Intra Prediction Mode-based Video Steganography by Minimizing the Distortion on the Texture

Qiankai Nie [a], Xuba Xu [b]

College of Information Science and Technology, Jinan University, Guangzhou 510632, China

[a]qiankai.nie@gmail.com, [b]xuxuba36@hotmail.com

Abstract

In this paper, an effective video steganography method based on intra prediction texture is proposed. Messages are embedded during intra-prediction of video coding without causing large embedding effects. According to the statistical model of the intra-prediction mode, we find that when modifying the IPM, the inherent spatial correlation of the prediction mode is destroyed. It motivates us to use texture features as a key factor in constructing distortion functions. In addition, the mapping rules based on edge strength can effectively improve the embedding efficiency. We use the syndrome-trellis codes (STC) as the actual embedding implementation. The experimental results show that compared with the steganography method based on IPM, the proposed steganography scheme has higher undetectability. It is also superior to these schemes on video quality.

Keywords

Steganography, H.264/AVC, Video, Intra Prediction Mode, Mapping Rule.

1. Introduction

With the popularization of the Internet, the degree of informatization has been greatly improved, and the frequent occurrence of security issues has caused people's attention. For video steganography, steganography based on H.264/AVC is the focus of cybersecurity research. Scholars from all over the world have done a lot of work and achieved many results. Video steganography has the characteristics of strong concealment, high embedded capacity, etc. It has been widely used in military espionage activities. This paper will think about new ideas, focus on the research based on the minimization of the distortion framework model, and propose a scheme with higher security and coding efficiency.

Xu et al. [1] proposed a hidden scheme based on the right amount of motion and used the amount of exercise that exceeds the preset threshold as the candidate carrier to be embedded, and used the LSB replacement technology [2] to replace the exercise with the larger weight to achieve the purpose of embedding. In 2014, Su Yuting et al. [3] proposed the embedding of information by using the rhomb code to modulate the motion vector. This scheme can improve the embedded capacity and the code rate has little effect.

The Framework of distortion minimization [4] puts the design of the steganographic scheme into two parts. The first part is to construct a distortion function that can meet and reflect the characteristics of the carrier, and can reasonably represent the distortion effect caused by the embedded secret information. This is a quantification process. The second part is to design the actual data embedded coding scheme. In combination with the distortion function, the embedded distortion can be minimized in the actual embedding process. Many data embedding coding schemes have been proposed today, among which the STC（Syndrome-Trellis Codes） coding schemes proposed by Filler et al. [4] have been widely used because of their high efficiency and flexibility.Cao et al. [5] defined the perturbation function by considering the local optimality of the motion vector (MV). Wang et al. [6] combined the motion features of video content, the local optimality of MVs, and the statistical distribution to define an efficient perturbation function.

In this paper, we will be based on the texture edge of the feature cost function. Combining the mapping rules based on edge strength, the embedded efficiency is further improved. Finally, using STC as the actual embedding method minimizes the distortion caused by embedding.

2. Distortion minimization for steganography

Minimizing overall embedding distortion is an acceptable way to improve steganographic security. Unlike pixel manipulation in image steganography, the steganalysis method based on prediction mode performs embedding during intra prediction processing. The specific distortion minimization framework is as follows.

In the encoding process, the carrier sequence is represented by $\mathbf{x} = (x_1, \ldots, x_n)$, and x_i is an integer, such as a mode value within the block. In addition, a non-negative additive distortion $\rho(x_i, x_i')$ is defined in advance for each carrier element, where x_i' represents the i-th element in the hidden vector $\mathbf{x}' = (x_1', \ldots, x_n')$. Assuming that the $\alpha \cdot n$ bits secret message e is to be embedded with n payload α, the STC acts as the kernel of the framework to perform the embedding operation:

$$\bar{\mathbf{x}} = Emb(\mathbf{x}, \mathbf{e}) = \arg \min_{\mathbf{x}' \in \phi(\mathbf{e})} D(\mathbf{x}, \mathbf{x}') \text{, and } \mathbf{Hx}^T = \mathbf{e} \qquad (1)$$

where $\bar{\mathbf{x}}$ is the stego embeded with secret message, $\phi(\mathbf{e})$ is the coset corresponding to syndrome **e** and the overall distortion $D = \sum_{i=1}^{n} \rho_i$. At the extractor, the operation is simply computing $\mathbf{H}\bar{\mathbf{x}}$.

3. Proposed Distortion Funtion

3.1 Mode Predition

Compression technology is used to efficiently compress the original video into smaller compressed video, which facilitates transmission on the network and storage on hardware. Compression techniques mainly use the redundancy characteristics of video frames and the high degree of similarity between adjacent intra macroblocks. In this paper, 4x4 intra-block (I4B) is used as the carrier for data embedding, and the blocks within the 16x16 block are not used for encoding because the 16x16 block is used to describe the smooth region and is not suitable for embedding modifications. As the current popular video coding standard, H.264 includes 9 intra prediction modes based on prediction directions, as shown in FIG. 1

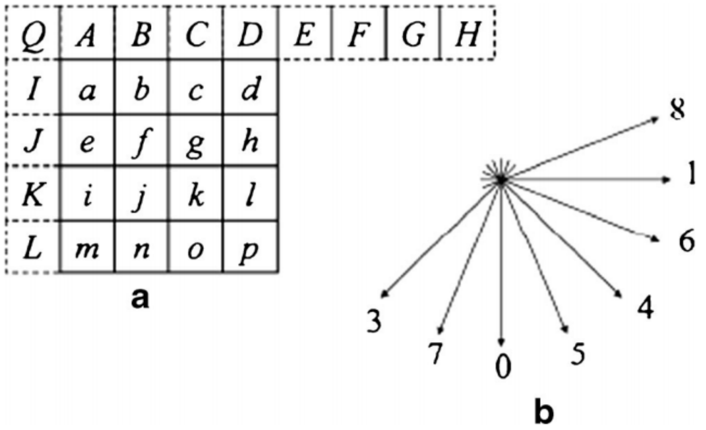
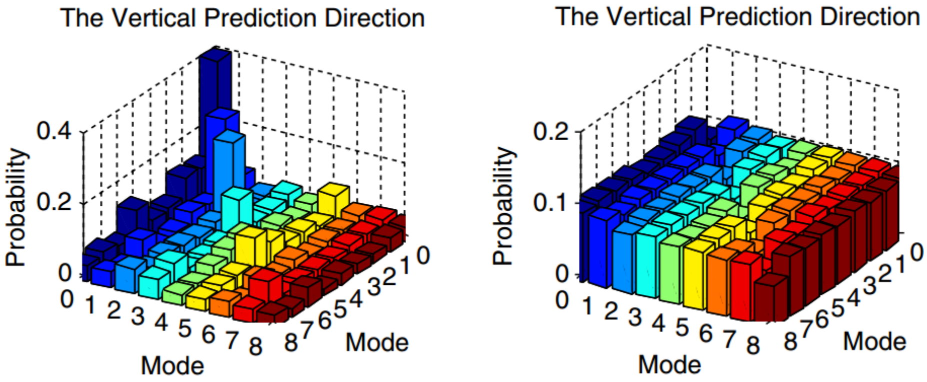


Fig. 1 a I4-MB prediction coding is conducted for samples a–p of a block using samples A–Q; b Eight "prediction directions" for I4-MB prediction

As a commonly used H.264 encoder, X264 uses the numerical values 0 to 8 to represent 9 prediction modes. As shown in FIG. 1 , in the nine prediction modes, the gray part represents a 4×4 prediction block. In the intra prediction process, the 4×4 prediction block performs prediction according to the left and upper neighboring luminance cells. The prediction mode determines the direction of the prediction block. For example, when using the prediction mode 0 (Vertical), the luminance unit of each column in the 4x4 prediction block is equal to the adjacent side luminance value in the vertical direction of the column. Among the more specific prediction modes 2 (DC), the value of the prediction block is equal to the neighboring edge (A+B+C+D+I+J+K+L+M)/7.

3.2 Texture Edge Features

In the intra-frame prediction-based steganography scheme, when the 4x4 prediction mode is modified, the prediction mode correlation in the entire frame is affected. Therefore, [7] analyzes information hiding and changes in intra-frame coding characteristics and finds that different 4×4 There is an inherent spatial correlation between the prediction modes of luminance blocks. Based on several different positional relationships between adjacent 4×4 prediction blocks in the spatial domain, a statistical model related to the prediction model is designed to quantitatively extract these related features.As shown in Fig. 2, (a) is a probability matrix diagram of the original prediction mode in the vertical direction, and (b) is a probability matrix diagram with information embedded in the same direction.



(a) (b)

Fig. 2 Vertical state transition probability matrix

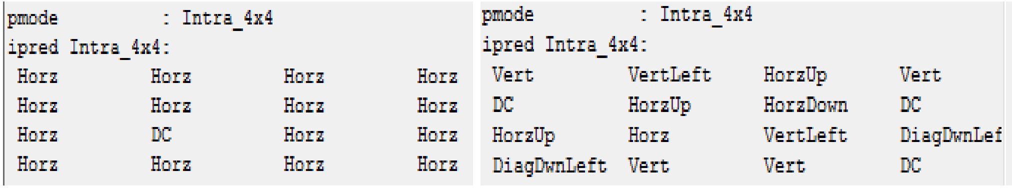
From the above, it can be seen that a classifier of a support vector machine can be used to construct an efficient steganalysis scheme. In image steganography, a high-security steganography scheme called HUGO (Highly Undetectable Steganography) [8] was proposed. The success of this solution lies in the fact that it is difficult to detect the texture and the edge region embedded information in the picture, which can effectively resist the SRM (Spatial RichModels) steganography analysis scheme based on the statistical model. Similarly, in the video steganography based on prediction mode, this chapter proposes a steganography scheme based on the texture edge feature for the first time, and utilizes the chaotic feature of the prediction mode in the texture edge to embed information.

(a) (b)

Fig. 3 (a) is the original video, and (b) is the number of prediction modes in each block.

Fig. 3(a) shows an H.264 video sequence named "Flowers", and Fig. 3(b) shows the number of prediction mode types for each 16x16 macroblock. NMT), while a 16x16 macroblock represents a local area. It can be seen that the “sky” in the image frame belongs to a smooth region, and the number of prediction mode types in these regions is small, which is grayish black in Fig. 3(b);Flowers and house belong to a complex texture region.The number of forecasting models is large and the display is gray.

Fig. 4 (a), (b) is a specific prediction of each 4x4 sub-block in a 16x16 macroblock, which are in a smooth region and a complex texture region, respectively. It can be seen that in the smooth region macroblock, the prediction modes of the respective 4x4 blocks are substantially the same, mainly because the selection of the prediction mode depends on the content of the image, and the texture of the smooth region is relatively single, and there is no need for multiple prediction modes for encoding. However, in the complex texture edge, the texture of the local area is not uniform, resulting in diversification of the prediction mode. Locally smooth regions are not suitable for embedding because the number of macroblock bits in these regions is small, and the image content is related to the Human Visual System (HVS), and humans are more sensitive to changes in the homogenous region.



```
pmode          : Intra_4x4
ipred Intra_4x4:
 Horz         Horz         Horz         Horz
 Horz         Horz         Horz         Horz
 Horz         DC           Horz         Horz
 Horz         Horz         Horz         Horz
```

```
pmode          : Intra_4x4
ipred Intra_4x4:
 Vert         VertLeft     HorzUp       Vert
 DC           HorzUp       HorzDown     DC
 HorzUp       Horz         VertLeft     DiagDwnLef
 DiagDwnLeft  Vert         Vert         DC
```

（a） （b）

Fig. 4 Specific prediction map for each 4x4 sub-block

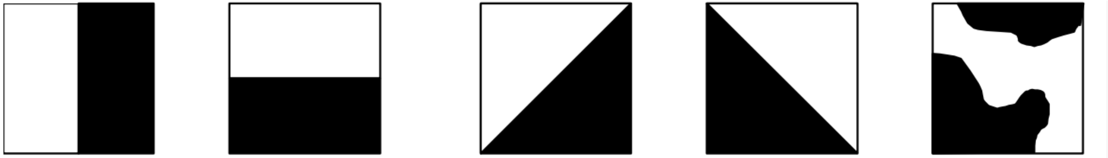
Through the analysis of the appeal, the number of prediction mode types (NMT) can be used as a key factor in constructing the cost function. During the intra-frame prediction period, replacing the optimal prediction mode **md** with the candidate mode **md'** inevitably disturbs the statistical distribution and the model. NMT-based embedding perturbation in the edge region can be defined as:

$$\mathrm{ED}_{i,t}=\frac{1}{\mathrm{NMT}_{i,t}\left(\mathrm{md}_{i,t}\right)}+\left|\frac{\mathrm{NMT}_{i,t}\left(\mathrm{md}_{i,t}^{'}\right)}{\mathrm{NMT}_{i,t}\left(\mathrm{md}_{i,t}\right)}-1\right| \quad (2)$$

Where $\mathrm{NMT}_{i,t}(\mathrm{md}_{i,t})$ represents the number of prediction mode types when the *i*-th prediction block is $\mathrm{md}_{i,t}$ in the macroblock in which the *i*-th prediction block is located in the *t*-th keyframe.

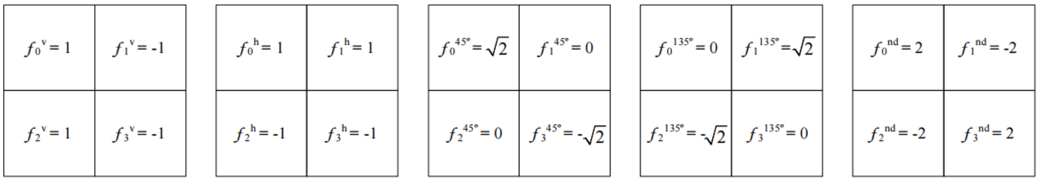
3.3 Mapping Rule

For the first time, [9] proposed an intra-prediction algorithm based on Dominant Edge Strength (DES). The algorithm divides the edges into five categories: vertical edges, horizontal edges, 45-degree oblique edges, 135-degree oblique edges, and non-directional edges, as shown in Figure 4-5. The detection of the edge strength of the intra 4x4 block is mainly based on the filtering operation. The 4x4 prediction block is first divided into 4 2x2 sub-blocks. Then, the average pixel value of each sub-block is successively filtered by five directions, as shown in Fig.5.



a) vertical edge b) horizontal edge c) 45 degree edge d) 135 degree edge e)non-directional edge

Fig.5 Five kinds of edge strength



a) vertical b) horizontal c) 45° diagonal d) 135° diagonal e) non-direction

Fig.6 Filter operation

Finally, the strength of the edge can be given by these five expressions:

$$S^{v} = |\sum a_n \times \int_n^{v}| \quad (3)$$

$$S^{h} = |\sum a_n \times \int_n^{h}| \quad (4)$$

$$S^{45^\circ} = |\sum a_n \times \int_n^{45^\circ}| \quad (5)$$

$$S^{135^\circ} = |\sum a_n \times \int_n^{135^\circ}| \quad (6)$$

$$S^{nd} = |\sum a_n \times \int_n^{nd}| \quad (7)$$

The edge strength and direction of a 4x4 prediction block depends on the maximum of the five formulas. These five edge strengths make up the general direction of the prediction mode, indicating prediction modes 0, 1, 3, 4 and 2, respectively.In the mapping rules, the differences in prediction directions are grouped into the same group, mainly considering that the difference in the prediction direction will result in a larger residual error, and the mutual replacement between the same group will not be replaced, thereby avoiding the replacement of the prediction with large residual error. mode. Therefore, the prediction mode can be divided into 3 groups, in which modes 0, 1, 3, 4 and 2 are in the same group as mode group 1; modes 6 and 7 which are different in the prediction direction are in the same group as the mode. Group 0; remaining prediction modes 5 and 8 are a group, which is mode group 2. The grouping situation is as follows:

$$gp_0 = \{md_6, md_7\}$$

$$gp_1 = \{md_0, md_1, md_2, md_3, md_4\}$$

$$gp_2 = \{md_5, md_8\}$$

3.4 The Definition of Distortion Function

The cost function is based on the core of minimizing the embedded distortion scheme. The quality of the cost function directly affects the security of the steganographic scheme. Therefore, steganography scheme proposed in this paper constructs the cost function based on edge features.

When the optimal prediction mode $md_{i,t}$ is replaced by $md'_{i,t}$, the complete cost function of this scheme can be given by a mathematical expression:

$$\phi_{i,t}(md_{i,t}, md'_{i,t}) = ED_{i,t} \quad (8)$$

For I-frames, the optimal prediction mode obtained using intra-prediction encoding of the standard H.264 can be expressed as:$md_{i,t}, md_{i,t}, \ldots, md_{i,t}$, $i \in \{0, 1, \ldots, N_t\}$, where N_t represents the number of I4B in the t-th keyframe. Further, the optimal group $\mathcal{G}_t = \{g(i,t)|md_{i,t} \in gp_{g(i)}, g(i,t) \in \{0,1,2\},$ $i = 1,\cdots,N_t\}$ is obtained according to the mapping rule. The cost $\rho_{i,t}^{+}$, $\rho_{i,t}^{-}$of each I4B block is calculated using the optimal set $\mathcal{G}_t$, which is the price used by STC plus "1" and minus "1", respectively. According to the principle of minimizing distortion embedding $\rho_{i,t}^{+}$ and $\rho_{i,t}^{-}$ constructing the cost function together, the mathematical expression is:

$$\rho_{i,t}(g(i,t), g'(i,t)) = \{\rho_{i,t}^{+}, \rho_{i,t}^{-}\} \quad (9)$$

where

$$\rho_{i,t}^{+} = \phi_{i,t}(md_{i,t}, md_{i,t}^{+})$$

$$\rho_{i,t}^{-} = \phi_{i,t}\left(md_{i,t}, md_{i,t}^{-}\right)$$

The selection of the prediction mode conforms to the mapping rule. At the same time, the addition of the "1" and minus "1" operations of the mode group is based on the STC principle and is actually operated:

$$md_{i,t}^{+} = \arg\min\ \phi\left(md_{i,t}, \Delta_{i,t}^{+}\right) \tag{10}$$

$$\Delta_{i,t}^{+} \in gp_{g'(i,t)}, g'(i,t) = g(i,t) + 1$$

$$md_{i,t}^{-} = \arg\min\ \phi\left(md_{i,t}, \Delta_{i,t}^{-}\right) \tag{11}$$

$$\Delta_{i,t}^{-} \in gp_{g'(i,t)}, g'(i,t) = g(i,t) - 1$$

Therefore, the total distortion D_t of all I4B in the tth keyframe can be minimized by steganographic embedding. D_t can be calculated as the distortion function by the following formula:

$$D_t(\mathcal{G}_t, \mathcal{G}_t') = \sum_{i=1}^{N_t} \rho_{i,t}\left(g(i,t), g'(i,t)\right) \tag{12}$$

Where N_t represents the number of I4B in the tth keyframe, $\mathcal{G}_t$ represents the optimal group set, and $\mathcal{G}_t'$ represents the candidate group matrix. The variable ρ is an additive distortion function and can fully express the modification effect of the tth keyframe i pattern groups.

4. Evaluation

4.1 Experiment Setup

It compares the performance between the proposed method and the yang's[10] and Bouchama's[11] method under the same embedding rate. The experiment considered four kinds of embedding rates (ERs) ranging from 0.125 to 0.5, representing the Bits Per Prediction Mode (bppm) for each prediction mode, and setting the two bit rates (BR) of the compressed sequence to 0.5 Mbit. /s and 1Mbit/s, where the embedding rate (ER) represents the average embedded secret bit length of I4B, and its calculation formula is:

$$\alpha = \frac{L}{N_t} \tag{13}$$

Where L represents the length of the embedded binary message in the t-th frame and N_t is the total number of I4B in the tth keyframe.

The experimental steps are as follows:

(1) In a single compression, 1500 video I-frames are extracted from 250 sub-sequences. By adjusting the initial frame parameters and the maximum interval parameters, the subsequences are compressed. Performing 5 compressions in sequence yields 7500 I-frames, of which 5000 I-frames are randomly added to the training frame, and the remaining 2500 are added to the test frame.

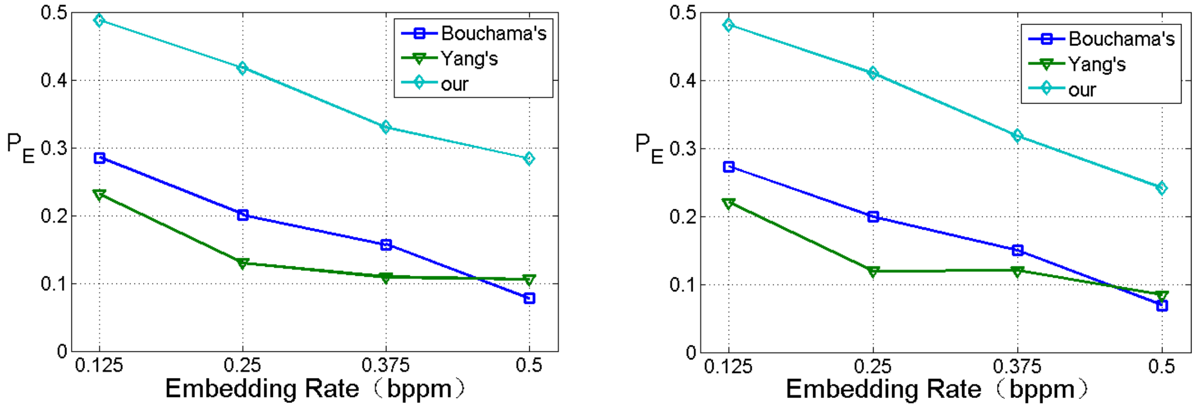
(2) According to the given four embedding rates {0.125, 0.25, 0.375, 0.5}, an 01-bit string is randomly generated as emulation information for embedding operation, in which STC encoding uses triple information embedding. In training frames and test frames, four embedding rates are added to 5000 and 2500 portable I frames, respectively.

(3) According to different embedding rates, the IPMC and IPMSC analysis algorithms are used to perform feature extraction on 5000+2500=7500 initial I-frames and portable I-frames respectively.

IPMC extracts 5000 sets of 13-dimensional feature training sets, and IPMSC extracts 5000 sets of 9081s. Dimension feature training set.

(4) Each 5000 sets of IPMC features and 50,000 sets of IPMSC features are sent to the LIBSVM classifier for training. After the test model is generated, the feature test set is sent to the classifier to obtain four different embedding rate corresponding performance indicators.

4.2 Steganalysis Result



(a) (b)

Fig.7 Detection error rate P_E for different embedding rates when (a) 0.5 Mbit / s bit rate and (b) 1.0 Mbit / s bit rate in IPMC detection

Fig.7 and Table 1 show the test results of the IPMSC analysis model. We can see that our proposed solution is better than Bouchama's and Yang's in terms of each embedding rate. With the increase of the embedding rate, the security of each scheme has gradually declined. Because the mapping rules of the proposed scheme are based on the spatial prediction direction, and the cost function is constructed based on the edge features, it can effectively resist steganalysis based on spatial correlation. Therefore, the proposed scheme always maintains its advantages. It can be concluded that the cost function combined with edge features is effective and can improve the steganographic algorithm's ability to combat steganalysis.

Table 1 Performance against IPMSC features

ER (bppm)		0.125		0.250		0.375		0.500	
BR (mbps)		0.5	1.0	0.5	1.0	0.5	1.0	0.5	1.0
P_E (%)	Bouchama's	28.5	24.1	20.1	19.9	15.7	15.0	7.8	6.9
	Yang's	23.1	22.0	13.0	11.9	10.9	12.0	10.6	8.4
	方案2	48.8	48.1	41.7	41.0	33.0	31.8	28.4	26.8

5. Conclusion

By analyzing the spatial correlation of prediction modes, a steganography scheme based on edge features is proposed to reduce the damage to spatial correlation. This scheme embeds the embedding as much as possible in the complex edge region of the image and can preserve the structure of the smooth region well. By introducing a mapping rule based on edge strength, the efficiency of embedding is further improved, and while ensuring security and visual quality, the increase in bit rate is also effectively avoided.

The edge features of the proposed scheme are performed in the original domain. The next research work mainly considers the introduction of filters and the feature extraction in the DCT domain.

References

[1] Xu C, Ping X, Zhang T. Steganography in Compressed Video Stream[C]// International Conference on Innovative Computing, Information and Control. IEEE, 2006:269-272.

[2] Jung K H, Yoo K Y. Steganographic method based on interpolation and LSB substitution of digital images[J]. Multimedia Tools & Applications, 2014, 74(6):1-13.

[3] Xu C, Ping X, Zhang T. Steganography in Compressed Video Stream[C]// International Conference on Innovative Computing, Information and Control. IEEE, 2006:269-272.

[4] Filler T, Judas J, Fridrich J. Minimizing Additive Distortion in Steganography Using Syndrome-Trellis Codes[J]. IEEE Transactions on Information Forensics & Security, 2011, 6(3):920-935.

[5] Cao Y, Zhang H, Zhao X, et al. Video Steganography Based on Optimized Motion Estimation Perturbation[C]// ACM Workshop on Information Hiding and Multimedia Security. ACM, 2015:25-31.

[6] Wang P, Zhang H, Cao Y, et al. A Novel Embedding Distortion for Motion Vector-Based Steganography Considering Motion Characteristic, Local Optimality and Statistical Distribution[C]// ACM Workshop on Information Hiding and Multimedia Security. ACM, 2016:127-137.

[7] Li S, Deng H, Tian H, et al. Steganalysis of prediction mode modulated data-hiding algorithms in H.264/AVC video stream[J]. annals of telecommunications - annales des télécommunications, 2014, 69(7-8):461-473.

[8] Pevný T, Filler T, Bas P. Using High-Dimensional Image Models to Perform Highly Undetectable Steganography[J]. Lecture Notes in Computer Science, 2010, 6387:161-177.

[9] Wang J F, Wang J C, Chen J T, et al. A novel fast algorithm for intra mode decision in H.264/AVC encoders[C]// IEEE International Symposium on Circuits and Systems, 2006. ISCAS 2006. Proceedings. IEEE, 2006:4 pp.

[10] Yang G, Li J, He Y, et al. An information hiding algorithm based on intra-prediction modes and matrix coding for H.264/AVC video stream[J]. AEUE - International Journal of Electronics and Communications, 2011, 65(4):331-337.

[11] Bouchama S, Hamami L, Aliane H. H.264/AVC Data Hiding Based on Intra Prediction Modes for Real-time Applications[J]. Lecture Notes in Engineering & Computer Science, 2012, 2200(1).