Video Steganalysis Based on Subtractive Probability of Optimal Matching Feature

Yanzhen Ren
Computer Science Department
Key Laboratory of Aerospace Information Security and trusted computing Ministry of Education
Wuhan University
Wuhan, China
renyz@whu.edu.cn

Liming Zhai
Computer Science Department
Key Laboratory of Aerospace Information Security and trusted computing Ministry of Education
Wuhan University
Wuhan, China
zhailiminghn@163.com

Lina Wang
Computer Science Department
Key Laboratory of Aerospace Information Security and trusted computing Ministry of Education
Wuhan University
Wuhan, China
lnwang@whu.edu.cn

Tingting Zhu
Information security department
Naval university of engineering
Wuhan, China
zlztt1688@sina.com.cn

ABSTRACT

This paper presents a novel motion vector (MV) steganalysis method. MV-based steganographic methods exploite the variability of MV to embed messages by modifying MV slightly. However, we have noticed that the modified MVs after steganography cannot follow the optimal matching rule which is the target of motion estimation. It means that steganographic methods conflict with the basic principle of video compression. Aiming at this difference, we proposed a steganalysis feature based on Subtractive Probability of Optimal Matching(SPOM), which statistics the MV's Probability of the Optimal matching (POM) around its neighbors, and extract the classification feature by subtracting the POM of the test video and its recompressed video. Experiment results show that the proposed feature is sensitive to MV-based steganography methods, and outperforms the other methods, especially for high temporal activity video.

1. INTRODUCTION

Steganography is a technology of sending secret messages under the covert channel of normal objects, such as digital images and videos. Steganalysis is the art to attack steganography by identifying suspected cover objects. In recent years, with the widespread of networked multimedia applications, such as IPTV, Video Conference, and Video on Demand, etc., compressed video streams which can easily achieve a large hiding capacity become a new covert channel for steganography. A lot of video steganography algorithms and tools emerged[9, 5, 8, 16, 1, 2, 4]. Unlike traditional video steganography, which uses the spatial or frequency coefficients of video frames for embedding secret messages, the MV-based schemes adopt MVs as the information carrier to achieve covert communication. There are two benefits of MV-based steganography method. The first is that the statistical characteristics of the spatial or frequency coefficients of video frames will not be changed, thus the methods is secure to the traditional steganalysis methods. The second is that the degradation of the visual quality will be very little. Owing to these two reasons, several MV-based steganographic algorithms have been proposed in the last few years (e.g., [8, 1, 2, 4, 16, 5]).

Kutter *et al.*[8] firstly proposed a MV-based steganlography method, selected MVs whoes magnitudes are non-zero and embedded secret message by modifying the LSBs of those MVs' horizontal or vertical components. Xu *et al.*[1] suggested embedding the data in the MVs which magnitudes are above a predefined threshold. Fang and Chang[2] designed a method adjusting MVs' phase angles by modifying MV's magnitude. In order to reduce the negative effects of coding efficiency. Aly[4] selected the candidate subset of motion vectors which have smaller prediction errors to hide information. Cao *et al.*[16] choose the optimal or suboptimal MVs according to the embedded information, and wet-paper coding algorithm was adopted to enhance the security of steganography. Jing *et al.*[5] choose a small part of motion vectors to modify to resist the steganalysis method in[19]. All MV-based steganography methods above share some features in common, i.e., they first select a subset of MVs following a predefined selection principle, then do the embedding operations by modifying the MV's magnitude to meet consistent with the secret information.

To attack the MV-based steganography, Su *et al.*[19] firstly proposed a MV-steganalystic method based on the statisti-

IH&MMSec'14, June 11–13, 2014, Salzburg, Austria.

http://dx.doi.org/10.1145/2600918.2600938 .

cal analysis of relative properties between neighboring MVs. A feature classification technique is adopted to determine the existence of hidden messages. Cao *et al.*[17] considered the tendency of MV's reversion, designed a recompression calibration-based approach to estimate the cover video, and proposed a steganalysis method which adopts MV reversion-based features derived from the differences between the original and the calibrated videos. Deng *et al.*[18] presented a MV recovery algorithm based on local-polynomial kernel regression model, and proposed statistical features based on calibration distance histogram for steganalysis. The methods in [19] and [18] are based on the correlation which exists between the neighboring MVs, but for videos which have high temporal activity, the correlated characteristic is not obvious, and the steganalysis method will work ineffectively.

Most of the steganalysis features proposed in the literatures(e.g., [17, 18, 19]) are all based on the statistic characters of the MV's magnitude. However, MV is an intermediate variable in video compression, which magnitude is not unique, thus the steganalysis features based on MV's magnitude are not stable. In this paper, we change the direction to look for the steganalysis feature. We have noticed that those MVs modified by MV-based steganographic methods don't follow the optimal matching rules which are the aim of motion estimation in general video compression. It means that steganographic methods conflict with the basic encoding principles. Based on this idea, we proposed a feature by measuring MVs' optimal matching characteristics to complete steganalysis.

The rest of the paper is organized as follows: In Section II, we introduce the basic concepts of motion estimation in video compression, and present the optimal matching characteristics of MVs in cover and stego video. The implementation of our proposed steganalysis method is described in Section III. In Section IV,comparative experiments are carried out to show the performance of the proposed method. Finally, concluding remarks are given in Section V with some considerations about future research.

2. THE OPTIMAL MATCHING CHARACTERISTICS OF MVS

2.1 Motion Estimation

Motion vector is generated by motion estimation (ME) in video compression. Most video encoding standards, such as MPEG-2, MPEG-4, H.263, H.264, H.265, etc., implement video compression by removing both the spatial and temporal redundancy[12]. The goal of motion estimation is to reduce the temporal redundancy of video, and its basic idea is to predict the current coded frame by one or more prior coded frames due to the high correlation of neighbored video frames. The schematic diagram of motion estimation is shown in Figure 1. To encode the current macro-block (MB) of size $m \times m$, the encoder adopts one prior coded frame R as the reference, and search for the best matching block R_B within a searching area R_s as B's reference block. As result, the motion vector mv_B represents the spatial displacement offset between B and R_B, and the Prediction Error (PE) which is the pixel difference between B and R_B is further coded and transmitted.

To search for the best matching block, Full search (FS) algorithm gives the global optimum solution by exhaustively

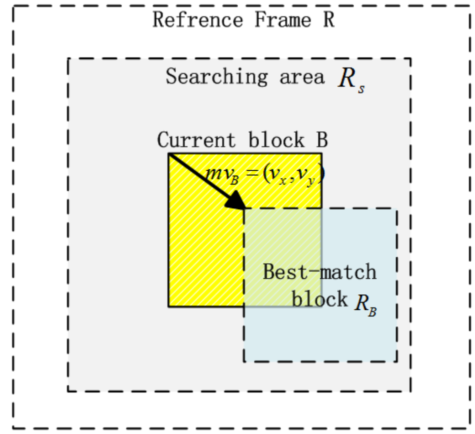


Figure 1: Schematic Diagram of Motion Estimation

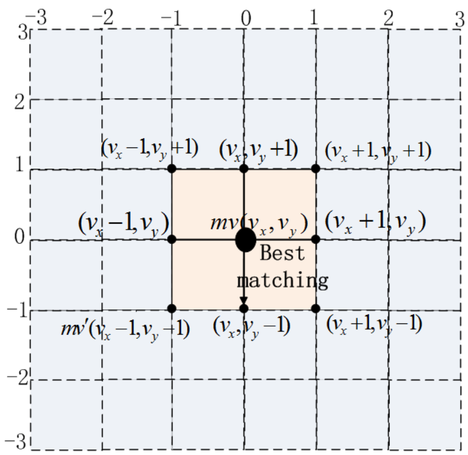
testing all the candidate blocks within the searching area. To lower the huge computational complexity of FS, many fast block-matching algorithms (BMAs) have been proposed and applied to the video compression standards, such as Three-step Search (TSS), Four-step Search(FSS), Diamond Search(DS), etc.[6]. These fast BMAs exploit different searching patterns and searching strategies to find the optimum motion vector which drastically reduce the number of search points as compared with the FS algorithm. To measure the prediction error between the target and candidate blocks, several matching criteria have been used, such as Mean Squared Error (MSE), Sum of Absolute Differences (SAD), Mean Absolute Difference (MAD)[14], etc. Where SAD is commonly used and computed as:

$$SAD(B, R_B) = \sum_{1 \le i,j \le m} |B(i,j) - R_B(i,j)| \qquad (1)$$

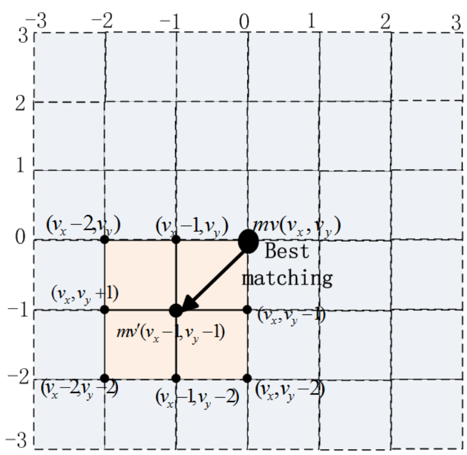
Where $B(i,j)$ and $R_B(i,j)$ are luminance values of current coded block B and reference block R_B. The purpose of BMAs is to find an optimal matching block which makes SAD or MAD to be minimum.

2.2 The optimal matching characteristics of MVs in cover and stego video

For cover video which is compressed by normal video encoder and not modified by MV-based steganography methods, its MVs are generated by motion estimation and its reference block should be the best matching block among its neighbors. For stego video which MVs are modified to embed secret message, the MV will shift from the optimal matching location. In order to maintain the visual quality and avoid distortion drift, the steganography methods need to modify the reference block accordingly, thus the new reference block will not be the best matching block among its neighbors. Figure 2 shows an example of MV steganography to explain the opinion simply. The original MV is $mv(v_x, v_y)$, which is the best matching MV within its neighbors. When mv is modified to $mv^{'}(v_{x-1}, v_{y-1})$ by steganography method shown in Figure 2(b), we can see that the new MV $mv^{'}$ will not be the best matching MV within its neighbors.



(a) Original MV and its neighbors



(b) Modified MV and its neighbors

Figure 2: Example of MV steganography

The optimal matching properties of MVs in cover video and stego video will be analyzed in this section. Without loss of generality, given an inter-Macro Block (MB) $B(i, j)$ in a video frame, while (i, j) are the coordinates of B's left-upper point. B's MV in compressed stream is $mv(v_x, v_y)$, where v_x is the horizontal value of MV, and v_y is the vertical value of MV. The corresponding reference MB is $R_B(i + v_x, j + v_y)$ in reference frame R. To measure the optimal matching properties of MVs, we will give some definitions as follows:

Definition 1. MV's neighborhood Set $N_{n\times n}$. For $mv(v_x, v_y)$, given a neighborhood range $n \times n$, where $n = 2k + 1, k \geq 1$, given a MV set

$$\mathtt{C} = mv(v_x + d_x, v_y + d_y) | -k \leq d_x \leq k, -k \leq d_y \leq k$$

$\mathtt{C}$ is called mv's neighborhood Set $N_{n\times n}$. Figure 3 shows the example of $N_{5\times 5}$.

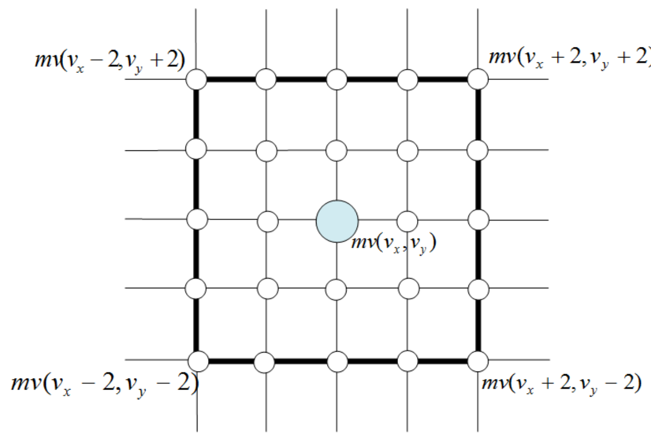


Figure 3: The example of MV neighborhood Set $N_{5\times 5}$

Definition 2. Optimal matching. For $mv(v_x, v_y)$ and its reference block R_B, given a MV neighborhood set $N_{n\times n}$. For arbitrary $mv_i \in N_{n\times n}$, its corresponding reference MB is R_i ($1 <= i <= n^2$), and the set of neighborhood reference MB is R_{round}. If $SAD(B, R_B) = MIN_{R_i \in R_{round}}(SAD(B, R_i))$, where function $MIN_r(x)$ means the minimum value of x in range r, then we called mv is the optimal matching, otherwise, mv is not optimal matching.

Definition 3. Probability of optimal matching (POM). Given a video segment V, the number of MVs in V is N_{mv}, and the number of MVs which are optimal matching is N_{om}, then the POM of V is $P_V = N_{om}/N_{mv}$.

Definition 4. Calibrated Video. Calibration is a procedure of video recompression. Given a compressed video segment V, decompressed V to spatial domain and gained the YUV sequence Vseq, then compress Vseq again without any embedding to get the compressed video segmen $\hat{V}$. $\hat{V}$ is named as the calibrated video of V.

Definition 5. Subtractive probability of optimal matching (SPOM). Given a test video segment V, $\hat{V}$ is the calibrated video of V. The POM of V is P_{test}, and the POM of $\hat{V}$ is P_{cab}, then the SPOM of V is P_{err}, which is the subtraction of P_{cab} and P_{test}, i.e. $P_{err} = P_{cab} - P_{test}$.

For cover video, the MVs are selected by motion estimation to search for the best matching MV. If MVs are selected by Full search algorithm, all MVs should be optimal matching, and the POM of those cover videos should be one. However this situation is too idealistic. All the video encoders used actually employ fast Block Matching Algorithms (BMAs) to search MV, and a little amount of points are checked. In this situation, POM of those cover video cannot reach one. We had done experiment on nine standard video sequences to test POM of cover videos and stego videos, the sequences are selected from Xiph[15] and described in Table 1. Table 2 shows POM of cover videos which are encoded by X264[13] and JM18.0[10] respectively on different neighborhood Set. Table 3 shows POM of the stego videos which all MVs are modified randomly by 1/4 pixels, and the videos are encoded by X264[13] and JM18.0[10]. The result in Table 2 shows that POM of cover videos compressed by different encoder are all larger than 50%, and The result in Table 3 shows that POM of stego videos are smaller and lower than their cover videos obviously. It can be found that POM should be a good feature to distinguish cover video and MV-based stego video.

Table 1: Description of Video Sequence

Video Sequence	Frame size	Frame Number	Camera motion	Object motion
Bus	352x288	150	Panning	Translation(fast)
Flower	352x288	250	Panning	Translational
Foreman	352x288	300	Panning	Translation(fast)
Coastguard	352x288	300	panning	Translation (slow)
Container	352x288	300	panning	Translation (slow)
Mother	352x288	300	neglible	Translation (slow)
Stefan	352x288	300	Panning	Translational
Soccer	352x288	600	Panning(fast)	Translation(fast)
Football	352x288	360	Panning(fast)	Translation(fast)

Table 2: POM of Cover Videos

Video Sequence	X.264 encoder			JM18.0 Encoder		
	$N_{3\times3}$	$N_{5\times5}$	$N_{9\times9}$	$N_{3\times3}$	$N_{5\times5}$	$N_{9\times9}$
Bus	62.46%	61.04%	60.64%	61.75%	61.53%	60.46%
Flower	64.82%	63.65%	63.35%	62.83%	61.79%	61.65%
Foreman	57.25%	54.18%	53.31%	58.77%	57.65%	57.24%
Coastguard	63.50%	62.47%	63.50%	61.46%	60.79%	60.15%
Container	67.46%	65.75%	65.72%	66.59%	65.77%	64.61%
Mother	52.47%	52.16%	51.99%	55.65%	54.87%	54.13%
Stefan	64.45%	63.78%	62.97%	61.87%	61.43%	60.99%
Soccer	51.62%	51.54%	51.24%	53.86%	52.71%	52.15%
Football	53.75%	53.25%	52.47%	55.65%	54.73%	54.02%
Average	**59.75%**	**58.65%**	**58.35%**	**59.83%**	**59.10%**	**58.38%**

For the actual steganography methods described in literatures [e.g.,[8, 1, 2, 4, 16, 5]], the modified rate of MVs cannot be 100%, thus the POM of the stego videos cannot be so small as Table 3. For example, the methods in [8] selected MVs which magnitude is non-zero and embedded secret message by modifying LSBs of those MVs' horizontal or vertical components. Even if all MVs are non-zero, and the modified rate of the method in [8] is nearly 50% when each MV is embeded. Figure 4(a) shows the POM of the videos on different embedding rate by methods described in [8], while embedding rate means the proportion

Table 3: POM of Stego Videos (All MVs are Modified)

Video Sequence	X.264 encoder			JM18.0 Encoder		
	$N_{3\times3}$	$N_{5\times5}$	$N_{9\times9}$	$N_{3\times3}$	$N_{5\times5}$	$N_{9\times9}$
Bus	2.98%	2.44%	2.37%	1.85%	1.37%	1.29%
Flower	2.88%	1.98%	1.90%	1.73%	1.44%	1.09%
Foreman	3.67%	3.10%	2.96%	2.57%	2.28%	1.97%
Coastguard	3.88%	3.33%	2.99%	2.84%	2.58%	2.29%
Container	2.62%	2.48%	2.06%	2.54%	2.46%	2.02%
Mother	3.35%	3.10%	2.82%	2.99%	2.49%	2.11%
Stefan	2.66%	2.22%	2.03%	1.86%	1.60%	1.16%
Soccer	3.16%	2.95%	2.65%	2.79%	2.37%	2.07%
Football	1.46%	1.03%	1.01%	0.96%	0.85%	1.76%
Average	**2.96%**	**2.51%**	**2.31%**	**2.24%**	**2.00%**	**1.75%**

of the number of embedding bits and the number of non-zero MVs. When embedding rate is zero, it means that the video is cover video, and when embedding rate is 100%, it means all the non-zero MVs in the video are embedded by LSB method and almost half non-zero MVs should be modified. From Figure 4(a) we can find that POM of varied cover videos are different greatly, but for one specific video, the POM on different embedding rate is liner to embedding rate nearly. Therefore, if we can get the POM of the video's cover video, we can steganalysis the video accurately.

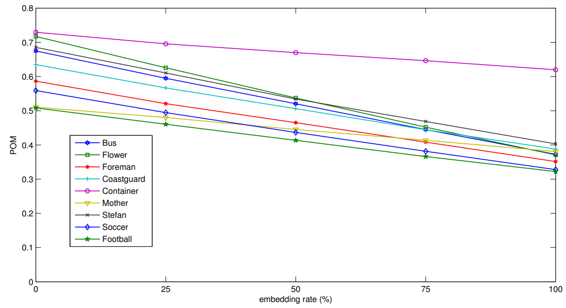
Calibration [7] is a well known image steganalytic concept which estimates the macroscopic properties of the cover from the stego image. For compressed video, MV is an intermediate referenced data, the video frame is resumed by MV and its corresponding PE jointly. The recovered frame of MV-based stego video has little distortion from its original cover video, and the MVs of the recompressed video are selected by motion estimation, thus POM of calibrated video can be taken as the estimation of its cover video. This idea had been verified by the experiment results shown in Figure 4(b), where the x-axis is the calibrated videos which are recompressed by videos on varied embedding rate (cover videos' embedding rate is zero). Figure 4(b) shows that POM of calibrated videos on different embedding rate are almost near to its cover video's.

For a test video, we can adopt its SPOM feature between the test video and its calibrated video to check whether the video is steoganographied. Figure 4(c) shows SPOM of the test video sequences in Table 1, we can find that SPOM of cover video is almost near to zero, and SPOM of stego video is linear to its embedding rate. Therefore SPOM can classify the cover video and stego video effectively. Based on this feature, we propose our steganalyzer in next section and compare its performance with the other methods in section IV.

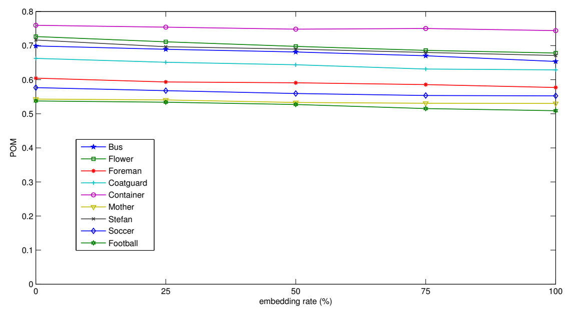
3. PROPOSED STEGANALYZER

Based on the fact that MV's POM in stego video is lower than its original cover video's, and the POM of calibrated video is almost similar to cover video, we proposed a MV-based steganlyzer which classify the video by the SPOM feature of test video. The framework of our steganlyzer is shown in Figure 5, and the procedure is described as follows.

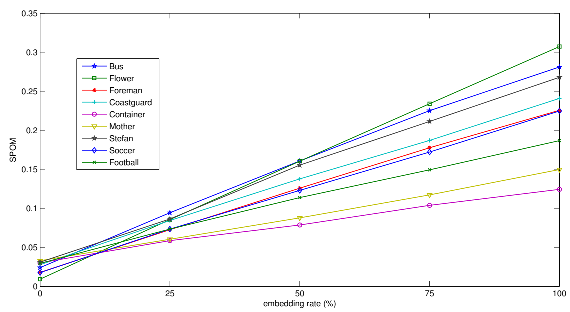
For a given test video V, firstly, calculate the POM of V P_{test} as described in Section II. Then Decode V to spatial



(a) POM of video sequences



(b) POM of calibrated video sequences



(c) SPOM of video sequences

Figure 4: The POM and SPOM feature of standard video sequences

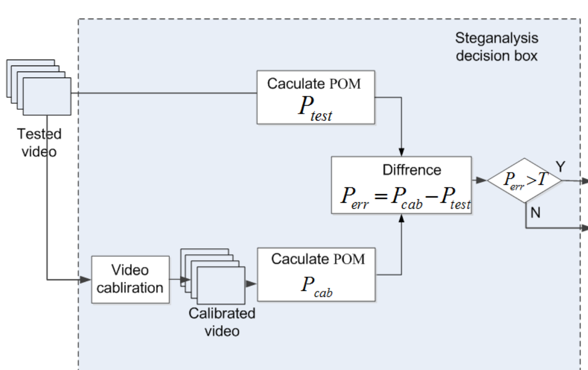


Figure 5: Proposed framework for steganalysis

domain and encode it again to get the calibrated video $\hat{V}$, and calculate the POM value of $\hat{V}$'s P_{cab}. Calculate the subtraction of P_{cab} and P_{test} to get V's steganalysis feature P_{err}, $P_{err} = P_{cab} - P_{test}$. Finnaly, given an experienced threshold T, if $P_{err} > T$, then judge the test video as stego video, otherwise, the test video is classified as cover video.

4. EXPERIMENTS

To evaluate the performance of our proposed feature, we apply the proposed steganalytic algorithm to the standard video sequences, and compared its performance with the methods state-in-art. Besides, the performance under high temporal activity videos are also evaluated.

4.1 Experiment Setup

4.1.1 Test Sequences

A video database of 47 CIF video sequences in 4:2:0 YUV format is used for experiments which are selected from Xiph[15] and TNT[11], with a special emphasis on the dynamic components in the videos, such as camera motion and moving objects. The frame rate of the video sequences is 30 fps. Since the length of sequences is varied and most of them have 300 frames, then we divide each sequence into non-overlapping sub-sequences with 100 frames and the total number of sub-sequences sums up to 162. The experiments of these sub-sequences are carried on compressed video sequences in H.264 compressed by X264 [13] encoder, and decoded by FFmepg H.264 decoder[3].

4.1.2 Steganographic Methods

Our experiments focus on attacking three MV-based steganographic methods, i.e. Xu *et al.*'s [1], Aly's [4], Cao *et al.*'s [16] methods, and which are referred to as Xu, Aly, and Cao. These targets are implemented using X264 [13] encoder. As the message bits are embedded into MVs, the embedding strength in the experiment is measured by the average embedded bits per inter-frame (bpf). The MV's modified step is 1/4 pixel and embedded in a random embedding range from -3/4 to 3/4 pixels.

4.1.3 The threshold T

the value of T in our method is gainned by training method. 1/3 video sequences (15 YUV sequence consisting of 52 subsequence) are randomly selected. All subsequences are compressed by X264 encoder with standard settings to produce the class of cover videos. At the same time, all subsequences are embedded with the method in [1] at different embedding strength, including 50 bpf, 100 bpf, 200 bpf. When T = 0.036, the true negative rate and true positive rate of the method can be reached a good tradeoff, both of them can be reached to 92.80%, thus we adopt T as 0.036 in the experiment.

4.1.4 The detected range of the Neighborhood

We have made experiments on $N_{3\times3}$, $N_{5\times5}$, $N_{9\times9}$, the result shows that the increasing of the detected range have got small upgrade for detection accuracy. From table 2 and table 3 we can learn that the range of the Neighborhood has little affect on the POM of the test videos.Thus in this experiment, we adopt $N_{5\times5}$, it can get a good tradeoff between detected accuracy and computational complexity.

4.2 Performance Results

In this experiment, besides our proposed features, Su *et al.*'s [19], Cao *et al.*'s [17], and Deng *et al.*'s[18] steganalytic features are leveraged for comparison, and they are referred to as Su2011, Cao2012, and Deng2012. We evaluate the performance using true positive (TP) and true negative (TN), where TP is the occurrence that a stego video is classified as stego, TN is the occurrence that a cover video is calssified as cover. The performances of the steganalyzers with varied embedding strengths are tested, and the corresponding results are recorded in Table 4.

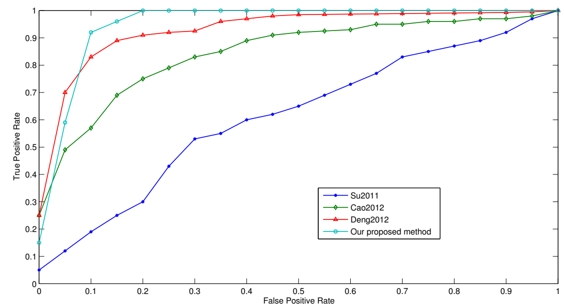
Table 4: Performance Comparsion within Su2011's, Cao2012's, Deng2012's and Our Proposed Method

	bpf	***Su2011***		***Cao2012***		***Deng2012***		**Our proposed**	
		TP	***TN***	***TP***	***TN***	***TP***	***TN***	***TP***	***TN***
Xu	**50**	58.4	54.6	88.5	86.7	88.3	94.9	92.8	92.8
	100	76.1	76.2	90.7	88.1	87.5	93.4	100	92.8
	200	84.5	95.3	95.4	89.7	92.1	92.7	100	92.8
	300	97.5	95.6	99.3	95.2	95.7	92.1	100	92.8
Aly	**50**	59.1	52.3	79.5	80.2	90.7	87.6	94.2	92.8
	100	72.3	63.6	84.6	81.2	91.3	91.7	100	92.8
	200	81.3	77.4	91.5	85.9	94.8	92.1	100	92.8
	300	94.5	93.1	98.1	91.3	96.5	93	100	92.8
Cao	**50**	53.4	52.8	76.9	78.5	90.5	86.4	91.5	92.8
	100	65	55.4	79.8	75.8	92	87.9	100	92.8
	200	79.1	76.9	88.1	85.1	94.7	91.8	100	92.8
	300	90.7	89.3	93	90.7	93.1	90.6	100	92.8
Average		**76**	**73.5**	**89.2**	**85.7**	**92.3**	**91.2**	**98.2**	**92.8**
Feature dimension		**12**		**15**		**3**		**1**	

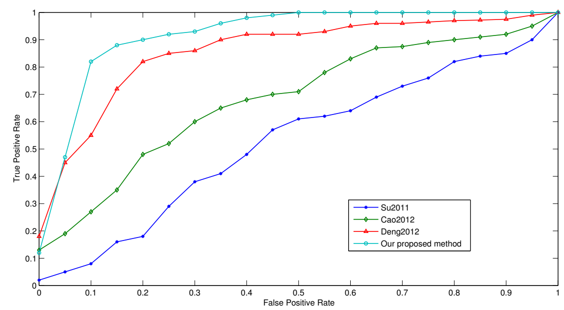
In comparison with the detection performance of Su2011's, Cao2012's and Deng2012's steganalyzers, our proposed features have better performance. Since that the classification in our method is judged by comparing with one fixed threshold T, thus, TN of our method is the same in varied embedding strength, and TP for varied test sample is different. The average TP of our proposed method is 98.2%. When embedding strength is higher than 100bpf, TP of our proposed method is 100%. The dimension of steganalysis features in Su2011's, Cao2012's and Deng2012's is 12, 15, and 3 respectively, in our method, only 1 feature is needed, it means that the feature we proposed is sensitive to MV-based steganography methods.

The detector receiver operating characteristic (ROC) curves of these seganalyzers for different steganography are shown in Figure 6, where the embedding strength is 50bpf. Figure 6(a) is the ROC for stego video by Xu's method, and Figure 6(b) is the ROC for stego video by Cao's method. We can find that the performance of our method is better than the comparing methods for the same steganography method.

The computational complexity of the proposed method is not very high. To extract the feature, each video should be recompressed once, and each inter-frame should be decoded and the POM should be calculated for each MV, The computational complexity of the method is less than the complexity to recompress the video twice. The experiment results show that the averge time consumed to get the SPOM of a CIF video of 300 frame is almost 20 seconds. To improve the computational efficiency, some new high-performance computing technologies, such as parallel computing or cloud computing can be adopted.



(a) Steganography method is Xu's



(b) Steganography method is Cao's

Figure 6: ROC curves of steganalyzers using Su2011's, Cao2012's and Deng2012's and our proposed features on 50bpf

4.3 Discussion

The classification features of literature methods are based on the correlation among adjacent MVS, thus the detection accuracy strongly depends on the video's content. For example, the methods in [19] is based on the spatial correlation and temporal correlation of the motion vectors in video streams. Experiments results show that the correlation strength is relevant to the video content. When there has fast-moving object in the scene, the correlation in spatial domain is strong. However, when there is a low motion object in the scene or the adjacent scenes has little difference, the correlation in temporal domain is stronger than spatial domain. In a video sequence with many scenes, it is difficult to evaluate the differences in strength of correlation. For some high temporal activity videos, such as Highway, Tempete, etc. [15], the correlation among adjacent MVs is weak, and the performance of literature methods will be dropped.

For our proposed method, the classification feature is based on the basic principle of MV's generation. The optimal matching characteristic is constant for all kinds of compressed video. When MV is modified, its optimal matching characteristic will be destroyed, even for the video of weak correlation among adjacent MVS, our method is still effective. To prove this opinion, we have done experiments on high temporal activity videos to compare the performance of our proposed method and the literature methods.

Four video sequences are selected from Xiph[11] and TNT[14] in the 4:2:0 YUV, and the detailed description of the video sequences is given in Table 5.

Table 5: Description of the High Temporal Activity Video Sequences

Video Sequence	*Frame size*	*Number of frames*	*Camera motion*	*Object motion*
Soccer	352x288	300	Panning(fast)	Translation(fast)
Highway	352x288	360	Panning	Translation(fast)
Tempete	352x288	260	Zooming	Translation(fast)
Ice	352x288	480	Static	Translation(fast)

The cover and stego videos of Table 5 are compressed by X264[13], and the stego videos are produced by the steganography of Xu's [1] to embed random messages, the embedding strength is 300bpf. The experiment results are shown in Table 6. Since that TP of the comparing steganalysis methods are all 100% in this situation, we only show TN of each methods. From Table 6, we can find that the average of TN in our method is 92.6%, others' are lower than 42.1%.

Table 6: TN of Su2011's, Cao2012's, Deng2012's and Our Proposed Method(300bpf, Target Method is Xu's)

Video Sequence	**Su2011**	**Cao2012**	**Deng2012**	**Our proposed**
	TN	TN	TN	TN
Soccer	18.2	27.6	32.6	92.8
Higway	25.6	31.7	44	94.5
Tempete	28.7	39.7	46.2	91.7
Ice	26.2	35.1	45.7	91.3
Average	**24.7**	**33.5**	**42.1**	**92.6**

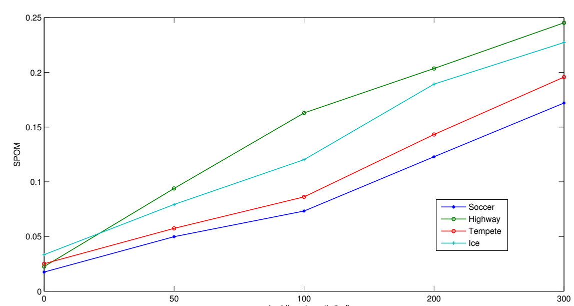


Figure 7: SPOM of high temporal activity video sequences

Figure 7 shows the SPOM features of the video sequences in Table 6 under varied embedding strength. For an embedding strength of 300bpf, the value of each video's SPOM is larger than T obviously, and TN of our proposed method depends only on the embedding strength. When embedding strength is under 25 bpf, the performance of our method needs further improvement.

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a novel MV-steganalytic method which is based on the optimal matching characteristic in compressed video by calculating the feature of MVs' Subtractive Probability of Optimal Matching (SPOM). Experiment results show that the feature is stable and sensitive to distinguish the cover and MV-based stego video. There are three advantages of the feature. Firstly, based on the common principle of MV's generation, the feature can be applicable for a variety of video compression standards which select the optimal matching MV to reduce the video's temporal redundancy, including MPEG-2, MPEG-4, H.263, H.264, H.265, etc. Secondly, the feature is independence on the correlation of adjacent MV, thus the feature has less relationship with the content of the video. Especially for high temporal activity videos, the feature has better classification capability than literature methods. Thirdly, the feature is comparing the relationship between the MV and its neighbors, instead of comparing the value of the MV and recovered MV (eg. [17, 18]), thus it has little dependence on the recovery accuracy of calibration algorithm, and especially suitable for the video encoding standards with variable block size motion estimation, such as H.264, H.265.

However, the detection performance of our proposed method under very low embedding strength (for example: 25bpf) is not satisfactory and should be improved further. Meanwhile, although the method can gain a good performance aginst the MV-based steganographic method in literatures, it only takes account of MV's POM feature for a local range. For some future steganographic methods which can guarantee locally optimal matching character, the proposed method should be improved, and the megered feature of local feature and global feature will be successed. Therefore, in our future work, we will combine the SPOM feature with the correlation character of the adjacent MVs to attack the steganographic methods which will be emerged in future.

6. ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Foundation of China (NSFC) under the grant No. 61373169, 61272451 and 61272453.

7. REFERENCES

[1] C.Xu, X.Ping, and T.Zhang. Steganography in compressed video stream. *IEEE First International Conference on Innovative Computing, Information and Control* (2006), 269–272.

[2] D.Fang, and L.Chang. Data hiding for digital video with phase of motion vector. *Proc. Int. Symposium on Circuit and Systems (ISCAS)* (2006), 1422–1425.

[3] FFmpeg. *FFmpeg 264 decoder*. http://ffmpeg.org/. *@Online*.

[4] H.A.Aly. Data hiding in motion vectors of compressed video based on their associated prediction error. *IEEE Transactions on Information Forensics and Security 1* (6 2011), 14–18.

[5] H.Jing, X.He, Q.Han, and X.M.Niu. Motion vector based information hiding algorithm for h.264/avc against motion vector steganalysis. *4th Asian Conference on Intelligent Information and Database Systems* (2012), 78–83.

[6] H.Y.C.Tourapis, and A.M.Tourapis. Fast motion estimation within the h.264 codec. *International Conference on Multimedia and Expo 3* (2003), 517–520.

[7] J.Fridrich. Feature-based steganalysis for jpeg images and its implications for future design of steganographic schemes. *Proc.IH'04, Lecture Notes in Computer Science 3200/2005* (2004), 67–81.

[8] M.Kutter, F.Jordan, and T.Ebrahimi. Proposal of a watermarking technique for hiding/retrieving data in compressed and decompressed video. *Technical Report M2281, ISO/IEC document, JTCI/ SC29/ WG11* (1997).

[9] Stego, M. *MSU :: MSU Graphics Media Lab*. http://compression.ru/video/stego-video/index-en.html. *@Online*.

[10] Suehring, K. *H.264/AVC JM Reference Software Download*. http://iphome.hhi.de/suehring/tml/download/. *@Online*.

[11] TNT. *Tnt.uni Video Test Media*. ftp://ftp.tnt.uni-hannover.de/pub/svc/testsequences/. *@Online*.

[12] T.Wiegand, G.J.Sullivan, G.Bjontegaard, and A.Luthra. Overview of the h.264/avc video coding standard. *IEEE Trans. CircuitsSyst. VideoTechnol 13* (2003), 560–576.

[13] VideoLAN. *VideoLAN x264, the best H.264/AVC encoder*. http://www.videolan.org/developers/x264.html. *@Online*.

[14] Xiaozhong, X., and Yun, H. Improvements on fast motion estimation strategy for h.264/avc. *IEEE Transactions on Circuits and Systems for Video Technology 18*, 3 (2008), 285–293.

[15] Xiph. *Xiph.org :: Derf's Test Media Collection*. http://media.xiph.org/video/derf/. *@Online*.

[16] Y.Cao, X.F.Zhao, and D.G.Feng. Video steganography with perturbed motion estimation. *13th International Conference on Information Hiding* (May 2011).

[17] Y.Cao, X.F.Zhao, and D.G.Feng. Video steganalysis exploiting motion vector reversion-based features. *IEEE Signal Process Lett.19* (2012), 35–38.

[18] Y.Deng, Y.Wu, and L.Zhou. Digital video steganalysis using motion vector recovery-based features. *Applied Optics 51*, Issue 20 (2012), 4667–4677.

[19] Y.T.Su, C.Q.Zhang, and C.T.Zhang. A video steganalytic algorithm against motion-vector-based steganography. *Signal Processing* (2011), 1901–1909.