Journal of Computers Vol. 34 No. 2, April 2023, pp. 175-189
doi: 10.53106/199115992023043402013

Research on Video Encryption Technology Based on Cross Coupled Map Lattices System

Hao-Qiang Xu*, Jian-Dong Liu

College of Information Engineering, Beijing Institute of Petrochemical Technology,
Beijing, China
azdhr1632@163.com

Received 23 May 2022; Revised 11 September 2022; Accepted 28 October 2022

Abstract. The traditional video encryption algorithm only encrypts video images, which has the problems of an extended time-consuming algorithm and poor format retention. To improve the efficiency of video encryption, this paper proposes a multi-link selective video encryption algorithm based on the Cross Coupled Map Lattices system by combining H.264/AVC video coding structure. The algorithm reduces the amount of encrypted data while ensuring encryption security to satisfy the needs of video encryption security and real-time performance. The encryption algorithm's security and visual encryption effect are analyzed subjectively and objectively. The experimental results show that the encryption scheme has an excellent visual encryption effect and strong attack resistance, the encryption time consumption is low, and the video format remains unchanged. It can be applied to real-time video encryption occasions such as video conferences.

Keywords: video encryption, H.264/AVC, video coding, Cross Coupled Map Lattices system

1 Introduction

In recent years, with the rapid development of the Internet and multimedia technology, as well as the decreasing cost of hardware for video recording, video data has gradually replaced images, text, and other traditional multimedia data as the most widely used digital information in network communications. It is because of the popularity of video data applications and the economic value of the information carried, which inevitably causes the phenomena of illegal access, malicious tampering, information interception, data theft, and illegal copying of video information to increase. Therefore, the development and design of video encryption algorithms have excellent application value and research prospects.

The chaos model has excellent cryptographic properties such as determinism, pseudo-randomness, and initial value sensitivity and has been widely used in traditional multimedia encryption, such as image and text. In recent years, it has also made significant development in video encryption. Wei [1] proposed a combined video encryption algorithm for multiple image frames based on CNN hyper-chaotic system and logistic map. Yang [2] proposed a video stream encryption scheme based on a cat map improved logistic chaos system combined with the group encryption algorithm. Zang [3] proposed a new three-dimensional discrete chaotic system, using the system-generated three-dimensional chaotic matrix generated by the system that can be used to achieve direct encryption of video frames. Although the above image video encryption methods have enhanced the efficiency of video encryption through multi-image frame combination techniques, group encryption, and multi-dimensional matrix operations, they have not fully considered the temporal and spatial correlation between video frames, that is, the pixel correlation and redundant information of video data. The presence of pixel correlation and redundant information is a prerequisite for video information compression coding. Even though the above algorithms enhance the encryption efficiency of frame contents, the compression and encoding process is significantly lengthened during the compressed transmission of data due to the destruction of pixel correlation and redundant information, which in turn prevents the real-time encrypted transmission of video information.

The selective video encryption algorithm is a class of video encryption techniques combined with video coding, and the encryption object is the critical syntax element in the video coding process. The algorithm uses indirect encryption to affect the expression of pixel points in the decoded video sequence, which has the characteristics of high encryption efficiency and good real-time performance. It has become a new research hotspot in

* Corresponding Author

the field of video encryption. Ahn [4] used a one-dimensional logistic map as the encryption key to encrypt the intra-frame prediction patterns of video coding, improving video encryption and coding efficiency. However, one-dimensional logistic mapping, as a low-dimensional mapping, has a small scrambling space. The single encrypted intra-frame prediction pattern approach also lacks adequate protection for P and B frames in videos. Rohit [5] used two one-dimensional logistic maps cross-coupled with each other as a pseudo-random series generator and encrypted the motion vector according to the congruence rule, which thoroughly corrupted the contour information of the video object. The cross-coupling enhances the complexity of chaotic sequences, but only the double mapping iterates over each other, making it weak against differential attacks. Encrypting only motion vectors also cannot cope with special attacks such as motion vector recovery (MVR), resulting in a lack of security in the algorithm. Karmakar [6] designed a class of sparsity tracking algorithms containing 5D hyper-chaotic DNA encryption modules to replace the DCT transform in video coding, misrepresenting video pixel point information. However, the higher system complexity of 5D hyperchaotic DNA encryption prolongs the encoding time. Encrypting the DCT degrades the compression performance of the video, and the encoded video stream also suffers from format retention problems.

The above literature shows that the comprehensive performance of single-link selective encryption is poor and cannot meet the system's security requirements. Furthermore, the design of the pseudo-random sequence generator has a significant impact on the encryption performance of the algorithm. The low-dimensional chaotic scrambling ability is insufficient and lacks security. In contrast, due to the high computational complexity, the high-dimensional chaotic system will dramatically impact the encoding time. Therefore, this paper proposes a multi-link selective video encryption algorithm based on the Cross Coupled Map Lattices system in conjunction with H.264/AVC video coding protocol.

Introducing the Cross Coupled Map Lattices model and encrypting multiple video coding sessions of the H.264/AVC video coding protocol allows the system to generate chaotic sequences with high complexity and security by simple computation, thus achieving improved security of video encryption while ensuring the basic invariance of coding efficiency and video format retention.

The main contributions of this paper are as follows:

1) The Cross Coupled Map Lattices model is introduced, its spatio-temporal behavior is briefly analyzed, and the effect of the "return diffusion" behavior of the cross-coupled image lattice model on the generation of chaotic sequences is discussed. Finally, a pseudo-random sequence generator is designed based on the model to generate pseudo-random sequences with high complexity and security quickly.

2) Analyze the coding structure characteristics of H.264/AVC video coding protocol in intra-frame prediction, inter-frame prediction, entropy coding, and other links, and select multiple key syntax elements as encryption objects to realize multi-link encryption of the video coding process. On the one hand, it avoids the impact of encryption operation on video compression coding efficiency and format retention, and on the other hand, it solves the problem of poor security of single-link selective encryption.

3) Visual encryption effectiveness and security analysis are performed to test the novelty of the proposed video encryption algorithm. Comparing the performance parameters of the proposed encryption algorithm with the prior art results, we observe that the proposed algorithm has a better visual encryption effect, higher encryption efficiency, and security.

The remaining of this paper is organized as follows. Section 2 introduces the Cross Coupled Map Lattices System and its corresponding pseudo-random sequence generation algorithm. Section 3 analyzes the coding structure characteristics of H.264 and presents the video encryption scheme proposed in the article. Section 4 discusses the results of visual encryption performance and anti-attack performance evaluation of the proposed video encryption algorithm. Section 5 concludes this paper.

2 Cross Coupled Map Lattices Pseudo-random Sequence Generation Algorithm

2.1 Cross Coupled Map Lattices System

The Cross Coupled Map Lattices system (CCML) [7] is a typical model for spatiotemporal chaos systems and is formulated as:

$$x_{n+1}(i)=\begin{cases}\dfrac{1}{(1+\varepsilon)}f(x_n(i))+\dfrac{\varepsilon}{2(1+\varepsilon)}(x_n(i-1)+x_n(i+1)), & i \text{ is even numbers}\\ \dfrac{1}{(1+\varepsilon)}f(x_n(i))+\dfrac{\varepsilon}{2(1+\varepsilon)}(x_{n+1}(i-1)+x_{n+1}(i+1)), & i \text{ is odd numbers}\end{cases}. \quad \textbf{(1)}$$

Where $x_n(i)$ denotes the n moments i-th discrete lattice point, i=1, 2, 3, ..., L; L is the lattice point size; ε is a coupling coefficient that satisfies $0 < \varepsilon < 1$; there is a boundary condition $x_n(0) = x_n(L)$, $x_n(L+1) = x_n(1)$.

Using tent map as a nonlinear function of the CCML model, it can be expressed as follows:

$$F_\alpha : x_i = \begin{cases}\dfrac{x_{i-1}}{\alpha}, 0 < x_{i-1} \le \alpha\\ \dfrac{1-x_{i-1}}{1-\alpha}, \alpha < x_{i-1} < 1\end{cases}. \quad \textbf{(2)}$$

The tent mapping obeys a uniform distribution and is chaotic on the nonlinear parameter $\alpha \in (0,1)$, which has a wider range of parameters and better statistical properties than the classical logistic mapping.

The Cross Coupled Map Lattices model maintains a two-way diffusion between the time and space domains. Let the even number of lattice points advance half a step in the time domain and plot the reaction-diffusion state of the lattice points to obtain the Cross Coupled Map Lattices Spatio-temporal state diagram, as shown in Fig. 1.

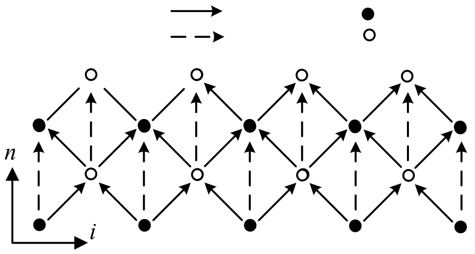


Fig. 1. Cross Coupled Map Lattices Spatio-temporal state diagram

As seen from Fig. 1, although the cross-coupled image lattice artificially specifies the existence of different iterations for odd and even lattice points, the odd-even lattice points evolve in the same way on the continuous time domain. At the same time, any later lattice $x_{n+1}(i)$ contains the diffusion contribution of adjacent lattice $x_{n+1}(i+1)$ and $x_{n+1}(i\text{-}1)$ to its lattice state $x_n(i)$, which is called “return diffusion”. The existence of “return diffusion” makes the Cross Coupled Map Lattices produce multiple diffusion in a reaction process, which greatly improves the system’s security without affecting the computational complexity.

2.2 Pseudo-random Sequence Generation Algorithm

The Cross Coupled Map Lattices system can generate pseudo-random sequences with strong randomness. However, the chaotic sequence values are all in the real number field and cannot be used directly for video encryption. Take the real number chaotic sequence generated by the Cross Coupled Map Lattices system discretized to obtain usable integer pseudo-random sequences, as shown in Fig. 2.

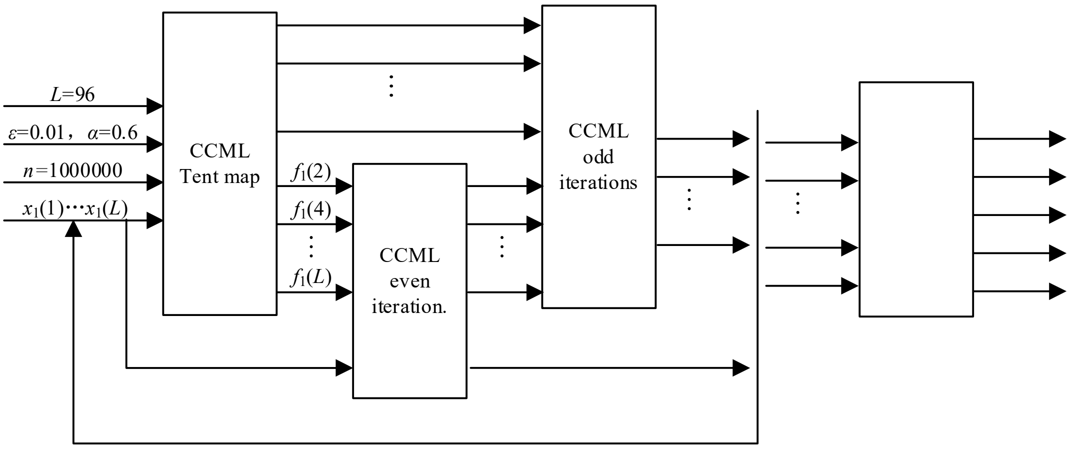


Fig. 2. CCML pseudo-random sequence generation

Step 1: Model initialization. Set the lattice size $L = 96$, coupling coefficient $\varepsilon = 0.01$, nonlinear parameter α=0.6, and the sequence length of 1 million. Use the random function to generate the sequence initial values $\{x_1(1), x_1(2), \ldots, x_1(96)\}$.

Step 2: Tent map sequence generation. Substitute sequence initial values $\{x_1(1), x_1(2), \ldots, x_1(96)\}$ into equation (2) to generate the tent map sequence $\{f_1(1), f_1(2), \ldots, f_1(96)\}$.

Step 3: Even iteration of the CCML model. Substitute odd lattice point sequence initial values $\{x_1(1), x_1(3), \ldots, x_1(95)\}$and tent map sequences $\{f_1(2), f_1(4), \ldots, f_1(96)\}$ into equation (1) to generate even lattice point sequence $\{x_2(2), x_2(4), \ldots, x_2(96)\}$.

Step 4: Odd iteration of the CCML model. Substitute even lattice point sequence $\{x_2(2), x_2(4), \ldots, x_2(96)\}$and tent map sequence $\{f_1(1), f_1(3), \ldots, f_1(95)\}$ into equation (1) to generate odd lattice point sequences $\{x_2(1), x_2(3), \ldots, x_2(95)\}$.

Step 5: Repeat iterations. End this iteration process, integrate the results of Step 3 and Step 4, and substitute them back to Step 3 to start the next iteration, repeat the process n times. The output sequence starts from n = 101, which discarded the results of the first 100 iterations to avoid the transient effect of spatio-temporal chaos that affects the randomness of sequences.

Step 6: Real sequence discretization. The sequence of real numbers $\{x_n(1), x_n(2), \ldots, x_n(96)\}$ is integer quantized by equation (3) to generate the single-byte integer sequence $\{X_n(1), X_n(2), \ldots, X_n(96)\}$.

$$X_n(i) = floor[x_n(i) \times 10^{13}] mod 2^8 \ . \tag{3}$$

Where *floor()* is the downward rounding function and mod means modulo operation, each operation will generate an integer X_n(i) in the range (0, 2^8).

Step 7: Select the pseudo-random sequences required for encryption. Use the random function to select five pseudo-random sequences {*key*1, *key*2, *key*3, *key*4, *key*5} from the integer sequence generated in Step 6 for video encryption.

3 Video Encryption Algorithm

3.1 Encrypted Location Analysis

The H.264/AVC standard was developed by the Joint Video Team (JVT) in 2003. On the basis of the original coding standards, new or improved coding methods such as adaptive intra-frame prediction and 1 / 4 pixel motion compensation are added to improve the quality and efficiency of video compression quality. In H.264/AVC, the encoder its first intra-frame prediction and inter-frame prediction to predict the video data for encoding, then performs conversion and quantization operations to reduce the amount of encoded data, uses a "Z" scan to rearrange the quantized residuals, and finally achieves the transformation of encoded data to the bitstream by entropy coding. Selective video encryption is a class of video encryption techniques combined with video coding by embedding encryption algorithms into the video compression and the coding process to destroy the commercial value of video and improve video processing efficiency. However, the wrong embedding position of the encryption algorithm will lead to problems such as a significant reduction in compression efficiency or the inability to decode the compressed stream. Therefore, the encryption location of the selective video encryption algorithm should meet the following conditions: (1) the encrypted video stream cannot leak plaintext information and can resist common security attacks. (2) The encryption algorithm does not affect the video format and can decode the compressed video file correctly. (3) The encryption operation has low computational complexity and does not affect the encoding time. (4) The encryption process does not affect the compression efficiency and has little impact on the bit rate. To satisfy the above four requirements, the encryption algorithm in this paper encrypts only the key syntactic elements within the intra-frame prediction, inter-frame prediction, and entropy encoding sessions. Fig. 3 is the Selective encryption algorithm framework.

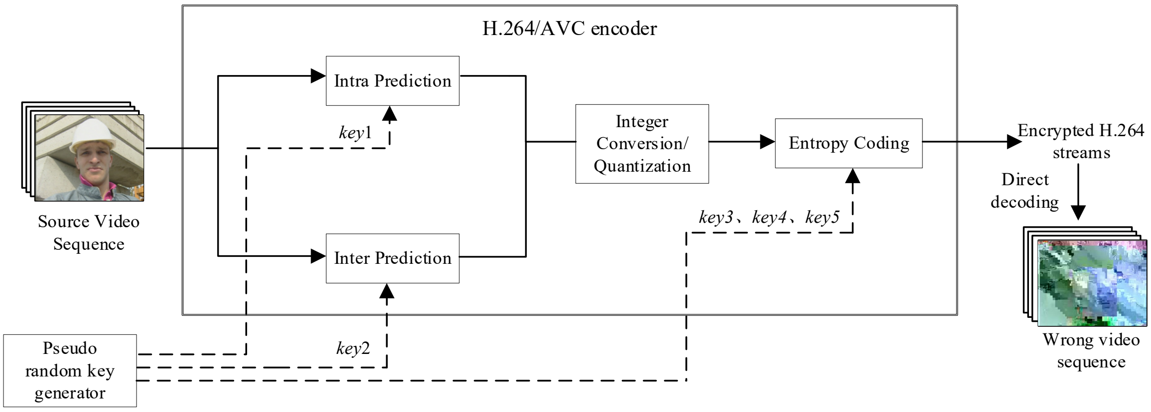


Fig. 3. Selective encryption algorithm framework

3.2 Intra-frame Predictive Mode Encryption

In video coding, intra-frame prediction predicts macroblocks from the same image to eliminate spatially redundant information between pixels. The intra-frame prediction mode (IPM) contains the prediction information between pixels of keyframes, defines the reference direction of macroblock inter-prediction, and is an essential basis for decoding and reconstructing keyframes. Encrypting it will cause error diffusion, which in turn disrupts the frame reconstruction of subsequent P and B frames, thus achieving holistic encryption of the video.

The algorithm in this paper focuses on the encryption of 4×4 luminance block prediction patterns within the I-frame macroblock in the video sequence. The H.264 encoding protocol uses variable-length encoding to write IPM [8]. The encoder of H.264 first compares the predicted and actual values of the prediction pattern based on the calculation of the encoding conditional operator prev_intra4×4_pred_mode_flag, then encodes the IPM selectively into the syntax element rem_intra4×4_pred_ mode. To prevent the stream length from being affected by the encryption operation, the encryption algorithm encrypts the IPM only when the predicted and actual values of the prediction pattern are different, which means the coding condition operator prev_intra4×4_pred_mode-_flag = 0, and the encryption method is as follows:

$$rem\_IPM\_INF = IPM\_INF \oplus key1\ . \tag{4}$$

Where *rem_IPM_INF* is the encrypted intra-frame prediction pattern; *key*1 is an arbitrary pseudo-random sequence generated by the Cross Coupled Map Lattices model; *IPM_INF* is the actual intra-frame prediction pattern; *key*1 and *IPM_INF* are encrypted using an XOR operation.

3.3 Motion Vector and Motion Compensation Encryption

Inter-frame prediction is a predictive coding based on the temporal correlation of the frames before and after the video sequence, mainly through motion estimation and motion compensation algorithms, to eliminate the video time redundancy information. Motion estimation refers to calculating the reference frame macroblock position, the motion vector (MV), relative to the current macroblock by searching the block matching algorithm in the spatial domain. In video coding, the encoder only encodes the predicted difference between the MV of the current macroblock and the adjacent macroblocks, so only the motion vector differences (MVD) need to be encrypted to achieve the purpose of dislocating the video motion information [9].

The algorithm in this paper focuses on encrypting the motion vector prediction residuals of P-frame macroblocks of video sequences. In the H.264 coding protocol, the encoder uses a tree-structured chunking algorithm to split the macroblock into chunks of different sizes for motion estimation. MVDs of different size chunks have different energy and occupied encoding time. MVDs of large-size chunks have more energy but occupy shorter encoding time, while MVDs with small-size chunks have the opposite. If the encryption algorithm encrypts the MVD information of each chunk directly, it will directly affect the stream length and encryption efficiency of compressed video coding. And the encryption of MVD symbol bits can avoid the influence of chunk size on encryption efficiency, and the encryption method is as follows:

$$en\_MVD\_sign = MVD\_sign \oplus key2 \ . \tag{5}$$

Where *en_MVD_sign* and *MVD_sign* are the MVD symbol bits before and after encryption; *key*2 is an arbitrary pseudo-random sequence generated by the Cross Coupled Map Lattices model; *key*2 and *MVD_sign* are encrypted using an XOR operation.

3.4 Entropy Encoding Syntax Element Encryption

The final step in video coding is entropy encoding. Entropy coding follows the following principles: symbols with a higher probability of occurrence are given shorter code lengths, and symbols with a lower probability of occurrence are given larger code lengths. Entropy coding can use the probability distribution of symbol coding to effectively reduce the redundant information of video source symbols and thus reduce the stream length. The video coding protocol of H.264 has two types of entropy coding. This paper chooses to encrypt the entropy coding method represented by Exponential Columbus Coding and Context-Adaptive Variable-Length Coding (CAVLC).

The Exponential Columbus Coding encodes all syntax elements in the H.264 protocol except for the residual data, and its coding structure is [*M zero*][1][*INFO*]. Where *INFO* is the encoded information value of *M* bits, and *M zero* is composed of *M* consecutive zeros. Set the value to be coded as *code_Num*, then *M* and coding information value *INFO* can be expressed as:

$$M = floor(\log_2(code\_Num+1)) \ . \tag{6}$$

$$INFO = code\_Num + 1 - 2M \ . \tag{7}$$

floor() is a downward rounding function. Since M zero is continuous zero values, it is used to calculate the length of decoded values in the decoding stage and cannot be encrypted. Therefore, only the encoded information value INFO is encrypted, and the encryption method is as follows:

$$en\_INFO = INFO \oplus key3 \ . \tag{8}$$

Where *en_INFO* and *INFO* are the exponential Columbus encoded information values before and after encryption; *key*3 is an arbitrary pseudo-random sequence generated by the Cross Coupled Map Lattices model; *key*3 and *INFO* are encrypted using an XOR operation.

CAVLC encodes the residual data in the H.264 protocol. The residual data generated after the prediction, transformation, and quantization of video sequences are sparse, and most of them are 0. The encoder represents the continuous zero string by the travel coding of CAVLC and stores the important residual change information in the non-zero coefficients. Encryption of the trailing coefficient symbol values and non-zero coefficient amplitude symbols of the CAVLC can affect the correct representation of the residual information. The encryption method is as follows:

$$en\_T1\_sign = T1\_sign \oplus key4 \ . \tag{9}$$

$$en\_TCLevel\_sign = TCLevel\_sign \oplus key5 \ . \tag{10}$$

Where *T1_sign* and *en_T1_sign* are the trailing coefficient symbols before and after encryption; *TCLevel_sign* and *en_TCLevel_sign* are the non-zero coefficient amplitude symbol bits before and after encryption; *key*4 and *key*5 are both arbitrary pseudo-random sequences generated by the Cross Coupled Map Lattices model; *key*4 and *key*5 adopt XOR operation with *T1_sign* and *TCLevel_sign* respectively to achieve encryption.

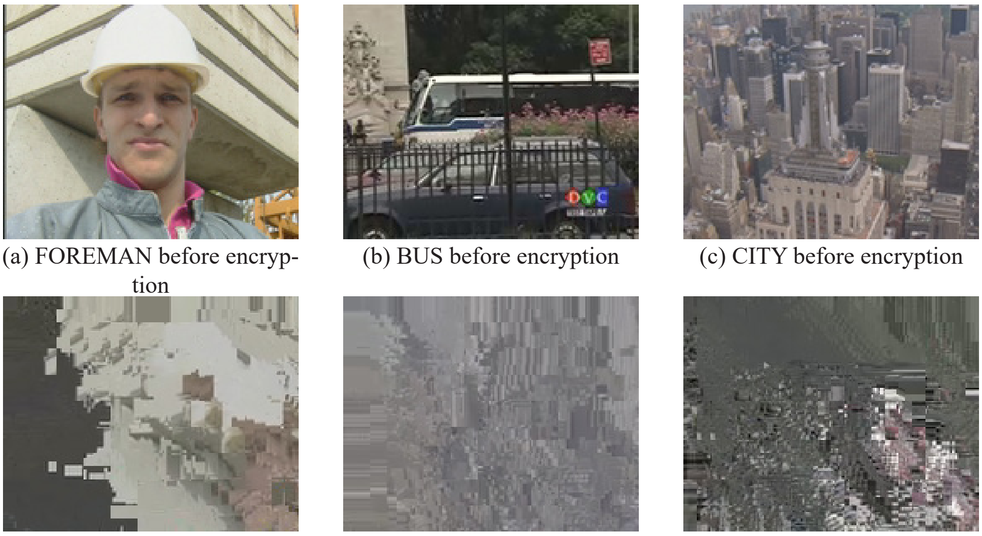
4 Encryption Effect and Security Analysis

4.1 Experimental Operating Environment

To objectively evaluate the performance of the video encryption algorithm in this paper and ensure real and reliable test results. We use the official H.264 VCEG test platform JM8.6, set the video encoding grade as the basic grade, the encoding structure as IPP......, the encoding mode as CAVLC encoding, the sampling format as 4:2:0, and the quantization parameter as 28. The hardware environment tested is Intel(R) Core(TM) i7-4720HQ 2.60GHz with 12GB of RAM. The CIF format video sequences with different characteristic attributes of Foreman, News, Crew, City, Football, Bus, Ice, Soccer, and Mobile were selected from the standard video test sequences for encryption operations and analyzed for encryption effectiveness and attack resistance.

4.2 Visual Encryption Effect Analysis

The visual encryption effect analysis is the most intuitive and clear quality standard of video encryption. Choose FOREMAN, BUS, and CITY as three sequences for the encryption test, and the effect before and after encryption of the tenth frame of the video sequence is shown in Fig. 4.



(a) FOREMAN before encryption (b) BUS before encryption (c) CITY before encryption

(d) FOREMAN after encryption (e) BUS after encryption (f) CITY after encryption

Fig. 4. Visual effect before and after encryption

From Fig. 4, the source video sequence FOREMAN contains a background still and subtle changes in the character's facial expressions, as shown in (a). The encrypted video sequence produces a significant chromatic shift in the horizontal and vertical edges, and the subtle changes in expressions are amplified, accompanied by range shifts and jitter. The source video sequence BUS has a fast-moving vehicle and a slow-moving background following the camera, as shown in (b). After encryption, the video sequence is affected by the error diffusion and motion vector change, the chromaticity information of the screen is damaged extensively, the vehicle's outline moving at high speed is completely unrecognizable, and only the jitter in the vertical and horizontal directions can be seen. The source video sequence CITY is a complex textured building shot at a high altitude moving slowly, as shown in (c). After encryption, only a small amount of texture chromaticity information can be observed in the lower right corner of the video, and the rest of the video has a large degree of dithering and chromaticity loss.

The above experimental results show that the video encryption algorithm in this paper has a good visual encryption effect for video sequences with different attributes, especially for videos with strong motion attributes, which can meet the video encryption requirements of different occasions.

4.3 Peak Signal to Noise Ratio Analysis

Peak Signal to Noise Ratio (PSNR) is the most widely used quality metric. It can be used to measure the signal fidelity of both the original and encrypted video frames and is calculated as follows:

$$MSE = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(x_{ij}-y_{ij})^2 \quad . \tag{11}$$

$$en\_INFO = INFO \oplus key3 \quad . \tag{12}$$

Where M and N are the numbers of corresponding rows and columns of the video frame; x_{ij} and y_{ij} are the original pixel value and the decoded pixel value of the ith row and jth column of the video frame; MSE is the mean square error, which can measure the distortion level of the video frame.

Table 1. Video sequence Y-component PSNR analysis

Video sequence	Original value	Literature [11]	Literature [12]	Article results
BUS	33.9943	15.8600	13.6400	7.8598
CITY	34.6686	18.6500	18.7700	14.5976
CREW	36.4798	16.3000	16.6500	10.0603
FOOTBALL	34.9076	15.3500	15.5900	12.1766
FOREMAN	35.7396	13.4900	13.6400	9.3099
ICE	38.7281	14.9800	14.8300	12.4926
MOBILE	33.5390	9.4400	9.4900	9.2326
NEWS	36.7843	-	-	7.1311
SOCCER	35.2564	15.9300	16.1100	12.8723
Mean	35.5664	15.0000	14.8400	10.6370

For a given compiled code system and fixed video content, the change in PSNR is an indicator of quality change. When the video content remains constant and the bit rate increases, the *PSNR* increases monotonically by a similar magnitude to the video subjective quality test results. Generally speaking, the video quality is excellent when the *PSNR* of the video frames is greater than 40dB, and the video content is unrecognizable when it is less than or equal to 13dB. Among them, the PSNR loss rate of the luminance component (Y component) has a much higher impact on the subjective quality of the video than the *PSNR* loss rate of the chrominance component [10]. Table 1 shows the comparative test results of the average PSNR values of the Y-component for the first 50 frames of the nine test sequences. The second column in the table shows the PSNR values of the standard coding video luminance components, whose mean values are around 36 dB, and the video quality is good. Literature [11] and literature [12] designed a three-factor encryption scheme and a logistic chaotic system encryption scheme for the entropy coding link of video coding, respectively. The PSNR of the luminance component of the video sequence after encryption by these two types of schemes is significantly reduced, and the average value is close to 15dB, which has basically satisfied the demand for video encryption. However, the PSNR values of the luminance components of all video sequences corresponding to the algorithms in this paper are less than 15 dB and lower than those in the literature [11] and [12], especially the PSNR values of video sequences with strong motion attributes such as BUS, CREW, FOREMAN, drop more than 30% compared with the results of the two papers. It shows that the video encryption algorithm in this paper can affect the signal fidelity of the video to a large extent, causing a high degree of distortion of the video data, which affects the normal viewing of the video content.

4.4 Structural Similarity Analysis

Structural similarity (*SSIM*) [13] is another image fidelity metric that uses structural information and structural distortion concepts to quantify the input-related behavior when the video frame information is distorted. Compared with the traditional signal metric *PSNR*, *SSIM* is more sensitive to distortion behaviors such as jitter and mean shift and thus more closely matches the subjective perception of the image by the visual system. *SSIM* is calculated as:

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x{}^2+\mu_y{}^2+C_1)(\sigma_x{}^2+\sigma_y{}^2+C_2)}\ . \tag{13}$$

Where x and y are the original encrypted video images and the encrypted video images; μ_x, μ_y are the mean values; σ_x, σ_y, and σ_{xy} are the standard deviations; C_1 and $C2$ are the constants.

The value of *SSIM* ranges from 0 to 1 and is positively correlated with the subjectively perceived quality of the video frame image. The higher the quality of the video frame image, the closer the value of *SSIM* is to 1. Table 2 shows the comparison test results of the average SSIM values for the first 50 frames of the eight standard video sequences.

Table 2. Video sequence SSIM analysis

Video sequence	Original value	Literature [11]	Literature [12]	Article results
BUS	0.9532	0.0922	0.1381	0.0185
CITY	0.9443	0.0747	0.0899	0.1090
CREW	0.9466	0.1740	0.2075	0.0988
FOOTBALL	0.9307	0.0922	0.1381	0.0539
FOREMAN	0.9566	0.1171	0.1218	0.0738
ICE	0.9805	0.1654	0.1745	0.0702
MOBILE	0.9778	0.0270	0.0449	0.0815
NEWS	0.9693	-	-	0.1281
SOCCER	0.9122	0.1322	0.1604	0.0942
Mean	0.9524	0.1094	0.1344	0.0809

From the data in Table 2, it can be seen that the experimental results of SSIM values of the encryption algorithm in this paper for video sequences with different attributes have different variations from the results in the literature [11] and literature [12]. Among them, video sequences with high contrast and complex textures like CITY and MOBILE have relatively high SSIM values. In contrast, those with strong motion attributes such as BUS, FOOTBALL, and FOREMAN have relatively low SSIM values. Analyzing the measurement criteria of SSIM, we can see that the encryption effect of the encryption algorithm in this paper for video sequences with strong motion attributes reflects more structural distortion such as jitter and image block effect caused by motion vector changes, while the encryption effect for video sequences with strong contrast reflects non-structural distortion such as error diffusion and contrast changes caused by intra-frame prediction patterns. Overall, the mean value of SSIM after encryption by the encryption scheme in this paper reaches 0.0809, which is lower than the experimental results in the literature [11] and literature [12], which indicates that in terms of subjective perception, the encryption algorithm in this paper can meet the basic requirements of video encryption.

4.5 Reduced Reference Objective Evaluation Analysis

To better evaluate the change rate of relevant features of video sequences before and after video encryption, the Reduce Reference (RR) method for video quality is introduced. Unlike such full-reference objective evaluation methods as *PSNR* and *SSIM*, the RR objective evaluation method first divides the video into multiple adjacent spatial-temporal regions (S-T regions), then performs an objective evaluation of video quality by extracting relevant feature parameters within the regions. Compared to the full-reference objective evaluation method, it better reflects the overall change in video quality in the time domain. The most typical RR objective evaluation method is the generic video quality measure (NTIA-VQM) proposed by Pinson and Wolf [14], which is defined as:

$$\begin{aligned}\text{NTIA-VQM} = &-0.2097 si\_loss + 0.5969 hv\_loss + 0.2483 hv\_gain \\ &+0.0912 chroma\_spread - 2{,}3416 si\_ain + 0.0431 \\ &ct\_ati\_gain + 0.0076 chroma\_extreme\end{aligned} \quad . \tag{14}$$

Where *si_loss* is the decline or loss of spatial information; *hv_loss* is the diagonal edge shift in horizontal and vertical directions; *chorma_spread* is the two-dimensional color distribution change; *si_gain* is the quality improvement resulting from edge sharpening and enhancement; *ct_ati_gain* is the spatial detail and motion information appearing in the S-T region; *chorma_extreme* for severe local color damage.

NTIA-VQM quantitatively measures the change in video feature behavior in terms of overall video characteristics, with a negative correlation between the video subjective quality test results. The larger the NTIA-VQM is, the more jitter, blur, and uneven motion exist in the video, and the more serious the visual degradation is. Conversely, the smaller the value of NTIA-VQM, the closer the common characteristics between encoded video and source video, and the better the video quality. Table 3 lists the NTIA-VQM values for normal and cryptographic decoding of the first 50 frames of the nine standard test video sequences. The experimental results show that the mean NTIA-VQM value of the standard coding process is close to 0.04, indicating that the video coding process will cause limited quality degradation of the source video sequence. The NTIA-VQM values of this paper's encryption algorithm increase by a factor of nearly 30 relatives to the standard coding NTIA-VQM values, and each value exceeds 1, which is beyond the parameter range of 0 to 1 usually used to evaluate the performance of video encoder compiled code, a detection result that only occurs in extremely distorted video scenes. NTIA-VQM's significant variation and numerical overflow show that this paper's encryption algorithm has a good effect on the characteristic parameters of the encrypted video. In the case of abnormal decryption, blurring, block distortion, uneven motion, noise, and error blocks will occur, and the overall encryption effect of video is evident, which can effectively stop the leakage of video information on the time domain.

Table 3. Video sequence NTIA-VQM analysis

Video sequence	Original value	Article results
BUS	0.0325	1.2191
CITY	0.0397	1.0714
CREW	0.0392	1.1847
FOOTBALL	0.0268	1.0514
FOREMAN	0.0271	1.2156
ICE	0.0749	1.1472
MOBILE	0.0268	1.2424
NEWS	0.0268	1.2424
SOCCER	0.0764	1.1051
Mean	0.0411	1.1644

4.6 Coding Time and Compression Efficiency Analysis

Since the transmission and access of video data nowadays have real-time requirements, the design of encryption and decryption algorithms should not cause significant delays in the transmission and access of video data. For selective video encryption algorithms, two critical factors that affect the delay time of video data transmission are encoding time and compression efficiency. When the video transmission speed is kept constant, the long encoding time will affect the video input process, and the low compression efficiency will increase the channel bandwidth occupied by the video stream, both of which will increase the data transmission time [15]. For nine different types of video sequences for the encoding test, record the encoding time and compressed file size required for standard encoding and encrypted encoding of 50 frames of video, as shown in Table 4 and Table 5.

Table 4. Video sequence encoding time analysis

Video sequence	Unencrypted code (s)	Encrypted code (s)	Percent change (%)
BUS	85.224	86.933	1.97%
CITY	81.605	81.446	-0.20%
CREW	82.325	82.555	0.28%
FOOTBALL	84.495	84.668	0.20%
FOREMAN	79.731	79.457	-0.34%
ICE	79.282	78.611	-0.85%
MOBILE	85.897	84.333	-1.86%
NEWS	75.783	76.416	0.84%
SOCCER	82.296	81.909	-0.47%
Mean	81.849	81.814	0.00%

Table 5. Video sequence compression efficiency analysis

Video sequence	Unencrypted code (bit)	Encrypted code (bit)	Percent change (%)
BUS	727056	728560	0.207
CITY	244536	245008	0.193
CREW	453608	455408	0.397
FOOTBALL	1118992	1120080	0.097
FOREMAN	172832	174440	0.930
ICE	314200	314352	0.048
MOBILE	851504	850760	-0.087
NEWS	118448	119000	0.466
SOCCER	397288	396432	-0.215
Mean	488718	489337	0.2262

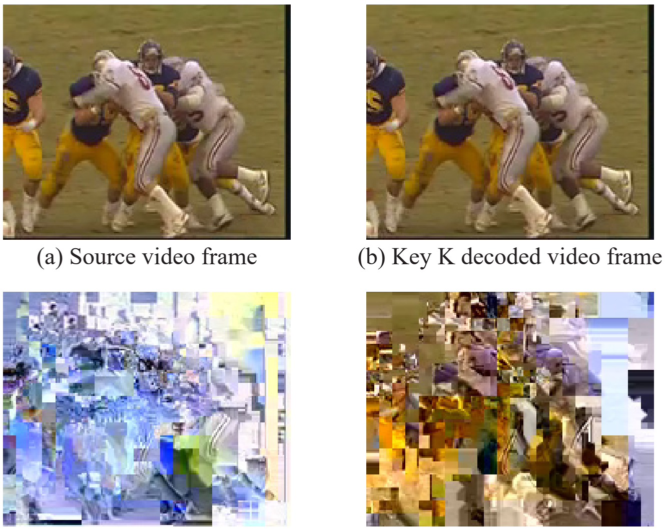
From the data in Table 4, we can conclude that the video compression coding time of the encryption algorithm used in this paper does not change significantly compared to the plaintext coding time. The coding time increase for video sequences like BUS, where high contrast backgrounds and high moving speed subjects exist, is also less than 2% of the overall coding time. There is also a situation where the encryption operation smoothes the textures and thus reduces the encoding time for video sequences with complex textures. As can be seen from Table 5, the encryption algorithm has a specific effect on file size before and after video encoding. Different types of video sequence compression efficiency vary, but the overall change rate is less than 1%. Encryption operation has a negligible effect on code flow.

4.7 Key-space Analysis

Key-space refers to the size range of encryption keys. The larger the key-space of the encryption algorithm, the stronger its resistance to brute force attack. To ensure the security of video encryption, the encryption algorithm's key-space should be less than 2^{128} [16]. This paper algorithm sets the lattice size of the Cross Coupled Map Lattices system model L=96; the quantized key length is 8 bits; the coupling coefficient ε and the nonlinear parameter α coupling parameters are floating-point numbers of 23 valid bits ranging from 0 to 1. Therefore, the key-space is $2^{(96\times8+2\times23)}$=2814, which can resist the brute force attack. Ninety-six initial values of lattice points and any change of the two control parameters will change the resulting chaotic pseudo-random sequence.

4.8 Key Sensitivity Analysis

Good video encryption schemes should have high key sensitivity, which can distort the decoded video content when only a small key change in the key is produced, thus better resisting attacks. To analyze the key sensitivity of the encryption algorithm in this paper, select the video sequence FOOTBALL as the ciphertext object. Define the initial key as K, and encrypt the video by the random sequence generated by K to get the compressed stream file C. Change any value of a bit in key K to the modified key K'. Fig. 5 shows the 25th frame image of the source video and the 25th frame image of the video decoded by Key K, Non-Key, and Key K' for stream file C respectively.



(a) Source video frame

(b) Key K decoded video frame

(c) Non-key decoded video frame

(d) Key K' decoded video frame

Fig. 5. FOOTBALL key sensitivity test

From Fig. 5, it can be seen that the video frame b) obtained by decoding the encrypted, compressed stream file C using the key K is basically the same as the source video frame a), and the stream file is decrypted correctly. On the contrary, the non-key decoded video frame c) has brightness and contrast variations, and the image occurs more dithering and image block effect phenomenon, from which no information about the source video image can be obtained. The video frame d) generated by decryption using the key K' has a certain degree of luminance information leakage. However, a large amount of distortion, such as dithering and image block effect, disrupts the leaked luminance information sufficiently that the exact image content is still not discernible. The overall image correlation coefficient of d) and a) is calculated, and its calculation result is 0.1744, which indicates that there is no clear linear correlation between the change key video sequence and the source video sequence, and the source video content cannot be deciphered by way of changing the key. The above experimental results show that the algorithm in this paper has high cryptographic key sensitivity and can resist common known plaintext attacks.

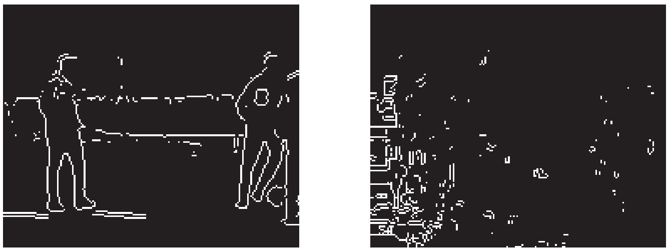
4.9 Edge Detection Analysis

The edge is a collection of points in an image with drastic changes in brightness. The video sequence edges reflect video entities' contour and motion change characteristics, and the accurate measurement of edge information will disclose the plaintext information of video data. The edge detection attack first uses graying to process the video cipher image, then detects the residual edge information, and finally draws the edge detection image to analyze cipher profile information of the current frame cipher image and adjacent frame cipher image. If the video encryption is poor, the edge information detected from the ciphertext will leak the contour information, texture features, and even the motion vector parameters of the entities in the original video sequence before and after the frames.

Using the Sobel operator performs edge detection of video frame images to obtain edge detection maps of plaintext and ciphertext images of SOCCER video frames, as shown in Fig 6. From the experimental results, it can be seen that after the encryption algorithm encrypts the video information in this paper, the detected edge information of the ciphertext image is mostly the irregular contour information associated with the image distortion of the edge offset and dithering phenomenon, from which the contour and texture features of the plaintext image cannot be obtained. The edge difference ratio (EDR) is introduced to quantify the edge detection resistance of the encryption algorithm, which is defined as [17]:

$$\text{EDR}=\frac{\sum_{i,j=1}^{N}\left|P(i,j)-\overline{P}(i,j)\right|}{\sum_{i,j=1}^{N}\left|P(i,j)+\overline{P}(i,j)\right|}\ . \qquad \textbf{(15)}$$

Where $P(i, j)$ is the pixel value of plaintext edge detection; $\overline{P}(i, j)$ is the pixel value of ciphertext edge detection; (i, j) is the location of the edge detection pixel point. EDR value from 0 to 1, and the value closer to 1 the edge of the ciphertext image contains less information about the plaintext image, the stronger the algorithm resists edge detection. Eight video sequences, such as BUS, are edge detected, and their corresponding EDR values are calculated, as shown in Table 6. From the data in Table, the EDR values are relatively low except for CITY video sequences due to the presence of jitter of high texture features and ciphertext offset and the phenomenon of mutual overlap of block effect contour information. The EDR values of the rest of the tested sequences are at 0.94, which indicates encrypted video sequences leak limited plaintext information, and the video encryption algorithm can resist edge detection attacks.



(a) SOCCER original frame (b) SOCCER encrypted frame

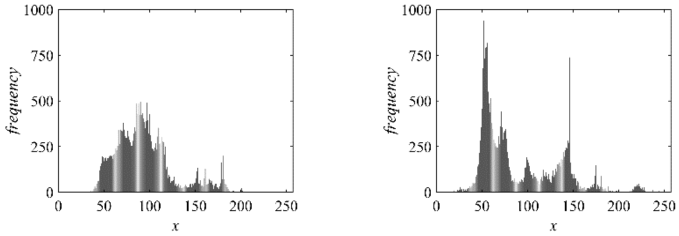
Fig. 6. SOCCER edge detection images

Table 6. Video sequence edge difference ratio calculation results

Video sequence	BUS	CITY	CREW	FOREMAN	SOCCER	MOBILE
EDR	0.9427	0.925	0.9453	0.9496	0.9395	0.945

4.10 Statistical Attack Resistance Analysis

The statistical attack is an attack method to decipher the cryptosystem by analyzing ciphertext and plaintext statistical patterns. This paper's encryption object is the corresponding syntax element in the video encoding process. Unlike traditional encryption methods, this encryption result is reflected in the output stream in a discontinuous and non-intuitive way, and an attacker cannot analyze the statistics of the output stream to obtain the corresponding plaintext and key information. Fig. 7 shows the statistical histogram of the plaintext image and the corresponding encrypted image at frame 10 of the CREW video sequence, which can be used to analyze the variation of statistical patterns between the plaintext and ciphertext.



(a) CREW original frame (b) CREW encrypted frame

Fig. 7. CREW statistical histogram analysis

From Fig. 7, the histograms of the encrypted image and the plaintext image of the original video sequence are utterly different from each other in terms of distribution status, peak point position, or shape characteristics, and there is no specific statistical law change between the two images, so the video encryption algorithm in this paper can resist statistical attacks.

4.11 Comparative Analysis

To further evaluate the performance of the video encryption algorithm in this paper, a comparative analysis with related studies was conducted in terms of the cryptosystem, key-space, and compression efficiency impact, respectively, and the analysis results are shown in Table 7. The algorithm design of contemporary video encryption needs to consider the security of video encryption and the impact of encoding time change rate and compression efficiency on content dissemination. In Table 7, both this paper and the literature [18] use the spatiotemporal chaotic system for the sequence generator design, which enables parallel generation of pseudo-random sequences compared to the RC4 stream cipher and AES packet cipher used in the literature [19-22], thus minimizing the impact on the encoding time. The literature [21] ensures the invariance of the compression efficiency by encrypting the entropy coding single-link of H.264, relatively the algorithm in this paper still controls the impact of the algorithm on the compression efficiency to a low level based on the implementation of multi-link encryption. It is worth mentioning that the proposed algorithm in this paper has a larger key-space and higher security compared with other research algorithms. The comparative analysis results show that the video encryption algorithm designed in this paper has the best comprehensive capability and can best meet the dual requirements of security and real-time for video encryption.

Table 7. Comparative analysis of encryption algorithms

Encryption algorithm	Password system	Key space size	Time change rate	Bit change rate
Peng [18]	CML	177	0%	6.48%
Hong [19]	RC4	128	0.46%	3.41%
Wei [20]	RC4	128	1%	2.14%
Asghar [21]	AES	128	0.2%	0%
Liu [22]	AES	128	0.58%	0.63%
this paper algorithm	CCML	814	0%	0.22%

5 Conclusion

This paper proposes an multi-link selective video encryption algorithm based on the Cross Coupled Map Lattices model in conjunction with H.264/AVC video coding protocol. The video compression performance evaluation index and algorithm anti-attack performance analysis results show that the encrypted video content is severely distorted and can resist common attack methods. The introduction of the Cross Coupled Map Lattices model generates more complex chaotic sequences at a lower computational complexity, improving encryption efficiency and security. Analyzing the characteristics of H.264 video coding structure and selecting multiple specific key syntax elements for encryption overcomes the problems of insufficient security and reduced compression efficiency that existed in the past single encryption and dramatically improves the all-around performance of video encryption. Although the comprehensive performance of the algorithm's encryption in this paper is excellent, there are inevitable fluctuations in the compression efficiency and encoding time of the videos with different attributes. The operation of directly quantized the Cross Coupled Map Lattices model real sequences also affects the uniformity and difference properties of the sequences to some extent.

References

[1] C.-J. Wei, G.-D. Li, Encryption algorithm of video images combining hyper-chaotic system and Logistic mapping, Computer Engineering 48(5)(2022) 263-271.

[2] L.-J. Yang, S.-C. Xie, J.-Z. Zhang, Color video stream encryption algorithm based on multi-chaotic system, Video Engineering 40(12)(2016) 7-11.

[3] H.-Y. Zang, J. Yang, G.-D. Li, Video encryption based on chaotic array system: working with image directly, International Journal of Information and Communication Technology 16(1)(2020) 84-97.

[4] J. Ahn, H.-J. Shim, B. Jeon, I. Choi, Digital video scrambling method using Intra Prediction Mode, in: Proc. 2004

Pacific-Rim Conference on Multimedia (PCM), 2004.

[5] R.S. Malladar, S.R. Kunte, Selective video encryption using the cross coupling of one-dimensional logistic maps, International Journal of Computer Network and Information Security 13(5)(2021) 40-54.

[6] J. Karmakar, A. Pathak, D. Nandi, M.K. Mandal, Sparse representation based compressive video encryption using hyper-chaos and DNA coding, Digital Signal Processing 117(2021) 103143.

[7] W.-M. Yang, Spatiotemporal Chaos and Coupled Map Lattices, Second ed., Shanghai Scientific and Technical Education Publishing House, Shanghai, 1995 (Chapter 6).

[8] H. Xu, X.-J. Tong, Z. Wang, M. Zhang, Y. Liu, J. Ma, Robust video encryption for H.264 compressed bitstream based on cross-coupled chaotic cipher, Multimedia Systems 26(4)(2020) 363-381.

[9] Y.-J. Song, Z.-L. Zhu, W. Zhang, H. Yu, Efficient protection using chaos for Context-Adaptive Binary Arithmetic Coding in H.264/Advanced Video Coding, Multimedia Tools and Applications 78(14)(2019) 18967-18994.

[10] J.D. Gibson, Book Reviews: Digital coding of waveforms: Principles and applications to speech and video- N.S. Jayant and P. Noll, Proceedings of the IEEE 75(4)(1987) 526-527.

[11] Y. Wang, M. O’Neill, F. Kurugollu, A tunable encryption scheme and analysis of fast selective encryption for CAVLC and CABAC in H.264/AVC, IEEE Transactions on Circuits and Systems for Video Technology 23(9)(2013) 1476-1490.

[12] J.-J. Li, C.-Y. Wang, X. Chen, Z. Tang, G. Hui, C.-C. Chang, A selective encryption scheme of CABAC based on video context in high efficiency video coding, Multimedia Tools & Applications 77(10)(2018) 12837-12851.

[13] O. Sugimoto, R. Kawada, M. Wada, S. Matsumoto, Objective measurement scheme for perceived picture quality degradation caused by MPEG encoding without any reference pictures, in: Proc. of Visual Communications & Image Processing 2001, 2000.

[14] M.-H. Pinson, S. Wolf, A new standardized method for objectively measuring video quality, IEEE Transactions on Broadcasting 50(3)(2004) 312-322.

[15] H. Li, T. Xiezhang, C. Yang, L. Deng, P. Yi, Secure Video Surveillance Framework in Smart City, Sensors 21(13)(2021) 4419.

[16] D. Lambić, Security Analysis and Improvement of the Pseudo-random Number Generator Based on Piecewise Logistic Map, Journal of Electronic Testing 35(4)(2019) 519-527.

[17] A. Sallam, O. Faragallah, E.M. El-Rabaie, HEVC selective encryption using RC6 block cipher technique, IEEE Transactions on Multimedia 20(7)(2018) 1636-1644.

[18] F. Peng, X.-W. Zhu, M. Long, An ROI privacy protection scheme for H.264 Video based on FMO and Chaos, IEEE Transactions on Information Forensics and Security 8(10)(2013) 1688-1699.

[19] S. Hong, M. Han, The study of selective encryption of motion vector based on the S-Box for the security improvement in the process of video, Multimedia Tools and Applications 71(3)(2014) 1577-1597.

[20] Z. Wei, Y. Wu, X. Ding, R.H. Deng, A scalable and format-compliant encryption scheme for H.264/SVC bitstreams, Signal Processing Image Communication 27(9)(2012) 1011-1024.

[21] M.-N. Asghar, M. Ghanbari, M. Fleury, M.J. Reed, Sufficient encryption based on entropy coding syntax elements of H.264/SVC, Multimedia Tools and Applications 74(23)(2015) 10215-10241.

[22] S. Liu, S. Rho, W. Jifara, F. Jiang, C. Liu, A hybrid framework of data hiding and encryption in H.264/SVC, Discrete Applied Mathematics 241(2018) 48-57.