MATEC Web of Conferences 25, 03003 (2015)
DOI: 10.1051/matecconf/20152503003

Research of Video Steganalysis Algorithm Based on H265 Protocol

Kaicheng Wu
School of Mathematics and Computer Science, Jianghan University, Wuhan, Hubei, China

ABSTRACT: A new generation of video coding standard H265 has raised a public concern by relevant scholars from all walks of life since April 2010. High-definition video is characterized by huge data size, complex encoding structure, high real-time performance and so on. Therefore, to design encryption and decryption program according to the actual needs and combined with application scenarios is imperative. Video steganalysis algorithm is a key content of video steganalysis (VSA), and VSA is a key technology of video decryption.

This paper researches LSB matching VSA based on H265 protocol with the research background of 26 original Video sequences, it firstly extracts classification features out from training samples as input of SVM, and trains in SVM to obtain high-quality category classification model, and then tests whether there is suspicious information in the video sample. The experimental results show that VSA algorithm based on LSB matching can be more practical to obtain all frame embedded secret information and carrier and video of local frame embedded. In addition, VSA adopts the method of frame by frame with a strong robustness in resisting attack in the corresponding time domain.

Keywords: video steganalysis; LSB matching; median filter; characteristic; secret information

1 INTRODUCTION

With the rapid development of high and new technology in the Internet and streaming media, multimedia communication has become an indispensable part for people to exchange information. Recently, the Joint Collaborative Team on Video Coding (JCT-VC) proposes a new generation of video coding standard HEVC. The first edition of HEVC standard was completed in January 2013, which was called as H265 by ITU-T. Steganography and steganalysis is one of themes of network information confrontation. Steganography embeds the secret information into the text, image, video and other digital carrier in disguise to achieve unknown covert communication. Steganalysis is divided into active and passive steganalysis. The passive steganalysis aims at determining whether the video contains the secret information, while the active steganalysis aims at estimating the secret information itself, that is, calculating the length of secret information, hiding place or parameters used in the process of steganalysis. To explore the principle of H265 video steganalysis, this paper proposes a video steganalysis algorithm based on LSB matching, so as to provide the theoretical basis for the development of high-quality multimedia communication technology.

Many people have made efforts on the video steganalysis. Weidong Zhong, et al. (2012) proposed a real-time video steganalysis method, which obtains an estimated value of video frame by a sliding window with the size of L+1, extracts the corresponding DCT and Markov characteristics, and tests steganographic video by the use of neural network, support vector machine and multiple classification methods[1]. Yifeng Sun, et al. (2010) proposed a video steganalysis detecting algorithm based on motion estimation, which researches the impact of non-motion estimation on embedded information through the changes of error of mean square, finding that the motion vector is sensitive to steganography. The smaller the block is, the more sensitive it is to the steganography [2]. Changyong Xu, et al. (2010) analyzed the impact of noise superimposition on temporal correlation and spatial correlation of the video sequence, and proposed a steganalysis algorithm based on space-time correlation by the use of impact of metric steganography of four-directional difference histogram for each frame in the video sequence on the spatial correlation, and use of metric steganography of adjacent frame difference histogram on the temporal correlation [3].

Based on previous research, this paper proposes a video LSB matching steganalysis method through constructing regional correlation diagrams, and adopts this method for video steganalysis of H265 protocol, so as to provide theoretical groundwork for the development of China's multimedia communication technology.

2 OVERVIEW OF H265 CODEC

People have an endless pursuit on video resolution. Because of this demand, the high-definition video increasingly emerges, but the butt joint of the current video CODEC standard and high-definition video has derivation [4]. To achieve seamless joint, the first session of JCT-VC was held in Germany in April 2010. JCT-VC refers to a video compression standard organization. In this session, a new generation of video

Article available at http://www.matec-conferences.org or http://dx.doi.org/10.1051/matecconf/20152503003

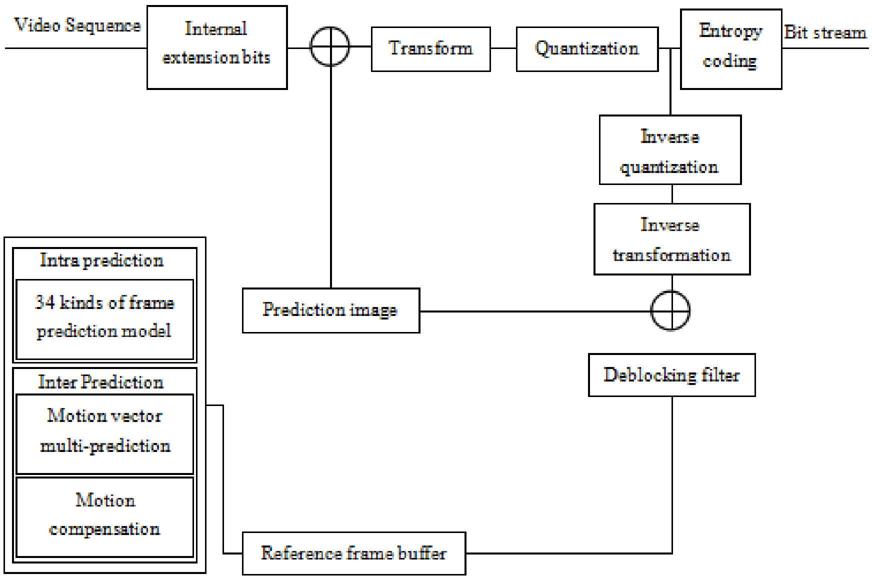

Figure 1. Framework of mixed video encoder with H265 standard

Table 1. List of basic situation of H265 model configuration scheme

HE	LC
Tree structure coding unit	Tree structure coding unit
(Luminance component 8×8 to 64 × 64)	(Luminance component 8×8 to 64 × 64)
Prediction unit	Prediction unit
Tree structure transformation unit (Three layers of depth)	Tree structure transformation unit (Two layers of depth)
Conversion unit is any consecutive three layers from 4 × 4	Conversion unit is any consecutive three layers from 4 × 4
to 32 × 32 (always square)	to 32 × 32 (always square)
Intra-frame prediction (up to 35 kinds of pattern)	Intra-frame prediction (up to 35 kinds of pattern)
Brightness pixel interpolation filter based on the discrete	Brightness pixel interpolation filter based on the discrete
cosine transform	cosine transform
(1/4 pixel 8-tab filter)	(1/4 pixel 8-tab filter)
Brightness pixel interpolation filter based on DCT trans-	Brightness pixel interpolation filter based on DCT trans-
form	form
(1/8 pixel 4-tap filter)	(1/8 pixel 4-tap filter)
Encoding unit Skip mode, combined advanced motion	Encoding unit Skip mode, combined advanced motion
vector prediction of prediction unit	vector prediction of prediction unit
Self-adaption binary arithmetic coding based on syntax	Low complexity encoding(LCEC)
elements	X
(SBAC)	Extended conversion accuracy (4 bits)
Increase the depth of the internal bits (4 bits)	Deblocking filter
X	X
Deblocking filter	
Self-adaption recursive filter	

coding standard is proposed: HEVC (High Efficiency Video Coding), namely, H265, and a prediction model (TMuC) [5],[6],[7] is established, then a new generation of video coding standard H265 emerges.

The design of H265 standard aims at improving coding efficiency and transmission system integration degree and data loss robustness, as well as enforceability of parallel processing architecture. Video coding layer of H265 still applies for the way of mixing performance of H264 video compression standard rules. As shown in Figure 1, the framework of the mixed video encoder with H265 standard is as follows.

H265 encoder contains two kinds of encoding schemes [8]: the High Efficiency (HE) encoding scheme and the Low Complexity (LC) encoding scheme. The specific configuration situation of HE and LC encoding schemes is shown in Table 1.

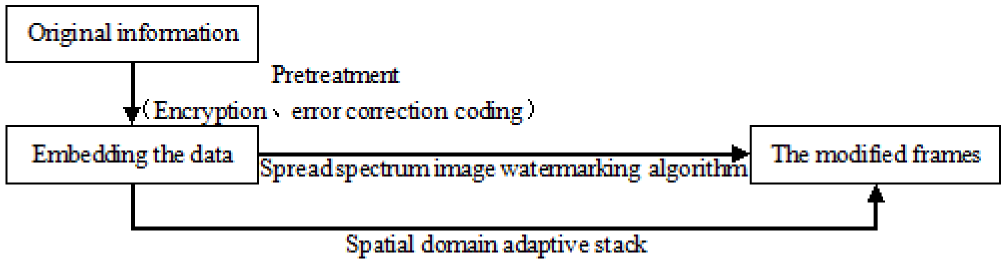

Figure 2. Schematic diagram of MSU steganography principle

3 VIDEO STEGANALYSIS METHOD

3.1 *Steganography principle*

Since embedding of secret information is distributed in different frequency bands of video, which makes the secret information in the individual frame is difficult to be detected. If the spread spectrum technology is used, it may make such secret information signal amplified and easy to detect information. Three ways of embedding the secret information are as follows:

$$v_i' = v_i + \alpha x_i \tag{1}$$

$$v_i' = v_i(1 + \alpha x_i) \tag{2}$$

$$v_i' = v_i\left(e^{\alpha x_i}\right) \tag{3}$$

Where: v_i in the formula (1), (2) and (3) represents coefficient of the original carrier image; x_i represents the embedded information sequence; α represents the embedded strength coefficient, and v_i' represents the steganography coefficient after embedding.

MSU Stego Video [9] is a kind of software for adding and processing information secret, which is a kind of new software appearing in combination with video content in the current network. Such software has a high robustness in terms of the types of information hiding, which can basically hide any type of files in the AVI video files. Su, et al [10] construct a video test frame sequence with a special significance by the use of a single gray value image with several frames, and then add secret information for frame content in each frequency band of Video by the use of MSU Video steganography software, and also compare and analyze the different image signals between secret carrier image and original image, so as to make clear of VSA mode features.

MSU is a kind of improved spread spectrum watermarking algorithm. Such algorithm has a very strong robustness to some extent. It is basically consistent with distribution rules of hidden information data in a single gray value image. In addition, such algorithm can also resist the impact of various standard video compression coding systems by rational selection of the embedded parameters. MSU steganography principle is shown in Figure 2.

3.2 *LSB matching steganalysis algorithm*

In different types of files, that is, secrete information, it is necessary to imbed the secret files into the target video (carrier embed point) in a way of secret information, but both bits are often not compatible. Based on the above incompatible situation, this paper uses the LSB matching steganalysis algorithm. Such algorithm adds and subtracts 1 for pixel value at this point based on the stochastic criteria. By the use of matching criteria based on related regions and to clarify the correlation T between the regions, such algorithm represents the correlation, $T \in [-255, 255]$ through calculating the difference value and mean value between the pixel value and center pixel value in eight neighborhoods. The smaller the absolute value of such value is, the higher the correlation of the region is. Its calculation is shown in formula (4):

$$T(i,j) = \frac{1}{8}\left(\sum_{u=i-1}^{i+1}\sum_{v=j-1}^{j+1} x_{u,v} - 9x_{i,j}\right) \tag{4}$$

Using the above formula (4), , we can obtain regional correlation T and absolute value $|T|$ in each pixel point of video for video pretreatment, so that the video constituted by pixel points turns into a video frame constituted by the value of regional correlation. This paper names it as RC (Regional Correlation) Diagram.

Based on the above matching criteria, the principle of LSB matching steganography can be described by the formula (5). Where: x_i represents the pixel value of the embedded points; m_i represents the bits with embedded secret information; m_i represents the pixel value after embedding secret information:

$$x_i' = \begin{cases} x_i & m_i = (x_i \bmod 2) \\ x_i - 1 & m_i \neq (x_i \bmod 2) \,\&\&\, (x_i = 255 \| T > 0) \\ x_i + 1 & m_i \neq (x_i \bmod 2) \,\&\&\, (x_i = 0 \| T < 0) \end{cases} \tag{5}$$

Assume that the size of video frame is $m \times n$ and define that $h(d)$ is a histogram of RC Diagram. Its calculation is shown in formula (6):

$$h(d)=\sum_{i=1}^{m}\sum_{j=1}^{n}\delta(T(i,j),d) \tag{6}$$

The regional correlation of the carrier video frame is relatively strong, while LSB matching steganography will weaken such correlation, so that the value of T is 0, and the number value of 1 decreases. However, when $T \geq 2$, the value will have a corresponding increase. By the use of t_1, t_2 classification, Feature 1 and Feature 2 are shown in formula (7) as t_1, t_2, and the two features of the video frame after steganography will increase. The Feature 3 as shown in formula (8) that t_3 represents the centroid of the characteristic function in the histogram, N is DFT length, and $H(k)=\mathrm{DET}(h(d))$.

To make the trend displayed by the histogram of RC Diagram more obvious, there is a need in smoothing original image. The process of such smoothing is the elimination of burr in image, such as setting up a low pass filter.

$$t_1=\frac{1}{h(0)+h(1)};\quad t_2=\left(\frac{\sum_{d=2}^{6}h(d)}{h(0)+h(1)}\right)\times\frac{\sum_{d=0}^{6}\times h(d)}{\sum_{d=0}^{6}h(d)} \tag{7}$$

$$t_3=\frac{\sum_{k=0}^{N/2}|H(k)|\times k}{\sum_{k=0}^{N/2}|H(k)|} \tag{8}$$

Features 4, 5, 6, 7, 8 and 9 are respectively represented by $M,\delta^2(d),r,U,e,\mu_3(d)$. Where: M represents the mean value of RC Diagram; $\delta^2(d)$ represents the variance; r represents the magnitude of variance, which is used to reflect the smoothness of RC Diagram; U represents the magnitude of consistency; e represents the information entropy of RC Diagram; $\mu_3(d)$ represents the third moment of RC Diagram, degree of skewness of the histogram. The calculation of above characteristic is shown in formula (9):

$$\begin{cases} M=\sum_{d=0}^{L-1}\left[\frac{h(d)}{m\times n}\times d\right]; r=\frac{\delta^2(d)}{1+\delta^2(d)}; U=\sum_{d=0}^{L-1}\left[\frac{h(d)}{m\times n}\right]^2 \\ \delta^2(d)=\sum_{d=0}^{L-1}\left[(d-M)^2\times\frac{h(d)}{m\times n}\right] \\ e=-\sum_{d=0}^{L-1}\left[\log_2\left(\frac{h(d)}{m\times n}\right)\times\frac{h(d)}{m\times n}\right] \\ \mu_3(d)=\sum_{d=0}^{L-1}\left[(d-M)^3\times\frac{h(d)}{m\times n}\right] \end{cases} \tag{9}$$

The calculation as two-dimensional histogram of RC Diagram is described in formula (10). x takes four directions: horizontal direction $(\Delta m=0,\Delta n=1)$, vertical direction $(\Delta m=1,\Delta n=0)$, opposite angle $(\Delta m=1,\Delta n=1)$ and anti-opposite angle $(\Delta m=1,\Delta n=-1)$. Where: $h^2(d_1,d_2)$ represents the mean value of four directions in two-dimensional histogram of RC Diagram.

$$\begin{cases} h_x^2(d_1,d_2)=\sum_{i=1}^{m}\sum_{j=1}^{n-1}\delta(T(i,j)=d_1,T(i+\Delta m,j+\Delta n)=d_2) \\ h^2(d_1,d_2)=\frac{1}{4}\times\left[h_h^2(d_1,d_2)+h_v^2(d_1,d_2)+h_d^2(d_1,d_2)+h_{ad}^2(d_1,d_2)\right] \end{cases} \tag{10}$$

Feature 10 and Feature 11 can be calculated by the formula (10). Its values are represented by t_{10} and t_{11}. Its calculation is shown in formula (11):

$$\begin{cases} t_{10}=\left[\sum_{d_1=0}^{1}\sum_{d_2=0}^{L-1}h^2(d_1,d_2)+\sum_{d_1=0}^{L-1}\sum_{d_2=0}^{1}h^2(d_1,d_2)\right]^{-1} \\ t_{11}=\frac{\sum_{k_1=0}^{\frac{N}{2}}\sum_{k_2=0}^{\frac{N}{2}}\left[H^2(k_1,k_2)\times(k_1+k_2)\right]}{\sum_{k_1=0}^{N}\sum_{k_2=0}^{N}H^2(k_1,k_2)}, H^2(k_1,k_2)=\mathrm{DFT}(h^2(d_1,d_2)) \end{cases} \tag{11}$$

3.3 *Classification methods of selected model in support vector machine*

This paper adopts the support vector machine (SVM) showing significant advantages in dealing with small samples, nonlinear and high-dimensional pattern recognition problems as a classifier [11]. SVM is a learning method based on statistical theory. The core idea of this method is explained as follows:

1) Based on minimization of structural risks, to control structural risks of learning machine through VC dimension of the minimum function set, it may have strong generalization ability.

2) To obtain scientific control of VC dimension, it can be achieved by the use of maximizing class interval.

3) To avoid solving nonlinear mapping for solving inner product, coring technology can be used to effectively achieve the goals. SVM seeks for a function based on functional Mercer theorem, so as to make the inner product of the sample space correspond to inner product of the transformation space.

In order to get better test results, it is necessary for the classifier to select appropriate estimated parameters. In training classifier of the support vector machine, the selection of estimated parameters mainly considers the following three aspects:

1) Compromise between the decision function and training sample, that is, computational cost C.

2) Selection of mapping function.

3) Kernel function: $K(x_i,x_j)=\Phi(x_i)\times\Phi(x_j)$

This paper carries out classification of carrier and secret video by the use of libsvm classification procedures provided by Lin ChihJen, and uses the radial basis function as the kernel function. Before classification and according to the formula (12), the features

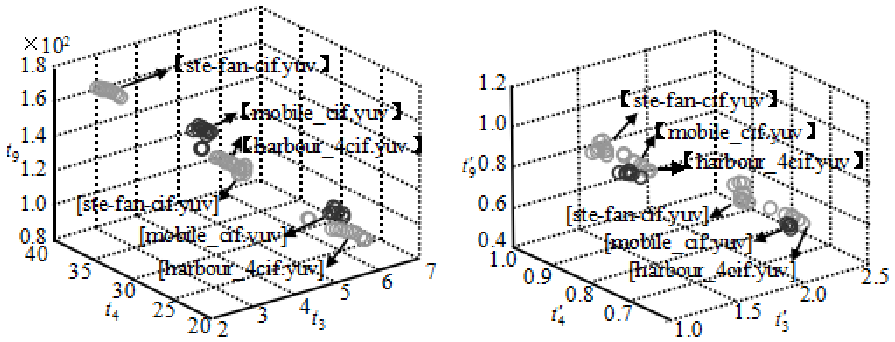

Figure 3. Effect diagram of the feature classification (Left - Direct extraction; right - extraction after median filtering)

Table 2. List of classification results of different embedding rates when all frames are embedded with information

Single frame embedding rate	Accuracy	False alarm rate	Omission rate
100.00%	99.95%	0.09%	0.00%
75.00%	99.93%	0.14%	0.00%
50.00%	99.91%	0.05%	0.14%
25.00%	99.88%	0.23%	0.00%
10.00%	97.44%	2.77%	2.35%

Table 3. List of classification results of different embedding rates when a part of frame is embedded

Single frame embedding rate	Frame embedding rate			
	80.00%	60.00%	40.00%	20.00%
	Accuracy	Accuracy	Accuracy	Accuracy
00.00%	91.51%	97.52%	94.40%	90.20%
75.00%	100.00%	99.90%	99.87%	99.73%
50.00%	100.00%	99.78%	100.00%	99.32%
25.00%	100.00%	100.00%	99.26%	98.43%
10.00%	93.81%	92.76%	79.74%	66.88%
Mean value of detection rate	97.06%	97.99%	94.65%	90.91%

used for classification shall be normalized in the range of [-1, + 1]:

$$t_i'' = \frac{2t_i' - 3\min(t_i') - \max(t_i')}{\max(t_i') - \min(t_i')} \qquad (12)$$

$\min(t_i'), \max(t_i')$ in formula (12) respectively represent the minimum and maximum values of Feature i .

4 EXPERIMENTAL RESULT AND ANALYSIS

The experimental object of this paper is the video sequence. A total of 26 original video sequences are selected, including all kinds of video sequences of multiple kinds of velocity. The frame size of each sequence is 100 frames. There are three types of frame size, namely: 4CIF, CIF and QCIF. The experiment adopts the triple cross validation, and randomly selects training samples and testing samples from the sample library with the proportion of 2:1, and firstly selects the classification features in the training samples as SVM input, implements training, and obtains optimal classification model, and then uses the trained classifier to detect whether the video sample to be tested contains suspicious information. To get an accurate classification results, the experiment is repeated for 50 times according to the above process, and it obtains the average results.

4.1 *Classification results*

Median filter can remove the distinction between various disparate video carriers that respectively extracts 11 features of the original carrier and Stego Video, and then compare with 11 features obtained after the median filtering. The original feature is represented by $t_i(i = 1,2,\cdots,11)$; the feature after filtering is represented by $t_i'(i = 1,2,\cdots,11)$. To display the difference target more intuitive, the research just verifies the same features 3, 4, 9 that extracted from three video sequences. Each of them produces a three-dimensional scatter diagram as shown in Figure 3.

(1)Video1.ste-fan-cif.yuv;(2)Video2.mobile_cif.yuv; (3) Video 3.harbour_4cif.yuv.

[XX] in Figure 3 is XX Stego Video, [XX] is XX carrier Video. The effect of feature classification in the figure shows that directly extracted three-dimensional classification features have partial overlapping in ste-fan-cif.yuv carrier Video and harbour_4cif.yuv Stego Video, while three-dimensional classification features extracted from the median filter can clearly distinguish carrier and Stego Video. Thus, this paper adopts the median filter to remove the effect of differences between different video carriers, so as to serve for the high-quality classification performance.

4.2 *Classification performance analysis of frame embedding in different ways*

The secret information is embedded in each frame in Video sequences, with the same embedding rate of single frame. The embedding rate is respectively 100%, 75%, 50%, 25% and 10%. The test results of the Stego Video are shown in Table 2.

As shown in Table 2, the relationship trend curve of embedding rate, accuracy, false alarm rate and omission rate of single frame is shown in Figure 4 and Figure 5:

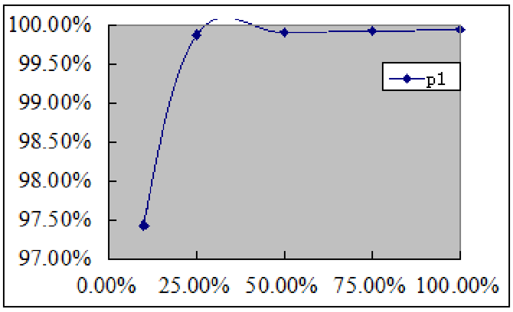

Figure 4. Trend chart of the impact of different embedding rates of single frame on accuracy

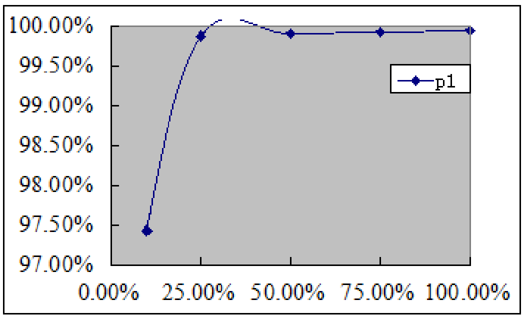

Figure 5. Trend diagram of the impact of different embedding rates of single frame on false alarm rate and omission rate

The classification effect of VSA on embedding by all frames is very good, and the false alarm rate and omission rate are relatively low. With the decrease of the embedding rate, the detection rate also has a slight decrease. When the embedding rate is 10%, a high detection rate can also be gained.

A part of video is randomly selected for embedding. The number of secrete frames can be respectively selected as 80%, 60%, 40% and 20%. The embedding rate of single frame in each video sequence can randomly select one from the embedding rate p of 75%, 50%, 25% and 10%. Its classification results are shown in Table 3.

As shown in Table 3, when various kinds of embedding rates are available in a video sequence, LSB matching VSA algorithm proposed in this paper has a higher detection rate for the frame with a higher embedding rate or non-steganography frame. For the frame with a lower embedding rate, it is prone to be misjudged as a video frame without the secret information.

With the decrease of Stego Video frame ratio in the video sequence, the misjudgment rate has a corresponding increase. But on the whole, LSB matching VSA algorithm has a very high mean value of the detection rate with a high availability based on the situation of steganographic mixed embedding of Video LSB matching.

5 CONCLUSION

Based on the summarization of H265 CODEC, this paper sets forth the principle of VSA, and proposes a LSB matching VSA algorithm, and gives out a kind of computational algorithm of the classification of features value, and carries out video steganalysis analysis and experiment on 26 original video sequences by the use of a classification method used for selecting models by the support vector machine. The experimental result shows that:

1) Analysis of the carrier and Stego Video can be realized through constructing RC Diagram and extracting relevant statistical characteristics, and removing the difference of features in different video sequences by the median filter.

2) The classification effect of LSB matching VSA algorithm on embedding by all frames is very good, and the false alarm rate and omission rate are relatively low. With the decrease of the embedding rate, the detection rate also has a slight decrease. When the embedding rate is 10%, a high detection rate can also be gained.

3) When various kinds of embedding rates are available in a video sequence, LSB matching VSA algorithm proposed in this paper has a higher detection rate for the frame with a higher embedding rate or non-steganography frame.

4) LSB matching VSA algorithm has a very high mean value of the detection rate with a high availability based on the situation of steganographic mixed

embedding of Video LSB matching.

In conclusion, the video steganalysis algorithm based on LSB matching has a very strong robustness. The further research shall focus on LSB matching steganography in the detection field and other video steganalysis algorithms, so as to expand an applicable scope of such algorithm.

REFERENCES

[1] Weidong Zhong, Junqing Wu, Gengrui Wu & Haibin Yang. 2012. Invisible video steganalysis method based on space-time redundancy statistics. *Application Research of Computer*. 29 (10): 3846-3850.

[2] Yifeng Sun & Fenlin Liu. 2010. Video steganalysis method based on motion estimation. *Pattern recognition and artificial intelligence*. 23 (6): 759-765.

[3] Changyong Xu & Xijian Ping. 2010. Video steganalysis method based on space-time correlation. *Journal of China Image and Graphics*. 15 (9): 1331-1337.

[4] Fang Guo. 2013. *Research of video security CODEC program based on standard H265*. Chengdu: Southwest Jiaotong University.

[5] Taoran Lu, Xiaoan Lu, Qian Xu, Yunfei Zheng, Joel Sole, & Peng Yin. 2011. A Video Coding Analyzer for Next-Generation Compression Standards. *IEEE International Conference on Consumer Electronics* (ICCE).

[6] Ken McCann, Samsung, & ZetaCast. 2010. Tool Experiment 12: Evaluation of TMuC Tools [R], ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) document JCTV-B312, July.

[7] Frank Bossen. 2010. AHG report: Software development and TMuC software technical evaluation [R], ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) document JCTVC-B003, July.

[8] Sze V, Budagavi M & Chandrakasan A. 2009. Massively Parallel SBAC [R], ITU-Y SGI 6/Q. 6VCEG-AL21. [S.1.]: JCT-VC.

[9] WANG Wei-hong, & FARID H. 2007. Exposing digital forgeries in video by detecting duplication. *Proc of the 9th Workshop on Multimedia & Security*. New York: ACM Press: 35-42.

[10] SU Yu-ting, WANG Li-li, & ZHANG Chun-tian. 2007. A new digital video steganalysis algorithm against motion vector. *Proc of WiCOM*: 164-167.

[11] Rong Li, Shiwei Ye, & Zhongzhi Shi. 2002. SVM-KNN classifier: a new method of improving SVM classification accuracy. *Journal of Electronics*. 30 (5): 745-748.