Hindawi
Security and Communication Networks
Volume 2022, Article ID 1154098, 14 pages
https://doi.org/10.1155/2022/1154098

Research Article

Coverless Video Steganography Based on Audio and Frame Features

Chunhu Zhang, Yun Tan, Jiaohua Qin, and Xuyu Xiang

College of Computer Science and Information Technology, Central South University of Forestry & Technology, Changsha 410004, China

Correspondence should be addressed to Yun Tan; tantanyun@hotmail.com

Received 8 December 2021; Accepted 24 February 2022; Published 4 April 2022

Academic Editor: Beijing Chen

The coverless steganography based on video has become a research hot spot recently. However, the existing schemes usually hide secret information based on the single-frame feature of video and do not take advantage of other rich features. In this work, we propose a novel coverless steganography, which makes full use of the audio and frame image features of the video. First, three features are extracted to obtain hash bit sequences, which include DWT (discrete wavelet transform) coefficients and short-term energy of audio and the SIFT (scale-invariant feature transformation) feature of frame images. Then, we build a retrieval database according to the relationship between the generated bit sequences and three features of the corresponding videos. The sender divides the secret information into segments and sends the corresponding retrieval information and carrier videos to the receiver. The receiver can use the retrieval information to recover the secret information from the carrier videos correspondingly. The experimental results show that the proposed method can achieve larger capacity, less time cost, higher hiding success rate, and stronger robustness compared with the existing coverless steganography schemes based on the video.

1. Introduction

In today's era with frequent information leakage and theft, the safe transmission of confidential information is extremely significant. Information hiding technologies can help to solve the problem of secure transmission and effective recovery of secret information. Traditional steganography schemes mainly embed secret information by changing the specific features of the carrier [1–4]. However, due to the modification of carriers, this kind of algorithms has the risk of being detected by steganalysis. The coverless steganography sets up a specific mapping relationship with the characteristics of carriers to hide secret information. It has better concealment performance than traditional information hiding algorithms since its carrier has not been changed.

The existing coverless steganography schemes are mainly divided into two categories: coverless steganography based on text [5–8] and coverless steganography based on image [9–12]. The coverless text steganography usually uses the unique features of text (such as word frequency and keywords) to hide secret information, which was first proposed by Chen et al. [13]. They vectorized and segmented the Chinese secret information and obtained the retrieval information corresponding to the secret information from the Chinese text retrieval database to hide secret information. By using statistical features of text, Zhang et al. [14] selected the normal texts containing the secret information retrieved from the text database to hide secret information. Different from the aforementioned methods, Wang et al. [15] used the data characteristics of nonrepetitive and diverse lyrics generated by GAN to hide secret information, which had good perceptibility and embedding rate. The coverless image steganography was first proposed by Zhou in 2015 [16]. The key point of this kind of algorithms is to extract the specific features of the image, such as the texture, colour, and shape, to establish a specific mapping relationship to hide the secret information. Zhou et al. [16] divided the secret information into bit sequences and then sent the specific carrier images and auxiliary information matched with the bit sequences to the receiver. Zheng et al. [9] divided the carrier images into blocks and used the direction information of feature points of scale-invariant feature transform (SIFT) to hide the

segmented secret bit sequences. Similarly, Zhou et al. [10] used the directional gradient histogram (HOG) of non-overlapping image blocks to hide the secret bit sequences. Due to the powerful ability of deep neural network to extract features [17], some researchers introduced it to coverless steganography. Liu et al. [18] used DWT to transform images and the DenseNet network to recommend carrier images. Luo et al. [19] used the labels of the object in the carrier image to generate hash bit sequences and hid secret information by sending the carrier image containing multiple objects. The receiver used Faster RCNN [20] to extract the labels of the objects in the carrier images to recover the secret information. In addition to using the features of the image to map hash bits, some researches hid secret information based on image generation. Wu et al. [11] set up a mapping relationship between secret information and texture image and used the synthesis process of texture image to hide secret information. Chen et al. [12] divided the natural images into multiple image blocks, every one of which can represent 1-bit secret information, and retrieved the corresponding image blocks to synthesize the carrier image according to the secret information. The generative adversarial networks had caused many technological changes in the field of computer vision [21], and its ability to generate real natural images had been widely recognized. Li et al. [22] used the encoder to extract the content vector of the secret image and input it into the generative model to generate a real and natural carrier image under the penalty mechanism. Yu et al. [23] used the vectorized secret information to directly control the generative model to generate the carrier image and introduced the attention mechanism to correct the image distortion and background anomaly, so as to improve the concealment. However, when more secret information bits need to be hidden, the coverless steganography based on image or text needs to transmit a large number of carriers, which will undoubtedly arouse suspicion by external attackers and increase the risk of secret information being attacked.

Compared with image and text, there are more features that could be extracted from video to hide secret information, such as the frame image features, the temporal features between frames, and the audio features. Therefore, coverless steganography based on the video does not need to transmit too many carriers when more secret information bits need to be hidden. At the same time, due to the wide use of portable multimedia devices such as smart phones, the number of short videos is large enough on the Internet. The daily spread of video makes it an ideal covert communication carrier. These advantages provide a basis for the development of coverless steganography based on the video. However, there are a few coverless steganography methods based on the video. Tan et al. [24] calculated the directional optical flow feature of the adjacent frame images and obtained the robust histograms of oriented optical flow (RHOOF), then mapped the hash bits according to the discrimination relationship of each component of the histogram. The optical flow information of the adjacent frame images in this scheme is sensitive to random noise, so its robustness needs to be improved. Pan et al. [25] first performed framing processing on the video to extract valid frame images and then used the semantic information extracted from the frame images by MobilenetV2 [26] to generate hash sequences. However, this method only used a single-frame image feature of video, and its hiding capacity and hidden success rate were relatively low. At the same time, this method took a long time to train the MobilenetV2 network and the trained model also took a long time to generate a byte, which undoubtedly weakened the practicability of this scheme. Zou et al. [27] used deep neural network to extract the hash codes of frame images and set up mapping rules to improve the capacity of the scheme. However, the hash codes of the frame images were directly generated by the neural network, and the robustness of this network was poor, resulting in the weak anti-interference ability to noise.

In order to make more effective use of the features of carrier video and improve the hiding capacity and robustness, a coverless video steganography scheme based on audio and frame features is proposed in this work. First, the frame images and audio components of the carrier video are extracted. Then, three features of the two components are mapped into bit sequences, and the retrieval database is established according to the mapping relationship. The sender divides the secret information into bit sequences of equal length and then searches the retrieval information and carrier videos in the retrieval database. The retrieval information and carrier videos will be sent to the receiver. Then, the receiver can recover the secret information according to the mapping rules. The contributions of this paper are as follows:

(1) A novel coverless video steganography scheme is proposed based on audio and frame features, which makes full use of the features of frame images and audios of the carrier videos.

(2) The feasibility of audio features for coverless steganography is investigated, which has not been fully studied in existing researches. The short-term energy and DWT coefficient of audio are used to hide secret information. The experimental results show good performance.

(3) The robustness, capacity, cost time, and hiding success rate are analysed and tested. The proposed method achieves good improvements compared with the existing video-based coverless steganography schemes.

The rest of the paper is arranged as follows. Preliminaries are shown in Section 2. The proposed coverless video steganography is described in Section 3. The experimental results and analysis are shown in Section 4. Finally, conclusions are drawn in Section 5.

2. Preliminaries

2.1. Short-Term Energy of Audio Signal. Since the continuous change of the audio signal with time can be characterized by a nonstationary random process and has short-term correlation, short-term analysis is generally used for audio

signal processing. The signal is divided into frames first to ensure the local stability. Then, windows are added to keep the signal continuous, as shown in Figure 1.

Assuming the audio signal is $X(n)$, the i-th frame of signal is obtained after windowing by

$$Y_i(n) = w(n) \times X\left((i-1) \times i_{nc} + n\right), 1 \le n \le L_f, 1 \le i \le f_n, \tag{1}$$

where $w(\cdot)$ is the window function with the width of w_{len}, f_n is the total number of frames, L_f is the frame length, i_{nc} is the frameshift length, and $Y_i(n)$ represents the n-th signal value of the i-th frame of the audio signal. The short-term energy can reflect the strength of the audio signal, which can be obtained by

$$E(i) = \sum_{n=0}^{L-1} Y_i^2(n), \quad 1 \le i \le f_n, \tag{2}$$

where L represents the length of the audio signal.

2.2. Discrete Wavelet Transform. Discrete wavelet transform (DWT) [18] is a transform method whose process is as follows:

$$\psi_{m,n}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) = \frac{1}{\sqrt{a_0^m}}\psi\left(\frac{t-nb_0a_0^m}{a_0^m}\right), \tag{3}$$

$$W_f(m,n) = \int_{-\infty}^{+\infty} f(t)\psi_{m,n}^*(t)\mathrm{d}t. \tag{4}$$

Equation (3) is the discrete wavelet function of DWT, and m and n are integers; a_0 is a constant greater than 1, and b_0 is a constant greater than 0; the different values of a and b are affected by m, n, a_0, and b_0, and the difference of these two parameters is related to the selection of discrete wavelet function $\psi_{m,n}(t)$. Equation (4) represents the process of DWT. $f(t)$ represents the audio signal in time domain, t represents time, and $*$ represents complex conjugate value of discrete wavelet function $\psi_{m,n}(t)$.

After DWT, the audio signal can output low-frequency and high-frequency components. The low-frequency component contains the most energy of the audio signal, whereas the high-frequency component mainly contains detailed information of speech signal such as the impact of noise. As shown in Figure 2, after each DWT, the length of low-frequency information is halved, and the contour is more obvious and stable. Therefore, the low-frequency component of DWT has good stability and robustness, which can be used for information hiding.

2.3. Scale-Invariant Feature Transformation. Scale-invariant feature transformation (SIFT) has the feature of scale invariance, which is not affected by the variation of light, noise, and visual angle. Because of its excellent stability and robustness, it can be applied to information hiding [12]. The steps of SIFT feature detection are as follows:

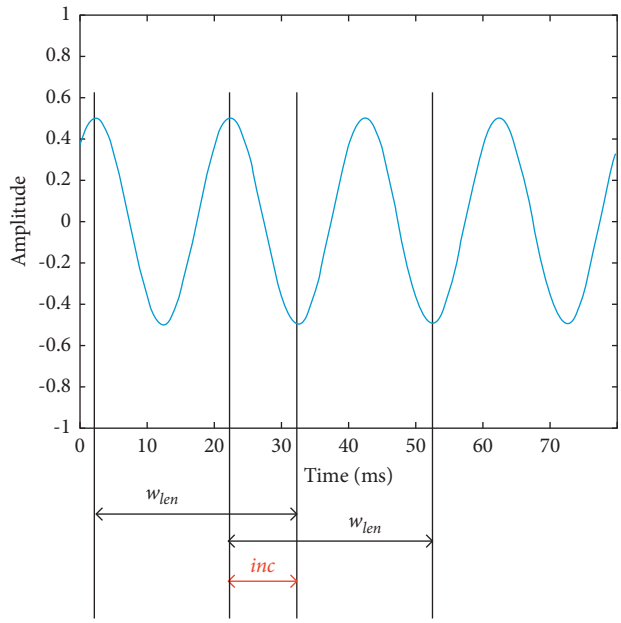

Figure 1: Framing process of the audio signal.

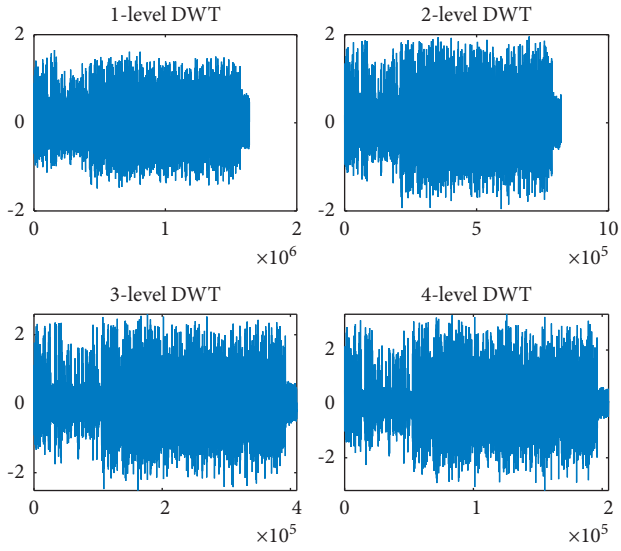

Figure 2: Low-frequency information after DWT.

(1) Detect the extremum of scale space. Gaussian difference function is used to search for potential feature points with constant scale and direction for all images in scale space.

(2) Locate the feature points. Based on the stability principle, a fine model is fitted to determine the position of the final feature points accurately.

(3) Determine the directions of the feature points. Each feature point is assigned one or more directions based on the gradient direction of the local image. Because the image data are converted to the set direction, position, and scale features on which the

subsequent operation is also based, it provides invariance for SIFT features.

(4) Describe the feature points. The gradient of the local image in the domain of each feature point is transformed into another representation, which could help resist a certain degree of image local distortion and illumination change.

3. Our Proposed Method

In this paper, we propose a coverless video steganography based on audio and frame features whose framework is shown in Figure 3. In our method, we first process video to get the audio and image components. Then, three features of these two components are extracted: SIFT feature, short-term energy feature, and DWT coefficient feature are used to generate hash bit sequences by different robust mapping methods. After that, the retrieval database is built according to the position information. At the sender, the secret information is divided into bit sequences, which are used to search corresponding retrieval information and videos in the retrieval database. The retrieval information and videos are sent to the receiver. After receiving the retrieval information and videos, the bit sequences are obtained by calculating the corresponding features in the video according to the retrieval information, such that the secret information can be recovered by these bit sequences.

3.1. Mapping of Bit Sequence. The mapping method of bit sequence is related to the robustness and accuracy of coverless steganography, so it is the core part of the algorithm. Our method includes three features, which are extracted from audio and image, respectively. Three mapping schemes of bit sequence are described as follows.

3.1.1. Mapping Based on Short-Term Energy of the Audio. The mapping based on short-term energy of audio is as follows:

(1) Process the audio signal $X(n)$. According to equation (1), the audio signal $X(n)$ could be divided into frames and windowed to get f_n frames of the audio signal $Y_i(m)$. Here, we set the frame length $w_{\text{len}} = 200$ and the frameshift $i_{nc} = 80$. Then, we calculate the short-term energy of each frame audio signal according to equation (2).

(2) Segment the energy of the audio signal. According to the principle that $L_0 = 180$ frames of total energy is used to map 1-bit information, the short-term energy $E(i)$ is segmented to get h_0 segments of short-term energy $En(h)$ according to equation (5), the first $8 \times N_0$ segments of which is used to map to generate bit sequences.

$$En(h) = \sum_{i=1}^{L_0} E(i), h_0 = \text{floor}\left(\frac{f_n}{L_0}\right) N_0 = \text{floor}\left(\frac{h_0}{8}\right), 1 \le h \le h_0. \tag{5}$$

(3) Generate the hash sequences. The 8 slices of short-term energy segments are selected from $En(h)$ in sequence, and the mean value is taken as the threshold K. According to the relationship between the short-term energy and the threshold value K, we obtain the bit sequence B_1, as shown in Figure 4. We obtain the hash sequence $\widetilde{B}_1$ by bit reversal.

$$B_1(j) = \begin{cases} 1, \text{if } En(j) \ge K \quad K = \dfrac{\left(\sum_{m=1}^{8} En(j)\right)}{8} \\ 0, \text{if } En(j) < K \quad 1 \le j \le 8 \end{cases}. \tag{6}$$

3.1.2. Mapping Based on DWT Coefficients of the Audio. We use the stable low-frequency information obtained by DWT to generate robust bit sequences as follows:

(1) Perform DWT on the audio signal. We perform DWT on the audio signal $X(n)$ continuously three times and output the absolute value of the low-frequency information U whose length is l.

(2) Process the coefficient of low-frequency information U. We use $L_1 = 2750$ values of low-frequency information U to map 1-bit information, and we can get $h_1 = \text{floor}(l/L_1)$ low-frequency coefficients Zc.

$$Zc(j) = \sum_{i=(j-1)\times L_1+1}^{j\times L_1} U(i), 1 \le j \le h_1. \tag{7}$$

(3) Generate the hash sequences. By comparing the numerical relation of the adjacent DWT coefficients Zc, we obtain the bit sequence H of length $h_1 - 1$:

$$H(j) = \begin{cases} 1, \text{if } Zc(j) > Zc(j+1) \\ 0, \text{if } Zc(j) \le Zc(j+1) \end{cases} 1 \le j \le h_1 - 1. \tag{8}$$

(4) Output the bit sequences in bytes. In this bit sequence H, the byte sequence B_2 is output byte by byte in sequence, as shown in Figure 5. We obtain the byte sequence $\widetilde{B}_2$ by bit inverting B_2.

(5) Repeat step 4 h times to output $2 \times h$ byte sequences.

$$h = \text{floor}\left(\frac{(h_0 - 1)}{8}\right). \tag{9}$$

3.1.3. Mapping Based on SIFT Feature of Frame Images. After extracting the frame images $I(m)$ from the video, the SIFT feature is used to generate a hash bit sequence from each frame image. Different from Zheng's method [12], we map the bit sequence by counting the number of SIFT feature points of image subblocks. Particularly, we use the frame image feature mapping to generate a bit sequence and then reverse it by bit to get a new bit sequence to further increase the capacity. The mapping steps are as follows:

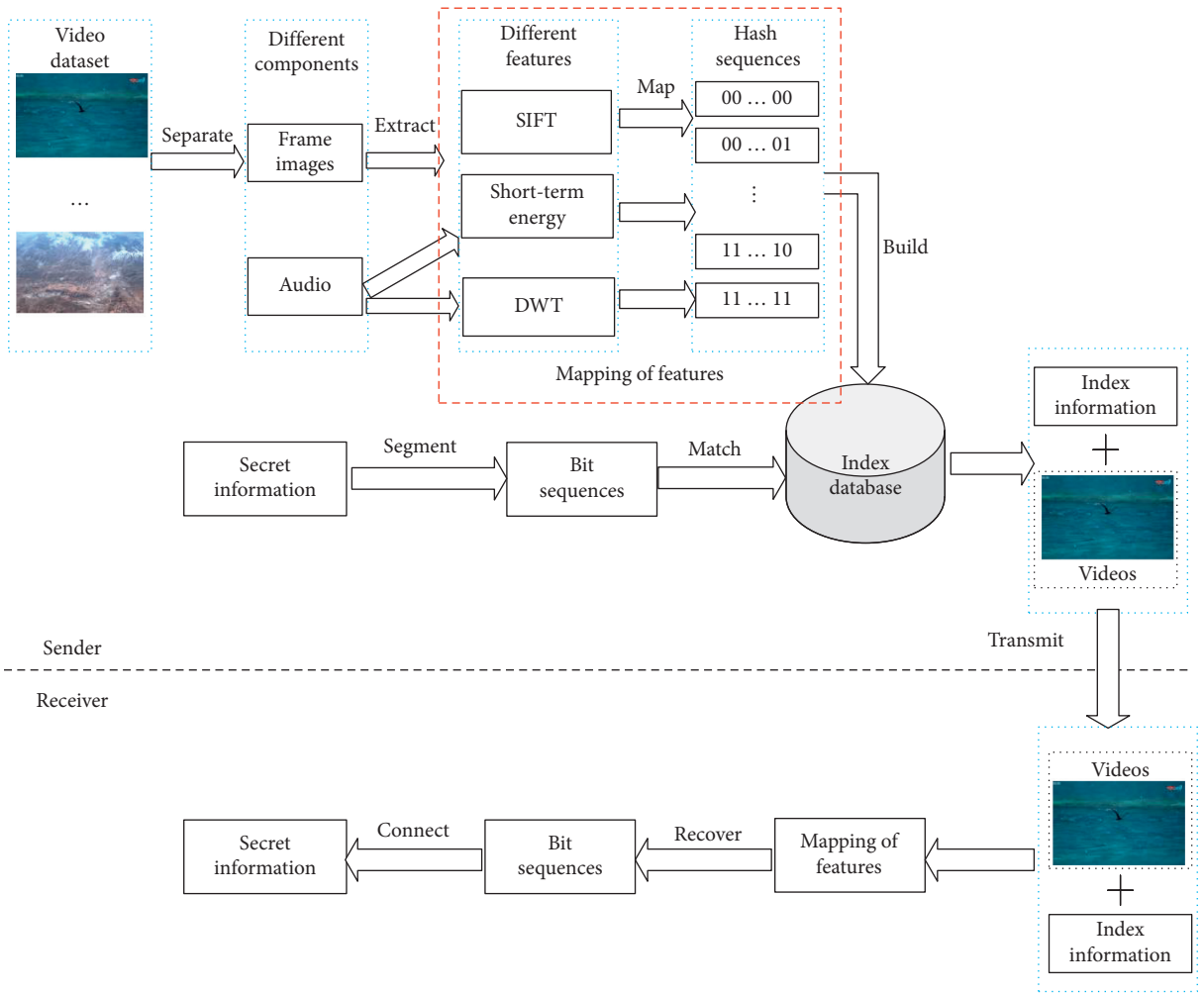

Figure 3: The framework of the proposed coverless video steganography scheme.

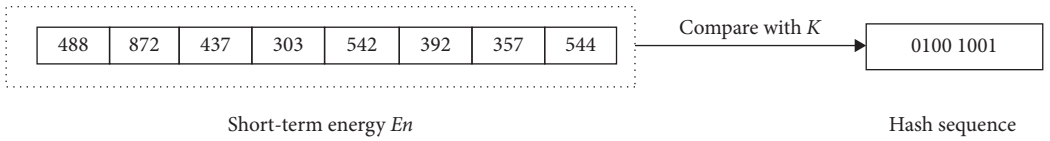

Figure 4: Generation of hash sequence B_1.

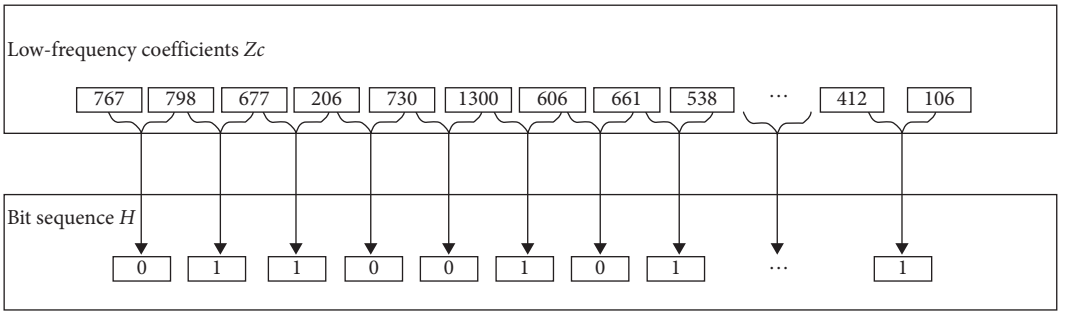

Figure 5: Generation of hash sequence B_2.

(1) Process the frame image. For a frame image, we transform it to greyscale first, uniform its size for 512 × 512, and divide it into 3 × 3 blocks.

(2) Generate and count SIFT feature points. We perform SFIT transformation on this frame image to obtain the location information of SIFT feature points. Then, we count the number of SIFT feature points $S(i)$ of image subblocks, $1 \le i \le 9$.

(3) Generate the hash sequences. By comparing the number of SFIT feature points of different image subblocks, we obtain the hash bit sequence B_3, as shown in Figure 6. And then, we could obtain the hash bit sequence $\bar{B}_3$ by bit inverting B_3.

$$B_3(i) = \begin{cases} 1, \text{if } S(i) < S(i+1) \\ 0, \text{otherwise} \end{cases}, 1 \le i \le 8. \quad (10)$$

(4) Repeat steps 1, 2, and 3 until the bit sequences of all images are generated.

3.2. Establishment of Retrieval Database. The retrieval database could help sender search the carriers corresponding to the secret information, so that it is an important part of algorithm. The establishment process is as follows:

(1) Extract two components of the video. We use Arabic numerals to mark the position of this video, which will be used to mark the video ID of the subsequent hash sequences, and then extract the frame images $I(m)$ and audio $X(n)$ from it.

(2) Extract different features. We extract the SIFT features of the frame images $I(m)$ and mark their feature ID with 0. Then, the short-term energy features and DWT coefficient features of the audios $X(n)$ are extracted, and the feature ID is marked as 1 and 2, respectively.

(3) Generate hash sequences and update the retrieval information. We obtain hash sequences using three mapping ways mentioned above. At the same time, we append 0 at the end of the feature ID if the hash sequence is generated by feature mapping directly, otherwise 1.

(4) Repeat steps 1 to 3 until 256 types of different byte sequences are mapped, and the retrieval database is established, as shown in Figure 7. The algorithm of the establishment of the retrieval database is described in Algorithm 1.

It can be seen from Figure 7 that a byte sequence may have multiple corresponding retrieval information. Therefore, the sender can randomly select one of the multiple retrieval information of the byte sequence as the corresponding retrieval information, so that the same byte sequence of secret information has multiple different mapping items. It can make the auxiliary information transmitted by the sender, have more variability, increase the cracking difficulty of external attackers, and enhance the complexity of our method.

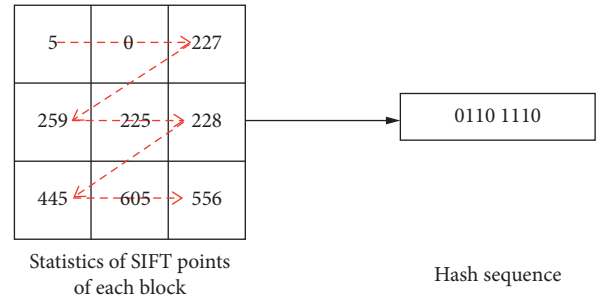

Figure 6: Generation of hash sequence B_3.

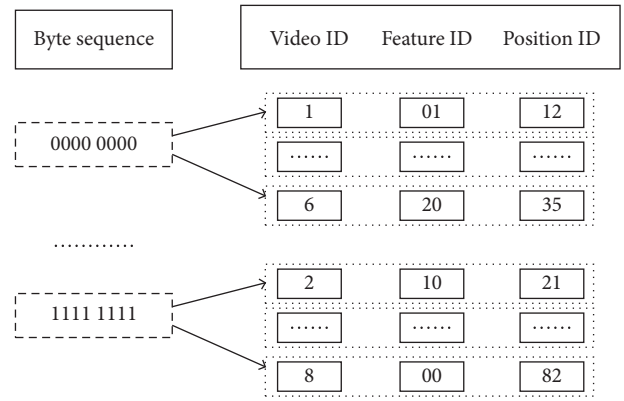

Figure 7: Retrieval database of videos.

3.3. Transmission of Secret Information. The specific process of secret information transmission is as follows:

(1) Construct retrieval database of videos and obtain the carrier videos V.

(2) For the secret information S of length L_s, segment every 8 bits (1 byte) and fills the tail with some auxiliary information to get the byte sequence $P = \{P_1, P_2, \ldots\ldots, P_{L_p}\}$:

$$L_P = \begin{cases} \text{floor}\left(\frac{L_s}{8}\right) + 1, & \text{if mod}(L_s, 8) = 0, \\ \text{floor}\left(\frac{L_s}{8}\right) + 2, & \text{if mod}(L_s, 8) \neq 0. \end{cases} \quad (11)$$

If $\text{mod}(Ls, 8) \neq 0$, we pad 0 to the end of S to form a byte sequence and add 1 byte at the same time, which represents the number of 0 padded; if $\text{mod}(Ls, 8) = 0$, the sender pads a byte 0000 0000 to indicate that the original secret information has not been padded.

(3) Get the retrieval information C_i according to P_i from the retrieval database.

(4) Repeat step 3 until all the bytes in P have been matched to get the retrieval information $C = \{C_1, C_2, \ldots\ldots, C_{L_p}\}$.

(5) Send the retrieval information C and carrier videos V to the receiver. If both sides of the communication

have an encryption protocol, the retrieval information C can be encrypted. The algorithm of transmission of secret information is described in Algorithm 2.

3.4. Recovery of Secret Information. The specific process of secret information at receiver is as follows:

(1) The receiver can recover the bit sequence P_i according to the retrieval information C_i and the mapping method described in Section 3.2.

(2) Repeat step 1 until all the remaining search information of C has been matched to obtain the byte sequence $P = \{P_1, P_2, \ldots, P_{L_p}\}$.

(3) Recover secret information S according to P. If the last byte of P is 0000 0000, the last byte is directly removed to recover the secret information S. If the last byte is not 0000 0000, then according to its corresponding decimal value, the padded "0s" of the last two bytes are removed to get the original secret information S. The algorithm of secret information recovery is described in Algorithm 3.

4. Experimental Results and Analysis

The experimental environment is as follows: AMD Ryzen 5 3600 6-Core Processor CPU at 3.59 GHz, 16 Gb RAM, and NVIDIA GeForce RTX 2070 (mobile) 8G, whose driver version is 27.21.14.5671. The test software is MATLAB 2020a.

As far as we know, Pan et al. [25] proposed the first coverless video steganography method in 2020. Therefore, we will conduct performance comparison experiments with Pan's scheme on the same data set, containing some public videos we obtained from the Internet using crawler technology. The video data set consists of 240 short videos in the format of MP4. Most of these are standard definition videos, and a few are high-definition videos. Among them, the longest duration of video is about 5 minutes. The themes of this video data set include news brief, music videos, entertainment broadcast, video clip, and documentary clip. In the test of audio features, we extract the audio components of video, in which the sampling rate Fs is 44100, and remove the weak signal values—signals with an absolute value less than at the beginning and the end of the audio components to reduce noise influences. In the test experiment of the frame image features, we select some videos in the video data set for experimental testing of frame images. The partial data sets are shown in Figure 8.

Tan et al. [24] proposed a coverless steganography scheme based on optical flow analysis of video in 2021. In order to compare the latest scheme, we select the public data set UCF101 used in the paper [24] for robustness experiment comparison. UCF101 is a public video data set, the content of which is various actions and scenes. According to Tan's settings, we randomly select videos of different actions and scenes. The size of these files is about 200~800 kb and the duration of these videos is about 2~10 seconds, as shown in Figure 9.

4.1. Capacity. The hidden capacity is an important indicator of the information hiding algorithm. A coverless steganography algorithm with a large capacity can help reduce the number of carriers needed for transmission. Our scheme uses the frame image and audio components of video to establish the mapping relationship with the secret information, so the hiding capacity of our algorithm should be discussed in combination with the frame image and audio.

4.1.1. Capacity of the Audio. In this paper, we segment the audio signal and use the feature mapping of each segment of the audio signal to generate bytes. In the short-term energy feature, we first divide the audio signal into frames, then use 8×180 frames of the audio signal to map 1-byte sequence and use bit inversion to get another byte sequence. According to equation (1), the 1-second audio signal can be divided into f_n frames when the sampling rate is Fs and we can obtain

$$200 \times f_n - 80(f_n - 1) = Fs. \tag{12}$$

Then, equation (12) is transformed as

$$f_n = \frac{Fs - 80}{120}. \tag{13}$$

Therefore, a x-second video can be mapped to C_1 bytes.

$$C_1 = 2 \times \text{floor}\left(\frac{x \times f_n}{8 \times 180}\right) = 2 \times \text{floor}\left(\frac{x \times (Fs - 80)}{172800}\right). \tag{14}$$

We set $n_1 = (Fs - 80)/172800$, and then, we can simplify C_1.

$$C_1 = 2 \times \text{floor}(x \times n_1). \tag{15}$$

In the feature of DWT coefficient, we perform four times of DWT on the audio signal to get stable low-frequency information. The length of the audio becomes $1/2^4$ of the original audio, and the value of the 1-second audio signal also changed from Fs to $Fs/2^4$ after four times of DWT transform. We use 8×2750 DWT coefficients to map 1 byte, so a video of x seconds can be mapped to generate C_2 bytes.

$$C_1 = 2 \times \text{floor}\left(\frac{x \times Fs}{16 \times 8 \times 2750}\right) = 2 \times \text{floor}\left(\frac{x \times Fs}{352000}\right). \tag{16}$$

We set $n_2 = Fs/352000$, and then, we can simplify C_2.

$$C_2 = 2 \times \text{floor}(x \times n_2). \tag{17}$$

Therefore, the hidden capacity of the x-second audio is $C_1 + C_2$ bytes when the audio sampling rate is Fs. It can be seen from the above that the size of the audio feature capacity is related to the audio duration x, the sampling rate Fs, and the parameters L_0 and L_1. In fact, in order to balance the robustness and capacity of the scheme, we conduct robustness tests on the values of L_0 and L_1 in Section 4.2.1 and finally determined the values of these two parameters.

4.1.2. Capacity of the Video. For a frame image, our algorithm uses frame image feature mapping and bit inversion

Input: Video database $v = \{v_1, v_2, \ldots, v_a\}$
Output: Retrieval database $R = \{R_1, R_2, \ldots, R_{256}\}$, Carrier videos $V = \{V_1, V_2, \ldots, V_b\}$
(1) For $i = 1$ to a
(2) Obtain the video ID: $ID_S = i$
(3) Extract frame images from video v_i: $I(m)$ = ExtracImg (v_i)
(4) For $j = 1$ to Length $(I(m))$
(5) Generate hash sequence: Hash_j = CalSIFT (I_j)
(6) Update retrieval database: R = update (Video ID, Feature ID, Position ID)
(7) End for
(8) Extract audio from video v_i: $X(n)$ = ExtracAud (v_i)
(9) Generate hash sequence of the short-term energy features: Hash_{STE} = CalSTE $(X(n))$
(10) For $j = 1$ to Length $(\text{Hash}_{\text{DWT}})$
(11) Update retrieval database: R = update (Video ID, Feature ID, Position ID)
(12) END for
(13) Generate hash sequence of the DWT coefficient features: Hash_{DWT} = CalDWT $(X(n))$
(14) For $j = 1$ to Length $(\text{Hash}_{\text{DWT}})$
(15) Update retrieval database: R = update (Video ID, Feature ID, Position ID
(16) END for
(17) $V_i = v_i$
(18) If $\forall r \in R$, $r \neq$ Null then
(19) Return R, V
(20) End if
(21) End for

Algorithm 1: Establishment of the retrieval database.

Input: Video database $v = \{v_1, v_2, \ldots, v_a\}$, Secret information $S = \{S_1, S_2, \ldots, S_{L_S}\}$
Output: Carrier videos $V = \{V_1, V_2, \ldots, V_b\}$, Retrieval information $C = \{C_1, C_2, \ldots, C_{L_P}\}$
(1) Construct retrieval database R and obtain carrier videos V
(2) Segment the secret information: S' = segment (S)
(3) Padding the bytes sequence: P = pad (S')
(4) For $i = 1$ to L_P
(5) Search in the index information C_i corresponding to P_i
(6) End for
(7) Send the retrieval information C and carrier videos V to the receiver

Algorithm 2: Transmission of secret information.

Input: Carrier videos $V = \{V_1, V_2, \ldots, V_b\}$, Retrieval information $C = \{C_1, C_2, \ldots, C_{L_P}\}$
Output: Secret information $S = \{S_1, S_2, \ldots, S_{L_S}\}$
(1) Receive retrieval information C and carrier videos V.
(2) For $i = 1$ to L_P
(3) Obtain the byte sequence P_i according to C_i and mapping method
(4) End for
(5) Remove the padding bytes at the end of the byte sequence P: S' = Remove (P)
(6) Connect byte sequence S' to restore secret information sequence: S =Connect (S')

Algorithm 3: Recovery of secret information.

operations to generate 2 bytes, whereas Pan's method can only map 1 byte when the robustness is optimal, but Tan's method can map 4 bytes.

For a x-second video, whose audio sampling rate is Fs and the frame images that can be extracted are M, we use the total number of bits mapped on a certain carrier to measure the hidden capacity. The results are shown in Table 1.

It can be seen that the number of bits generated per frame image in our scheme is consistent with that of Zou's scheme, twice that of Pan's scheme, but half that of Tan's

Figure 8: The samples of our database.

Figure 9: The samples of database UCF101.

Table 1: Capacity comparison of different methods.

Method	Capacity (bits/image)	Capacity (bits/audio)	Capacity (bits/video)
Ours	16	$\mathbf{8 \times (C_1 + C_2)}$	$8 \times (C_1 + C_2 + 2 \times M)$
Tan's [24]	**32**	—	$\mathbf{32 \times M}$
Pan's [25]	8	—	$8 \times M$
Zou's [27]	16	—	$16 \times M$

scheme. However, the other three schemes cannot use audio to map bits, but our method can map and generate $8 \times (C_1 + C_2)$ bits using the features of audio. Ideally, when the video length is long enough, the number of hash bits mapped by audio of our scheme is sufficient and the capacity of the solution can exceed the capacity of Tan's scheme.

4.2. Robustness. In the process of transmission, the carrier videos will be affected by noise or external attacks. Therefore, the robustness is an important indicator. We use external noise or attack to affect frame images and audios of the carrier video, and evaluate the robustness according to the similarities and differences between the byte sequence $P_1(t)$ recovered from the retrieval information and the byte sequence $P_0(t)$ obtained from the original secret information segmentation, which is calculated as

$$R_{\text{block}} = \frac{\sum_{t=1}^{L_p} Ac(t)}{L_p}, Ac(t) = \begin{cases} 1, & \text{if } P_0(t) = P_1(t) \\ 0, & \text{otherwise} \end{cases}, \quad (18)$$

where L_P is the byte number of sequences $P_0(t)$ and $P_1(t)$.

The paper [24] used the bit error rate to measure the robustness of the algorithm, which was calculated as

$$R_{\text{block}} = \frac{\sum_{t=1}^{L_p} Ac(t)}{L_p}, Ac(t) = \begin{cases} 1, & \text{if } P_0(t) = P_1(t) \\ 0, & \text{otherwise} \end{cases}, \quad (19)$$

where $p_0(t)$ and $p_1(t)$ represent t-th bit of sequence $P_0(t)$ and $P_1(t)$, respectively, and L_c is the total bit number.

4.2.1. Robustness Based on Audio Features. The audio is mainly affected by external Gaussian white noise. In this section, we use seven kinds of Gaussian noise under different SNRs to test the robustness of two audio features. We test the experiment 20 times, remove the maximum and minimum values of the results, and take the average in the rest of results.

We compare the robust performance of short-term energy features with different frame numbers L_0. Our method uses the accumulated value of 180 frames of short-term energy of the audio signal as the mapping. Table 2 shows the experimental test results of the robustness of short-term energy feature with different frame numbers L_0. It can be seen that the robustness of short-term energy feature increases slowly with the increase of the number of frames. Moreover, the capacity will decrease if increasing L_0. Therefore, in order to balance robustness and capacity, we set $L_0 = 180$ for better performance.

We compare the robustness performance of DWT coefficient with different segment numbers L_1 in Table 3. Our method uses the cumulative sum of 2750 segment values of U as the mapping. It can be seen that there is no big difference in the antinoise performance of DWT coefficient with different segment numbers L_1. And there is a negative correlation between the size of L_1 and the capacity of the algorithm. In order to balance robustness and capacity, we set $L_1 = 2750$ for better performance.

We compare the robust performance of DWT coefficient feature with different times c of DWT in Table 4. It can be seen that at the beginning, with the increase of the number of DWT, the robustness is improved, but if c is greater than 4, the robustness decreases. Therefore, we set $c = 4$ for better performance.

We compare the robust performance of DWT coefficient feature with different wavelet basis functions in Table 5. It can be seen that when the wavelet basis function is rbio3.1, the comprehensive robustness of DWT coefficient is better; therefore, we set the wavelet basis function as rbio3.1.

Table 2: R_{block} of short-term energy with different frame numbersL_0.

L_0	Gauss noise						
	snr = 0	snr = 3	snr = 5	snr = 10	snr = 15	snr = 20	snr = 30
180	0.8607	0.9070	0.9276	0.9600	0.9775	0.9867	0.9952
360	0.8894	0.9264	0.9420	0.9674	0.9810	0.9895	0.9964
540	0.9010	0.9325	0.9493	0.9721	0.9845	0.9908	0.9968
720	0.9150	0.9436	0.9568	0.9772	0.9875	0.9924	0.9975
900	0.9154	0.9484	0.9584	0.9783	0.9882	0.9935	0.9983
1080	0.9187	0.9478	0.9592	0.9793	0.9886	0.9939	**0.9984**
1260	0.9229	0.9515	0.9615	0.9802	0.9891	0.9936	0.9981
1440	0.9288	0.9511	0.9640	0.9802	0.9880	0.9922	0.9979
1620	0.9323	0.9539	0.9650	0.9805	0.9891	0.9938	0.9978
1800	**0.9335**	**0.9557**	**0.9664**	**0.9824**	**0.9906**	**0.9945**	0.9982

Table 3: R_{block} of short-term energy with different segment numbersL_1.

L_1	Gauss noise						
	snr = 0	snr = 3	snr = 5	snr = 10	snr = 15	snr = 20	snr = 30
1375	0.8334	0.8773	0.9015	0.9423	0.9659	0.9800	0.9923
2750	0.8688	0.9061	0.9238	0.9567	0.9743	0.9849	0.9942
5500	0.8803	0.9160	0.9341	0.9654	0.9811	0.9898	0.9958
8250	**0.8880**	0.9208	**0.9379**	0.9663	0.9813	0.9882	0.9953
11000	0.8853	**0.9211**	0.9350	**0.9667**	**0.9816**	**0.9905**	**0.9967**

Table 4: R_{block} of DWT coefficient with different times c.

c	Gauss noise						
	snr = 0	snr = 3	snr = 5	snr = 10	snr = 15	snr = 20	snr = 30
1	0.8053	0.8606	0.8875	0.9390	0.9684	0.9818	0.9940
2	0.8397	0.8846	0.9105	0.9503	0.9727	0.9853	0.9953
3	0.8565	0.8980	0.9194	0.9559	**0.9758**	**0.9864**	**0.9956**
4	**0.8688**	**0.9061**	**0.9238**	**0.9567**	0.9743	0.9849	0.9942
5	0.8646	0.9024	0.9225	0.9547	0.9738	0.9845	0.9948
6	0.8351	0.8791	0.9021	0.9445	0.9678	0.9815	0.9935
7	0.7294	0.7963	0.8314	0.8990	0.9404	0.9644	0.9875

Table 5: R_{block} of DWT coefficient with different wavelet basis functions.

Wavelet	Gauss noise						
	snr = 0	snr = 3	snr = 5	snr = 10	snr = 15	snr = 20	snr = 30
db5	0.8574	0.8975	0.9182	0.9541	0.9749	**0.9864**	**0.9953**
db15	0.8527	0.8925	0.9132	0.9518	0.9724	0.9848	0.9946
coif1	0.8569	0.8982	0.9181	0.9541	0.9730	0.9850	0.9949
coif5	0.8540	0.8960	0.9148	0.9511	0.9714	0.9832	0.9935
fk4	0.8555	0.8956	0.9167	0.9510	0.9718	0.9835	0.9938
fk18	0.8547	0.8954	0.9162	0.9515	0.9726	0.9842	0.9947
sym2	0.8556	0.8957	0.9167	0.9519	0.9729	0.9843	0.9947
sym8	0.8578	0.8989	0.9168	0.9532	0.9726	0.9847	0.9944
dmey	0.8547	0.8970	0.9184	0.9535	0.9732	0.9841	0.9946
bior1.1	0.8498	0.8925	0.9149	0.9525	0.9710	0.9830	0.9941
bior3.1	0.5131	0.6424	0.7118	0.8323	0.9048	0.9465	0.9827
rbio3.1	0.8697	**0.9094**	**0.9270**	**0.9574**	**0.9758**	0.9859	0.9951
rbio3.3	**0.8709**	0.9053	0.9252	0.9559	0.9752	0.9854	0.9949
rbio3.5	0.8683	0.9054	0.9230	0.9554	0.9737	0.9854	0.9942
rbio4.4	0.8587	0.8992	0.9190	0.9545	0.9749	0.9858	0.9948
rbio5.5	0.8452	0.8895	0.9117	0.9496	0.9713	0.9834	0.9953
rbio6.8	0.8600	0.8986	0.9206	0.9555	0.9754	0.9854	0.9948

Table 6: R_{block} of two audio features.

Gauss noise	snr = 0	snr = 3	snr = 5	snr = 10	snr = 15	snr = 20	snr = 30
Robustness	0.8624	0.9068	0.9270	0.9599	0.9768	0.9867	0.9950

Table 7: R_{block} of different methods on our database.

Attack	Parameter	Pan's [25]	Zou's [27]	Ours
Salt and pepper noise	$\sigma = 0.001$	0.7335	0.7323	**0.8765**
	$\sigma = 0.005$	0.4776	0.4252	**0.7595**
Gauss noise	$\sigma = 0.001$	0.2269	0.5739	**0.6864**
	$\sigma = 0.005$	0.2267	0.2374	**0.5104**
	$\sigma = 0.01$	0.2250	0.1330	**0.4252**
Speckle noise	$\sigma = 0.01$	0.4291	0.5151	**0.6224**
	$\sigma = 0.05$	0.2949	0.2070	**0.4162**
JPEG compression	$Q = 10$	0.0886	0.1752	**0.5990**
	$Q = 70$	0.5863	**0.8030**	0.8029
	$Q = 90$	0.7221	**0.8897**	0.8601
Centred cropping	Ratio = 10%	0.2267	0.0467	**0.3672**
	Ratio = 20%	0.0853	0.0213	**0.1512**
Edge cropping	Ratio = 10%	0.2436	0.1909	**0.7410**
	Ratio = 20%	0.1472	0.0877	**0.5730**
Rotation	Rotation angle = 10°	0.0333	0.0131	**0.2351**
	Rotation angle = 15°	0.0333	0.0067	**0.1551**
Translation	(16, 10)	0.1741	0.1665	**0.5820**
	(40, 25)	0.0557	0.0627	**0.3238**
Mean filtering	Window size: 3×3	0.5958	0.5822	**0.6155**
	Window size: 5×5	0.3542	0.2788	**0.4847**
Gamma correction	Factor = 0.8	0.4395	0.4579	**0.7271**
Colour histogram equalization	None	0.0906	0.0465	**0.3441**

The overall robustness performance of the two audio features is shown in Table 6. We set the short-term energy feature parameter as $L_0 = 180$, the DWT coefficient feature parameter as $L_1 = 2750$, the wavelet basis function as rbio3.1, and the number of DWT as $c = 4$. It can be seen that the robustness of the audio feature mapping method is strong, which can reach 86% under the Gaussian noise with SNR of 0.

4.2.2. Robustness Based on Frame Image Features. In this section, we use a variety of geometric attacks and noise attacks with different parameters to test the robustness of different methods on our video data set. According to the setting of paper [25], the j of Pan's scheme is set to 9. The experimental results of different methods on our database are shown in Table 7.

It can be seen that the robust performance of our method is better than that of Zou's and Pan's method under most external image attacks. Because these two schemes use the neural network to directly extract the features of the frame images, the influence of the pixel values has a great impact on the output results of the network, resulting in the weak anti-interference ability to noise. In particular, the pixel matrix of the frame image will be quite different from the original matrix if the video encounters geometric attack. Affected by the prior knowledge of the training set, the extracted features of the neural network may be quite different from the original features. With the scale invariance of SIFT, our method can stably extract the feature points of the frame images and has good robustness to noise attack and geometric attack.

In order to compare the experiment with the state-of-the-art scheme [24], we use equation (19) to compare the robustness experiments with Pan's scheme and Tan's scheme on the data set UCF101. According to the setting of paper [24], the bin number N is set to 8, and the subblock number S is set to 4; according to the setting of paper [25], j is set to 9. The results are shown in Table 8. We can see that most of the experimental data of our scheme is stronger than the other two schemes, especially the anticompression performance. Compared with other antinoise performance, anticompression performance is particularly important for coverless steganography based on video. Because carrier video generally undergoes a compression step before sending, which often damages the video content.

4.3. Efficiency Analysis. The complexity and efficiency will affect the feasibility and practicability of the steganography scheme. The cost of our scheme is mainly related to the map of hash bits and three features, because it involves the calculation and mapping of three features. We measure the efficiency of the schemes based on the time

Table 8: R_{bit} of different methods on database UCF101.

Attack	Parameter	Tan's [24]	Pan's [25]	Ours
Salt and pepper noise	$\sigma = 0.001$	**0.9986**	0.9559	0.9763
	$\sigma = 0.005$	**0.9923**	0.9063	0.9504
	$\sigma = 0.01$	**0.9877**	0.8731	0.9186
Gauss noise	$\sigma = 0.001$	0.7005	0.7889	**0.9139**
	$\sigma = 0.005$	0.6485	0.7889	**0.8292**
	$\sigma = 0.01$	0.6198	**0.7821**	0.7473
Speckle noise	$\sigma = 0.001$	0.8235	0.9150	**0.9466**
	$\sigma = 0.005$	0.8098	0.8698	**0.8829**
	$\sigma = 0.01$	0.8000	**0.8431**	0.7952
Compressed MPEG-4 file with H.264 (.mp4 file)	None	0.9589	0.9172	**0.9594**
Compressed motion JPEG 2000 file (.mj2 file)	None	0.8476	0.9676	**0.9790**

Table 9: Time cost comparison of different methods.

Method	Tan's [24]	Pan's [25]	Zou's [27]	Ours
Time cost	0.7416 s/B	1.3769 s/B	1.2994 s/B	**0.1755 s/B**

required to hide a byte, and the unit is "s/B." From the results in Table 9, it can be seen that the time required in our scheme is the least, which is about one quarter of the time cost of Tan's method and about one seventh of the time cost of Zou's and Pan's methods. Therefore, the cost of our scheme is the lowest, which undoubtedly enhances the feasibility of our scheme.

4.4. Hiding Success Rate. Information hiding algorithm should not only consider the capacity of the method but also pay attention to the hiding success rate, which can be expressed by the number of different bytes that a video can hide. Hiding success rate can reflect the effectiveness of the algorithm, and its calculation formula is shown in

$$S = \frac{Q}{2^w}, \tag{20}$$

where Q is the total number of bit sequences generated by multiple videos and $w = 8$ in this experiment.

We use 85 videos in the video data set to test our method and Pan's method, and the results are shown in Figure 10. The hiding success rate of our method is always higher than that of Pan's method, and only 9 videos are enough to map 256 types of different bit sequences. This is because we use three features and bit inversion operation, and thus, a video can generate a variety of hash sequences. The hiding success rate of Pan's method can only approach 99% with 85 videos, which means that the redundancy of bit sequences generated by multiple videos is high and a large number of videos are needed to map all kinds of bit sequences.

4.5. Security Analysis. The coverless video steganography based on audio and frame features proposed in this paper has multiple securities as follows:

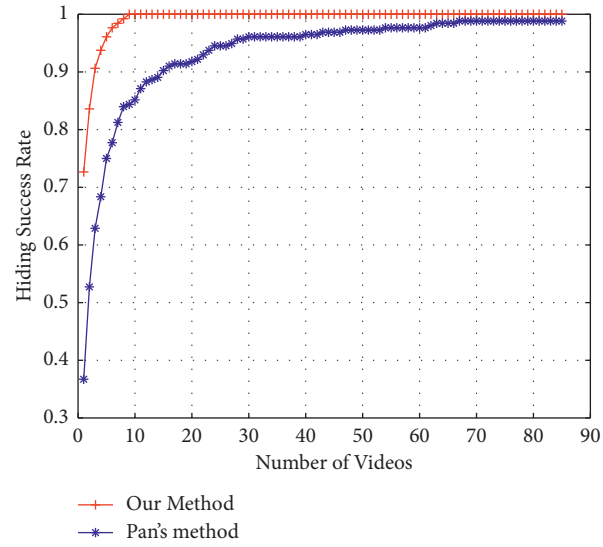

Figure 10: Comparison of hiding success rate.

(1) We use three features of video to map the hash bit sequences and hide the secret information, rather than modifying the carrier video. Therefore, this method could resist steganalysis tools, which could ensure the security of secret information.

(2) The carrier video used by our method is from the abundant short videos on the Internet, which could greatly reduce the attention of the outside world to the secret communication, so as to improve the security of communication.

5. Conclusion

A coverless video steganography based on audio and frame features is proposed in this work, which makes full use of short-term energy feature, DWT coefficient feature, and SIFT feature of video to map hash bit sequences and hide secret information. The experimental results show that, compared with the existing coverless video steganography, our method has larger capacity, less time cost, higher success rate of hiding, and stronger robustness to most external

attacks. In the future, we will try to further improve the robustness and capacity.

Data Availability

The video database we built can be obtained upon request to the corresponding author. The UCF101 data used to support the findings of this study are available at https://www.crcv.ucf.edu/data/UCF101.php.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant 62002392), the National Natural Science Foundation of Hunan (Grants 2020JJ4140 and 2020JJ4141), the Science Research Projects of Hunan Provincial Education Department (Grant 19B584), the Degree & Postgraduate Education Reform Project of Hunan Province (Grant 2019JGYB154), and the Postgraduate Excellent teaching team Project of Hunan Province (Grant [2019] 370-133).

References

[1] B. Wang, W. Kong, N. Li, N. Neal, W. Xiong, and N. N. Xiong, "A dual-chaining watermark scheme for data integrity protection in Internet of things," *Computers, Materials & Continua*, vol. 58, no. 3, pp. 679–695, 2019.

[2] D. R. Vinay and B. J. Ananda, "A novel secure data hiding technique into video sequences using RVIHS," *International Journal of Computer Network and Information Security*, vol. 13, no. 2, pp. 53–65, 2021.

[3] X. Zhong, P. C. Huang, S. Mastorakis, and F. Y. Shih, "An automated and robust image watermarking scheme based on deep neural networks," *IEEE Transactions on Multimedia*, vol. 23, pp. 1951–1961, 2020.

[4] S. Ren, T. Zhang, M. Wang, and K. Shahzad, "Identifiable tampering multi-carrier image information hiding algorithm based on compressed sensing," *IEEE Access*, vol. 8, Article ID 214992, 2020.

[5] X. Chen and S. Chen, "Text coverless information hiding based on compound and selection of words," *Soft Computing*, vol. 23, no. 15, pp. 6323–6330, 2019.

[6] X. Chen, S. Chen, and Y. Wu, "Coverless information hiding method based on the Chinese character encoding," *Journal of Internet Technology*, vol. 18, no. 2, pp. 313–320, 2017.

[7] Z. Zhou, Y. Mu, N. Zhao, Q. M. J. Wu, and C.N. Yang, "Coverless information hiding method based on multi-keywords," in *Proceedings of the International Conference on Cloud Computing and Security*, pp. 39–47, Nanjing, China, July, 2016.

[8] Z. Zhou, Y. Mu, C.N. Yang, and N. Zhao, "Coverless multi-keywords information hiding method based on text," *International Journal of Security and Its Applications*, vol. 10, no. 9, pp. 309–320, 2016.

[9] S. Zheng, L. Wang, B. Ling, and D. Hu, "Coverless information hiding based on robust image hashing," in *Proceedings of the International Conference on Intelligent Computing*, pp. 536–547, Nanjing, China, July, 2017.

[10] Z. Zhou, Q. J. Wu, C.-N. Yang, X. Sun, and Z. Pan, "Coverless image steganography using histograms of oriented gradients-based hashing algorithm," *Journal of Internet Technology*, vol. 18, no. 5, pp. 1177–1184, 2017.

[11] K. Wu and C. Wang, "Steganography using reversible texture synthesis," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 130–139, 2014.

[12] X. Y. Chen, A. Q. Qiu, X. M. Sun, S. Wang, and G. Wei, "A high-capacity coverless image steganography method based on double-level index and block matching," *Mathematical Biosciences and Engineering: MBE*, vol. 16, no. 5, pp. 4708–4722, 2019.

[13] X. Chen, H. Sun, Y. Tobe, Z. Zhou, and X. Sun, "Coverless information hiding method based on the Chinese mathematical expression," in *Proceedings of the International Conference On Cloud Computing And Security*, pp. 133–143, Nanjing, China,, August, 2015.

[14] J. Zhang, H. Huang, L. Wang, H. Lin, and D. Gao, "Coverless text information hiding method using the frequent words hash," *International Journal on Network Security*, vol. 19, no. 6, pp. 1016–1023, 2017.

[15] C. Wang, Y. Liu, Y. Tong, and J. Wang, "GAN-GLS: generative lyric steganography based on generative adversarial networks," *Computers, Materials & Continua*, vol. 69, no. 1, pp. 1375–1390, 2021.

[16] Z. Zhou, H. Sun, R. Harit, X. Chen, and X. Sun, "Coverless image steganography without embedding," in *Cloud Computing and Security*, pp. 123–132, Springer, Switzerland, 2015.

[17] B. Chen, W. Tan, G. Coatrieux, Y. Zheng, and Y.Q. Shi, "A serial image copy-move forgery localization scheme with source/target distinguishment," *IEEE Transactions on Multimedia*, vol. 23, pp. 3506–3517, 2021.

[18] Q. Liu, X. Xiang, J. Qin, Y. Tan, J. Tan, and Y. Luo, "Coverless steganography based on image retrieval of DenseNet features and DWT sequence mapping," *Knowledge-Based Systems*, vol. 192, no. 1, Article ID 105375, 2020.

[19] Y. Luo, J. Qin, X. Xiang, and Y. Tan, "Coverless image steganography based on multi-object recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 7, pp. 2779–2791, 2021.

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, no. 1, pp. 91–99, 2015.

[21] B. Chen, X. Liu, Y. Zheng, G. Zhao, and Y.Q. Shi, "A robust GAN-generated face detection method based on dual-color spaces and an improved Xception," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 99, p. 1, 2021.

[22] Q. Li, X. Wang, X. Wang, and Y. Shi, "CCCIH: content-consistency coverless information hiding method based on generative models," *Neural Processing Letters*, vol. 53, no. 6, pp. 4037–4046, 2021.

[23] C. Yu, D. Hu, S. Zheng, W. Jiang, M. Li, and Z.Q. Zhao, "An improved steganography without embedding based on attention GAN," *Peer-to-Peer Networking and Applications*, vol. 14, no. 3, pp. 1446–1457, 2021.

[24] Y. Tan, J. Qin, X. Xiang, C. Zhang, and Z. Wang, "Coverless steganography based on motion analysis of video," *Security and Communication Networks*, vol. 2021, Article ID 5554058, 16 pages, 2021.

[25] N. Pan, J. Qin, Y. Tan, X. Xiang, and G. Hou, "A video coverless information hiding algorithm based on semantic segmentation," *EURASIP Journal on Image and Video Processing*, vol. 2020, no. 1, pp. 1–18, 2020.

[26] M. Sandler, A. Howard, M. Zhu, and L.C. Chen, "MobileNetV2: inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, Seattle, WA, USA, June, 2018.

[27] L. Zou, W. Wan, B. Wei, and J. Sun, "Coverless video steganography based on inter frame combination," *Communications in Computer and Information Science*, vol. 1386, pp. 134–141, 2021.