ph.D. degree in electrical engineering and computer science from the Technical University of Berlin, Germany, in 2005. He is currently an Assistant Professor in information technology with Zagazig University.