A HEVC Video Steganalysis Algorithm Based on PU Partition Modes

Zhonghao Li[1], Laijin Meng[1], Shutong Xu[1], Zhaohong Li[1, 2, *], Yunqing Shi[3] and Yuanchang Liang[1]

Abstract: Steganalysis is a technique used for detecting the existence of secret information embedded into cover media such as images and videos. Currently, with the higher speed of the Internet, videos have become a kind of main methods for transferring information. The latest video coding standard High Efficiency Video Coding (HEVC) shows better coding performance compared with the H.264/AVC standard published in the previous time. Therefore, since the HEVC was published, HEVC videos have been widely used as carriers of hidden information.

In this paper, a steganalysis algorithm is proposed to detect the latest HEVC video steganography method which is based on the modification of Prediction Units (PU) partition modes. To detect the embedded data, All the PU partition modes are extracted from P pictures, and the probability of each PU partition mode in cover videos and stego videos is adopted as the classification feature. Furthermore, feature optimization is applied, that the 25-dimensional steganalysis feature has been reduced to the 3-dimensional feature. Then the Support Vector Machine (SVM) is used to identify stego videos. It is demonstrated in experimental results that the proposed steganalysis algorithm can effectively detect the stego videos, and much higher classification accuracy has been achieved compared with state-of-the-art work.

Keywords: Video steganalysis, PU partition modes, data hiding, HEVC videos.

1 Introduction

With the development of digitization, the digital video has gradually become the mainstream video. However, it brings security problems including copyright protection, identity information authentication and so on. Therefore, the information hiding technology has been developed. Information hiding is a method of hiding confidential information into an innocent looking carrier without evoking any suspicion. The message carrying and clean objects are called stego and cover respectively. Among existent carriers such as music, pictures, documents and videos, videos have been more and more

---

[1] School of Electronic and Information Engineering, Beijing JiaoTong University, Beijing, 100044, China.

[2] Guangdong Provincial Key Laboratory of Information Security Technology, Guangzhou, 510000, China.

[3] Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, 07102, USA.

[*] Corresponding Author: Zhaohong Li. Email: zhhli2@bjtu.edu.cn.

frequently used because a large amount of information is contained in videos which makes it uneasy to detect stego videos after information hiding.

As a new generation of coding standard, High Efficiency Video Coding (HEVC) videos have drawn much more attention since it was put forward. Compared with the previous generation of coding standard H.264/AVC, HEVC is more complex and has been widely used in not only high-definition (HD) digital videos but also ultra high-definition (Ultra HD) digital videos. The common method of modifying DCT/DST coefficients [Lin, Chung, Chang et al. (2013)] was applied to HEVC videos by Chang et al. [Chang, Chung, Chen et al. (2014)] firstly, which had a high embedding capacity at low bit rate. Xu et al. [Xu, Wang and Wang (2012)] took measures to modify the intra prediction modes, which could ensure a certain embedding capacity as well as improve the visual quality of the video. In the aspect of modifying inter prediction modes, Li et al. [Li, Wang, Liu et al. (2016)] proposed a new information hiding algorithm based on motion vector space encoding. Xie et al. [Xie, Yang, Li et al. (2018)] modified the PU partition modes of HEVC to ensure high visual quality. What's more, the embedding capacity of Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)] was greatly improved at the same time.

In order to avoid embedded information obtained by other people, the security of information hiding is very significant. Steganalysis is a technique for detecting stego media, which has become a hot topic of information security in recent years. However, most of the steganalysis algorithms only focus on images, hence it is very urgent and important to develop steganalysis method for videos. Among the existing steganalysis techniques for videos, Jainsky et al. [Jainsky, Kundur and Halversion (2007)] developed an algorithm for digital video steganalysis, named MoViSteg for Motion-based Video Steganalysis that exploited the temporal correlation among individual image frames to enhance steganalysis performance. Wu et al. [Wu, Liu, Huang et al. (2014)] took the joint distribution of motion vector (MV) differences as features. Kong et al. [Kong, Wang and Wang (2014)] constructed the transition probability matrix of intra prediction mode for original videos and recompressed videos and used it as a steganalysis feature, which had a high detection rate of different types of carrier videos with low embedding rate. Nie et al. [Nie, Xu, Feng et al. (2018)] presented a novel intra prediction mode-based video steganography by minimizing the embedding distortion defined according to SAD. Sheng et al. [Sheng, Wang and Huang (2017)] proposed a steganalysis algorithm based on the change of PU partition modes of cover videos and stego videos. The prediction modes vector of I pictures was extracted, and the transition probability matrix of the prediction modes was calculated. Then, the new vector was composed of raster scan sequence as the classification feature.

Inspired by the steganalysis method of Sheng et al.'s [Sheng, Wang and Huang (2017)], we noticed that Xie et al.'s information hiding algorithm [Xie, Yang, Li et al. (2018)] took measures to hide information by modifying PU partition modes and different PU partition modes are used to represent different binary information. So, we suppose that the quantity of PU partition modes would be changed after data hiding. Thus, an information steganalysis algorithm based on the probability of each PU partition mode in P pictures (PoPUPM) is proposed in this paper. After feature optimization, the 25-dimensional steganalysis feature is further reduced to be 3-dimensional. And the

detection accuracy of stego videos generated by Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)] is over 96% using the steganalysis algorithm proposed in this paper, which is much higher than that of Sheng et al.'s algorithm [Sheng, Wang and Huang (2017)].

The rest of this paper is organized as follows. Section 2 introduces the structure of PU partition modes in HEVC. Section 3 discusses the steganalysis feature based on PU partition modes of P pictures. Section 4 gives the proposed method in detail. Section 5 gives the experimental results. Finally, a summary is given in Section 6.

2 Basics of PU partition in HEVC

HEVC is the latest video coding standard published by Video Code Expert Group (VCEG) and Moving Pictures Expert Group (MPEG). Compared with H.264/AVC, HEVC has the same coding structure that contains Video Coding Layer (VCL) and Network Abstraction Layer (NAL) [Sheng, Wang and Huang (2017)]. But it is the innovative point of HEVC that the quadtree structure to partition aim images are used for prediction and transform coding.

HEVC divides the video into many groups of pictures (GOPs), and every group includes the same quantity of continuous frames. Based on quadtree algorithm, every picture will be subdivided into a quantity of square code tree unit (CTU) with the same size. CTU can also be partitioned iteratively into smaller code unit (CU). Each CU has its further partitioned into transform unit (TU) and prediction unit (PU).

How a CU is divided into different PU modes is based on the CU prediction mode. As shown in Fig. 1, PU partition modes for intra prediction and inter prediction are different, where N depends on the size of CU. For intra prediction, a coding block (CB) of size $2N \times 2N$ can be split into one or four prediction blocks (PBs). And for another prediction mode, inter prediction, a CB can be split into two PBs symmetrically or asymmetrically. There are totally 25 possible PU partition modes listed in Tab. 1, which are marked as indexes 1-25 respectively.
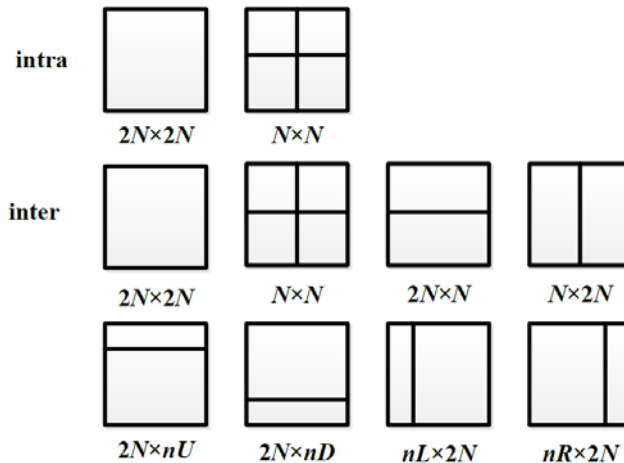

Figure 1: The partition of PU in intra and inter prediction mode

Table 1: Indexes of PU Partition Modes in P-picture

| ID	Size	ID	Size	ID	Size	ID	Size
1	4×4	8	16×12	15	32×24	22	64×48
2	8×8	9	16×4	16	32×8	23	64×16
3	4×8	10	12×16	17	24×32	24	48×64
4	8×4	11	4×16	18	8×32	25	16×64
5	16×16	12	32×32	19	64×64		
6	16×8	13	32×16	20	64×32		
7	8×16	14	16×32	21	32×64		

3 Feature analysis of stego videos

As illustrated in Section 2, for a specific CU of 16×16 or larger size, it has only two kinds of PU partition modes for intra prediction, but there are eight PU partition modes for inter prediction in P-pictures. Consequently, Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)] hides information by changing the PU partition modes for inter prediction in P pictures.

3.1 Xie et al.'s steganography method

The main steps of Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)] are as follows.

At first, the division depth and the optimum PU partition modes of each CU structure are recorded during the HEVC encoding process by default. Afterwards if a CTU includes 16×16 or 32×32 CUs, and the PU partition modes of these CUs are classified to the three groups shown in Fig. 2, then the PU partition modes are supposed to be modified according to the to-be-embedded binary information.
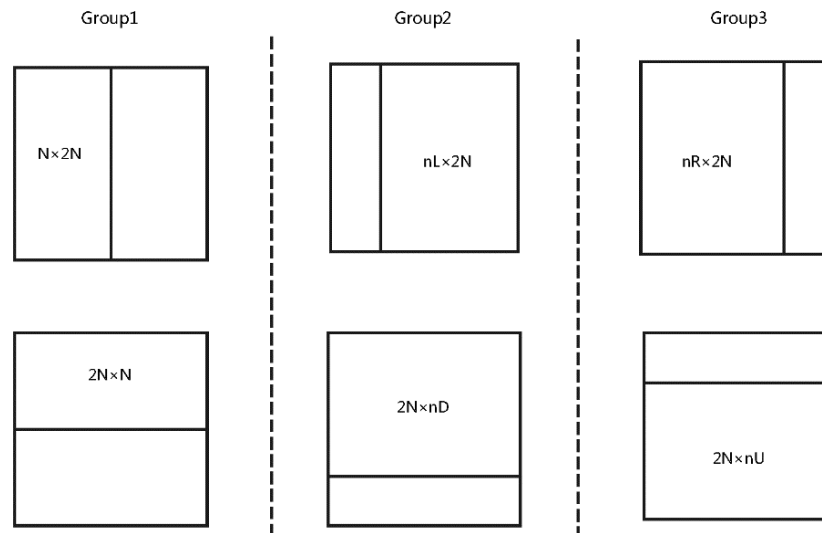
Figure 2: Three groups of different PU Partition Modes for 16×16 and 32×32 CU sizes

An example can be used to make the process more specific. It can be assumed that the Group 1, Group 2 and Group 3 represent binary bits 00, 10 and 11, respectively, and in the HEVC encoding process, the achieved PU partition mode for a 32×32 CU is the horizontally symmetrical type in Group 1. If the to-be embedded bits are 10 or 11, the PU partition mode will be modified to be the horizontal partition mode in Group 2 or Group 3. Otherwise, the PU partition mode will keep itself.

As demonstrated in Fig. 3, Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)] almost has no impact on visual quality of stego videos, and the PSNR decreases no more than 1% even though the bit rate is approximately 15 Mbps.
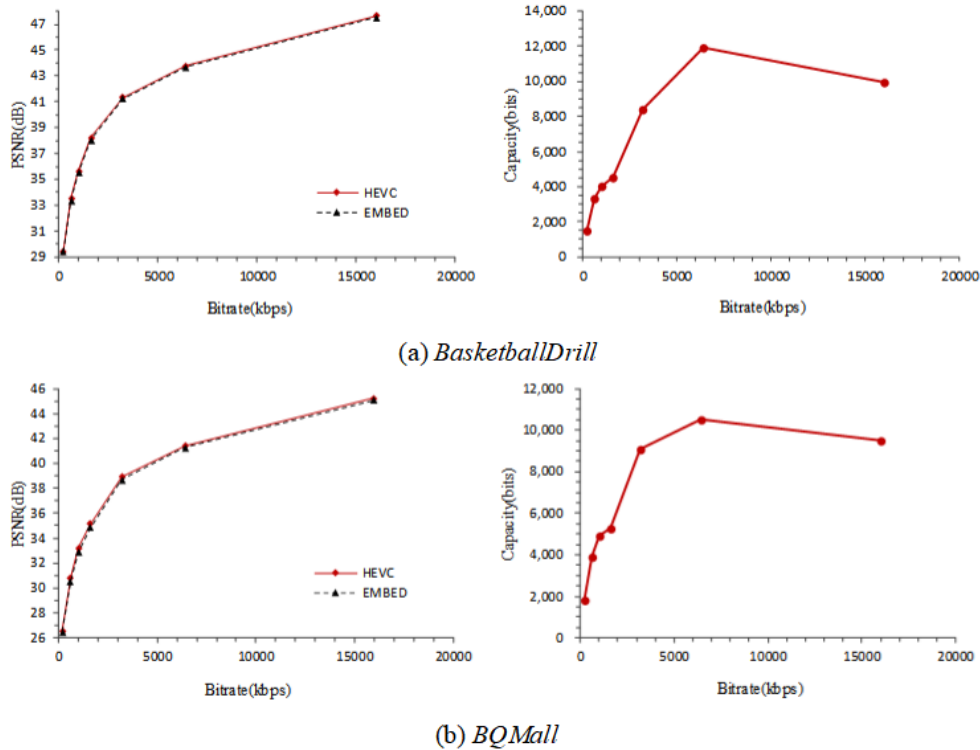
(a) *BasketballDrill*

(b) *BQMall*

Figure 3: PSNR and capacity performance of Xie et al.'s algorithm

Even if Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)] has great advantages on visual quality and bit rate compared to other data hiding algorithms, it is the security issue that data hiding algorithms should also take into consideration. However, the security issue has not been discussed in Xie et al.'s paper [Xie, Yang, Li et al. (2018)].

What's more, since the optimal PU partition modes are modified to be other specified PU partition modes in Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)], the statistical distribution of the PU partition modes is supposed to be changed in stego videos. Hence in the following section, the PU partition modes of stego videos generated by Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)] will be analyzed, and we have conducted a more in-depth study of the security of the algorithm.

3.2 Target feature analysis and selection

To figure out the statistical feature in stego videos, the quantity of 25 different PU modes in P pictures of stego videos and cover videos has been extracted respectively. Based on the data extracted, the rate of change of PU modes in the cover videos before and after embedding information has been calculated from Eq. (1).

$$\text{ROC} = \left| \frac{N_s - N_c}{N_c} \right| \tag{1}$$

Where in Eq. (1), N_c and N_s are the quantity of each PU partition mode in P pictures of a cover video and a stego video respectively. For illustration, the two videos 'Ducks' and 'Basketball Drive' are used as sample sequences. The sample sequence 'Ducks' is in resolution 720P and has 80 frames, and 'Basketball Drive' is in 1080P and has 50 frames.

For experimental setting, HM 16.15 is used to encode sample sequences to cover videos, and Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)] is used to generate stego videos. The GOP size is 4(IPPP).

It can be seen from Fig. 4 that the quantity of each of the 25 PU modes has changed after information hiding. However, the ROC of 18 in 25 PU partition modes are smaller than 20%. It is only ROC of the typical 3 PU partition modes (8×4, 4×8, 8×16) that has decreased by more than 50% in various resolutions (720P, 1080P) as well as various bit rates (4 M, 8 M, 10 M, 12 M, 30 M, 50 M).
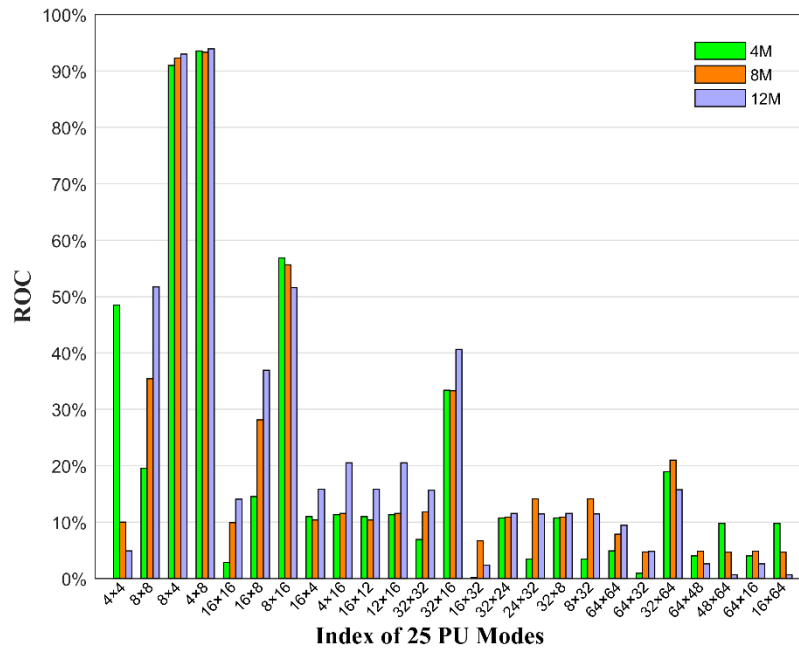
(a) Ducks 720P
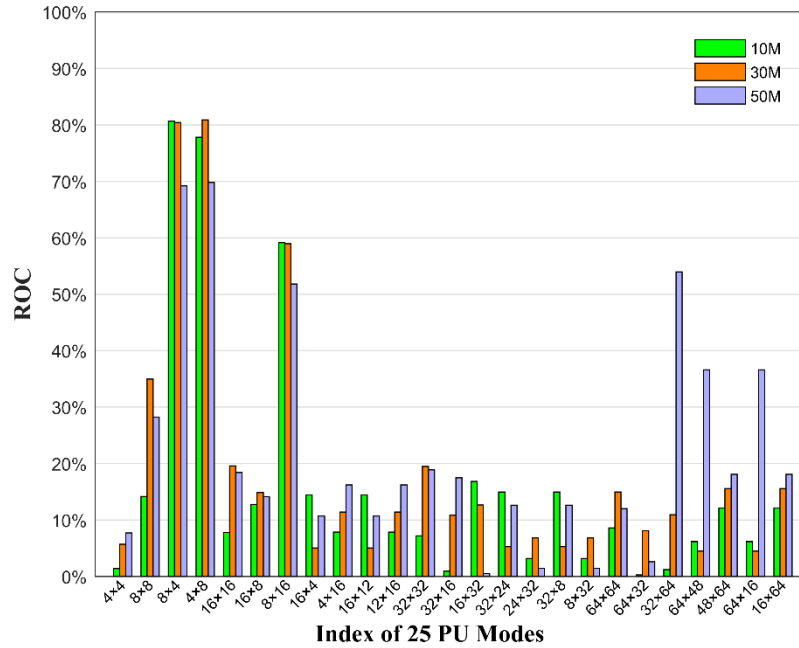
(b) Basketball Drive 1080P

Figure 4: ROC of PU Partition Modes at different video resolutions

To illustrate this statistical phenomenon of PU modes demonstrated above, the reason can be explained according to the HEVC encoding algorithms and architectures.

A CU block can be taken as an example. When a CU is subdivided into multiple PU modes, it is after extending all the possible PU partition modes that the optimal prediction parameters for each PU partition mode are determined. As a result, we typically increase the bit rate required for signaling the selected PU modes, but decrease the resulting rate-distortion cost (RD cost). In summary, different subdivisions into PUs used for inter-prediction are closely related to trade-offs between distortion and bit rate.

In the HEVC encoding process by default, the CUs are subdivided into smaller PUs such as 8×4 and 4×8, which can minimize the RD cost as well as make full use of the bit rate. However, in Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)], the encoding parameter bitrate is fixed, and no other encoding parameters and configurations have been modified, except for the PU partition modes for CUs in size of 16×16 and 32×32. In other words, the optimal PU partition modes have been modified to be other modes in Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)], which results in more prediction redundancy. Hence the fixed bit rate required for transmitting the prediction redundancy is not enough. To compensate for the contradiction between distortion and bit rate generated from Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)], the quantity of PU modes in small sizes such as 8×4 and 4×8 in P pictures will be converted to be larger PU modes. It is only in this way that the prediction redundancy can be reduced to an acceptable level which the fixed bit rate can transmit. Therefore, we suppose that the distribution of 25 PU partition modes in P pictures can be used as a classification feature to detect stego videos from cover videos.

What's more, it is demonstrated in Fig. 4 that in various resolutions (720P, 1080P) and various bit rates (4 M, 8 M, 10 M, 12 M, 30 M, 50 M), it is just the ROC of PU modes in size of 8×4 and 4×8 that are larger than 75%, and ROC of PU modes in size of 8×16 that is larger than 50%. To transform the feature data in the high-dimensional (25 dimensions) space to a space of fewer dimensions, how the distribution of small PU partition modes in size of 8×4, 4×8 and 8×16 are selected as the target feature will be introduced in the following section.

4 Proposed method

In general, embedding information into a cover video will cause some modification of components in it. Based on the analysis in Section 3, it has been found that PU partition modes in P pictures have been changed in stego videos generated by Xie et al.'s stegonagraphy algorithm [Xie, Yang, Li et al. (2018)]. Hence, the proposed method adopts statistical distribution of PU partition modes in P pictures as the classification feature. Fig. 5 shows the detailed diagram of the proposed method.
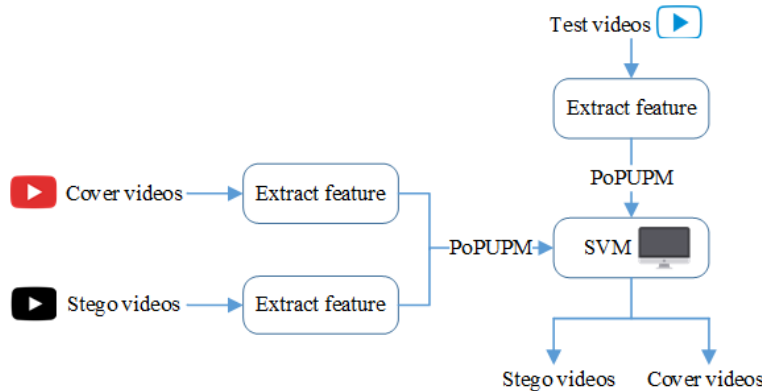

Figure 5: Detailed diagram of the proposed method

Firstly, extract all the PU partition modes from P-pictures of cover videos and stego videos. Secondly, the probability statistical distribution is chosen as the proposed feature according to Eq. (2).

$$P_i = \frac{N_i}{\sum_{i=1}^{25} N_i} \tag{2}$$

Where i ranges from 1 to 25 as there are 25 modes of PU partition modes in P pictures. N_i represents the total quantity of the ith PU partition mode in a video sequence, afterwards we can get the 25-dimensional classification feature P_i which is the probability of PU partition modes (PoPUPM-25D for abbreviation). Furthermore, it is indicated in Section 3 that the outstanding PU partition modes (8×4, 4×8, 8×16) can be adopted as a 3-dimensional feature (PoPUPM-3D).

After feature extraction, the Support Vector Machine (SVM) is trained using the PoPUPM data. Finally, the trained SVM classifier can be applied to detect stego videos.

5 Experimental results

For sample sequences, following videos shown in Tab. 2 are used as test sequences. Videos in resolution of 720P are divided into several parts, and each part has 80 frames. For 1080P videos, each sequence is divided into 10 parts, and each part has 50 frames. In total, we get 33 videos in 720P and 30 videos in 1080P.

For experimental setting, HM 16.15 is used to encode sample sequences to cover videos, and Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)] is used to generate stego videos. The configuration of encoding is shown in Tab. 3.

Table 2: Sample sequences in experiments

Resolution	Name of sequences
1280×720	Park joy, Ducks, Vidyo1, Johnny, Four people
1920×1080	Ducks take off, Basketball Drive, Cactus,

Table 3: The main configuration parameters of HM

Parameters	Configuration	
	720P	1080P
GOP size	4(IPPP)	4(IPPP)
Frames to be encoded	80	50
Bitrate	4 M/8 M/12 M	10 M/30 M/50 M
The others use default setting		

In many fields, SVM is an excellent and convenient tool to classify a large amount of information. Here, all the features extracted from cover videos and stego videos are sent to SVM classifier. For each experimental procedure, random 5/6 of all the sample sequences are selected as training sequences, while the rest are testing sequences. As for SVM classifier, polynomial is adopted as the kernel function, validation function is used to calculate the optimal Gamma and Cost for the kernel. This experimental procedure will be repeated for 20 times, and the average accuracy is adopted as classification accuracy.

Tabs. 4(a)-(b) are the classification accuracy of PoPUPM-25D, and it shows that stego videos can be identified with an extremely high accuracy approaching 100% for both 720P and 1080P videos. Tabs. 5(a)-(b) are the classification accuracy of PoPUPM-3D. Even though it decreases a little bit compared to that using PoPUPM-25D, it remains a high level which is above 96%. It is demonstrated in Tabs. 4-5 that in various resolutions and bitrates, the selected feature is always quite effective regardless of PoPUPM-25D or PoPUPM-3D.

Table 4: The classification accuracy with PoPUPM-25D

(a) 720P video set		(b) 1080P video set	
Bitrate	Classification accuracy	Bitrate	Classification accuracy
4M	99.1%	10M	100.0%
8M	100.0%	30M	100.0%
12M	100.0%	50M	100.0%

Table 5: The classification accuracy with PoPUPM-3D

(a) 720P video set		(b) 1080P video set	
Bitrate	Classification accuracy	Bitrate	Classification accuracy
4M	96.6%	10M	99.8%
8M	99.5%	30M	100%
12M	99.7%	50M	100%

Moreover, we built three types of mix-bitrate video groups to simulate conditions of detecting stego videos in practical applications. The first video set is mixed by all the 720P videos with different bitrates (4 M, 8 M, 12 M). The second video set is mixed by all the 1080P videos (10 M, 30 M, 50 M). The third video set is mixed by both the 720P and 1080P videos.

Table 6: The classification accuracy of mix-bitrate video sets with PoPUPM-25D

Resolution	Classification accuracy
720P	97.0%
1080P	93.6%
720P&1080P	95.4%

Table 7: The classification accuracy of mix-bitrate video sets with PoPUPM-3D

Resolution	Classification accuracy
720P	99.7%
1080P	100%
720P&1080P	99.9%

Then the three mixed videos sets are sent to SVM classifier respectively using the same experimental setting above. The results are shown in Tabs. 6-7. As demonstrated in Tab. 6, the classification accuracy with PoPUPM-25D remains approximately 100%. For classification accuracy with PoPUPM-3D in Tab. 7, every classification accuracy decreases by only a few proportions of 3 or 4 percentages. On the other hand, it can be seen in Fig.4 that the ROC of PoPUPM-3D feature of 720P videos is higher than that of 1080P videos, and that is why the classification accuracy of PoPUPM-3D for 1080P mix-bitrate videos

decreases 3.4% more than that for fixed-bitrate 720P videos. In conclusion, the proposed method can identify unknown-bitrate videos with high classification accuracy.

Furthermore, the latest HEVC video steganalysis algorithm [Sheng, Wang and Huang (2017)] for prediction patterns is adopted to detect Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)] and compared with the proposed algorithm. In Sheng et al. [Sheng, Wang and Huang (2017)], the rate of change in quantity R_M is defined by Eq. (3),

$$R_M = |M - M'|/M \tag{3}$$

Where M and M' are the numbers of different sizes of PU before and after recompression respectively. Similarly, the rate of change in occupancy ratio R_P is defined by Eq. (4),

$$R_P = |P' - P|/P \tag{4}$$

Where P and P' are the occupancy ratios of different sizes of PU before and after recompression respectively. Then R_M and R_P of PU partition modes in sizes of 4×4, 8×8, 16×16 for are used as the classification feature. With the help of the source code from the author Sheng et al. [Sheng, Wang and Huang (2017)], we reproduce Sheng et al.'s algorithm [Sheng, Wang and Huang (2017)] to detect Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)]. The experimental configuration is the same as the proposed algorithm. Then the detection accuracy results are shown in Tab. 8.

Table 8: The classification accuracy of Sheng et al.'s algorithm

(a) 720P video set		(b) 1080P video set	
Bitrate	Classification accuracy	Bitrate	Classification accuracy
4M	49.3%	10M	55.9%
8M	52.5%	30M	46.5%
12M	49.4%	50M	51.9%

It can be seen in Tab. 8 that the classification accuracy of Sheng et al.'s algorithm is about 50%, which indicates that Sheng et al.'s steganalysis algorithm [Sheng, Wang and Huang (2017)] is no longer effective as in detecting video stegonagraphy algorithms. [Xu, Wang and Wang (2012); Xu, Wang and Xu (2015)] Compared with Sheng et al.'s algorithm, the classification accuracy of the proposed algorithm is much higher and the dimensions of the proposed feature are lower.

6 Conclusion

In this paper, a HEVC video steganalysis algorithm based on the statistical distribution of PU partition modes in P pictures is proposed. And the SVM classifier is used to discriminate the cover videos and stego videos. After feature optimization, the 25-dimensional feature is transformed to be 3 dimensional. Experiments are carried out on video sequences with various resolutions and bitrates. The results demonstrate that the detection accuracy to Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)] is over 96% for the videos in fixed bitrate. After mixing videos with different bitrates, the detection accuracy is still over 93%. In the future, more information data hiding algorithms will be introduced to

test the proposed steganalysis algorithm. Besides, in view of the security problem of Xie et al.'s algorithm [Xie, Yang, Li et al. (2018)], an improved information hiding algorithm could be developed to avoid the steganalysis algorithm proposed in this paper.

References

Chang, P. C.; Chung, K. L.; Chen, J. J.; Lin, C. H.; Lin, T. J. (2014): A DCT/DST-based error propagation-free data hiding algorithm for HEVC intra-coded frames. *Journal of Visual Communication and Image Representation*, vol. 25, no. 2, pp. 239-253.

Jainsky, J. S.; Kundur, D.; Halversion, D. R. (2007): Towards digital video steganalysis using asymptotic memoryless detection. *Proceedings of the 9th International Workshop on Multimedia and Security*, pp. 161-168.

Kong, W.; Wang, H.; Wang, K. (2014): Steganalysis of intra-prediction mode modulated information hiding algorithms based on transition probability matrix for H.264/AVC video. *Journal of Sichuan University (Natural Science Edition)*, vol. 51, no. 6, pp. 1183-1191.

Li, S.; Wang, L.; Liu, P.; Huang, Y. (2016): A HEVC information hiding approach based on motion vection space encoding. *Chinese Journal of Computers*, vol. 39, no. 7, pp. 1450-1463.

Lin, T. J.; Chung, K. L.; Chang, P. C.; Huang, Y. H.; Liao, H. Y. et al. (2013): An improved DCT-based perturbation scheme for high capacity data hiding in H.264/AVC intra frames. *Journal of Systems and Software*, vol. 86, no. 3, pp. 604-614.

Nie, Q.; Xu, X.; Feng, B.; Zhang, L. (2018): Defining embedding distortion for intra prediction mode-based video steganography. *Computers, Materials & Continua*, vol. 55, no. 1, pp. 59-70.

Sheng, Q.; Wang, R.; Huang, M. (2017): A prediction mode steganalysis detection algorithm for HEVC. *Journal of Optoelectronics Laser*, vol. 28, no. 4, pp. 433-440.

Wu, H.; Liu, Y.; Huang, J.; Yang, X. (2014): Improved staganalysis algorithm against motion vector based video steganography. *IEEE International Conference on Image Processing*.

Xie, W.; Yang, Y.; Li, Z.; Wang, J.; Zhang, M. (2018): An information hiding algorithm for HEVC videos based on PU partitioning modes. *International Conference on Cloud Computing and Security*, pp. 252-264.

Xu, D.; Wang, R.; Wang, J. (2012): Prediction mode modulated data-hiding algorithm for H.264/AVC. *Journal of Real-Time Image Processing*, vol. 7, no. 4, pp. 205-214.

Xu, J.; Wang, R.; Xu, D. (2015): An information hiding algorithm for HEVC based on angle differences of intra prediction mode. *Journal of Software*, vol. 10, no. 2, pp. 213-221.