ATLANTIS PRESS

Advances in Computer Science Research, volume 80

International Conference on Computer Science, Electronics and Communication Engineering (CSECE 2018)

A Novel Video Encryption Method Based on Faster R-CNN

Lijuan Duan[1,2,3,*], Dongkui Zhang[1,2,3], Fan Xu[1,2,3] and Guoqin Cui[4]

[1]Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[2]Beijing Key Laboratory of Trusted Computing, Beijing 100124, China
[3]National Engineering Laboratory for Critical Technologies of Information Security Classified Protection, Beijing 100124
[4]State Key Laboratory of Digital Multi-media Chip Technology, Vimicro Corporation, Beijing, 100191, China
[*]Corresponding author

***Abstract*—In order to improve the generalization performance of video encryption and reduce the amount of data in video en-cryption, this paper proposes a video encryption on regions of interest (ROI) method based on Faster R-CNN by combining machine learning with information security. The method trains a Faster R-CNN model using the ROI dataset firstly, and then uses the model to extract ROI in the video. Different encryption algorithms are used to encrypt ROI and non-ROI in the video respectively. To overcome the shortcomings of encryption algorithms that can only be used for a specific coded video, a special video encryption method is proposed to encrypt the video with different video coding structure and has better generalization performance. Compared with the encryption method in the video coding process, this method considers the content information of the video fully and has better performance. It can be concluded through experiments that the encryption method in this paper has the characteristics of higher security and less calculation.**

Keywords—video encryption; faster R-CNN; the ROI of video

I. INTRODUCTION

In the age of information, with the rapid development of computer and Internet technologies, people can post messages and obtain information on the Internet anytime, anywhere. Large amounts of data are transmitted through the network and more than 50 million servers work on the network over the world. Zettabytes of data are produced each year, which contain a large amount of video data [5][7]. Video in the national defense, education, monitoring, entertainment and other fields have been widely used, so data security on the internet cannot be ignored. Unauthorized access and the openness of the network lead to more and more serious data security problems. Especially, video security issues have become more serious and aroused more attentions [10]. Video encryption is an effective data encryption strategy to improve video security. Video encryption protects the original video information and improves the security of video information. Researchers have done a lot of research on video encryption and put forward a lot of video encryption methods. Video encryption methods are mainly divided into complete encryption and partial encryption algorithm.

Video complete encryption algorithm is to encrypt the whole video data with encryption algorithm in order to achieve the purpose of protecting video information. Think of video data as a series of data streams and then encrypt the video data streams with traditional encryption algorithms. However, traditional encryption algorithms such as AES and RSA can achieve good encryption effect on text data and unformatted data [13]. In [7], the author proposed a video encryption algorithm based on RSA. Because of the large quantity and strong correlation of video data, the information redundancy leads to too high complexity of video encryption and too long time consuming to meet real time encryption Request. Using image encryption algorithm to encrypt the video fully, the video is divided into a series of video frames, and then use the image encryption algorithm to encrypt every frame of video. It can reduce the amount of data to be encrypted in the video, but does not take the information redundancy between video frames into account, which results in higher encryption complexity [10]. Video complete encryption method does not consider the video data format and ignores the correlation between video frames, resulting in higher video encryption complexity, large amount of data to be encrypted and long time consuming. So it cannot meet the demands of real time encryption.

The video partial encryption algorithm encrypts the video in the coding process [6][11][13]. Encrypting video in the process of video encoding can reduce the amount of data to be encrypted and reduce the complexity of encryption. In [5], the author combines the stream cipher and the video cipher to encrypt the DCT transform coefficients. This method only encrypts the DCT transform coefficients, so it is not safe enough. In [10], the author encrypts motion vector difference (MVD), luma residual coefficients and chroma residual coefficients in the process of HEVC encoding. It can improve encrypted security to some extent. In [15], in order to protect the video data, the author proposed a video encryption method based on logistic chaos mapping, which encrypts the motion vector (MVD) and DCT variation coefficient the chaotic mapping in the process of HEVC encoding. In [13], the author proposed a video en-cryption method based on RGB three channel MPEG encoding, which achieved wonderful encryption effect on MPEG videos, but it has a weak generalization ability. In [14], Mamoona et al. proposed a video encryption method based on CABAC entropy coding and it has some limitations and can only encrypt H.264 and HEVC encoded video. The existing video encryption methods combined with video coding can reduce the complexity of encryption and increase the speed of encryption. However, they also have some limitations, and can only encrypt video in a specific encoding format, and have a weak generalization ability.

In order to overcome the shortcomings of existing video encryption algorithms, this paper proposed a region based video encryption method, which uses the Faster R-CNN to extract ROI in video frames. It can help to reduce the amount of video encrypted data. The video encryption method can encode a variety of video encoding and has a high generalization ability.

This article is organized as follows: Section 2 introduces the Faster-R-CNN network structure, training of ROI model, and extraction of ROI. Section 3 describes the encryption algorithm and the steps to encrypt the ROI. Section 4 is mainly to analyze the effect of video encryption through experiments.

II. PROPOSED METHOD

Video data often contains a lot of information, but people tend to focus only on some of the information they are inter-ested in and ignore some of the background information. In order to meet this demand, a video encryption method based on Faster R-CNN and ROI is proposed in this paper. Faster R-CNN is a multi-layer convolution neural network that achieves good results in the field of object detection and recognition. In this method, the Faster R-CNN is used to extract ROI in the video. The framework of video encryption system is shown in Fig. I. It is mainly divided into three parts: the training of the ROI model, the extraction of the ROI in the video and the encryption of the ROI. The detailed ideas of this method is as follows: Firstly, a ROI model is trained with the Faster R-CNN by using the data set of the ROI, and then the ROI in the video is extracted with the trained ROI model. Finally, the ROI is encrypted.

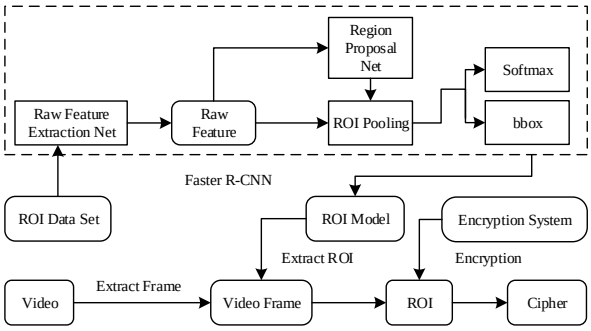

FIGURE I. THE FRAMEWORK OF VIDEO ENCRYPTION SYSTEM

A. Extraction of ROI

This section mainly introduces the extraction of the ROI in a video. In this paper, we used Faster R-CNN to extract the ROI in a video. The Faster R-CNN combines convolutional neural networks and machine learning [2]. It uses the region proposal network instead of the selective search algorithm to generate a suggestion window, and the region proposal network and the target detection network share the convolution layer features [3].

The structure of Faster R-CNN is shown in Figure II. Faster R-CNN consists of the convolutional neural network, ROI Pooling, Softmax, Bounding box regression and others. Convolution neural network is mainly used for the extraction of image feature[2][3]. In this paper, VGG16 is used to extract image features. Compared with other convolutional neural networks, VGG16 has simpler structure and superior performance. The Faster R-CNN with the region proposal network has better performance than the Faster R-CNN and R-CNN with the selective search algorithm in generation of the candidate box. The ROI pooling layer mainly performs the pooling operation on the candidate boxes generated by the RPN and generates a fixed-size feature map for each ROI.

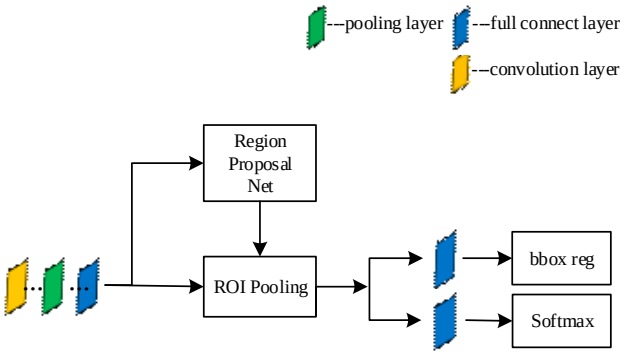

FIGURE II. THE STRUCTURE OF FASTER R-CNN

The dataset training the Faster R-CNN to generate the ROI model is from the WIDER FACE. More than 20,000 images were used to train and generate a ROI model on the GPU spending more than 10 hours. The effect of extraction of the ROI is shown in Fig. III. The model extracted the ROI in a) and extracted 5 ROI in b) accurately. It can be concluded that the ROI model in this paper is not only applicable to images containing a single ROI, but also to images containing multiple ROI.

a. Single ROI b. Multiple ROI

FIGURE III. EXTRAC ROI

B. Encryption Algorithms

After the ROI of the video is extracted using the ROI model trained, the videos are divided into the ROI and the non-ROI. In this paper, different encryption algorithms are used respectively to encrypt the non-ROI and the ROI in the video. The non-ROI of the video is encrypted by $GF(17)$ domain diffusion encryption algorithm based on plaintext [8], while the ROI of the video is encrypted by the encryption algorithm based on hyperchaos system and pixel information, which is more secure and complex. Using different encryption algorithms according to different contents of the video can improve the security of encryption and increase the difficulty of cracking and speed up encryption.

The finite field, also known as the galois field, is a field that contains only a limited number of elements. If *GF(p)* is a finite field, where p is a prime number, the addition is as shown in Equation 1, and the multiplication is as shown in Equation 2,

where x and y are the elements in the finite field. For the multiplication in $GF(p)$, only when p is a prime number, have the other elements in the finite field except 0 an inverse multiplication.

$$(x + y) mod\ p \tag{1}$$

$$(x \cdot y) mod\ p \tag{2}$$

TABLE I. MULTIPLICATION OPERATION OF GF(17) [8]

x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	0	2	4	6	8	10	12	14	16	1	3	5	7	9	11	13	15
3	0	3	6	9	12	15	1	4	7	10	13	16	2	5	8	11	14
4	0	4	8	12	16	3	7	11	15	2	6	10	14	1	5	9	13
5	0	5	10	15	3	8	13	1	6	11	16	4	9	14	2	7	12
6	0	6	12	1	7	13	2	8	14	3	9	15	4	10	16	5	11
7	0	7	14	4	11	1	8	15	5	12	2	9	16	6	13	3	10
8	0	8	16	7	15	6	14	5	13	4	12	3	11	2	10	1	9
9	0	9	1	10	2	11	3	12	4	13	5	14	6	15	7	16	8
10	0	10	3	13	6	16	9	2	12	5	15	8	1	11	4	14	7
11	0	11	5	16	10	4	15	9	3	14	8	2	13	7	1	12	6
12	0	12	7	2	14	9	4	16	11	6	1	13	8	3	15	10	5
13	0	13	9	5	1	14	10	6	2	15	11	7	3	16	12	8	4
14	0	14	11	8	5	2	16	13	10	7	4	1	15	12	9	6	3
15	0	15	13	11	9	7	5	3	1	16	14	12	10	8	6	4	2
16	0	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1

In this paper,$GF(17)$ domain multiplication is used as operation of diffusion encryption [8]. In order to prevent 0 element from appearing in multiplication in $GF(17)$ domain during encryption, the multiplication Operation of $GF(17)$ is shown in Table 1. For the first two pixels, the data is encrypted with equation 3, where and the subscript i denotes the position of the data and the subscript h denotes the upper four bits of the data and the subscript l denotes the lower four bits of the data and the operator $\times$ is a multiplication on the $GF(17)$ field . The encryption key k is generated with the overall information of the image and the data information of the pixel. Firstly, use the upper four bits of the encryption key to encrypt the upper four bits of the data. Secondly, use the lower four bits of the encryption key to encrypt the lower four bits of the data. Finally, generate the ultima encrypted ciphertext with the previous results.

$$\begin{cases} c_{i,h} = p_{i,h} \times k_{i,h} \\ c_{i,l} = p_{i,l} \times k_{i,l} \\ c_i = c_{i,h} \cdot 16 + c_{i,l} \quad i < 3 \end{cases} \tag{3}$$

When i ≥ 3, the data is encrypted with equation 4, and the upper four bits and the lower four bits of data are encrypted respectively. Then the first two ciphertexts adjacent to the data are diffused into the encrypted ciphertext of the data. Finally, the final ciphertext is generated with the upper four digit ciphertext and the lower four digit ciphertext.

$$\begin{cases} c_{i,h} = p_{i,h} \times k_{i,h} \times c_{i-1,h} \times c_{i-2,h} \\ c_{i,l} = p_{i,l} \times k_{i,l} \times c_{i-1,l} \times c_{i-2,l} \\ c_i = c_{i,h} \cdot 16 + c_{i,l} \quad i \geq 3 \end{cases} \tag{4}$$

The detailed procedure of decryption in this paper is as follows: Firstly, decrypt the plaintext of the first two data and then decrypt the remaining data using equation (5).

$$\begin{cases} p_{i,h} = c_{i,h} \div k_{i,h} \div p_{i-1,h} \div p_{i-2,h} \\ p_{i,l} = c_{i,l} \div k_{i,l} \div p_{i-1,l} \div p_{i-2,l} \\ p_i = p_{i,h} \cdot 16 + p_{i,l} \quad i < 3 \end{cases} \tag{5}$$

III. RESULT

This section introduces the experimental results and analysis of the results. The experimental program ran on the Tesla K40c GPU platform. In order to analyze the security and effect of the video encryption method proposed in this paper, we use three groups of videos to experiment. It can be seen from the experimental results that the ROI model in this paper can accurately extract the ROI in the video. Encrypting the ROI and non-ROI of the video with different encryption algorithms can reduce the amount of data encrypted and reduce the complexity of encryption.

A. *The Extraction of ROI*

The ROI in the three groups of videos are extracted with the trained ROI model. The experimental results are shown in the Fig. IV. It is the results of the extraction of the ROI in the first, 50th, 100th and 150th frames of the video. And the ROI in the three groups of video were extracted accurately.

FIGURE IV. RESULT OF EXTRACTION OF ROI

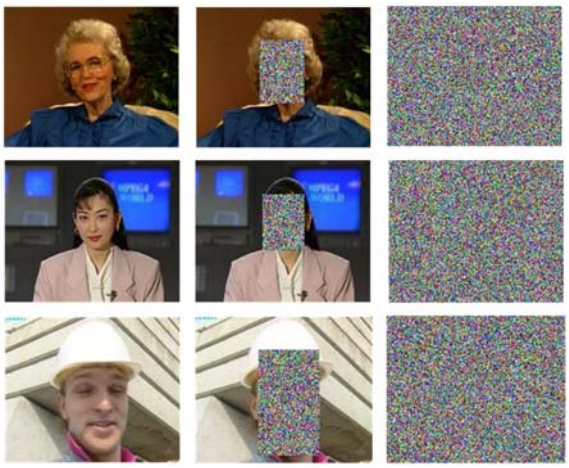

FIGURE V. ENCRYPTION OF THE VIDEO

The results of video encryption are shown in Fig. V. The ROI and non-ROI of the video are encrypted with different encryption algorithms. Both the ciphertext of the ROI and the non-ROI are noise-like and any information about images cannot been read from the ciphertext.

Figure VI shows the proportion of the ROI in every video frame. It can be seen that the ROI in every frame accounts for about 20%. So only encrypting the ROI in the video can almost reduce the amount of encrypted data by 80%, which can help to reduce the complexity of video encryption.

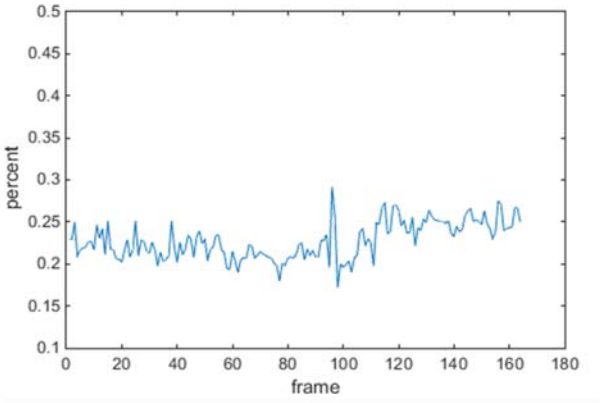

FIGURE VI. THE PERCENT OF ROI

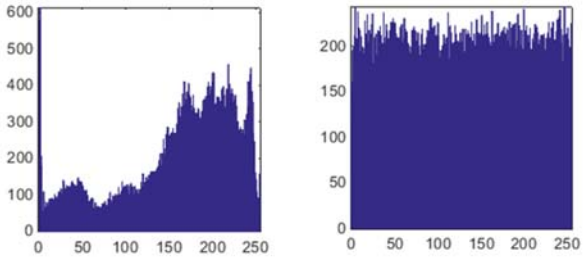

FIGURE VII. HISTOGRAMS OF THE PLAINTEXT AND THE CIPHERTEXT

B. Histogram Analysis

The histogram of the plaintext and the ciphertext is shown in Figure VII. It can see from the histogram of the plaintext that the pixel values are unevenly distributed, so some of image information can be read from it. But from the histogram of the ciphertext, it can be seen that the pixel values are evenly distributed, so any useful information cannot been read from it.

C. Correlation Analysis

The results of the correlation between adjacent pixels are shown in Figure VIII. It can be seen that there is a strong correlation between the adjacent pixels in the plaintext, while there is no correlation between adjacent pixels in the ciphertext because of the evenly distribution of pixel grey value.

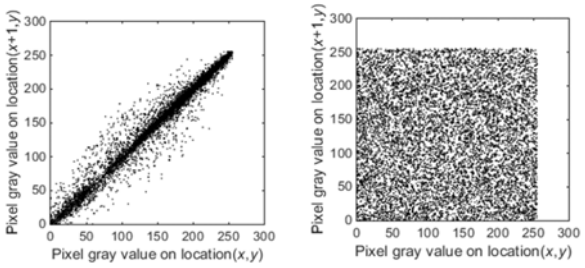

FIGURE VIII. CORRELATION BETWEEN ADJACENT PIXELS

The correlation coefficient is shown in Table II, the plaintext has a large correlation coefficient in three directions and the adjacent pixels have strong correlation. The correlation coefficient of ciphertexts in three directions is close to 0, which indicates that ciphertexts are not related between adjacent pixels.

TABLE II. CORRELATION COEFFICIENT

Image	Horizontal	Vertical	Diagonal
Plain	0.9862	0.9782	0.9623
Cipher	-0.0016	-0.0034	-0.0129

D. Key Sensitivity Analysis

This article make a small change to the key and then test the key's sensitivity. Table III shows the key sensitivity results. It can be seen that the NPCR and UACI of the three groups of test data are close to the theoretical value, so the encryption method has high key sensitivity.

TABLE III. KEY SENSITIVITY ANALYSIS

	grandma	akiyo	foreman	Theoretical
NPCR	99.6100	99.6452	99.6927	99.6094
UACI	33.4000	33.3395	33.3135	33.4635

E. Differential Attack Analysis

Differential attack is also called the choice of plaintext attack, the differential attack analysis results in Table IV, the three sets of data differential analysis of NPCR and UACI values are close to the theoretical value, indicating that the encryption method has high security, which can resist Differential attack.

TABLE IV. DIFFERENTIAL ATTACK ANALYSIS

	grandma	akiyo	foreman	Theoretical
NPCR	99.6633	99.7339	99.6756	99.6094
UACI	33.4409	33.7439	33.4504	33.4635

F. Encryption time Analysis

The analysis results of the encryption time are shown in Table V, T1 is the time of extracting the video region of interest. T2 is the time of encrypting the video region of interest. The region of interest in the video is encrypted by a fast scrambling encryption algorithm. T3 is the time of encrypting the non-ROI and T4 is the time of encrypting a video frame. It can be seen from Table 4 that the encryption method in this paper is less time consuming and faster than the video frame complete encryption.

TABLE V. DIFFERENTIAL ATTACK ANALYSIS

	T1(s)	**T2(s)**	**T3(s)**	**Our(s)**	**T4(s)**
grandma	0.2307	0.0470	0.0150	0.2927	0.3750
akiyo	0.2374	0.0460	0.0160	0.2994	0.6380
foreman	0.2267	0.1530	0.0050	0.3847	0.6500

IV. CONCLUSION

This paper presented a video encryption method on ROI of a video based on Faster R-CNN. We combined machine learning with video encryption and trained an extraction of ROI model with the dataset of ROI and Faster R-CNN. The ROI in the video was extracted effectively with the trained ROI model, then the ROI and non-ROI in the video were encrypted with different encryption algorithms. The method proposed in this paper can reduce the complexity of encryption and improve the encryption speed. And it can also deal with a variety of video encoding format, so it has a high generalization performance. And the encrypted video has the advantages of low correlation, high encryption sensitivity and high security.

ACKNOWLEDGMENT

This paper is partly supported by National Natural Science Foundation of China [grant numbers 61572004], the Beijing Municipal Natural Science Foundation [grant numbers 4152005], the Science and Technology Program of Tianjin [grant number 15YFXQGX0050], and the Science and Technology Planning Project of Qinghai Province [grant number 2016-ZJ-Y04].

REFERENCES

[1] X.Y. Wang, K. Guo. A new image alternate encryption method based on chaotic map. Nonlinear Dyn,2014(76),pp.1943–1950.

[2] J. Dai, K. He, and J. Sun. Convolutional feature masking for joint object and stuff segmentation.In CVPR,2015.

[3] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov. Scalable object detection using deep neural networks.In CVPR,2014.

[4] Siu-Kei Bing Zeng,Shuyuan Zhu. Perceptual encryption of H.264 videos:embedding sign-Flips into the integer-based transforms. IEE Trans. Inf. Forensics Secur,2014(9),9,pp 309-320.

[5] Saeed Bahrami,Majid Naderi, Encryption of Video Main Frames in the Field of DCT Transform Using A5/1 and W7 Stream Encryption Algorithms, Arabian Journal for Science and Engineering,2014(39),pp 4077-4088.

[6] S Mumthas,A Lijiya, Transform Domain Video Steganography Using RSA, Random DNA Encryption and Huffman Encoding, Procedia Computer Science,2017(115),pp 660-666.

[7] M.Garcia-Martinez, E.Camposs-Canton. Pseudo-random bit generator based on multi-modal maps. Nonlinear Dyn,2015(82),pp 2119-2131.

[8] Zhang Yong. Chaotic digital image encryption[M]. BeiJing: tsinghua university press,2016.

[9] Jovanovic B,Gajin S. An efficient mechanism of cryptographic synchronization within selectively encrypted H.265/HEVC video stream. Multimed Tools Appl,2017,pp 1-17.

[10] A.Alfalou,C.Brosseau,N.Abdallah, Simultaneous compression and encryption of color video images. Optics communications,2015(338),pp 371-379.

[11] Juan Chen,Fei Peng,Min Long, A Perceptual Encryption Scheme for HEVC Video with Lossless Compression, International Conference on Cloud Computing and Security,2017,pp 396-407.

[12] Xiaoyu Li,Chen Tang,Xinjun Zhu. Image/video encryption using single shot digital holography. Optics communications,2015(342),pp 218-223.

[13] Quist-Aphetsi Kester,Laurent NanaAnca,Christine Pascu, A Cryptographic Encryption Technique of MPEG Digital Video Images Based on RGB Layer Pixel Values, Advances in Intelligent Systems and Computing,2016,355.

[14] Mamoona N.Asghar,Rukhsana Kousar, Transparent encryption with scalable video communication: Lower-latency, CABAC-based schemes, Journal of Visual Communication and Image Reprsentation,2017(45),pp 122-136.

[15] Sallam, A.I., El-Rabaie, ES.M. Faragallah, Efficient HEVC selective stream encryption using chaotic logistic map,O.S. Multimedia Systems,2017,pp 1-19.

[16] Tew Yiqi,Wong KokSheik,Phan Raphael C. Region-of-interest encryption in HEVC compressed video.IEEE international conference on sonsumer electronics,2016.

[17] Xiao Di,Fu Qingqing,Xiang Tao. Chaotic image encryption of ROI. International journal of bifurcation and chaos,2016(26).

[18] Liu Yuling,Qu Xinxin,Guojiang. A ROI-based reversible data hiding scheme in encrypted medical images. Journal of visual communication and image representation,2016(39),pp 51-57.