

Received February 19, 2019, accepted February 23, 2019, date of publication March 4, 2019, date of current version May 22, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2902464

Hybrid Adaptive Video Steganography Scheme Under Game Model

KE NIU^{1,2}, JUN LI², XIAOYUAN YANG^{1,2}, SHUO ZHANG³, AND BO WANG⁴, (Member, IEEE)

¹Key Laboratory of Network and Information Security, People's Armed Police, College of Cryptographic Engineering, Engineering University of PAP, Xi'an 710086, China

²College of Cryptographic Engineering, Engineering University of PAP, Xi'an 710086, China

³College of Information and Communication, National University of Defense Technology, Xi'an 710106, China

⁴The State University of Technology at Buffalo, Buffalo, NY 14260, USA

Corresponding author: Xiaoyuan Yang (niukey@163.com)

This work was supported in part by the National Key R&D Program of China under Grant 2017YFB0802000, in part by the National Natural Science Foundation of China under Grant U1636114, Grant 61379152, Grant 61403417, and Grant 61872448, and in part by the Shaanxi Natural Science Foundation of China under Grant 2017JM6113 and Grant 2018JM6017.

ABSTRACT The steganographic model based on game theory is improved. A hybrid steganography model that satisfies the Kerckhoffs principle by combining the strategy adaptation of the game theory model and the content adaptation in the cover model is proposed. This model is used to design a high-security steganography algorithm for an H.264 video. The distortion function in UNIWARD algorithm is improved according to the H.264 video coding characteristics. The improved distortion function is used to generate the bias function under the game theory model, and the embedding probability of each position in the cover is calculated. The new distortion function under the hybrid model is generated by combining the embedding probability of the cover position and the distortion cost function under the improved UNIWARD algorithm and using it to embed information in the H.264 residual DCT coefficient block. The experiment proves that the algorithm has higher security than the UNIWARD algorithm. With the same analysis algorithm, the detection error rate is significantly improved compared to the UNIWARD algorithm. The algorithm has a large embedding capacity and good invisibility. When fully embedded, the capacity reaches 1.9% or more, the visual effect does not change significantly, and the PSNR value decreases within 2 dB. After the secret message is embedded in the information, the video bit rate is increased by 9%.

INDEX TERMS Video steganography, game theory model, hybrid adaptive, H.264 encoding.

I. INTRODUCTION

Steganography is the purpose of embedding secret information in the public multimedia cover to achieve secret communication with the receiver. The requirement is to hide the existence of information as much as possible. Therefore, the cover after embedding the information is compared with it before embedding. It is not only visually invisible, but more importantly, it is indistinguishable in statistical features.

The content adaptive information embedding algorithm has been widely used because of its high security [1]. A specific region is selected using a content-based distortion cost function. Then use optimized coding to embed information in the region for the purpose of minimum distortion. Generally, more information is embedded in the complex texture area. Embedding less or no information in simple texture regions,

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

such methods are better resistant to steganalysis. However, the problem with the content adaptive embedding method is that it does not meet the Kerckhoffs Principle [2]. Once the adaptive rule is known by the attacker, it will pose a great threat to the algorithm. It has been proved that in some cases, the leakage of adaptive rules will lead to the security of the algorithm is lower than that of uniform random embedding [3], [4], and there has been a steganalysis method specifically for content adaptive embedding [5]. Another problem is that the security of the steganographic algorithm is more dependent on the design of the distortion function in the framework of "distortion function + optimal coding", and the distortion function design is based on accurate statistical modeling of the cover. But the statistical modeling of the cover itself is very difficult, and in many cases it is not even possible to model accurately. In this case, the steganographic algorithm using this framework is obviously not optimal.

The purpose of information hiding is to hide the message secret in the multimedia cover and to make it impossible for the steganalyst to determine whether the multimedia cover contains information. If both the steganographer and the steganalyst are considered to be as both sides of the game, the game theory can be applied to information hiding. All participants in the game need to choose a strategy. In information hiding, the steganographer can design a certain method to select the location suitable for information embedding. The steganalyst can also focus on and analyze the information according to which position is more suitable for embedding information. Under this steganography method, the embedded position of the steganographer and the steganographic analysis position of the steganalyst are determined according to the other party. The security does not depend on the accurate modeling of the cover; both parties are also open, which is in line with the Kerckhoffs Principle. Therefore, the steganographic method under the game model can theoretically better overcome the problems existing in the content adaptive steganography framework.

The steganographic model based on game theory has certain research results. In [6], the game theory is introduced systematically into the digital information hiding technology. It is proved that there exist Nash equilibrium in both the steganographer and the steganalyst under the hybrid strategy, and the 2-bit information embedding is realized. In theory, the feasibility of information hiding technology based on game theory is demonstrated. Reference [7] further expands the work in [6], gives the Nash equilibrium conclusion on the n-bit information, and realizes the n-bit embedding. In theory, an information hiding model named “Strategic adaptive” based on game theory is proposed. Based on the conclusions of [6], [7], and [8] deduces the payment and Nash equilibrium of both sides of the game when embedding 2-bit information at each position, and expands the hidden capacity of the model. This paper proposes a “Strategic adaptive” and content adaptive hybrid hiding model. Based on this model, a high security steganography scheme for h.264 video that satisfies the Kerckhoffs Principle is designed.

The structure of this paper is as follows: The first section is the introduction. The second section is hybrid adaptive mode construction, including the establishment of the steganographic model and the correction of the model. The third section is the new distortion function design under the hybrid model, which mainly includes the hybrid model analysis and the application in video steganography and the new distortion function design. The fourth section is the design of video steganography algorithm, including analysis and detailed description of the algorithm. The fifth section is experimental data and analysis. The last section is the algorithm summary.

II. HYBRID ADAPTIVE MODEL CONSTRUCTION

A. GAME THEORY IN INFORMATION HIDING

Game theory is a theory about the interaction of strategies, that is, the theory of rational behavior in social situations. The

choice of each player in his or her actions must be based on his judgment of how other players will react [9], one of the most important research topics in game theory is the problem of solving Nash equilibrium points. After the strategy chosen by the players reaches the Nash equilibrium, any one of the players who unilaterally changes their strategies can only make their own income decline (or unchanged), and it is impossible to increase their own income, so in Nash equilibrium point, each player does not dare to act rashly, thus forming a balance.

When the steganographer uses the new method to select the embedding position for steganography in order to improve the security of the algorithm and resist the attacker’s steganalysis, the corresponding steganographer will also foresee this new location selection method and take a new strategy to attack. By analyzing the steganographic model from the perspective of game theory, we can find that the steganographer and the steganalyst in the model just form the two sides of the game: on the one hand, the steganographer chooses the embedded location in the cover to embed the secret information and tries not to be discovered by the steganalyst; on the other hand, the steganalyst should try to find the secret information from the steganographic cover. The benefit (payment) obtained by one party is the other party’s contribution. In the game theory, this situation is called the two-person zero-sum game.

B. MODEL ESTABLISHMENT

Based on the analysis of the information hiding game process, combined with the research conclusions of [7], the two-person zero-sum game model is summarized as follows:

1) The participants in this game are Alice and Eve. Alice is steganographer but Eve is steganalyst. Note that the $Cover = \{x_i | x_i \in \{0, 1\}, i = 1 \dots (n - 1)\}$, that is, $Cover$ is a sequence consisting of n bits, and the length of the secret information m is k .

2) Alice’s strategy action space is select k locations for embedding in a cover with n bits. Since there is no Nash equilibrium in pure strategy action space, it is assumed that Alice’s mixed strategy action space A is a probability distribution over n cover elements:

$$A_i = \{a_i = (a_1, a_2, \dots, a_{n-1}), |a_i \geq 0, i = 1, 2 \dots (n - 1), \sum_{i=0}^{n-1} a_i = k\} \quad (1)$$

where a_i represent the probability that Alice embeds message in position i .

3) The corresponding Eve’s mixed strategy action space is the probability distribution over n stego elements:

$$E_i = \{e_i | e_i \geq 0, i = 1, \dots, n - 1\} \quad (2)$$

e_i represents the probability of Eve interrogating the i -th position to obtain side information to determine whether the position is steganized [9].

4) A monotonically increasing function $f(i) \{i = 0, \dots, n - 1\} \in [1/2, 1]$ is defined to quantify the probability that the cover is most likely to take a certain value at the i -th position,

and can also measure the predictability of the i -th position value. Without loss of generality, define $f(i) = P(x_i = 1)$, indicating the probability that the i -th position takes a value of 1. When $f(i) = 1$, the probability that x_i takes 1 is 1, and the value of the position is determined, with the “minimum” unpredictability. When $f(i) = 1/2$, the probability that x_i takes 1 is 1/2, and its value is 0 or 1 is completely random, with “maximum” unpredictability.

Under the two-person zero-sum game model, [7] quantify the payoff of Eve and Alice as a function associated with $f(i)$ at each position, the probability $a(i)$ of Alice embedding k bit messages at the i -th position and the probability $e(i)$ of Eve interrogating at the i -th position to obtain side information can be expressed as (3) and (4):

$$a(i) = \frac{k}{\tilde{f}(i) \times \sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}} \quad (3)$$

$$e(i) = \frac{1}{\tilde{f}(i) \times \sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}} \quad (4)$$

where $\tilde{f}(i)$ is defined for notational convenience, $\tilde{f}(i) = f(i) - 1/2 \in [0, 1/2]$, called the bias function. $\tilde{f}(i)$ indicates the effect of a coefficient on the global statistical properties. When $\tilde{f}(i) = 0$, it means that the value of the coefficient is completely random between 0 and 1. Alice can embed 1-bit information on the coefficient without any risk under the “Strategic adaptive” embedding condition, which means that the modification of the coefficient has no effect on the global statistical characteristics. On the other hand, when $\tilde{f}(i) = 1/2$, it means that the value of the coefficient takes 0 with a probability of 100% (or 100% probability takes 1). If Alice embeds secret information on the coefficient, then Eve can be completely detected under the condition of “Strategic adaptive”, that is to say, the coefficient has the greatest influence on the global statistical characteristics. In order to avoid these two extreme cases, set $\tilde{f}(0) = \varepsilon$, $\tilde{f}(n-1) = 1/2 - \varepsilon$, where $\varepsilon \approx 0$ and $\varepsilon > 0$. Under the game theory model of [6]–[8], the bias function $\tilde{f}(i)$ is defined to describe the extent to which the cover is suitable for embedding, but does not give a specific form. In (3) and (4), the calculation of the bias function $\tilde{f}(i)$ is determined by the content of the cover, which combines the game-based Strategic adaptation with the cover-based content adaptation to generate a hybrid adaptive steganography model in this paper.

C. MODEL CORRECTION

In the above model, the sum of the embedding probabilities for each location in the cover is $\sum_{i=0}^{n-1} a(i) = k$, where k is the maximum message embedding length [7], i.e. a k -bit message is fully embedded in the cover. Since $k \geq 1$, the value of $a(i)$ may be greater than 1, which is inconsistent with the concept of probability theory. If you simply set $a(i)$ with a value greater than 1 to 1 [7], this approach creates two problems:

1) The actual embedding length is made smaller than the theoretical embedding length. Since the modification will

result in $a(i)$ value greater than 1 becoming smaller (modified to 1), such that $\sum_{i=0}^{n-1} a(i) = k$, the actual embedded information length is less than the theoretical length k of the message that can be embedded [8].

2) It is not optimal to select the embedding position according to the $a(i)$ value, such as $a(m) = 2$ and $a(n) = 3$, when their values are all set to 1, it indicates that the m and n positions have the same suitable embedding degree. But in fact the n position is more suitable for embedding ($a(n) > a(m)$).

In order to solve these two problems, the paper makes corrections to the probability model as follows:

Step1: Summing all the parts whose probability exceeds 1, that is, $Sum = \sum_{i=0}^{n-1} (a(i) - 1)$ where $a(i) > 1$, and set all of these probability values to 1.

Step2: Count the number of values in a that is less than 1, and record it as Num . Add $a(i)$ with a probability value less than 1 plus Sum/Num , that is, increase the excess portion evenly to a value less than 1.

Step3: If all the probabilities $a(i) \in (0, 1]$ after completing Step 1 and Step 2, the correction end, otherwise the steps Step 1 to Step 3 are repeated to perform the correction again until the values of all the probabilities are between 0 and 1.

Step4: Sort a with a probability value of 1 according to the original value to obtain a sequence (a_1, a_2, \dots, a_n) , where $a_i \geq a_{i+1}$. The correction value a'_i can be calculated according to (5):

$$a'_i = a_i + \varepsilon_i \quad (5)$$

where $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n \approx 0$, $\varepsilon_i \geq \varepsilon_{i+1}$.

Since $\varepsilon_i \approx 0$, the probability $a(i) \in (0, 1]$ can basically be considered. At the same time, for the probability $a(m)$ and $a(n)$ ($m \neq n$), their relative size relationship still has not changed, so it can still accurately reflect the suitability of message embedding in different locations. During the actual operation from step1 to step3, there may be cases where the value of $a(i)$ cannot be corrected to between 0 and 1. The algorithm stipulates that when the number of repetitions exceeds p steps, the correction is stopped, and the value greater than 1 is set to 1, then execute Step4.

III. DESIGN OF DISTORTION FUNCTION UNDER HYBRID MODEL

A. MODEL ANALYSIS AND APPLICATION

In the content adaptive model, the hidden person selects certain regions in the cover to embed information through specific adaptive rules, and the privacy rules are strictly kept secret. Since the analyst does not know the adaptive rules, it is unable to get embedded area of information and the analyst will select some areas according to experience or analysis of the entire cover. Under the steganalysis framework of “feature extraction + classifier recognition”, in order to not miss the detection of the embedded region, more and more feature types are extracted [10], and the dimension is

increasing. This high-dimensional feature not only causes difficulty in feature extraction, but also features that are not embedded in the information region in the cover, which may affect the detection effect and reduce the detection rate of the algorithm. However, once the adaptive rules are leaked, the analyst will know the embedded region in the cover, and only need to extract the features of the region in the steganographic analysis, which will effectively reduce the feature dimension and eliminate the interference of normal features, and greatly improve the detection rate of the algorithm. Therefore, the security based on the content adaptive steganography algorithm is based on the violation of the Kerckhoffs Principle, which poses a security risk to the algorithm.

The ‘‘Strategic adaptive’’ model based on game theory enables steganographer and the steganalyst to establish Nash equilibrium on the hybrid strategy set. The embedded strategy is open to both sides of the game and better satisfies the Kerckhoffs Principle. In this model [6], [7] the probability that the steganographer embeds information at the i -th position of the cover and the probability that the steganalyst interrogates the i -th position on the stego are $a(i)$ and $e(i)$, respectively. It is also represented by (3) and (4), but the bias function $\tilde{f}(i)$ is not give the specific generation method and is not associated with the cover content. Although it can better meet the requirements of statistical security in the single ‘‘Strategic adaptive’’ model, it does not consider the visual quality of the digital cover after embedding information, and the practicability is not strong [11].

According to the section II, the algorithm corrects the value range of $a(i)$ in our model, such that $a(i) \in (0, 1]$, which solves the problem that the actual embedded information length is smaller than the theoretical embedded information length and the optimal position cannot be selected. At the same time, the design of the bias function $\tilde{f}(i)$ is combined with the cover content. The value of the i -th position $\tilde{f}(i)$ is determined by the cover content, thus converting a single ‘‘Strategic adaptive’’ model into a hybrid model combining ‘‘Strategic adaptive’’ and content adaptation. This model better solves the statistical security and visual quality problems of the steganography algorithm. In this section, we use the hybrid model, combine the h.264 coding characteristics and the video content to establish the $\tilde{f}(i)$, and design a steganography algorithm suitable for H.264 video.

B. VIDEO DISTORTION FUNCTION DESIGN

The content adaptive hiding algorithm usually selects the embedding position based on minimizing the embedding distortion [12]. The typical ones of these algorithms are WOW [13], UNIWARD [14], HILL [15] etc., where the UNIWARD algorithm is applicable to the spatial domain and frequency domain, this paper uses the improved UNIWARD algorithm to calculate the bias function $\tilde{f}(i)$.

The distortion function is obtained according to the characteristics of the H.264 compression standard and the original

UNIWARD algorithm, as shown in (6) and (7):

$$D(X, Y) = \sum_{k=1}^3 \sum_{u=0}^{15} \sum_{v=0}^{15} \frac{|W_{uv}^{(k)}(T^{-1}(X)) - W_{uv}^{(k)}(T^{-1}(Y))|}{\sigma + |W_{uv}^{(k)}(T^{-1}(X))|} \quad (6)$$

$$D^{SI}(X, Y) = D(P, T^{-1}(Y)) - D(P, T^{-1}(X)) \\ = \sum_{k=1}^3 \sum_{u=0}^{15} \sum_{v=0}^{15} \left[\frac{|W_{uv}^{(k)}(P) - W_{uv}^{(k)}(T^{-1}(Y))|}{\sigma + |W_{uv}^{(k)}(P)|} - \frac{|W_{uv}^{(k)}(P) - W_{uv}^{(k)}(T^{-1}(X))|}{\sigma + |W_{uv}^{(k)}(P)|} \right] \quad (7)$$

where (6) and (7) are DCT domain embedded position distortion functions without side information and with side information, respectively. $W_{uv}^k(T^{-1}(X))$ and $W_{uv}^k(T^{-1}(Y))$ represent the wavelet transform coefficients of the cover and the stego video. u and v are the coefficient positions, $u, v \in \{0, \dots, 15\}$ k is the sub-band direction, and $k = 1, 2, 3$ corresponds to the three sub-bands LH, HL and HH in the wavelet transform. X, Y are the quantized DCT coefficients in the compressed video, T^{-1} representing inverse quantization and inverse DCT. $\sigma > 0$ is a stable constant used to avoid the case where the denominator is 0. P in (7) is the original video before compression.

Since wavelet transform and DCT are both nonlinear transforms, the change of a spatial domain data element will affect all wavelet coefficients or DCT coefficients after transformation; likewise, a wavelet coefficient or DCT coefficient change will affect all spatial domain data. Therefore, when calculating the position distortion according to (6) and (7), it is necessary to keep the values of all other positions in one transform unit unchanged, and the distortion cost value of the (u, v) position is $\rho_{u,v}(X, Y_{u,v})$, then:

$$\rho_{u,v}(X, Y_{u,v}) = \begin{cases} D(X, X_{\sim u,v}Y_{u,v}) & \text{with side information} \\ D^{SI}(X, X_{\sim u,v}Y_{u,v}) & \text{without side information} \end{cases} \quad (8)$$

In the (8), $X_{\sim u,v}Y_{u,v}$ represents a video in which only the $x_{u,v}$ in the cover is changed to $y_{u,v}$, and the remaining position elements remain unchanged. Sort $\rho_{u,v}(X, Y_{u,v})$, to get a one dimensional sequence $\rho(i)$, and construct a new bias function $\tilde{f}(i)$ from $\rho(i)$, according to (9).

$$\tilde{f}(i) = \begin{cases} \varepsilon & \tilde{f}(i) = 0 \\ \frac{1}{2} \times \frac{\rho(i) - MIN(\rho)}{MAX(\rho) - MIN(\rho)} & 0 < \tilde{f}(i) < 1/2 \\ 1/2 - \varepsilon & \tilde{f}(i) = 1/2 \end{cases} \quad (9)$$

where $MIN(\rho)$ is the minimum value of distortion cost, $MAX(\rho)$ is the maximum value of the distortion cost, and ε is a minimum positive number. Substituting $\tilde{f}(i)$ into (3) yields the embedding probability $a(i)$ for each embedding position under the game theory model.

Correct $a(i)$ to make $a'(i) \in (0, 1]$, $a'(i)$ indicates the corrected embedding probability of the i -th elements in the cover. When the embedding probability of a position is 0, it indicates that the position is least suitable for embedding information, and the corresponding distortion cost should be the maximum value; otherwise, the embedding probability of a position is 1, it indicates that the location is most suitable for embedding information, and its corresponding distortion cost is minimum [11]. Combining the embedded probability $a(i)$ with the original distortion cost function $\rho(i)$, a new distortion cost function is obtained according to (10). The original distortion cost function is modified by the embedded probability to obtain the new distortion cost function under the content adaptive and "Strategic adaptive" hybrid model.

$$\rho'(i) = \frac{\rho(i)}{a'(i)} \quad (10)$$

IV. VIDEO STEGANOGRAPHY ALGORITHM UNDER HYBRID MODEL

A. ALGORITHM ANALYSIS AND DESIGN

As a new generation video compression standard, H.264 has low transmission rate, high image quality, flexible compression mode, and good network adaptability. It has been widely used in video surveillance, network TV, HD video and other fields [16], [17]. Considering the characteristics of the H.264 standard, each 16×16 residual macroblock in the P frame (forward prediction frame) is used as an embedding position [18], [19]. To avoid the influence of video compression on the embedded information, the information is embedded in the quantized DCT coefficients.

In H.264, the video frame is DCT-transformed in units of 4×4 , so that the original distortion cost is calculated in units of 4×4 blocks using the improved UNWARD. The residual block is divided into 16 data blocks for wavelet analysis. Since the energy in the residual signal is small, the db4 wavelet base with smaller vanishing moment is selected in the wavelet analysis. Compared with the db8 wavelet base in the original algorithm, the execution speed of the algorithm is improved without degrading the performance. The 1D low-pass and high-pass filters used in the algorithm are shown in FIGURE 1.

The distortion cost function calculation in the algorithm can be divided into two categories: one type of H.264 cover has a corresponding original YUV video before compression, and this kind of cover is calculated by DCT domain distortion cost function with side information. The second type is the H.264 cover that cannot obtain the raw video. This type of video is calculated using the DCT domain distortion function without side information. After obtaining the distortion cost, the information is embedded in the quantized DCT coefficients of the residual macroblock by the STC [20].

B. ALGORITHM DESCRIPTION

The algorithm embeds information in the 16×16 residual DCT coefficients, and calculates the original distortion cost

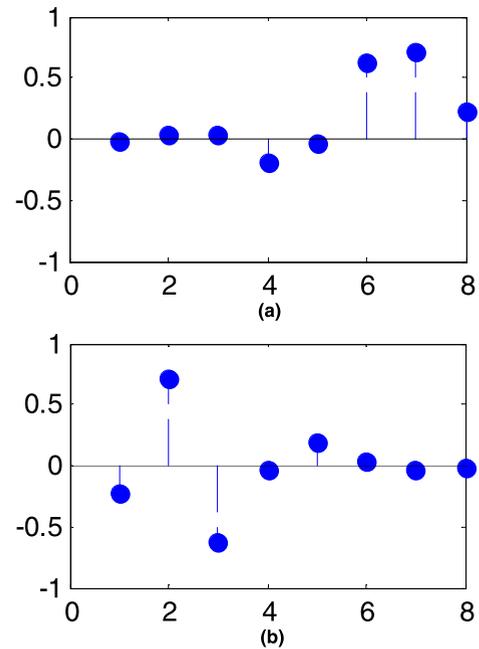


FIGURE 1. Daubechies wavelet filter bank. (a) db 4 wavelet decomp. low-pass. (b) db 4 wavelet decomp. high-pass.

function $\rho(i)$ according to whether the cover has the raw video selection $D(X, Y)$ or $D^{SI}(X, Y)$. The bias function is calculated by $\rho(i)$ to obtain the embedding probability $a(i)$ for each position, and a new distortion function $\rho'(i)$ is obtained by $\rho(i)$ and $a(i)$ together. Finally, STC coding is used to form stego H.264 video. The extraction process is to decode the stego video with STC.

Theoretically, blocks of any size can be used as embedded information units, but the information embedding process uses STC coding, which has the size requirement for the block of embedded information. When 4×4 is adopted, the STC code embedded information cannot be realized. At the same time, considering that the H.264 encoder is encoded by macroblock-by-macroblock, 16×16 is used as the embedding unit in this scheme. If a larger data block such as 32×32 , 64×64 is used, the STC encoding time will be lengthened.

The algorithm framework is shown in FIGURE 2.

The specific steps of the algorithm are as follows:

Step1: Decode the H.264 video cover to obtain one residual macroblock in the P frame. Calculate the number N of non-zero AC coefficients of the residual DCT block. If N is greater than or equal to the threshold T , perform Step 2, otherwise perform Step 7.

Step2: Determine the cover video type. For the cover without the raw video, according to (6) calculate the original distortion cost value in units of residual blocks, and vice versa, according to (7). The original distortion value of the i -th element is denoted as $\rho(i)$.

Step3: Substituting the maximum value $MAX(\rho)$ and the minimum value $MIN(\rho)$ of $\rho(i)$ and $\rho(i)$ into (9) to obtain the

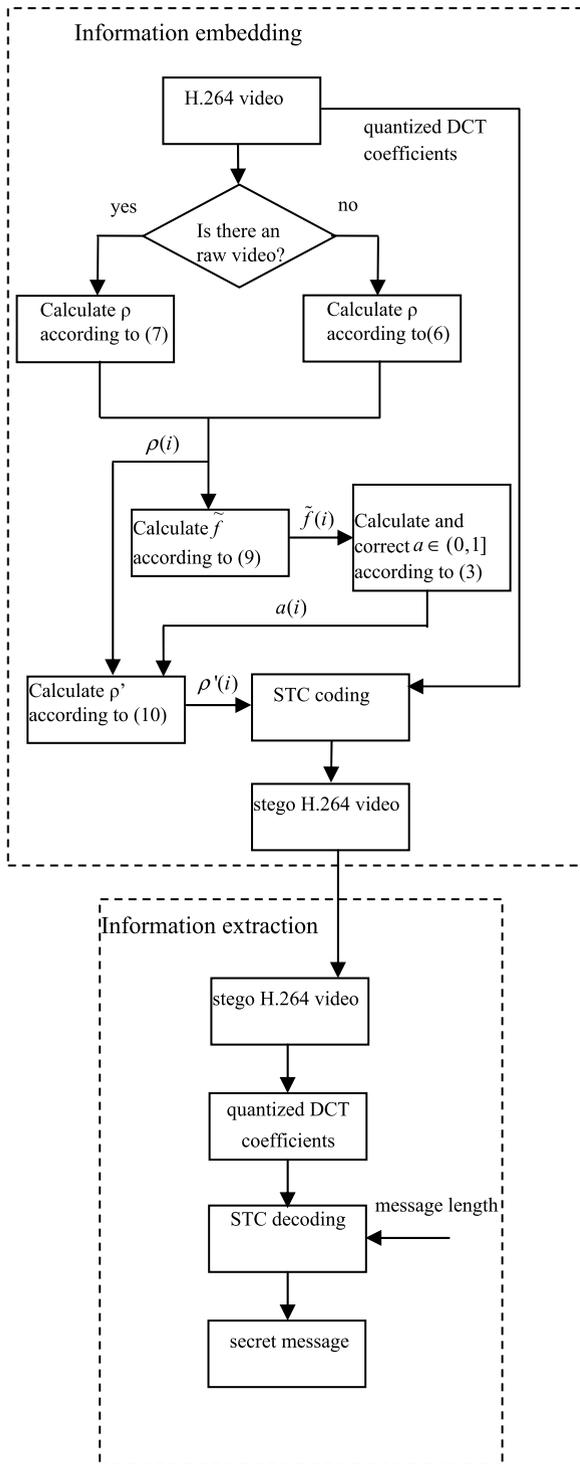


FIGURE 2. Framework of our scheme. (Contains information embedding and information extraction.)

bias function $\tilde{f}(i)$. According to $\tilde{f}(i)$ and (3), the embedding probability $a(i)$ at each position is calculated, and $a(i)$ is corrected by the section II to make $a(i) \in (0, 1]$.

Step4: From the corrected embedding probability $a(i)$ of the i -th position in the residual block and the i -th position

original distortion cost $\rho(i)$ calculated in Step 1, a new distortion function $\rho'(i)$ is obtained according to (10).

Step5: According to the new distortion cost value $\rho'(i)$ and the information to be embedded, the minimum distortion coding method STC is used to embed message in the quantized DCT coefficients to obtain a block of stego residual DCT coefficients.

Step6: Calculate the number N' of non-zero AC coefficients of the DCT block. If N' is less than the threshold T , it is an invalid embedded block, and the messages embedded in the invalid block needs to be re-embedded in the next macroblock.

Step7: Continue the H.264 encoding of the DCT coefficient block to obtain the H.264 code stream of the macroblock. Repeat Step1-Step7 until all messages are embedded or all macroblocks are embedded with message.

V. ALGORITHM SIMULATION

A. EXPERIMENTAL PARAMETER SETTING

In the experiment, the video clips of 7-segment QCIF format (176×144) and 4-segment CIF format (352×288) were used for simulation, as shown in TABLE 1 The Baseline Profile H.264 video codec is used for the test video using the X264 codec and the P264 decoder with a quality factor $QP = 28$. Since the experimental H.264 cover video has corresponding raw video, the algorithm uses SI-UNIWARD to calculate the original distortion function ρ , the stability factor σ takes 2^{-6} , and the non-zero AC coefficient threshold T in the DCT block takes 4. The constraint height $h = 6$ in the STC encoding.

B. EMBEDDED POSITION SELECTION

The embedding position in the algorithm is determined by the content-based distortion cost and the embedding probability under the game theory. FIGURE3 is a diagram of inverse transforming into a spatial domain after embedding information on the DCT domain of the foreman video's second residual frame. (a) is the luminance component of the cover video foreman's second frame; (b) is the residual data of the second frame in the H.264 compression process; (c) and (d) are diagrams of embedding information in residual DCT coefficients using SI-UNIWARD and the proposed algorithm, respectively; (e) is the embedded position of the DCT domain obtained by the SI-UNIWARD algorithm to the spatial domain; (f) is the embedded position of the DCT domain obtained by the hybrid model algorithm to the spatial domain.

It can be seen in FIGURE3.that the embedded position selected by the SI-UNIWARD algorithm is relatively scattered, which indicates that some locations in the relatively weak texture area are still considered to be suitable for embedding information under the content adaptive Principle. When the embedded position is known, these areas are easily statistically modeled to analyze the existence of information. The embedding position calculated by the algorithm proposed in this paper is more concentrated on the texture region.

TABLE 1. Experimental video clip.

Video sequence	Resolution	Number of frames	Video clip
Contaier_qcif.yuv	176×144	300	
Forman_qcif.yuv	176×144	300	
Missamerica_qcif.yuv	176×144	150	
News_qcif.yuv	176×144	300	
Saleman_qcif.yuv	176×144	449	
City_qcif.yuv	176×144	150	
Carphone_qcif.yuv	176×144	382	
Flower_cif.yuv	352×288	250	
Coastguard_cif.yuv	352×288	300	
Mobile_cif.yuv	352×288	300	
Stefan_cif.yuv	352×288	90	

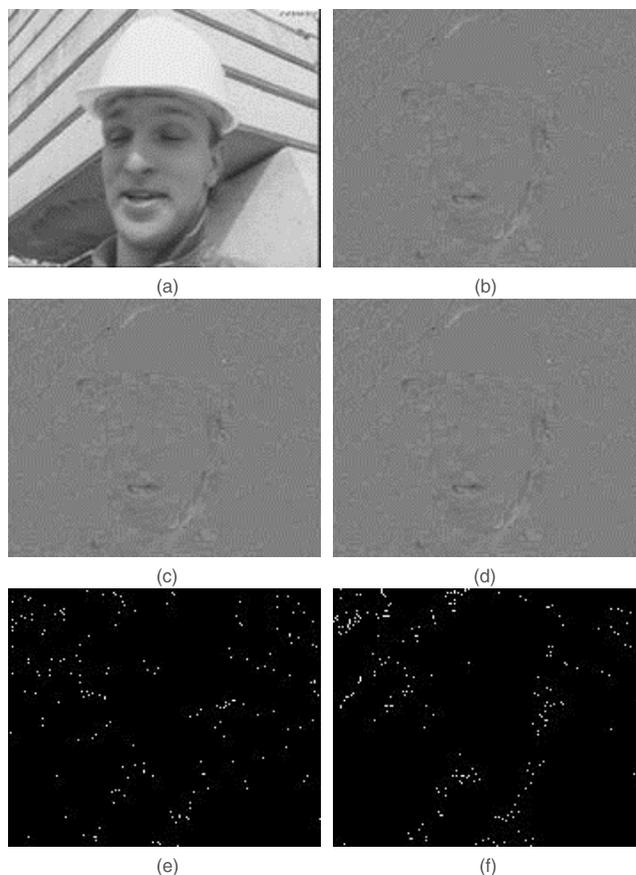


FIGURE 3. Embedding position chosen by UNIWARD and our algorithm. (a) Cover's luminance frame. (b) Residual frame. (c) Residual frame after embedding information with SI-UNIWARD algorithm. (d) Residual frame after embedding information with the proposed algorithm. (e) Embedded position selected by SI-UNIWARD algorithm. (f) Embedded position selected by the proposed algorithm.

Since it is very difficult to establish an accurate statistical model for the complex texture region, the analyst cannot easily obtain the analysis result even if he knows the possible embedding position of the information. In accordance with the requirements of the adaptive rule disclosure in game theory, the Kerckhoffs Principle is better satisfied.

C. EMBEDDED CAPACITY AND BIT RATE

The algorithm uses STC encoding to embed information, and the embedding rate needs to be pre-designed in STC encoding. Combining the characteristics of video coding, the embedding rate is determined according to the number of non-zero AC coefficients in each 16 × 16 residual DCT block. The embedding rate of each block is calculated by (11), where α_i is the embedding rate (bpp) of the i -th residual block, N_i is the number of non-zero AC coefficients in the i -th residual block. After the information is embedded, the optimal prediction in the original encoding process is destroyed, so the bitrate is increased. To measure the change of the bitrate, Define the bit rate increase rate BR_{var} and calculate BR_{var} according to (12). In (12), BR_{em} is the size of the video after embedding the information, and BR_{org} is the original video size. TABLE 2 shows the maximum embedding capacity and

bitrate of the video.

$$\alpha_i = \frac{N_i}{16 \times 16} \tag{11}$$

$$BR_{var} = \frac{BR_{em} - BR_{org}}{BR_{org}} \times 100\% \tag{12}$$

$$\beta = \frac{E}{S} \tag{13}$$

Calculate the embedding rate according to (13). In (13), β is the embedding rate, E is the embedded data size, and S is the size of stego video. It can be seen that the capacity of the algorithm is basically more than 1.9% of the video size, and the bitrate change is controlled within 9%. Video with complex texture and fast motion speed has a relatively high code rate increase, such as City_qcif.yuv and Mobile_cif.yuv video. This is due to the embedding of more information in these videos. In this case, the embedding rate can be appropriately controlled to reduce the change of the bitrate.

D. SECURITY AND INVISIBILITY

The algorithm is used to embed the information in the video and compare the histogram of the DCT block before and

TABLE 2. Capacity and bitrate of our algorithm.

Video sequence	Compressed video size (KB)		Embedding bit (KB)	Embedding rate (%)	Increased bitrate (%)
	Cover	Stego			
Container_qcif.yuv	84	88	1.76	2.00	4.76
Foreman_qcif.yuv	209	221	6.23	2.82	5.74
Miss-america_qcif.yuv	31	32	1.05	3.28	3.23
News_qcif.yuv	347	359	6.89	1.92	3.46
Salesman_qcif.yuv	121	125	2.39	1.91	3.31
City_qcif.yuv	99.9	107	3.89	3.64	7.11
Carphone_qcif.yuv	300	315	8.22	2.61	5.00
Flower_cif.yuv	1,892	2,005	58.21	2.65	5.97
Coastguard_cif.yuv	1,494	1,561	31.03	1.99	4.48
Mobile_cif.yuv	2,219	2,415	65.51	2.69	8.12
Stefan_cif.yuv	491	526	11.06	2.10	7.13

after embedding. Considering the 4×4 DCT transform in the h.264 encoding process, as shown in FIGURE4. It is the second frame of the foreman video's histogram of all 4×4 residual DCT block AC coefficients. (a) is an AC coefficient histogram of cover, and (b) is an AC coefficient histogram at full embedding. As can be seen from the figure, the algorithm better maintains the DCT coefficient histogram.

Considering that both SI_UNIWARD and the NPQ [21] are modified and embedded in the DCT coefficients of the compressed domain based on "content adaptation", the proposed algorithm is compared with them to further test the security of "hybrid adaptive model". The SI_UNIWARD is

embedded with STC code, the constraint height is $h = 6$, and the embedding rate is the same as the full embedding state of our proposed algorithm. In the NPQ algorithm, the message is embedded in the non-zero AC coefficients of all residual DCT blocks, setting the parameter $\mu = 0$, $\lambda_1 = \lambda_2 = 1/2$, which is the "safest" embedded state in the NPQ algorithm. Since both NPQ and SI_UNIWARD are image steganography algorithms, in the experiments in this paper, 16×16 residual macroblocks are regarded as "images" for information embedding. Three types of feature sets are used in the test, which are the joint feature set of SRM [22] and JRM [23], the ST_D [24] feature set and the MAX-GFR [25] feature set, and using the ensemble classifier [26] with Fisher linear discriminant as the base learner. The definition detection error rate is as shown in (14), where F is the detection error rate, NF is the false alarm rate, and PF is the miss rate. The detection error rate results are shown in TABLE 3. It can be seen that when the SRM and JRM joint feature set is used, the proposed algorithm security is higher than the NPQ algorithm, which is slightly higher than the SI_UNIWARD algorithm. When the ST_D feature set is used, the proposed algorithm is basically the same as SI_UNIWARD, slightly higher than the NPQ algorithm, because the ST_D feature set is mainly for spatial domain detection, the detection error rate is high for all algorithms. MAX-GFR is a class of attack features of selected channels [23], [27]. It is a weight allocation method in which the maximum change probability of corresponding correlation DCT coefficients is calculated as the weight of each filtered coefficient, and the final feature set is obtained by accumulating the weights in corresponding histogram statistical samples. It has a high detection rate for the traditional UNIWARD algorithm. It estimates the embedding probability of each cover through content adaptive strategy, and extracts and detects features for large probability embedding positions. The experimental results show that the algorithm in this chapter has better resistance to selecting channel attacks

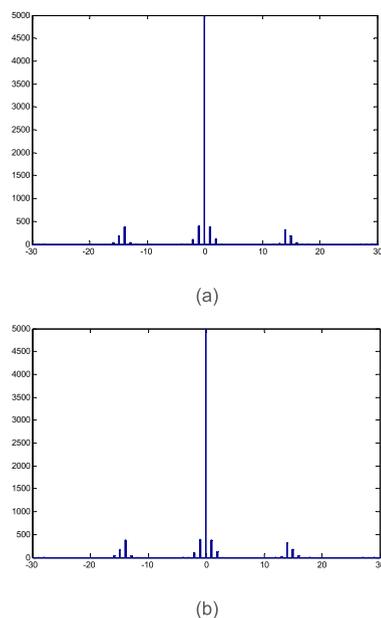


FIGURE 4. AC coefficient histogram. (a) is the cover video residual block AC coefficient histogram (b) is the stego video residual block AC coefficient histogram.

TABLE 3. Detection error rate under different feature sets.

Video sequence	SRM and JRM joint feature set			ST_D feature set			MAX-GFR set	
	NPQ (%)	SI_UNIW ARD(%)	Proposed (%)	NPQ (%)	SI_UNIW ARD(%)	Proposed (%)	SI_UNIWARD (%)	Proposed (%)
container_qcif.yuv	11.15	41.63	48.22	40.46	46.41	47.01	39.21	43.35
foreman_qcif.yuv	23.92	40.72	45.18	39.81	46.00	46.12	36.50	43.78
miss-america_qcif.yuv	19.20	38.41	44.41	36.09	41.92	42.72	32.47	39.11
news_qcif.yuv	26.74	44.13	49.58	44.47	49.02	49.13	43.82	42.19
salesman_qcif.yuv	12.72	41.19	48.63	41.14	48.27	48.11	38.45	40.20
City_qcif.yuv	10.10	39.93	45.10	42.76	47.94	48.72	39.75	41.42
carphone_qcif.yuv	14.41	43.74	47.25	39.22	43.18	44.51	33.64	38.43
flower_cif.yuv	18.39	46.80	48.73	42.10	48.12	48.00	37.73	40.98
Coastguard_cif.yuv	09.11	38.92	42.88	34.04	46.86	47.11	31.09	38.01
mobile_cif.yuv	12.01	40.11	43.52	38.92	45.33	45.21	34.14	40.37
Stefan_cif.yuv	16.74	42.58	47.71	43.53	47.93	48.41	37.36	41.56

TABLE 4. Erage PSNR value of each video.

Video sequence	Cover	Propose	Algorithm1	Algorithm2
container_qcif.yuv	40.30	39.55	39.09	38.44
foreman_qcif.yuv	39.70	38.52	38.40	38.23
Miss_america_qcif.yuv	40.55	40.39	39.79	38.71
news_qcif.yuv	41.42	41.12	40.50	39.69
salesman_qcif.yuv	39.81	39.40	38.48	37.69
City_qcif.yuv	43.10	42.60	42.36	41.49
carphone_qcif.yuv	38.73	38.70	37.83	37.18
flower_cif.yuv	36.79	35.94	36.42	35.59
Coastguard_cif.yuv	40.89	40.51	40.04	39.47
mobile_cif.yuv	35.73	33.98	33.70	33.34
stefan_cif.yuv	37.86	37.41	37.48	36.51

than SI-UNIWARD.

$$F = \frac{NF + PF}{2} \tag{14}$$

The codec of each frame in the video compression process is referenced to the previous frame (or several frames), and the modification of the reference frame will cause the codec error to accumulate, resulting in the video visual quality declining. The algorithm adaptively adjusts the embedding amount according to the non-zero AC coefficient of the residual block, and selects the texture complex region to use STC coding embedding information. At the same time, in order to prevent the cumulative effect of errors from increasing, the frequency at which I frames appear is controlled during h.264 encoding. The algorithm sets an I frame (Intra-coded frame) at least in 150 frames, which better maintains the visual quality of the video.

FIGURE 5. shows the effect of Saleman_qcif and City_qcif video clips before and after information embedding, and TABLE 4 shows the average PSNR value of each video in

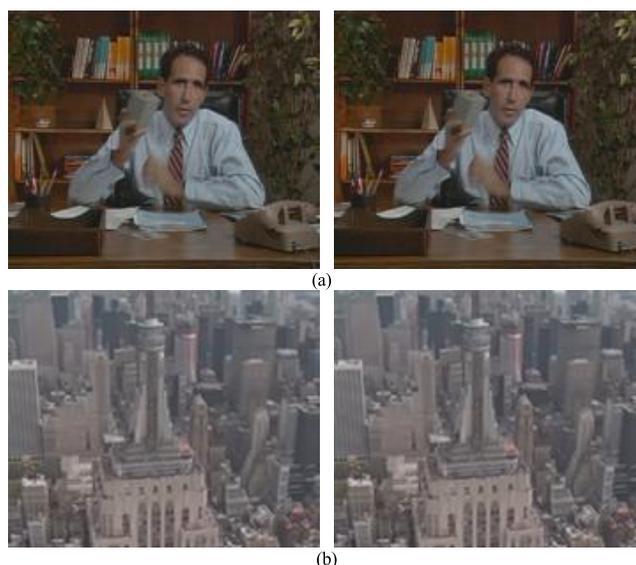


FIGURE 5. The effect of some video clips before and after information embedding, the left side is the original video frame, and the right side is the corresponding frame after the embedded information. (a) is the 60th frame of the Saleman_qcif, (b) is the 80th frame of the City_qcif.

our proposed and other algorithms. Algorithms 1 and 2 are the algorithms in documents [28] and [29], respectively.

VI. CONCLUSION

Under the framework of the steganographic algorithm of “optimized coding + distortion function”, the STC-based optimization coding is close to the theoretical upper limit, so designing a good distortion function is the key to improving the security of the steganography algorithm. The content-based adaptive distortion function has good resistance to steganographic analysis in the case of adaptive rule secrecy, but it is obviously violated the Kerckhoffs Principle in cryptography, which makes the steganography algorithm a security risk. The advantage of the proposed hybrid model

is to make full use of the characteristics of Nash equilibrium in game theory, so that the steganography algorithm under the model has the advantages of the content adaptive algorithm, and can theoretically satisfy the Kerckhoffs Principle. Based on the hybrid model, an H.264 video steganography algorithm is designed by improving the model to the video cover. A new practical steganographic framework of “optimized coding + distortion function + embedded probability function” is implemented. Experiments show that the algorithm has high security and further proves the effectiveness of the hybrid model. Especially when resisting the “channel selection” attack method (such as MAX-GFR), the security is significantly higher than the content adaptive UNIWARD algorithm. In the improved hybrid model steganography framework, “Strategic adaptive” and content adaptation are independent of each other. However, this kind of method needs to calculate the embedded probability function of each position, which has higher time complexity than the content adaptive algorithm. In the future work, different or new content adaptive algorithms can be applied to the new framework to further improve the performance of the steganographic algorithm under the hybrid model. The embedded framework proposed has good applicability. As long as it is combined with the content adaptive algorithm of the corresponding cover, the steganography algorithm with different types of cover can be designed. In addition to being applicable to video, it is also effectively applied to images and audio.

REFERENCES

- [1] Y. Zhang, D. Ye, J. Gan, Z. Li, and Q. Cheng, “An image steganography algorithm based on quantization index modulation resisting scaling attacks and statistical detection,” *Comput., Mater. Continua*, vol. 56, no. 1, pp. 151–167, Jul. 2018.
- [2] A. Kerckhoffs, “La cryptographie militaire,” *J. Sci. Militaires*, vol. 9, pp. 5–83, Jan. 1883.
- [3] R. Böhme, “An epistemological approach to steganography,” in *Proc. IH*, Darmstadt, Germany, Jun. 2009, pp. 8–10.
- [4] R. Böhme and A. Westfeld, “Exploiting preserved statistics for steganalysis,” in *Proc. IH*. New York, NY, USA: Springer-Verlag, 2004, pp. 82–96.
- [5] S. Tan and B. Li, “Targeted steganalysis of edge adaptive image steganography based on LSB matching revisited using B-spline fitting,” *IEEE Signal Process. Lett.*, vol. 19, no. 6, pp. 336–339, Jun. 2012.
- [6] P. Schöttle and R. Böhme, “A game-theoretic approach to content-adaptive steganography,” in *Proc. IH*. New York, NY, USA: Springer-Verlag, 2012, pp. 125–141.
- [7] B. Johnson, P. Schöttle, and R. Böhme, “Where to hide the bits?” in *Proc. GameSec*. Berlin, Germany: Springer, 2012, pp. 1–17.
- [8] J. Liu and G. M. Tang, “Game research on large-payload and adaptive steganographic counterwork,” *Acta Electron. Sinica*, vol. 42, no. 10, pp. 1963–1969, 2014.
- [9] J. Fridrich, “On the role of side information in steganography in empirical covers,” *Proc. SPIE*, vol. 8665, no. 4, pp. 280–289, 2013.
- [10] T. Pevný, T. Filler, and P. Bas, “Using high-dimensional image models to perform highly undetectable steganography,” in *Information Hiding* (Lecture Notes in Computer Science), vol. 6387. 2010, pp. 161–177.
- [11] J. Li et al., “A game-theoretic method for designing distortion function in spatial steganography,” *Multimedia Tools Appl.*, vol. 76, no. 10, pp. 12417–12431, 2016.
- [12] Q. Nie, J. Weng, X. Xu, and B. Feng, “Defining embedding distortion for intra prediction mode-based video steganography,” *Comput., Mater. Continua*, vol. 55, no. 1, pp. 59–70, 2018.
- [13] V. Holub and J. Fridrich, “Designing steganographic distortion using directional filters,” in *Proc. WIFS*, Tenerife, Spain, Dec. 2012, pp. 234–239.
- [14] V. Holub, J. Fridrich, and T. Denemark, “Universal distortion function for steganography in an arbitrary domain,” *EURASIP J. Inf. Secur.*, vol. 2014, no. 1, p. 1, 2014.
- [15] B. Li, M. Wang, J. Huang, and X. Li, “A new cost function for spatial image steganography,” in *Proc. ICIP*, Paris, France, Oct. 2014, pp. 4206–4210.
- [16] H. J. Bi and J. Wang, “H.264/AVC encoder principle,” in *Next-Generation Video Compression Coding Standard*, 1st ed. Beijing, China: Posts Telecom Press, 2009, pp. 98–102.
- [17] P. Liu et al., “Secure video streaming with lightweight cipher PRESENT in an SDN testbed,” *Comput., Mater. Continua*, vol. 57, no. 3, pp. 353–363, 2018.
- [18] Z. Pan, X. Yi, and L. Chen, “Motion and disparity vectors early determination for texture video in 3D-HEVC,” *Multimedia Tools Appl.*, pp. 1–18, Nov. 2018. doi: 10.1007/s11042-018-6830-7.
- [19] Z. Pan, J. Lei, Y. Zhang, and F. L. Wang, “Adaptive fractional-pixel motion estimation skipped algorithm for efficient HEVC motion estimation,” *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 1, pp. 12–19, Jan. 2018.
- [20] T. Filler, J. Judas, and J. Fridrich, “Minimizing additive distortion in steganography using syndrome-trellis codes,” *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 920–935, Sep. 2011.
- [21] F. Huang, J. Huang, and Y. Q. Shi, “New channel selection rule for JPEG steganography,” *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 4, pp. 1181–1191, Aug. 2012.
- [22] J. Fridrich and J. Kodovský, “Rich models for steganalysis of digital images,” *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 868–882, Jun. 2012.
- [23] J. Kodovský and J. Fridrich, “Steganalysis of JPEG images using rich models,” *Proc. SPIE*, vol. 8303, Jan. 2012, Art. no. 83030A.
- [24] Y. Su, F. Yu, and C. Zhang, “Digital video steganalysis based on a spatial temporal detector,” *KSI Trans. Internet Inf. Syst.*, vol. 11, no. 1, pp. 360–373, 2017.
- [25] Y. Zhang, C. Yang, X. Luo, F. Liu, J. Lu, and X. Song, “Steganalysis of content-adaptive JPEG steganography based on the weight allocation of filtered coefficients,” in *Proc. ICCSN*, Guangzhou, China, May 2017, pp. 1308–1312.
- [26] J. Kodovský, J. Fridrich, and V. Holub, “Ensemble classifiers for steganalysis of digital media,” *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 432–444, Apr. 2012.
- [27] T. D. Denemark, M. Boroumand, and J. Fridrich, “Steganalysis features for content-adaptive JPEG steganography,” *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 8, pp. 1736–1746, Aug. 2016.
- [28] S. Jangid and S. Sharma, “High PSNR based video steganography by MLC(multi-level clustering) algorithm,” in *Proc. ICICCS*, Madurai, India, Jun. 2017, pp. 589–594.
- [29] D. M. Firmansyah and T. Ahmad, “An improved neighbouring similarity method for video steganography,” in *Proc. CITSM*, Bandung, Indonesia, Apr. 2016, pp. 1–5.



KE NIU was born in Zhejiang, China, in 1981. He received the M.S. and Ph.D. degrees in cryptography from the Engineering University of PAP, Xi'an, China, in 2007 and 2018, respectively, where he is currently an Associate Professor with the Key Laboratory of Network and Information Security. His research interests include information hiding and multimedia security.



JUN LI was born in Hunan, China, in 1987. He received the B.S. degree in information research and security from the Engineering University of Chinese People Armed Police Force (PAP), Xi'an, China, in 2009, and the M.S. degree in cryptography from the Engineering University of PAP, Xi'an, in 2012, where he has been a Lecturer with the Key Laboratory of Network and Information Security. His research interests include information hiding and image processing.



XIAOYUAN YANG was born in Hunan, China, in 1959. He received the B.S. degree in applied mathematics and the M.S. degree in cryptography and encoding theory from Xidian University, Xi'an, China, in 1982 and 1991, respectively. He has been a Professor with the Key Laboratory of Network and Information Security, PAP. His research interests include cryptography and trusted computation.



SHUO ZHANG was born in Shandong, China, in 1988. She received the B.S. degree in CDMA from the Xi'an Communications Institute. She is currently pursuing the M.S. degree in military communication with the National University of Defense Technology. Her research interests include information hiding and communication command.



BO WANG received the B.S. degree in electronic and information engineering and the M.S. and Ph.D. degrees in signal and information processing from the Dalian University of Technology, China, in 2003, 2005, and 2010, respectively, where he was a Postdoctoral Research Associate with the Faculty of Management and Economics, from 2010 to 2012. He is currently a Visiting Scholar with The State University of New York at Buffalo. He has published over 50 journal and conference papers in digital multimedia forensics. His current research interests include the areas of multimedia processing and security, such as digital image processing and forensics. He is a member of the ACM. He is also the Session Chair of the Mlicom 2017. He currently serves as a Reviewer for *Signal Processing: Image Communication* and the *Journal of Electronic Imaging*.

• • •