electronics

Article

A Multi-Domain Embedding Framework for Robust Reversible Data Hiding Scheme in Encrypted Videos

Pei Chen [1,2], Zhuo Zhang [1,2], Yang Lei [1,2], Ke Niu [1,2,*] and Xiaoyuan Yang [1,2]

[1] Key Laboratory of Network and Information Security, The Chinese People Armed Police Force (PAP), Xi'an 710086, China
[2] College of Cryptography Engineering, Engineering University of PAP, Xi'an 710086, China
* Correspondence: niuke@163.com; Tel.: +86-029-13808593399

Abstract: For easier cloud management, reversible data hiding is performed in an encrypted domain to embed label information. However, the existing schemes are not robust and may cause the loss of label information during transmission. Enhancing robustness while maintaining reversibility in data hiding is a challenge. In this paper, a multi-domain embedding framework in encrypted videos is proposed to achieve both robustness and reversibility. In the framework, the multi-domain characteristic of encrypted video is fully used. The element for robust embedding is encrypted through Logistic chaotic scrambling, which is marked as element-I. To further improve robustness, the label information will be encoded with the Bose–Chaudhuri–Hocquenghem code. Then, the label information will be robustly embedded into element-I by modulating the amplitude of element-I, in which the auxiliary information is generated for lossless recovery of the element-I. The element for reversible embedding is marked as element-II, the sign of which will be encrypted by stream cipher. The auxiliary information will be reversibly embedded into element-II through traditional histogram shifting. To verity the feasibility of the framework, an anti-recompression RDH-EV based on the framework is proposed. The experimental results show that the proposed scheme outperforms the current representative ones in terms of robustness, while achieving reversibility. In the proposed scheme, video encryption and data hiding are commutative and the original video bitstream can be recovered fully. These demonstrate the feasibility of the multi-domain embedding framework in encrypted videos.

Keywords: multi-domain; reversible data hiding; robust; encrypted videos

Citation: Chen, P.; Zhang, Z.; Lei, Y.; Niu, K.; Yang, X. A Multi-Domain Embedding Framework for Robust Reversible Data Hiding Scheme in Encrypted Videos. *Electronics* **2022**, *11*, 2552. https://doi.org/10.3390/electronics11162552

Academic Editor: Jungong Han

Received: 1 July 2022
Accepted: 12 August 2022
Published: 15 August 2022

1. Introduction

Reversible data hiding [1] in encrypted domain (RDH-ED) is a technique that uses encrypted data as the carrier to reversibly embed information, and can still carry out correct data extraction and carrier decryption and recovery [2]. This technique is being increasingly used in the field of information security to ensure privacy and copyright protection [3]. Military data are often stored and transmitted in ciphertext. To authenticate access authentication, label information needs to be embedded into the encrypted data. In the medical field, multimodal medical images [4] are usually encrypted to prevent information on the patient's condition from leaking. For convenience of management, the information about individuals and conditions are embedded into encrypted images [5]. With the rapid development of the network, video information is widely disseminated in the Internet. At the same time, because there are many elements for data hiding in the video codec process, the reversible data hiding in encrypted videos (RDH-EV) has attracted the attention of many researchers [6–8].

Recently, the technology of reversible data hiding in encrypted images (RDH-EI) has seen great development, and can be roughly divided into three categories: vacating room after encryption [9–11], vacating room before encryption [12–14] and vacating room in

encryption [15–17]. However, due to the different video encoding structures, these schemes cannot be directly applied to encrypted videos. Therefore, there are relatively few schemes for RDH-EV. In [18], the intra-frame prediction mode (IPM) and the sign of motion vector difference (MVD) and the quantized discrete cosine transform (QDCT) coefficient are encrypted, and then the watermark is adaptively embedded into the amplitude of QDCT coefficients. The scheme can extract the watermark directly from the encrypted video and encryption and data hiding are commutative. In [19], a separable scheme is presented, that can embed data by codeword substitution in encrypted video streams. In the scheme, the IPM, MVD and QDCT are encrypted, while QDCT is used for embedding. The advantage of this scheme is that it will not increase the bit rate. The scheme presented in [20] is an improved version of that in [19]. Codewords with suffix length equal to 1 are used for data embedding by paired codeword substitution. Codeword redundancy is further used to effectively improve the embedding capacity. In [6–8], the authors have performed a series of works to continuously excavate the redundancy of QDCT codewords in HEVC so as to improve the embedding capacity. However, the common drawback of the above schemes is that the data hiding used is not reversible. In [21], IPM and the signs of MVD and QDCT are encrypted by stream cipher and then the information is embedded into the encrypted QDCT coefficients using histogram shifting (HS). The scheme is reversible and separable, but its robustness has not been experimentally tested. Compared with the scheme provided in [21], the scheme provided in [22] adds the scale factor for the embedding zone selection. This scheme can be expanded for different embedding capacity requirements, but the robustness is not tested. In [23], the signs for all IPM and MVD are encrypted and the sign of the QDCT coefficient is selectively encrypted. Then, the information is embedded into the QDCT coefficients with different priorities using HS, which improves the quality of the labeled video. Although the Bose–Chaudhuri–Hocquenghem (BCH) code is used in the scheme, the robustness is not improved. Because the embedding algorithm used in [23] is the traditional HS, it is difficult to resist a re-compression attack. In [24], the IPM, MVD and QDCT are encrypted. The information is embedded into the QDCT coefficients by two-dimensional HS, taking the two QDCT coefficients as an embedding point. Though this scheme effectively increases the video embedding capacity, it is also not robust. In [25], a separable and reversible scheme is proposed for H.265. The signs and amplitudes of MVD and the signs of QDCT are encrypted using RC4. The information is embedded into QDCT using the traditional HS and, therefore, the scheme is not robust.

It can be seen that the most current RDH-EVs do not take robustness into account. In addition, the above papers all encrypt the three elements of the video, but only an encrypted element for data embedding is applied, which does not make full use of the multi-domain characteristic of the encrypted video. Inspired by [26], by taking advantage of the characteristics of the various embeddable domains of video, a multi-domain embedding framework in encrypted video is proposed to achieve both reversibility and robustness. In the framework, two elements of video will be encrypted in different ways to form independent encrypted domain. One encrypted domain is used for robust embedding, the other is used for reversible embedding. In this paper, an anti-recompression RHD-EV is proposed to verity the feasibility of the framework. The data hiding in the proposed scheme is implemented by modulating the difference between two QDCT coefficients, which can be robust. In some application scenarios, the labeled video will be transmitted several times level by level after decrypting. For example, military information is generally reported and distributed level by level in internal systems. In this process, the video is likely to be inadvertently re-compressed, resulting in the loss of important label information. Therefore, for RDH-EV, it is necessary to have robustness against re-compression. The main contributions are as follows:

- The proposed framework for RDH-EV can achieve both robustness and reversibility.
- In terms of robustness, the proposed scheme outperforms the existing ones.
- Video encryption and data hiding in the proposed scheme are commutative.

The rest of the paper is organized as follows. In Section 2, the proposed framework is described and anti-recompression reversible data hiding in encrypted videos based on multi-domain embedding is proposed. The experimental results are shown in Section 3. Section 4 concludes the paper.

2. Proposed Framework and Scheme

In this section, the multi-domain embedding framework in encrypted videos and the anti-recompression RDH-EV scheme based on the framework are elaborated. The framework is shown in Figure 1. To achieve commutativity, data hiding and encryption usually avoid interference by modifying different attributes of the same element. In this framework, the element for robust embedding is encrypted through Logistic chaotic scrambling, which is marked as element-I. To further improve robustness, the label information will be encoded with BCH (7, 4, 1) [27,28]. Then, the label information will be robustly embedded into element-I by modulating the amplitude of element-I, in which the auxiliary information is generated for recovering the element-I lossless. The element for reversible embedding is marked as element-II, the sign of which will be encrypted by stream cipher. The auxiliary information will be reversibly embedded into element-II through traditional HS.

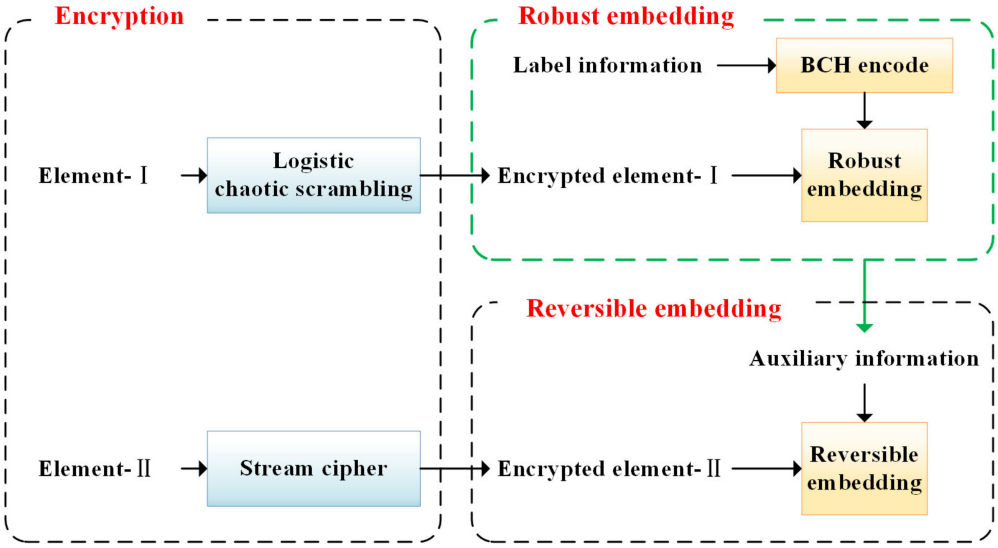

Figure 1. Framework of multi-domain embedding in encrypted video.

The scenario flow of the proposed scheme is shown in Figure 2. The scheme is mainly divided into three parts: video encryption, data embedding in encrypted video and data extraction and video recovery. In the video encryption phase, the video owner selectively encrypts elements of the original video bitstream using Logistic chaotic scrambling and the ZUC algorithm. In the data hiding phase, the data hider embeds the label information and the auxiliary information into different encrypted elements. The data hider can be a cloud service or the video owner, depending on the actual application scenarios. In the data extraction and video recovery phase, the authorized receiver extracts the information and fully recovers the original video bitstream.

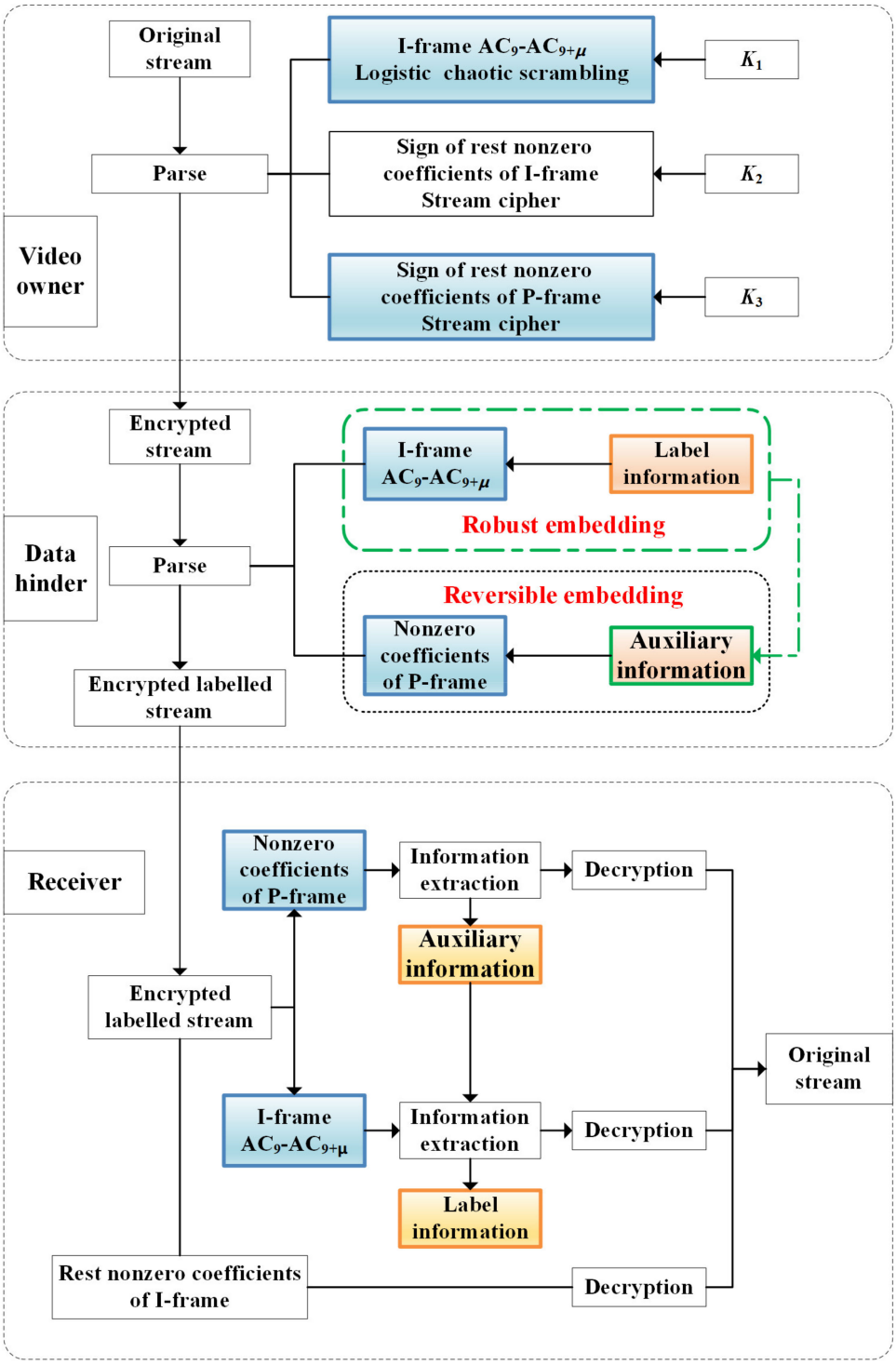

Figure 2. Scenario flow.

2.1. Video Encryption

Selective encryption selects only some syntax elements in the compression process or compressed bit stream for encryption, so as to ensure format compatibility [29,30]. In this paper, we will focus on a selective encryption scheme that can be applied to compressed H.264 bitstream. The QDCT coefficient of I-frame and the QDCT coefficient of P-frame are mainly encrypted to form two independent ciphertext fields.

As shown in Figure 2, partial QDCT of I-frame will be used for encryption and robust data hiding. It is not appropriate to encrypt the sign of QDCT coefficient, because the

proposed data hiding may change the sign of the QDCT coefficient, which will affect the normal description and make the scheme uncommutative. To keep the commutation of encryption and data hiding, Logistic chaotic scrambling is applied, which just permutes the location.

Chaotic systems have complex dynamic behavior and are widely used in the field of confidential communication. The one-dimensional Logistic mapping formula is as follows:

$$x_{n+1} = \lambda\, x_n(1 - x_n) \tag{1}$$

where $\lambda \in (0, 4]$ is Logistic parameter, and $x_n \in (0, 1)$. When $3.57 < \lambda \leq 4$, the map is in a chaotic state. The closer λ is to 4, the more uniformly the x range is distributed at (0,1).

When λ and x_0 are given, the random sequence $S = (s_1, \ldots, s_{l-1}, s_l)$ with length l can be generated according to Equation (1). Then, the index sequence $K = (k_1, \ldots, k_{l-1}, k_l)$ can be obtained according to Equation (2), which is the encryption key.

$$[S', K] = Sort(S) \tag{2}$$

In this paper, partial QDCT coefficients of I-frame are chosen as element-I. As shown in Figure 3, the partial AC coefficients, AC_9-$AC_{9+\mu}$ of a 4×4 QDCT block in non-first rows and non-first columns of an I-frame, are divided into a group, assuming $A_i = (a_1, \ldots, a_\mu, a_{\mu+1})$, where $i = 1, \ldots, n-1, n$; n is the number of QDCT blocks available in the I-frame for embedding; $\mu \in \{1, 3, 5\}$ is a robust embedding capacity parameter; and the default $\mu = 1$ in this paper. Suppose $M = (A_1, \ldots, A_{n-1}, A_n)$. Then M will be encrypted with key K.

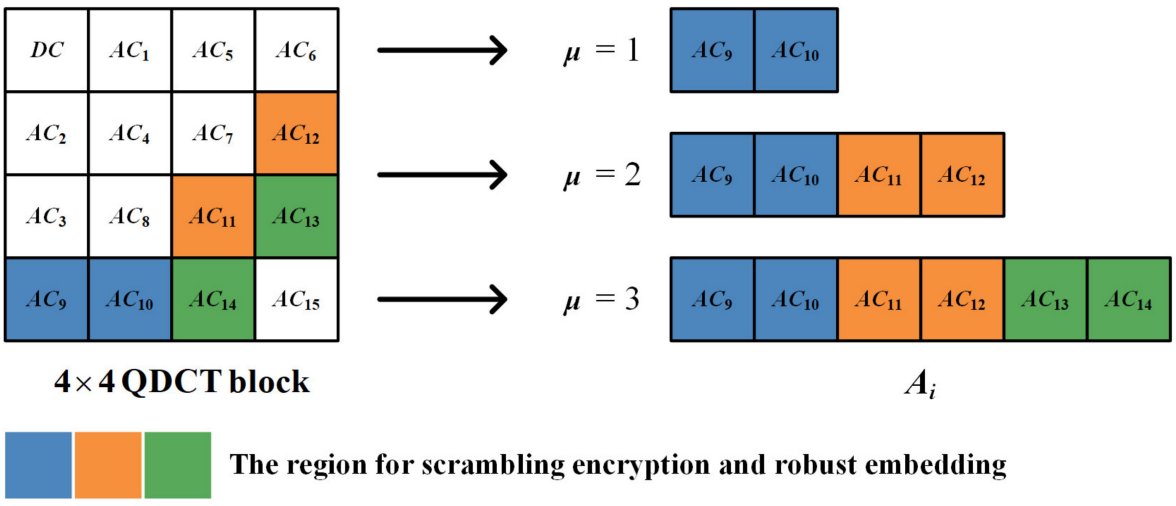

Figure 3. A set of AC coefficients.

The encryption method in the proposed scheme is as follows:

(1) Given λ and x_0, the scrambling key K_1 is obtained according to the Logistic chaotic scrambling algorithm. Then, M is encrypted with K_1, as shown in Figure 4. In this way, the encrypted domain for robust embedding is formed. Note that λ and x_0 will be reversibly embedded into the P-frame as part of the auxiliary information.

(2) The signs of the remaining non-zero QDCT coefficients of the I-frame are encrypted by stream cipher according to K_2. Specifically, the sign bits are XORed with the random sequence generated by the ZUC algorithm with the encryption key K_2.

(3) The signs of all non-zero QDCT coefficients of the P-frame are encrypted by stream cipher according to K_3. Specifically, the sign bits are XORed with the random sequence generated by the ZUC algorithm with the encryption key K_3. In this way, the encrypted domain for reversible embedding is formed.

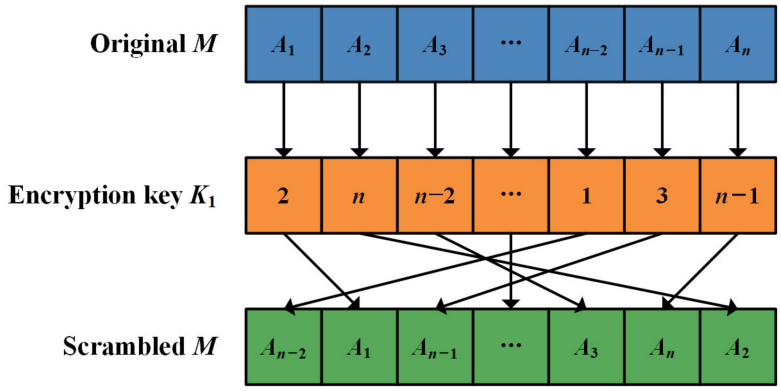

Figure 4. Scrambling encryption.

For security purposes, the IPM should also be encrypted. Details of the encryption can be found in [18–24], and will not be repeated in this paper.

2.2. Data Embedding in Encrypted Videos

1. Element selection

As described in Section 2.1, AC_9-$AC_{9+\mu}$ of I-frame are taken as the robust embedding points, which are marked as I-QDCT. Because the Logistic chaotic encryption only changes the position of QDCT coefficients, the amplitude of QDCT can be modified to a large extent without considering the change of its sign, so as to achieve strong robustness. The nonzero coefficient of P-frame is selected as the reversible embedding point, which is marked as P-QDCT. There is no doubt that element selection can be optimized further, but this paper mainly verifies the feasibility of the framework and does not focus too much on the details of processing.

2. Robust embedding in I-frame

AC_9-$AC_{9+\mu}$ can be divided into N non-overlapping coefficient pairs (b_1, b_2). One bit can be embedded into one coefficient pair, so N bits can be embedded into a 4×4 QDCT block. The information is embedded in sequence from back to front into 4×4 QDCT block. The specific embedding algorithm is as follows:

$$b_i' = \begin{cases} b_i - \left\lfloor \frac{d}{2} \right\rfloor - (1-2w)\zeta, & i = 1 \\ b_i + \left\lfloor \frac{d}{2} \right\rfloor + (1-2w)\zeta, & i = 2 \end{cases} \tag{3}$$

where ξ is the parameter that controls robustness and the default value is 1, $w \in \{0,1\}$ is information bit and d is the difference between two coefficients in a pair of coefficients.

After embedding, the difference d is modified as follow:

$$d' = b_1' - b_2' = \begin{cases} 2\zeta + z, & w = 1 \\ -2\zeta + z, & w = 0 \end{cases} \tag{4}$$

where $z = d \bmod 2$.

3. Reversible embedding in P-frame

In this stage, the auxiliary information is embedded into the nonzero coefficient of P-frame. The auxiliary information L includes Logistic parameter λ, initial value x_0, robustness control parameter ξ, robust embedding capacity parameter μ and the original difference d of each coefficient pair, which is very little overhead after Run-Huffman

encoding. Classic one-dimensional HS is used to embed the auxiliary information into P-QDCT, and its embedding method is as follow:

$$y' = \begin{cases} y + w \times \text{sign}(y), & |y| = 1 \\ y + \text{sign}(y), & \text{else} \end{cases} \tag{5}$$

where y is a single QDCT coefficient, y' is a modified coefficient, and sign() is the sign function.

4. Multi-domain embedding process

The proposed algorithm takes a group of pictures (GOP) as the basic embedding unit. For the whole video, the data embedding starts in the last GOP and the embedding is carried forward one by one. The main steps of the multi-domain embedding algorithm can be summarized as follows:

Step 1: Partially decode the encrypted video bitstream to obtain each GOP.

Step 2: Decode the encrypted I-QDCT and P-QDCT from the last GOP that has not been embedded.

Step 3: Divide the encrypted I-QDCT into non-overlapping coefficient pairs (b_1, b_2). The label information is encoded by the BCH code at first and then is embedded into the I-QDCT in a sequence from back to front according to Equation (3), in which the corresponding auxiliary information is generated.

Step 4: Embed the auxiliary information into the P-QDCT according to Equation (5) ina sequence from back to front.

Step 5: Go back to Step 2 until the label information is embedded completely.

Step 6: Encode the I-QDCT and P-QDCT to obtain the encrypted labeled video bitstream.

2.3. Information Extraction and Video Recovery

1. There are two steps to information extraction:

Step 1: Extract the auxiliary information L according to Equation (6), and restore P-QDCT according to Equation (7).

$$w = \begin{cases} 1, & |y| = 2 \\ 0, & |y| = 1 \end{cases} \tag{6}$$

$$y = y' - \text{sign}(y'), \text{ if } |y'| \geq 2 \tag{7}$$

Step 2: Extract the label information according to Equation (8), and use the auxiliary information L I to restore the I-QDCT according to Equation (9).

$$w = \begin{cases} 1, & d' > 0 \\ 0, & d' < 0 \end{cases} \tag{8}$$

$$a_i = \begin{cases} a_i' + \left\lfloor \frac{d}{2} \right\rfloor + (1 - 2w)\zeta, & i = 1 \\ a_i' - \left\lfloor \frac{d}{2} \right\rfloor - (1 - 2w)\zeta, & i = 2 \end{cases} \tag{9}$$

2. Video recovery is divided into the following two cases:

Case 1. The video is decrypted before the information is extracted. After the video decryption, for P-QDCT, the correct auxiliary information extraction can be carried out directly and the corresponding QDCT coefficient can be restored. In this case, it is necessary to use λ and x_0 in the auxiliary information to generate the key K_1, and then reverse the information extracted in the I-QDCT to obtain the correct label information.

Case 2. The information is extracted before the video is decrypted. In this case, it is only necessary to extract the information, restore the carrier in P-QDCT and I-QDCT successively, and then use the encryption key to decrypt the video to obtain the original video.

The above two situations show that the decryption and data extraction are commutative in the proposed scheme.

3. Experimental Results with Analysis

The effectiveness of the proposed scheme has been investigated through a series of simulation experiments. Section 3.1 introduces the video sequence used, the experimental runtime environment, the representative methods used for comparison and the objective evaluation metrics. The security of encrypted video is analyzed in Section 3.2. The robustness is deeply analyzed in Section 3.3. The embedding capacity is analyzed in Section 3.4. Section 3.5 reports the visual quality using both subjective visual performance and objective evaluation statistics. The bit rate increase ratio (BIR) and reversibility is analyzed in Section 3.6 and 3.7, respectively. Comparative analysis is given in Section 3.8. Further discussion and research are considered in Section 3.9.

3.1. Experimental Settings

1. Video sequences

For objectivity of the experimental results, eight well-known video sequences (i.e., foreman, carphone, salesman, news, container, coastguard and city) in QCIF format (176 × 144) are used for simulation. The video sequence can be accessed on the website of the YUV Video Sequence [31]. The eight selected video sequences are rich in variety, including rapid motion, slow motion, complex texture and simple texture. In this paper, the luminance component of the first 100 frames of video is used for experiments.

2. Experimental runtime environment and parameter setting

All simulation experiments in this section are conducted on a PC equipped with an Intel i7-8550U 4 GHz CPU and 8 GB memory. Simulations were run in MATLAB R2019a. The matlab implementation of H.264 is used for simulation, and can be accessed on the website of MathWorks [32]. The GOP is set as "IPPP" with length 20. The default quantization parameter (QP) is set as 28.

3. Representative methods used for comparison

Another five different representative schemes have been selected for comparison [21–25]. These five schemes were chosen to meet two criteria: first, these schemes are reversible and separable. Second, the five schemes can be implemented in H.264, although [25] is for HEVC. All of the selected schemes were implemented following the parameter-setting in the cited references. In [12], the parameter $\beta = 0$ or 1. In this experience, $\beta = 1$ is chosen for comparison.

4. Objective evaluation metrics

Four metrics are adopted for evaluation: peak signal-to-noise ratio (PSNR), structural similarity (SSIM), bit error rate (BER) and BIR.

PSNR is commonly used to measure the difference between two images. The larger the PNSR, the smaller the difference between the two images. Different from the absolute error measured by PSNR, SSIM is a perception model, which is more in line with the intuitive feeling of the human eye. The larger the SSIM, the more similar the two images are; values range from −1 to 1. BER in this paper is based on the number of the extracted bits with error divided by the number of all the embedded bits, which can effectively measure the robustness of the proposed algorithm. BIR is introduced to measure the variation of video bit rate, which can be calculated according to Equation (10).

$$\text{BIR} = \frac{BR\_em - BR\_ori}{BR\_ori} \times 100\% \tag{10}$$

where *BR_em* is the bitrate after the video is encrypted or embedded with information, and *BR_ori* is the original video bitrate.

3.2. Security Analysis of Encrypted Video

Selective encryption for video is designed to prevent unauthorized users from having complete and clear content. Selective encryption for video needs to meet cryptographic security and perceptual security [19]. In terms of cryptographic security, the ZUC algorithm and the Logistic chaos scrambling algorithm are used in the proposed scheme to ensure this aspect of security. Perceptual safety is mainly evaluated by subjective video quality and objective evaluation criteria (i.e., the PSNR and the SSIM).

The subjective results are shown in Figures 5 and 6. The video content cannot be distinguished from the figure, achieving protection of the video content. As shown in Table 1, the PSNR and the SSIM of each encrypted video sequence are provided. The PSNR of the encrypted video does not exceed 15 dB, and the SSIM does not exceed 0.25. Although the PSNR and the SSIM of the same video with different QP fluctuate after encrypting, the encryption effect can still meet the requirements of scrambling content. In addition, the histogram is compared for analysis. As shown in Figure 7, the histograms of encrypted frames are different from those of original frames, which shows encryption with a good performance. From both subjective and objective aspects, the video encryption in the proposed scheme can meet the requirement of perceptual security.

Figure 5. The original frames.

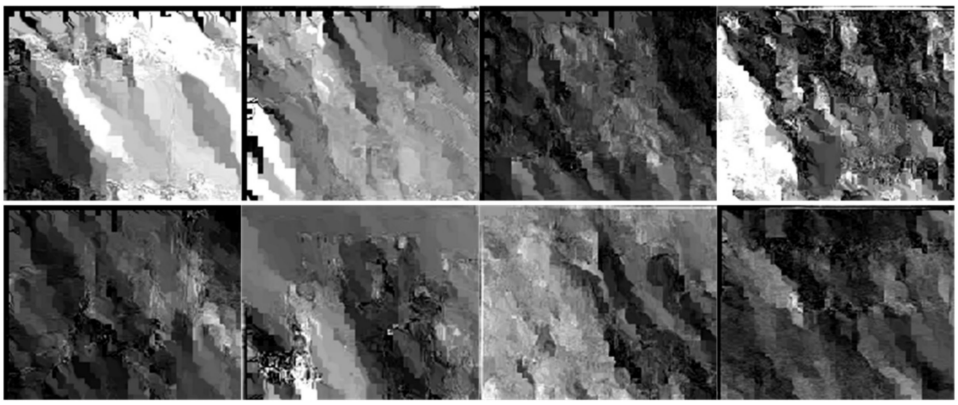

Figure 6. The corresponding encrypted frames.

Table 1. PSNR and SSIM after encryption.

Video Sequence	QP	PSNR/dB	SSIM
container	24	4.6496	0.1155
	28	5.4236	0.1430
	32	8.7477	0.0689
foreman	24	5.9519	0.0065
	28	6.9626	0.0280
	32	10.1921	0.2689
salesman	24	12.2562	0.0239
	28	13.8443	0.2343
	32	12.3003	0.1822
mobile	24	5.6959	0.0293
	28	5.9292	0.1186
	32	3.5338	0.0637
carphone	24	8.2273	0.0769
	28	8.8277	0.0030
	32	8.2801	0.1087
news	24	10.1027	0.0361
	28	10.3231	0.1732
	32	11.3371	0.2230
city	24	9.8430	0.0578
	28	9.4672	0.0079
	32	11.8376	0.1018
coastguard	24	8.2926	0.2086
	28	6.8150	0.0350
	32	3.4344	0.1214

3.3. Robustness Analysis

Robustness refers to the anti-interference ability of the data hiding algorithm. The robustness of a scheme is evaluated by its extraction accuracy (1-BER) under attacks. The higher the accuracy, the stronger the robustness. In this experiment, the representative algorithms [21–25] are selected for a comparative experiment. For consistency, all algorithms are tested in I-frame. Figure 8 compares the extraction accuracy (1-BER) of the proposed algorithm and the representative algorithms. As seen from the figure, the proposed scheme outperforms the compared schemes in terms of robustness. The proposed algorithm can achieve an accuracy more than 90% under low QP re-compression attacks, while the accuracies of the algorithms in [21–25] fluctuate around 50% in eight test sequences. Although the BCH code is used in [23], the algorithm cannot resist a re-compression attack, like the other four algorithms. This is because the embedding algorithms used in [21–25] are all traditional HS, with which the information needs to be extracted on accurate amplitudes (i.e., 1 and -1). Once one of the accurate amplitudes is quantized to 0 under the re-compression attack, it will affect the extraction of not only the current bit, but also the subsequent bits. As a result, most of the extracted information will be wrong and the BCH code will be unable to work. In the proposed algorithm, the information is extracted by distinguishing the positive and negative of d'. For example, although the coefficient pair (2, 1) containing bit '1' is quantized to (1, 0), the correct bit '1' can still be extracted according to Equation (10). Moreover, the embedding and extracting area is fixed, and the extraction error of any bit will not affect the extraction of other bits. Therefore, the proposed algorithm can better resist the re-compression attack.

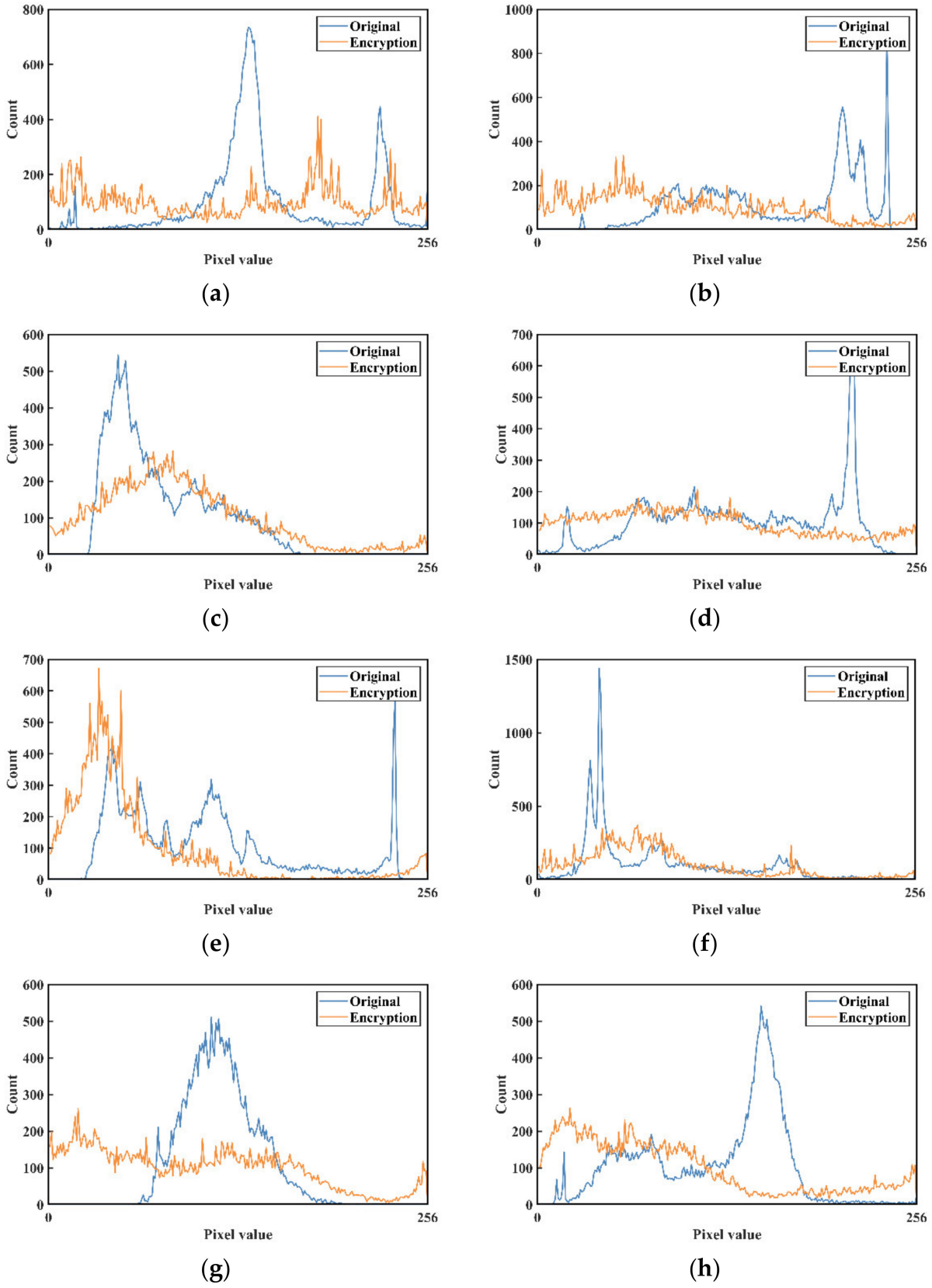

Figure 7. Histogram of original and encrypted frame. (**a**) container, (**b**) foreman, (**c**) salesman, (**d**) mobile, (**e**) carphone, (**f**) news, (**g**) city, (**h**) coastguard.

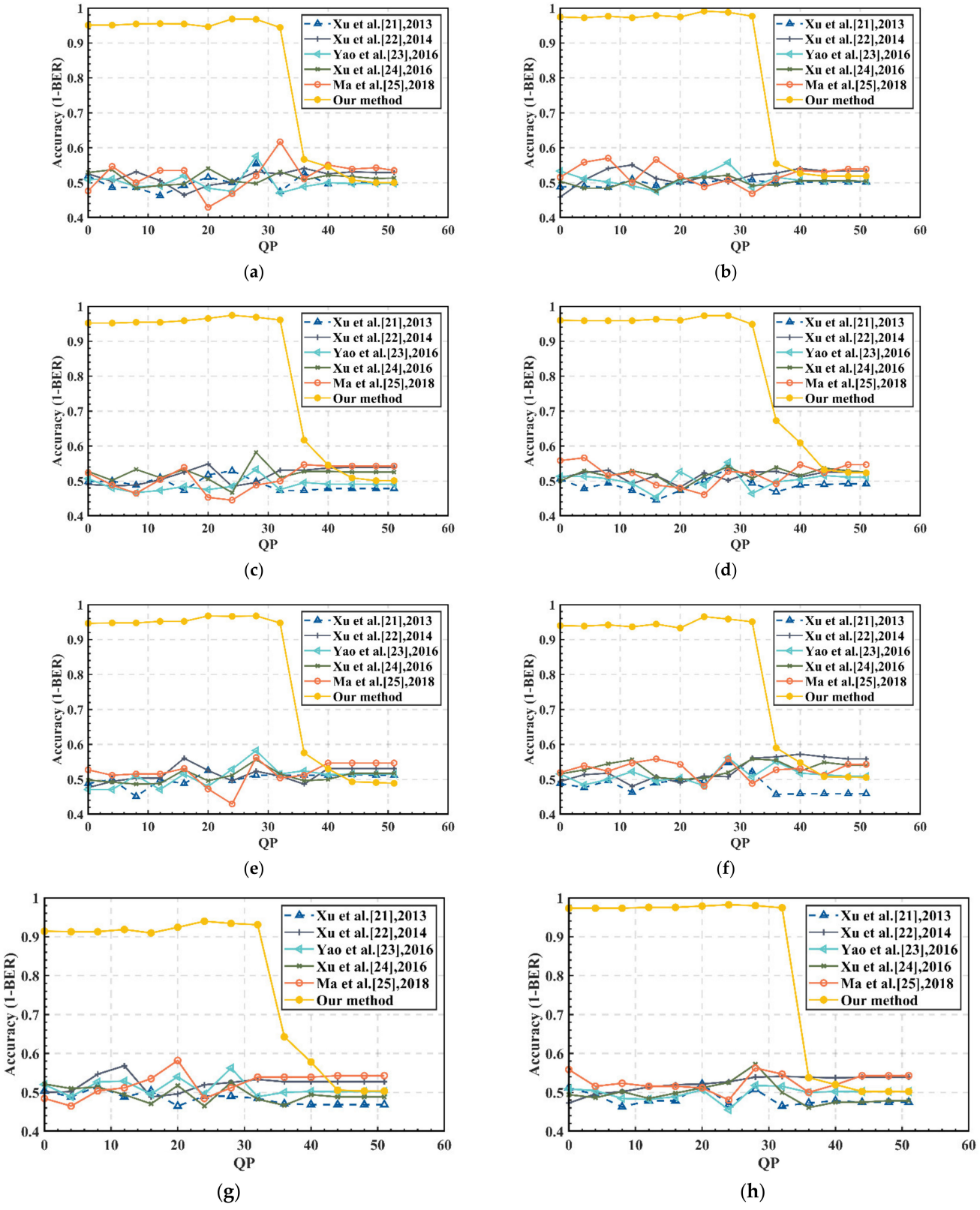

Figure 8. Extraction accuracy (1-BER) comparison between our method and the current methods [21–25] (original QP = 28). (**a**) container, (**b**) foreman, (**c**) salesman, (**d**) mobile, (**e**) carphone, (**f**) news, (**g**) city, (**h**) coastguard.

Table 2 displays the robustness of the proposed algorithm and the accuracy of label information extracted from the labeled video under different QP re-compression attacks.

Under the re-compression attack, some small QDCT coefficient pairs will be quantized to (0, 0) in the re-compression with large QP, so the information embedded in them may be lost. As can be seen from Table 2, when $\xi = 1$, it is difficult to resist the re-compression attack with QP $\geq$ 36. This is because the proposed algorithm mainly uses the medium and high frequency coefficients, most of which are small. When the re-compression QP is large and the parameter ξ is small, the labeled QDCT coefficient can be easily quantized to 0. When $\xi \geq 2$, the proposed algorithm can be effective against large QP re-compression attacks. Figure 9 displays the relationship between ξ and the ability to resist re-compression attacks. Obviously, the larger the ξ, the stronger the ability to resist re-compression attacks. That is because the larger the ξ, the larger the modified coefficient. A large coefficient is difficult to quantize to 0 so the information can be extracted correctly under a considerable degree of re-compression attacks. Figure 10 displays the extraction accuracy of foreman, salesman and carphone video sequences with different original encoding QP values under a QP = 40 re-compression attack. It can be seen that the larger the original encoding QP value, the larger the re-compression QP value that can be resisted. When the original encoding QP $\geq$ 32, it can effectively resist the QP = 40 re-compression attack. As seen from Figures 8–10, the proposed scheme not only outperforms other schemes in terms of robustness but also can adapt to the robustness requirements in different scenarios by setting ξ and original encoding QP.

Table 2. Accuracy of information extraction.

Video Sequence	ξ	Accuracy (1-BER)						
		QP = 16	QP = 20	QP = 24	QP = 28	QP = 32	QP = 36	QP = 40
container	1	0.9542	0.9464	0.9687	0.9676	0.9883	0.5669	0.5457
	2	0.9575	0.9486	0.9564	0.9620	0.9118	0.9676	0.8805
	3	0.9654	0.9676	0.9642	0.9575	0.9665	0.9308	0.9732
foreman	1	0.9787	0.9743	0.9910	0.9877	0.9765	0.5546	0.5267
	2	0.9843	0.9854	0.9754	0.9810	0.9520	0.9910	0.8973
	3	0.9832	0.9843	0.9854	0.9832	0.9888	0.9575	0.9966
salesman	1	0.9587	0.9654	0.9743	0.9687	0.9609	0.6171	0.5457
	2	0.9665	0.9709	0.9642	0.9698	0.9720	0.9888	0.8962
	3	0.9754	0.9754	0.9776	0.9787	0.9866	0.9676	0.9910
mobile	1	0.9631	0.9598	0.9732	0.9732	0.9486	0.6729	0.6093
	2	0.9709	0.9709	0.9698	0.9709	0.9620	0.9787	0.8928
	3	0.9665	0.9743	0.9720	0.9732	0.9654	0.9508	0.9843
carphone	1	0.9520	0.9676	0.9665	0.9676	0.9475	0.5758	0.5301
	2	0.9642	0.9665	0.9676	0.9709	0.9341	0.9799	0.8660
	3	0.9743	0.9754	0.9810	0.9743	0.9765	0.9475	0.9877
news	1	0.9441	0.9330	0.9654	0.9587	0.9508	0.5904	0.5479
	2	0.9598	0.9631	0.9654	0.9642	0.9397	0.9765	0.9006
	3	0.9654	0.9687	0.9687	0.9654	0.9709	0.9308	0.9821
city	1	0.9095	0.9241	0.9397	0.9341	0.9308	0.6428	0.5781
	2	0.9408	0.9430	0.9453	0.9453	0.9330	0.9587	0.8671
	3	0.9453	0.9430	0.9553	0.9441	0.9575	0.9363	0.9843
coastguard	1	0.9754	0.9787	0.9821	0.9799	0.9743	0.5379	0.5200
	2	0.9720	0.9698	0.9698	0.9776	0.9654	0.9877	0.8939
	3	0.9699	0.9721	0.9732	0.9699	0.9810	0.9509	0.9843

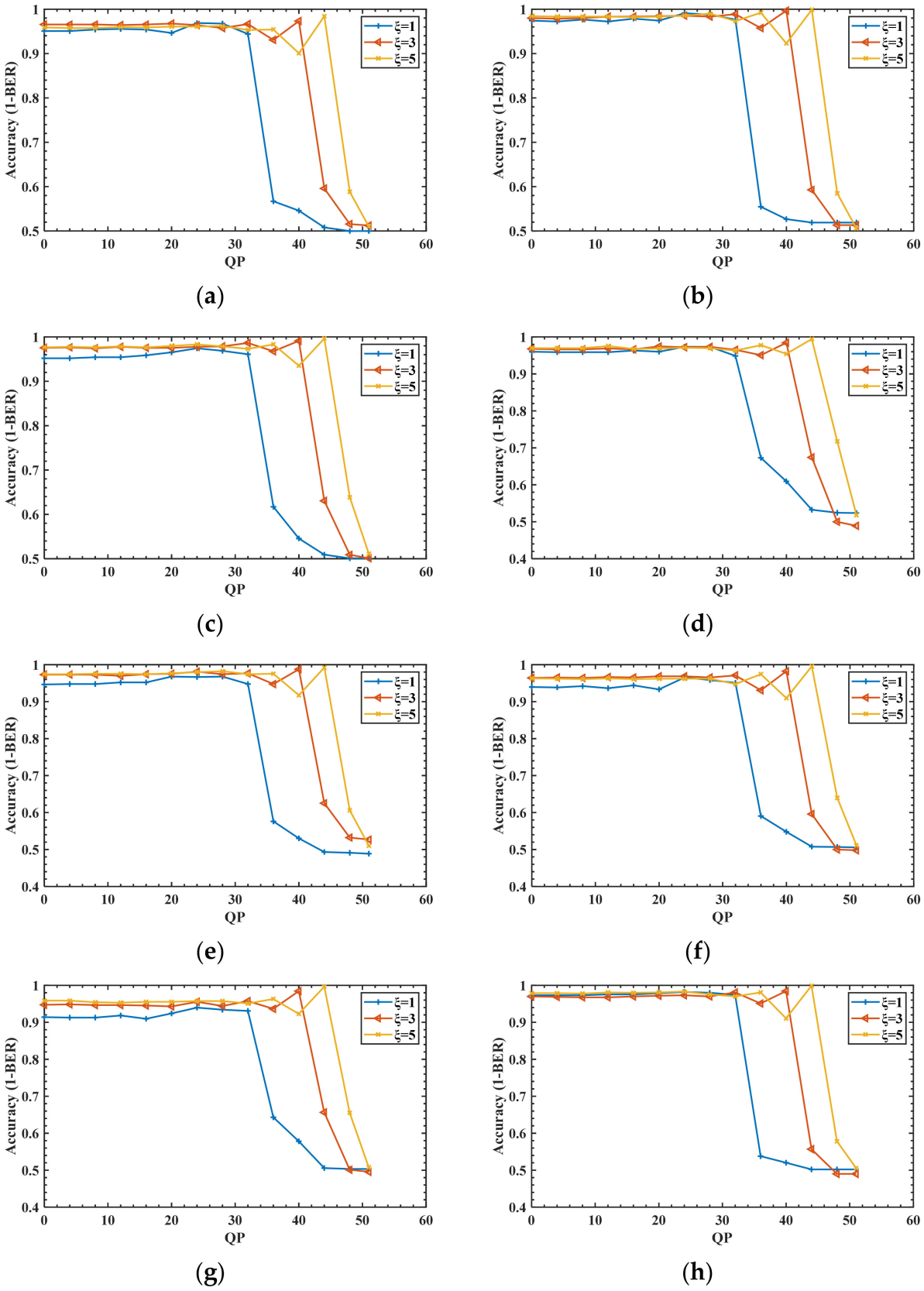

Figure 9. Accuracy of information extraction of different ξ. (**a**) container, (**b**) foreman, (**c**) salesman, (**d**) mobile, (**e**) carphone, (**f**) news, (**g**) city, (**h**) coastguard.

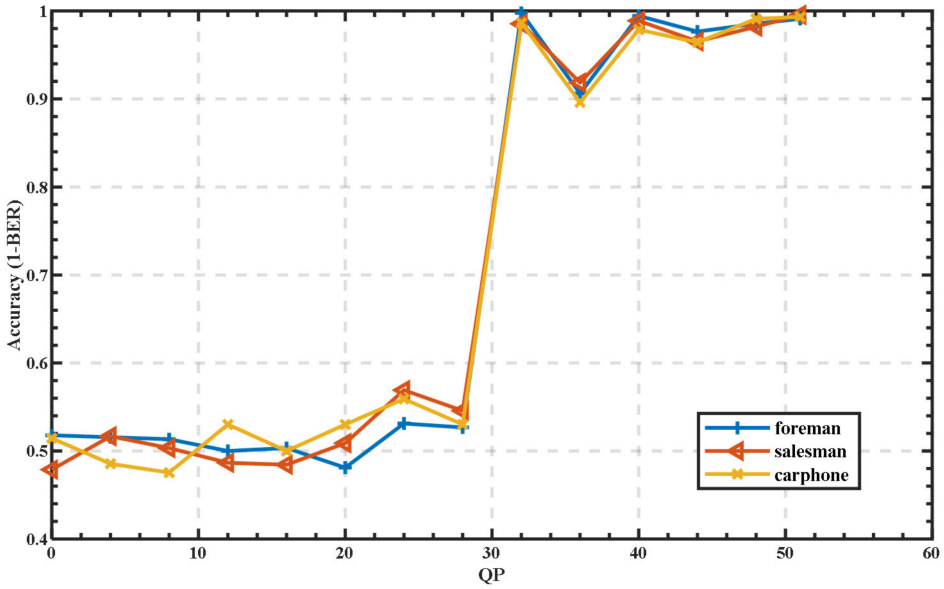

Figure 10. Accuracy of information extraction under QP = 40 re-compression attack.

3.4. Embedding Capacity

For the video, the embedding capacity requirement is not high because the information can be embedded in several frames to meet the embedding capacity requirement. For the proposed algorithm, the embedding capacity of a single frame can be calculated according to t Equation (11).

$$EC = \frac{h \times w \times (\mu + 1)}{size \times 2} \tag{11}$$

where h is the height of the frame, w is the width of the frame, $\mu \in \{1, 3, 5\}$ is the embedding capacity parameter, and *size* is the size of the QDCT block which is 4×4 in this paper.

For the QCIF format video, when $\mu = 1$, the embedding capacity is 1584 bits. When $\mu = 3$, the embedding capacity is 3168 bits. When $\mu = 5$, the embedding capacity is 4752 bits. Of course, in practice, more information is not embedded into a single frame. When the embedding capacity is higher, the video quality will obviously decrease accordingly, and a compromise option can be selected according to actual needs.

3.5. Visual Quality of the Decrypted Labeled Video

In some cases, authorized users will decrypt the encrypted video directly in order to browse the similar video quickly and the decrypted video still contains the label information. In this experiment, the visual quality of decrypted labeled videos will be evaluated. As shown in Figures 5 and 11, from the perspective of human eyes, it is difficult to detect the difference between the decrypted labeled video and the original video. Table 3 provides the PSNR and the SSIM of each video sequence with 256 bits of the label information. It can be seen that the video quality is reduced to varying degrees in the same video with different QP. For QP = 24, the average ΔPSNR and ΔSSIM are 4.02 dB and 0.0032, respectively. For QP = 28, the average ΔPSNR and ΔSSIM are 2.97 dB and 0.0049, respectively. For QP = 32, the average ΔPSNR and ΔSSIM are 3.33 dB and 0.0094, respectively. The ΔPSNR of the same decrypted labeled video with different QP fluctuates, but the ΔPSNR is not more than 5.48 dB at most. Judging from the SSIM which is more in line with human visual characteristics than the PSNR, as QP increases, the ΔSSIM increases in most video sequences. However, the ΔSSIM is small and not more than 0.0253 at most, illustrating that the proposed algorithm has good imperceptibility. Therefore, the proposed scheme can meet the needs of directly decrypting to obtain similar videos in the cloud environment.

Figure 11. The corresponding decrypted labeled frames.

Table 3. PSNR and SSIM of decrypted labeled videos.

Video Sequence	QP	PSNR/dB		ΔPSNR/dB	SSIM		ΔSSIM
		Original	Labeled		Original	Labeled	
container	24	40.70	37.89	2.81	0.9987	0.9976	0.0011
	28	37.85	34.87	2.98	0.9975	0.9951	0.0024
	32	34.65	31.18	3.47	0.9949	0.9886	0.0063
foreman	24	40.51	38.54	1.97	0.9990	0.9984	0.0006
	28	37.67	36.06	1.61	0.9981	0.9971	0.0010
	32	34.68	32.79	1.89	0.9962	0.9939	0.0023
salesman	24	39.90	34.47	5.43	0.9971	0.9894	0.0077
	28	36.66	32.78	3.88	0.9939	0.9844	0.0095
	32	33.47	30.47	3.00	0.9873	0.9738	0.0135
mobile	24	39.16	33.86	5.30	0.9989	0.9963	0.0026
	28	35.56	32.51	3.05	0.9974	0.9947	0.0027
	32	31.75	29.84	1.91	0.9937	0.9905	0.0032
carphone	24	41.06	38.88	2.18	0.9992	0.9987	0.0005
	28	38.08	36.17	1.91	0.9985	0.9977	0.0008
	32	34.98	29.88	5.10	0.9969	0.9902	0.0067
news	24	41.16	35.80	5.36	0.9990	0.9966	0.0024
	28	38.13	32.65	5.48	0.9980	0.9930	0.0050
	32	34.76	29.54	5.22	0.9958	0.9858	0.0100
city	24	41.75	37.28	4.47	0.9962	0.9886	0.0076
	28	36.93	33.14	3.79	0.9885	0.9712	0.0173
	32	33.48	30.71	2.77	0.9746	0.9493	0.0253
coastguard	24	39.34	34.69	4.65	0.9982	0.9949	0.0033
	28	36.09	35.03	1.06	0.9963	0.9953	0.0010
	32	32.98	29.72	3.26	0.9924	0.9842	0.0082

ΔPSNR = PSNR_original − PSNR_labeled, ΔSSIM = SSIM_original − SSIM_labeled.

3.6. Bit Rate Variation Analysis

During the encryption process, the encryption of the QDCT sign does not affect the bitrate due to the encoding characteristics. Scrambling encryption of the same area of different QDCT blocks will not seriously affect the overall distribution and the I-frame is less, so it will not significantly increase the bit rate. When embedding information, the embedding algorithm increases or decreases the amplitude of coefficient, and it is possible to modify the coefficient 0, so the bit rate will be increased. Table 4 provides the BIR caused by encrypting and embedding 256 bits of label information The bit rate changes differently in the same video sequence with different QP. The BIR caused by encryption fluctuates in the same encrypted video with different QP, but the BIR is no more than 1.14% at most. It

depends on the distribution of QDCT after scrambling encryption. As QP increases, the BIR caused by embedding increases. This is because with the increase in QP, the amplitude of QDCT becomes smaller and tends to 0 so that the probability of modifying the coefficient 0 increases. It is inevitable that with the increase in the embedding rate, the BIR will also increase. Therefore, a proper embedding rate need to be selected to control the BIR in some scenarios.

Table 4. BIR due to encryption and embedding.

Video Sequence	QP	BIR/%	
		Encryption	Embedding
container	24	0.89	0.92
	28	0.92	1.47
	32	0.69	2.26
foreman	24	0.24	0.48
	28	0.17	0.87
	32	0.08	1.52
salesman	24	0.42	0.36
	28	0.40	0.75
	32	0.31	1.62
mobile	24	0.26	0.10
	28	0.27	0.18
	32	0.24	0.39
carphone	24	0.47	0.59
	28	0.45	1.04
	32	0.34	1.70
news	24	1.14	0.72
	28	1.14	1.24
	32	0.89	1.95
city	24	0.24	0.44
	28	0.18	0.75
	32	0.14	1.41
coastguard	24	0.11	0.36
	28	0.08	0.64
	32	0.05	1.22

3.7. Reversibility Analysis

The reversibility of a scheme means that the data hiding it performs is reversible. Specially, the labeled video can be lossless when recovered after data extraction. The proposed scheme embeds the label information into the I-QDCT, generates the auxiliary information required for reversible extraction, and then uses HS to reversibly embed the auxiliary information into the P-QDCT. When extracting information, extract the auxiliary information in the P-frame and restore the P-QDCT at the same time. Then, while extracting the label information from the I-frame, restore the I-QDCT using the auxiliary information. The I-frame restoration process can only be carried out smoothly under the condition of successful extraction of auxiliary information, otherwise only the label information can be extracted from the I-frame and the carrier cannot be restored.

Figure 12 displays the original PSNR of video sequences and the PSNR of video sequences after information embedding and extraction. The PSNR of two situations corresponding to each video sequence are seen to be consistent. That is to say, the proposed scheme can fully recover the carrier and is reversible.

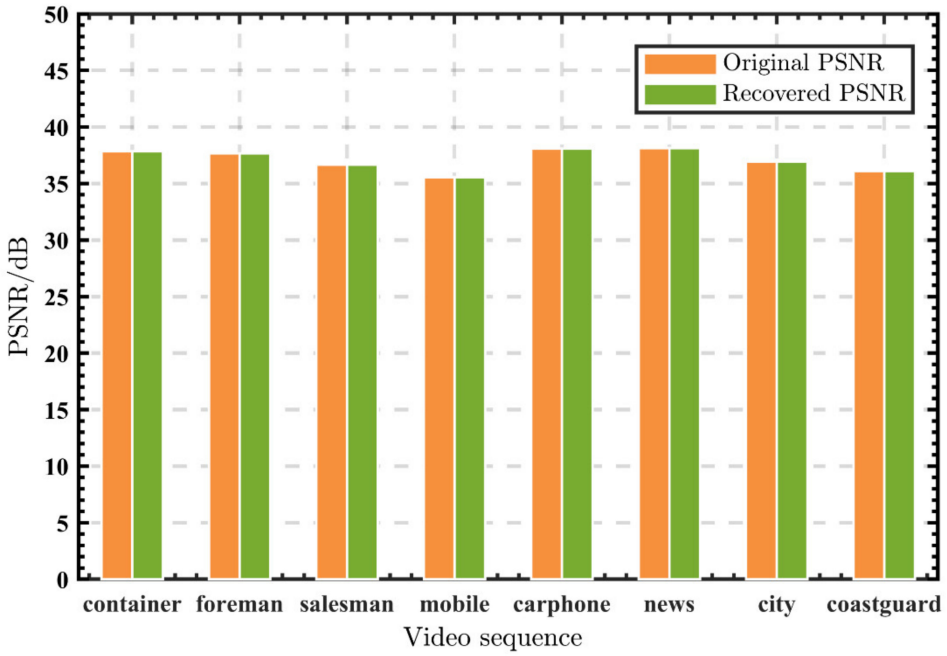

Figure 12. PSNR comparison.

3.8. Qualitative Analysis

As shown in Table 5, qualitative analysis of the related schemes is given. It can be seen that the existing schemes in [25–28] do not make full use of encrypted elements and do not take both reversibility and robustness into account. The existing schemes in [18–25] only utilize one encrypted element to embed information, while the proposed scheme utilizes two encrypted elements for embedding information. The scheme in [18] is robust but not reversible. The schemes in [19,20] are neither reversible nor robust. The schemes in [21–25] are reversible but not robust. The proposed scheme achieves robustness, reversibility and separability while maintaining format compliant.

Table 5. Qualitative analysis of the related schemes.

Scheme	Elements for Encryption	Data Embedding	Bit Rate Increase	Separability	Format Compliant	Reversibility	Robustness
[18]	IPM	No	Yes	Yes	Yes	No	Yes
	sign of MVD	No					
	sign of QDCT	Amplitude of QDCT					
[19]	IPM	No	No	Yes	Yes	No	No
	sign of MVD	No					
	sign of QDCT	Codeword *Level*					
[20]	IPM	No	No	Yes	Yes	No	No
	sign of MVD	No					
	sign of QDCT	Codeword *Level*					
[21]	IPM	No	Yes	Yes	Yes	Yes	No
	sign of MVD	No					
	sign of QDCT	Amplitude of QDCT					

Table 5. *Cont.*

Scheme	Elements for Encryption	Data Embedding	Bit Rate Increase	Separability	Format Compliant	Reversibility	Robustness
[22]	IPM	No	Yes	Yes	Yes	Yes	No
	sign of MVD	No					
	sign of QDCT	Amplitude of QDCT					
[23]	IPM	No	Yes	Yes	Yes	Yes	No
	sign of MVD	No					
	sign of QDCT coefficient	Amplitude of QDCT coefficient					
[24]	IPM	No	Yes	Yes	Yes	Yes	No
	sign of MVD	No					
	sign of QDCT coefficient	Amplitude of QDCT coefficient					
[25]	Sign of MVD	No	Yes	Yes	Yes	Yes	No
	amplitude of MVD	No					
	sign of QDCT	Amplitude of QDCT					
Proposed	IPM	No	Yes	Yes	Yes	Yes	Yes
	sign of QDCT	Amplitude of QDCT					
	position of QDCT	Amplitude of QDCT					

3.9. Further Discussion

1. Coefficient selection for scrambling and embedding

In this paper, AC_9–$AC_{9+\mu}$ were coarsely selected for scrambling encryption and robust embedding in order to verify the feasibility of the framework of multi-domain embedding in encrypted videos. There is no doubt that the selection of coefficients can be further optimized to improve robustness and imperceptibility. However, this was not the focus of this paper. Research can be carried out based on this in the follow-up work.

2. Element selection for robust embedding

In this paper, the QDCT coefficient of I-frame was selected for robust embedding because in the process of video bitstream transmission, ensuring the complete transmission of I-frame information is critical and should be prioritized. Therefore, to protect the label information comprehensively, the information is embedded into the I-frame, which, however, will affect the visual quality of the corresponding decrypted GOP. For the sake of visual quality, the elements of P-frame are obviously more appropriate. In this paper, we aimed to present the framework of multi-domain embedding in encrypted video, which can simultaneously achieve reversibility and robustness. As for which elements are selected for embedding, it mainly depends on whether the scheme focuses more on protecting label information or protecting video quality.

3. Robust embedding algorithm

A robust embedding algorithm is the key to determining the robustness of the scheme. At the same time, the modification caused by a robust embedding algorithm is an important factor affecting video quality. How to balance the robustness and imperceptibility of the algorithm is the focus of follow-up research based on the framework.

4. Conclusions

In this paper, a multi-domain embedding framework in encrypted videos was proposed, in which the multi-domain characteristic of encrypted video is ultilized to achieve both robustness and reversibility. In the framework, one element was encrypted by Logistic chaotic scrambling and the other element was encrypted by stream cipher. The label information was embedded into the former element, generating the corresponding auxiliary

information for lossless recovery of the former element. The auxiliary information was reversibly embedded into the latter element. An anti-recompression RHD-EV based on the framework was proposed to verity the feasibility of the framework. The robustness experiment showed that the proposed scheme outperforms the compared schemes. Under the low QP re-compression attack, the accuracy of data extraction in the proposed scheme was above 90%, while the accuracy of data extraction in compared schemes fluctuated around 50%. In the experiment of reversibility, the PSNR of the recovered video is consistent with the PSNR of the original video, which shows that the proposed scheme is perfectly reversible. The results of the two experiments in the paper demonstrate that compared with existing schemes, the scheme based on the framework can achieve both robustness and reversibility. In addition, encryption and data hiding are commutative for meeting different application scenarios.

Of course, the proposed framework can be further improved. For example, the Logistic chaotic scrambling algorithm will cause an increase in the bit rate. This problem can be solved by scrambling based on the whole QDCT block in the follow-up research. In addition, the proposed scheme based on the framework has its limitations, mentioned in Section 3.9. In future work, we will try to design a more robust embedding algorithm based on the framework to resist more attacks.

Author Contributions: Conceptualization, P.C.; Methodology, Z.Z.; Project administration, X.Y.; Validation, K.N.; Visualization, Z.Z.; Writing—original draft, P.C.; Writing—review and editing, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by the National Natural Science Foundation of China, grant number 61872384.

Data Availability Statement: Not applicable.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Mansour, R.F.; Parah, S.A. Reversible Data Hiding for Electronic Patient Information Security for Telemedicine Applications. *Arab. J. Sci. Eng.* **2021**, *46*, 9129–9144. [CrossRef]
2. Zhang, X. Reversible Data Hiding in Encrypted Image. *IEEE Signal Processing Lett.* **2011**, *18*, 255–258. [CrossRef]
3. Yin, Z.; Peng, Y.; Xiang, Y. Reversible Data Hiding in Encrypted Images Based on Pixel Prediction and Bit-Plane Compression. *IEEE Trans. Dependable Secur. Comput.* **2022**, *19*, 992–1002. [CrossRef]
4. Kong, W.; Miao, Q.; Liu, R.; Lei, Y.; Cui, J.; Xie, Q. Multimodal Medical Image Fusion Using Gradient Domain Guided Filter Random Walk and Side Window Filtering in Framelet Domain. *Inf. Sci.* **2022**, *585*, 418–440. [CrossRef]
5. Balasamy, K.; Suganyadevi, S. A Fuzzy Based ROI Selection for Encryption and Watermarking in Medical Image Using DWT and SVD. *Multim. Tools Appl.* **2021**, *80*, 7167–7186. [CrossRef]
6. Xu, D. Commutative Encryption and Data Hiding in HEVC Video Compression. *IEEE Access* **2019**, *7*, 66028–66041. [CrossRef]
7. Guan, B.; Xu, D.; Li, Q. An Efficient Commutative Encryption and Data Hiding Scheme for HEVC Video. *IEEE Access* **2020**, *8*, 60232–60245. [CrossRef]
8. Xu, D.; Guan, B. An Improved Commutative Encryption and Data Hiding Scheme for HEVC Video. *Multim. Tools Appl.* **2022**, *81*, 18105–18127. [CrossRef]
9. Zhang, X. Separable Reversible Data Hiding in Encrypted Image. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 826–832. [CrossRef]
10. Wang, X.; Chang, C.-C.; Lin, C.-C.; Chang, C.-C. Reversal of Pixel Rotation: A Reversible Data Hiding System towards Cybersecurity in Encrypted Images. *J. Vis. Commun. Image Represent.* **2022**, *82*, 103421. [CrossRef]
11. Yu, C.; Zhang, X.; Li, G.; Zhan, S.; Tang, Z. Reversible Data Hiding with Adaptive Difference Recovery for Encrypted Images. *Inf. Sci.* **2022**, *584*, 89–110. [CrossRef]
12. Qin, C.; Zhang, X. Effective Reversible Data Hiding in Encrypted Image with Privacy Protection for Image Content. *J. Vis. Commun. Image Represent.* **2015**, *31*, 154–164. [CrossRef]
13. Yin, Z.; Xiang, Y.; Zhang, X. Reversible Data Hiding in Encrypted Images Based on Multi-MSB Prediction and Huffman Coding. *IEEE Trans. Multim.* **2020**, *22*, 874–884. [CrossRef]
14. Yin, Z.; She, X.; Tang, J.; Luo, B. Reversible Data Hiding in Encrypted Images Based on Pixel Prediction and Multi-MSB Planes Rearrangement. *Signal Processing* **2021**, *187*, 108146. [CrossRef]
15. Huang, D.; Wang, J. High-Capacity Reversible Data Hiding in Encrypted Image Based on Specific Encryption Process. *Signal Processing Image Commun.* **2020**, *80*, 115632. [CrossRef]

16. Ke, Y.; Zhang, M.-Q.; Liu, J.; Su, T.; Yang, X. Fully Homomorphic Encryption Encapsulated Difference Expansion for Reversible Data Hiding in Encrypted Domain. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 2353–2365. [CrossRef]
17. Ke, Y.; Zhang, M.; Zhang, X.; Liu, J.; Su, T.; Yang, X. A Reversible Data Hiding Scheme in Encrypted Domain for Secret Image Sharing Based on Chinese Remainder Theorem. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 2469–2481. [CrossRef]
18. Lian, S.; Liu, Z.; Ren, Z.; Wang, H. Commutative Encryption and Watermarking in Video Compression. *IEEE Trans. Circuits Syst. Video Technol.* **2007**, *17*, 774–778. [CrossRef]
19. Xu, D.; Wang, R.; Shi, Y.Q. Data Hiding in Encrypted H.264/AVC Video Streams by Codeword Substitution. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 596–606. [CrossRef]
20. Xu, D.; Wang, R.; Shi, Y.Q. An Improved Scheme for Data Hiding in Encrypted H.264/AVC Videos. *J. Vis. Commun. Image Represent.* **2016**, *36*, 229–242. [CrossRef]
21. Xu, D.; Wang, R.; Shi, Y.-Q. Reversible Data Hiding in Encrypted H.264/AVC Video Streams. In *Proceedings of the Digital-Forensics and Watermarking—12th International Workshop, IWDW 2013, Auckland, New Zealand, 1–4 October 2013*; Revised Selected Papers; Shi, Y.-Q., Kim, H.-J., Pérez-González, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8389, pp. 141–152.
22. Xu, D.; Wang, R. Efficient Reversible Data Hiding in Encrypted H.264/AVC Videos. *J. Electron. Imaging* **2014**, *23*, 053022. [CrossRef]
23. Yao, Y.; Zhang, W.; Yu, N. Inter-Frame Distortion Drift Analysis for Reversible Data Hiding in Encrypted H.264/AVC Video Bitstreams. *Signal Processing* **2016**, *128*, 531–545. [CrossRef]
24. Xu, D.; Zhu, Y.; Wang, R.; Fu, J.; Chen, K. Two-Dimensional Histogram Modification for Reversible Data Hiding in Partially Encrypted H.264/AVC Videos. In *Proceedings of the Digital Forensics and Watermarking—15th International Workshop, IWDW 2016, Beijing, China, 17–19 September 2016*; Revised Selected Papers; Shi, Y.-Q., Kim, H.-J., Pérez-González, F., Liu, F., Eds.; Springer: Cham, Switzerland, 2016; Volume 10082, pp. 393–406.
25. Long, M.; Peng, F.; Li, H. Separable Reversible Data Hiding and Encryption for HEVC Video. *J. Real Time Image Processing* **2018**, *14*, 171–182. [CrossRef]
26. Wang, X.; Li, X.; Pei, Q. Independent Embedding Domain Based Two-Stage Robust Reversible Watermarking. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 2406–2417. [CrossRef]
27. Liu, Y.; Hu, M.; Ma, X.; Zhao, H. A New Robust Data Hiding Method for H.264/AVC without Intra-Frame Distortion Drift. *Neurocomputing* **2015**, *151*, 1076–1085. [CrossRef]
28. Liu, Y.; Li, Z.; Ma, X.; Liu, J. A Robust without Intra-Frame Distortion Drift Data Hiding Algorithm Based on H.264/AVC. *Multim. Tools Appl.* **2014**, *72*, 613–636. [CrossRef]
29. Dolati, N.; Shirazi, A.A.B.; Azadegan, H. A Selective Encryption for H.264/AVC Videos Based on Scrambling. *Multim. Tools Appl.* **2021**, *80*, 2319–2338. [CrossRef]
30. He, J.; Xu, Y.; Luo, W.; Tang, S.; Huang, J. A Novel Selective Encryption Scheme for H.264/AVC Video with Improved Visual Security. Signal Process. *Image Commun.* **2020**, *89*, 115994. [CrossRef]
31. YUV Video Sequences. Available online: http://trace.eas.asu.edu/yuv/index.html (accessed on 26 June 2022).
32. H.264 Baseline Codec. Available online: https://ww2.mathworks.cn/matlabcentral/fileexchange/39927-h-264-baseline-codec (accessed on 26 June 2022).