A Novel Embedding Distortion for Motion Vector-Based Steganography Considering Motion Characteristic, Local Optimality and Statistical Distribution

Peipei Wang
State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
wangpeipei@iie.ac.cn

Hong Zhang
State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
zhanghong@iie.ac.cn

Yun Cao*
State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
caoyun@iie.ac.cn

Xianfeng Zhao
State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
zhaoxianfeng@iie.ac.cn

ABSTRACT

This paper presents an effective motion vector (MV)-based steganography to cope with different steganalytic models. The main principle is to define a distortion scale expressing the multi-level embedding impact of MV modification. Three factors including motion characteristic of video content, MV's local optimality and statistical distribution are considered in distortion definition. For every embedding location, the contributions of three factors are dynamically adjusted according to MV's property. Based on the defined distortion function, two layered syndrome-trellis codes (STCs) are utilized to minimize the overall embedding impact in practical embedding implementation. Experimental results demonstrate that the proposed method achieves higher level of security compared with other existing MV-based approaches, especially for high quality videos.

Keywords

Steganography; video; motion vector; embedding distortion; H.264/AVC

*The corresponding author.

IH&MMSec 2016, June 20-23, 2016, Vigo, Spain

DOI: http://dx.doi.org/10.1145/2909827.2930801

1. INTRODUCTION

Steganography is the art and science of data hiding, which realizes covert communication by embedding secret messages into innocent-looking cover media, such as digital image, audio, video, *et al.*, without arousing eavesdropper's suspicion. Facilitated by advanced video compression and computer network technology, digital video has become one of the most influential media. Therefore, secret message delivery can utilize video transmission as the cloak.

As the particular parameter in compressed video, MV is widely considered to be the ideal covert information carrier for the following reasons. First, MVs are generated in motion estimation (ME) and further losslessly coded without quantization distortions, thus the visual quality degradation introduced by embedding data in MVs is relatively limited. Second, because MV is the key information expressing video content, compressed video always contains vast quantities of various MVs. The sufficient quantity ensures the high embedding capacity held by MV-based steganographic approaches. Besides, the broad range of values makes it's achievable to propose steganographic algorithms preserving MVs' statistical characteristics.

A series of steganographic methods have been developed in MV area. The early algorithms select a subset of MVs by following preset selection rules and then use simple LSB replacement to modify them. Kutter [1] selected non-zero MVs and embedded message bits by modifying the LSBs of their horizontal and vertical components. Xu *et al.* [2] suggested selecting MVs whose magnitudes are above a given threshold for modification. Other than magnitude-based selection rules, Aly [3] chose the candidate MVs according to their associated prediction errors. By applying mature coding techniques such as wet paper codes (WPCs) [4] and syndrome-trellis codes (STCs) [5, 6] to video steganogra-

phy, adaptive steganographic schemes with distortion functions have been presented in recent years. Cao *et al.* [7] selected the suboptimal MVs of small differences compared with their neighbors and utilized WPCs to embed information. In Yao *et al.*'s work [8], a distortion function for MV-based steganography was defined by considering the statistical distribution change and the prediction error change. And two-layered STCs [6] are used to minimize distortion for embedding process. In order to resist the steganalytic schemes based on MV's local optimality [9, 10], several approaches were proposed [11, 12]. Zhang *et al.* [11] designed a distortion based on video compression efficiency degradation. By selecting local optimal MVs in a designated area and further encoding using STCs with the least costs, the local optimality of modified MVs could be preserved. In [12], Cao *et al.* exploited the opportunity to optimize the ME perturbation using the loss caused by video compression process. With distortion scale calculated on optimal neighbors, data embedding was implemented using a double-layered (first channel: STCs, second channel: WPCs) coding structure.

Although the existing video steganographic schemes manage to elaborate effective distortion function and minimize the cost by adopting steganographic codes, they can not keep a high level of security. The following reasons can account for this. On the one hand, all of these steganographic approaches are designed for specialized purposes, thus when detected by other steganalytic methods, the level of security deteriorates. For example, Cao's [7] and Yao's steganographic methods [8] can not withstand local optimality based steganalysis such as [10, 9] and there is possibility that Zhang's [11] and Cao's steganographic schemes [12] can be detected by calibration based steganalysis such as [13, 14]. On the other hand, most distortion functions are defined using the embedding influence on single video's characteristic, which is dependent on the selected compressed video. Thus the distortion definition is not generally applicable for multifarious videos.

This paper aims at defining a distortion function for MV-based steganography by considering embedding influence from different aspects. Because rich motion regions have a large number of MVs with various-ranged values, motion characteristic of video content is considered in defining the distortion function. As one of the effective evidence exploited by steganalysis, the local optimality of MVs is also considered in distortion definition. Last but not the least, MVs' statistical distribution is also one important consideration. Particularly, in our proposed scheme, the distortion is dynamically allocated to these three factors by implementing adaption at every embedding location.

The rest of paper is organized as follows. In Section 2, we describe the framework of distortion minimization for MV-based steganograpy. In Section 3, the definition and analysis of distortion function is elaborated. The implementation of the video steganographic method is presented Section 4 and followed by the experimental results shown in Section 5. Finally, the conclusions and future works are given in Section 6.

2. DISTORTION MINIMIZATION FOR MV-BASED STEGANOGRAPHY

Minimizing the overall embedding distortion is an accepted approach to improve steganography security. Different from operations on pixels in image steganography, MV-based steganographic approaches perform embedding during ME process. Their specific framework of distortion-minimization is established as follows.

In MV-based steganography, n MVs are obtained during video compression, denoted by $\mathbf{MV} = \{mv_1, \ldots, mv_n\}$. By binary mapping $x_i = \mathscr{P}(mv_i)$ with a designed parity check function $\{\mathscr{P}(mv_i)|i = 1 \ldots n\} = \{0, 1\}$, the covert channel $\mathbf{x} = \{x_1, \ldots, x_n\}$ is established. Given the embedding rate γ, the $\gamma \cdot n$ -length binary message $\mathbf{m}$ can be embedded by turning $\mathbf{x}$ into $\mathbf{x}' = \{_1, \ldots, x'_n\}$ using steganographic codes, which satisfies

$$\mathbf{H}\mathbf{x}^T = \mathbf{m} \tag{1}$$

where $\mathbf{H}$ is the parity check matrix employed by the steganographic codes. As the result of embedding, the modified MV set $\mathbf{MV}' = \{mv'_1, \ldots, mv'_n\}$ is obtained and it satisfies $\mathscr{P}(mv'_i) = x'_i$.

On the assumption that the modification of MVs is mutally independent, the minimal distortion can be approached by embedding messages with STCs [5] and the embedding and extraction are formulated as

$$\widetilde{\mathbf{x}} = Emb(\mathbf{x}, \mathbf{m}) = arg \min_{\mathbf{x}' \in \mathscr{C}(\mathbf{m})} D(\mathbf{x}', \mathbf{x}) \tag{2}$$

$$Ext(\widetilde{\mathbf{x}}) = \mathbf{H}\widetilde{\mathbf{x}}^T = \mathbf{m} \tag{3}$$

where the overall distortion is computed by $D(\mathbf{x}', \mathbf{x}) = \sum_{i=1}^{n} \phi(x_i, x'_i)$, $\mathscr{C}(\mathbf{m})$ is the coset corresponding to syndrome $\mathbf{m}$ and $\widetilde{\mathbf{x}}$ is the binary bit stream embedded with message.

Based on STCs, two-layered ± 1 STCs [6] are proposed to improve higher security. In this paper, we define the distortion function for MVs and design a video steganographic method using two-layered STCs-based ± 1 embedding.

3. PROPOSED DISTORTION FUNCTION

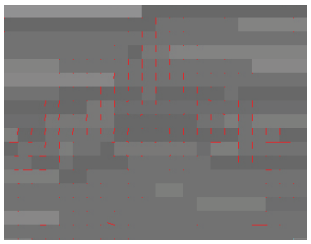
3.1 Motion Characteristic of Video Content

In image steganography [15, 16], data tends to be embedded in complex textural regions. Analogously, it is intuitive that modifications in rich motion areas are likely to raise less suspicions in video steganography. For a specific macroblock (MB), the richer its motion is, the more suitable it can be used for modification.

As the integral part of existing video coding standards, ME is designed to reduce the temporal redundancy between video frames. This is achieved by allowing blocks of pixels from currently coded frame to be matched with those from reference frame(s). As the result of ME, MV represents the spatial displacement offset between a block and its prediction. Therefore, it has been widely accepted that the MV indicates MB's motion to some extent [17, 18, 19]. In addition, in order to achieve better compression ratio, various quantization parameter (QP) value on MB basis is allowed in H.264/AVC (Advanced Video Coding) standard [20]. As referred in [21], QPs of the adjacent MBs will have very small change in values if they are actually the part of static background. Thus the MBs can be segmented as static background or foreground object according to their QP differences.

Figure 1(a) shows a frame of the H.264 stream named "Snatch.264". Figure 1(b) depicts the MV and QP values

(a)H.264 stream



(b)MV and QP values

Figure 1: MV and QP information of "Snatch.264" sequence.

of the same frame. The red line shows the magnitude and direction of MV and the gray block shows the value of QP. Darker the color of MB, higher is the QP magnitude and vice versa. It can be seen that both MV's magnitude and QP's difference can represent MB's motion characteristic. Thus the motion characteristic based distortion can be defined as follows.

Definition 1. (**Motion Characteristic Based Distortion, MCD**). For $mv_{i,j,t}$ which locates at (i,j) in the tth frame, its motion characteristic based distortion can be defined by

$$MCD_{i,j,t} = \frac{1}{|\boldsymbol{MV}_{i,j,t}| \cdot (|\triangle QP_{i,j,t}| + 1)} \tag{4}$$

$$|\boldsymbol{MV}_{i,j,t}| = \sqrt{mvx_{i,j,t}^2 + mvy_{i,j,t}^2} \tag{5}$$

where $\boldsymbol{MV}_{i,j,t}$ is the corresponding MV of $MB_{i,j,t}$, $|\boldsymbol{MV}_{i,j,t}|$ denotes the magnitude, and $|\triangle QP_{i,j,t}|$ is the absolute value of QP difference. Less distortion can be obtain if the MV's magnitude or the QP difference is larger, which means that modifications are inclined to make on MBs with rich motions. If the MV's magnitude is 0, it is prohibited for data embedding by defining its corresponding distortion as ∞.

3.2 Local Optimality of Motion Vector

As one of the inherent properties of MV, the local optimality has been used in MV-based steganalysis. In steganalytic methods such as [9, 10], it is believed that the extracted MVs from clean video are locally optimal while the modification operations of MV-based steganography will break the local optimality.

As shown in Figure 2(a), the MV $mv_{i,j,t}$ directly extracted from compressed video is also denoted by $mv_{i,j,t}^{0,0}$. By adding or subtracting one MV value, the potentially optimal MVs of $mv_{i,j,t}$ are obtained, which are denoted by $\boldsymbol{PMV}(mv_{i,j,t}) = \{mv_{i,j,t}^{x,y} | x,y \in \{1,0,-1\}\}$. Figure 2(b) shows their corresponding sums of absolute differences (SADs) $\boldsymbol{PSAD}(mv_{i,j,t}) = \{sad_{mv_{i,j,t}}^{x,y} | x,y \in \{1,0,-1\}\}$ between current block and its reference MB, which are computed by Eq. (6).

$$sad_{mv_{i,j,t}}^{x,y} = SAD(\boldsymbol{Pred}_{mv_{i,j,t}}^{x,y}, \boldsymbol{Rec}_{mv_{i,j,t}}) \tag{6}$$

where $\boldsymbol{Pred}_{mv_{i,j,t}}^{x,y}$ is the prediction block pointed by $mv_{i,j,t}^{x,y}$ and $\boldsymbol{Rec}_{mv_{i,j,t}}$ refers to the reconstructed block using extracted $mv_{i,j,t}$. Then we compare the $sad_{mv_{i,j,t}}^{0,0}$ with other SADs in the 3×3 structure. If it holds the minimum value, the associated $mv_{i,j,t}^{0,0}$ is identified as the local optimal MV.

Based on the assumption that the MVs directly obtained from the compressed video are locally optimal, Wang *et al.* [10] infer that MV-based steganographic approaches would destroy the local optimality of MVs. Holding the opinion that modifications shift MVs from the local optimal locations to non-optimal, they design the effective steganalytic feature based on adding or subtracting one MV value (AoSO) [10]. However, many recent studies [11, 12] have revealed it is possible that the modified MVs can preserved their local optimality, which confutes AoSO's theory. Under such case, the MVs after embedding could remain local optimal and the modifications in such steganographic methods may be made undetectable. Thus in local optimality preserved steganography, the keystone and difficulty is to find the set of substitutable candidates for every MV. The substitutable MVs (SMV) of $mv_{i,j,t}$ can be signified as

$$\boldsymbol{SMV}(mv_{i,j,t}) = \{smv_{i,j,t}^{k} | k = 1, \ldots, K\} \tag{7}$$

where K is the cardinality of $\boldsymbol{SMV}(mv_{i,j,t})$ and every element $smv_{i,j,t}^{k}$ in this set can pass the local optimality test.

One of the state-of-art methods utilizes the information distortion during lossy compression to construct $\boldsymbol{SMV}$ [12]. Because the video compression is an information-reducing process, many side-information like residual errors gets lost after transformation and quantization. As the calculated result based on reconstructed block, the $\boldsymbol{PSAD}(\widetilde{mv_{i,j,t}})$ at the decoder is usually different from the original $\boldsymbol{PSAD}(mv_{i,j,t})$ at the encoder side. Although analysis in [10] has proved that most $mv_{i,j,t}$s are still local optimal during decoding, the differences between two SAD matrices provide us with great chance to find $smv(mv_{i,j,t})$.

It is observed that some non-optimal neighbors of $mv_{i,j,t}$ are turned into local optimal ones due to lossy compression. Figure 3 shows the neighboring SAD matrices of $mv_{i,j,t}$ and its SMV $nmv_{i,j,t}$ at the encoder and decoder side respectively, where $nmv_{i,j,t}$ equals $(mvx_{i,j,t}+1, mvy_{i,j,t})$. At the decoder side, besides the original local optimal MV, its neighboring MV can also pass the local optimality test. Therefore, the SMV set of $mv_{i,j,t}$ consists of the MVs as follows.

$$\boldsymbol{SMV}(mv_{i,j,t}) = \{nmv_{i,j,t}^{k_n} | k_n = 1, \ldots, K_n\} \tag{8}$$

$$nmv_{i,j,t}^{k_n} \in \boldsymbol{Neighbors}_{i,j,t} \tag{9}$$

$$sad_{nmv_{i,j,t}^{k_n}}^{\sim 0,0} = min(\boldsymbol{PSAD}(nmv_{i,j,t}^{k_n})) \tag{10}$$

where K_n is the number of the SMVs of $mv_{i,j,t}$, $\boldsymbol{Neighbors}_{i,j,t}$

$mv_{i,j,t}^{-1,-1}$ $(mvx_{i,j,t}-1, mvy_{i,j,t}-1)$	$mv_{i,j,t}^{-1,0}$ $(mvx_{i,j,t}-1, mvy_{i,j,t})$	$mv_{i,j,t}^{-1,1}$ $(mvx_{i,j,t}-1, mvy_{i,j,t}+1)$
$mv_{i,j,t}^{0,-1}$ $(mvx_{i,j,t}, mvy_{i,j,t}-1)$	$mv_{i,j,t}^{0,0}$ $(mvx_{i,j,t}, mvy_{i,j,t})$	$mv_{i,j,t}^{0,1}$ $(mvx_{i,j,t}, mvy_{i,j,t}+1)$
$mv_{i,j,t}^{1,-1}$ $(mvx_{i,j,t}+1, mvy_{i,j,t}-1)$	$mv_{i,j,t}^{1,0}$ $(mvx_{i,j,t}+1, mvy_{i,j,t})$	$mv_{i,j,t}^{1,1}$ $(mvx_{i,j,t}+1, mvy_{i,j,t}+1)$

(a)The structure of $mv_{i,j,t}^{x,y}$.

$sad_{i,j,t}^{-1,-1}$	$sad_{i,j,t}^{-1,0}$	$sad_{i,j,t}^{-1,1}$
$sad_{i,j,t}^{0,-1}$	$sad_{i,j,t}^{0,0}$	$sad_{i,j,t}^{0,1}$
$sad_{i,j,t}^{1,-1}$	$sad_{i,j,t}^{1,0}$	$sad_{i,j,t}^{1,1}$

(b)The structure of $sad_{i,j,t}^{x,y}$.

Figure 2: The structures for local optimality computation: the extracted MV, potentially optimal MVs and their corresponding SADs.

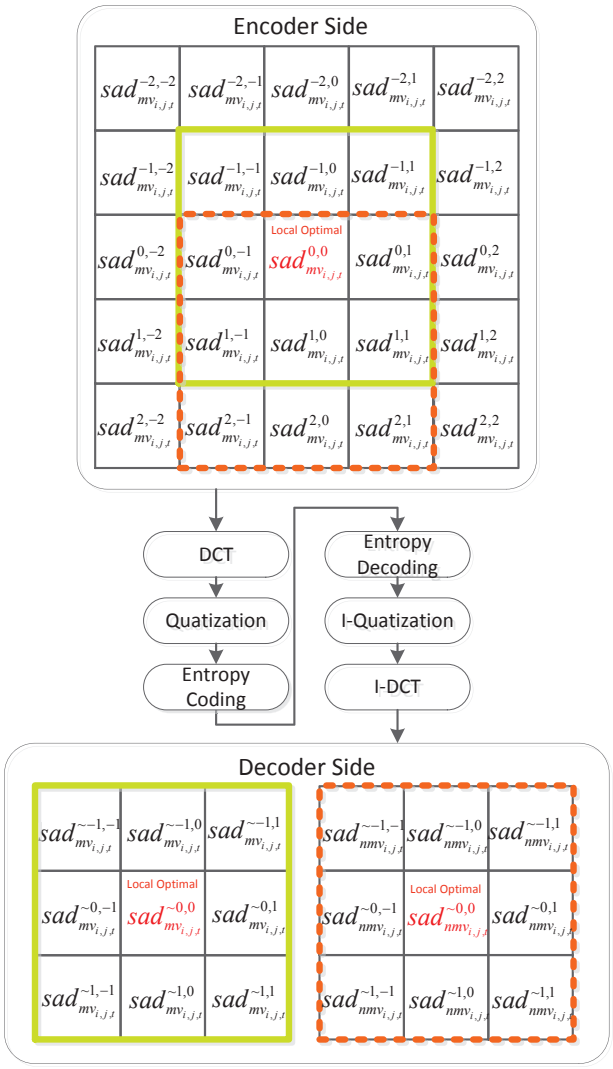


Figure 3: Illustration of SMV introduced by lossy compression.

$= \{(mvx_{i,j,t}+1, mvy_{i,j,t}), (mvx_{i,j,t}-1, mvy_{i,j,t}), (mvx_{i,j,t}, mvy_{i,j,t}+1), (mvx_{i,j,t}, mvy_{i,j,t}-1), (mvx_{i,j,t}+1, mvy_{i,j,t}+1), (mvx_{i,j,t}+1, mvy_{i,j,t}-1), (mvx_{i,j,t}-1, mvy_{i,j,t}+1), (mvx_{i,j,t}-1, mvy_{i,j,t}-1)\}$ is the set containing the neighboring MVs of $mv_{i,j,t}$, $sad^{\sim 0,0}_{nmv^{k_n}_{i,j,t}}$ denotes the corresponding SAD of $nmv^{k_n}_{i,j,t}$ obtained at the decoder. Thus the SMV set is constructed using the neighboring MVs whose local optimality can be preserved at the decoder side. If the MVs are replaced with the SMVs in this set, the modifications would resist the detection of AoSO feature.

Another recent method [11] is designed to select SMVs in a designated searching area. Different from obtaining MVs using default searching method, this approach collects all the SMVs by perturbing the ME process. And as the result of motion compensation (MC) in video decompression, the video quality and the MVs' local optimality can be ensured at the decoder side. As illustrated in Figure 4, the SMV set of $mv_{i,j,t}$ is composed of the MVs as follows.

$$\boldsymbol{SMV}(mv_{i,j,t}) = \{cmv^{k_c}_{i,j,t} | k_c = 1, \ldots, K_c\} \quad (11)$$

$$cmv^{k_c}_{i,j,t} \in \boldsymbol{SearchArea}_{i,j,t} \quad (12)$$

$$sad^{\sim 0,0}_{cmv^{k_c}_{i,j,t}} = min(\boldsymbol{PSAD}(cmv^{k_c}_{i,j,t})) \quad (13)$$

where K_c is the number of the SMV of $mv_{i,j,t}$, $\boldsymbol{SearchArea}_{i,j,t}$ denotes the designated searching area in reference frames for $MB_{i,j,t}$, $cmv^{k_c}_{i,j,t}$ is the qualified MV located in the searching area, and $sad^{\sim 0,0}_{cmv^{k_c}_{i,j,t}}$ denotes the corresponding SAD of $cmv^{k_c}_{i,j,t}$ at the decoder. In this case, the SMVs' local optimality can also be preserved and the modifications from original MVs to SMVs would also be made in undetectable way.

However, both of these two steganographic approaches have obvious limitations in real-time scene. As analyzed above, the first method finds the SMVs based on the lossy compression's disturbance to SADs. Thus the number of SMVs will decrease drastically if the information loss reduces during the video compression. Table 1 lists the proportions of MVs with different number of SMVs. It can be noticed that reduction of the SMVs' number indeed exists with the increment of bit-rate. Peculiarly, the videos with high bit-rate contain very little number of SMVs, which is not sufficient to support practical steganography. As to the second method, original MV is replaced with any of the qualified SMVs within the designated searching area. These operations lead to larger modifications of the MVs' horizontal or vertical components, which makes it relatively detectable when using statistical analysis or calibration-based steganalysis.

In this paper, an adaptive selection strategy is proposed to satisfy the application requirement of practical steganography. By adopting the advantages of above two SMV construction methods, this strategy can choose the approach of local optimality preservation for every MV adaptively. By utilizing this strategy, the local optimality based distortion can be defined as follows.

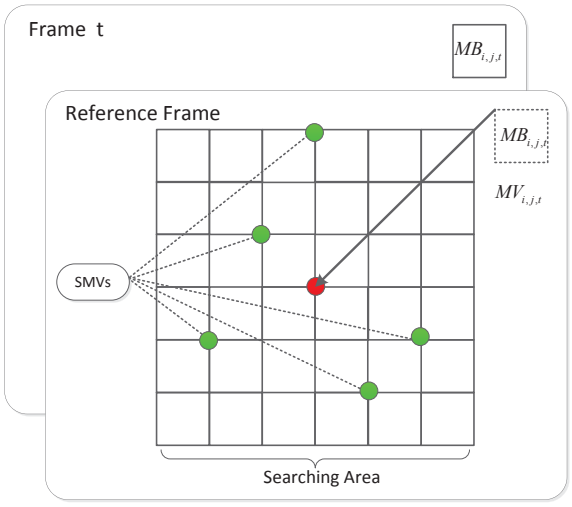


Figure 4: Illustration of SMVs located in the designated searching area.

Table 1: Proportion of MVs with different number of SMVs.

Sequence	Bitrate (Mbps)	1	2	3	4
Akiyo	**0.5**	15.32	21.59	19.02	30.58
	1	17.47	26.47	13.51	22.03
	3	21.68	8.45	5.03	4.76
	10	3.69	0.04	0.01	0.00
Coastguard	**0.5**	2.06	9.01	18.07	70.62
	1	9.02	17.31	24.42	47.03
	3	33.79	19.57	15.79	8.59
	10	23.62	1.03	0.05	0.00
Foreman	**0.5**	4.29	14.21	15.17	64.25
	1	15.09	22.17	14.42	41.15
	3	25.94	15.40	9.10	17.03
	10	19.97	3.63	1.49	2.75

Definition 2. (**Local Optimality Based Distortion, LOD**). The local optimality based distortion of $mv_{i,j,t}$ can be defined by

$$LOD_{i,j,t}=\begin{cases}(\frac{1}{K_n}\sum_{k_n=1}^{K_n}(J_{nmv_{i,j,t}^{k_n}}-J_{mv_{i,j,t}})^2)^{\frac{1}{2}} & if\ K_n\geq 1,\\ (\frac{1}{K_c}\sum_{k_c=1}^{K_c}(J_{cmv_{i,j,t}^{k_c}}-J_{mv_{i,j,t}})^2)^{\frac{1}{2}} & others.\end{cases} \tag{14}$$

where $J_{mv}=sad_{mv}+\lambda\cdot R_{mv}$, which is the Lagrangian cost function in RDO model, λ is the Lagrange parameter, and R_{mv} denotes the bits for coding mv. With distortion function defined by the $LOD_{i,j,t}$, the SMVs for every $LOD_{i,j,t}$ can be found out. Because the proposed strategy adaptively selects the applicable SMV construction method at every location, the local optimality of modified MVs is certain to be preserved. And it is expected that by considering the local optimality of MVs, the steganographic approach can resist the detection of local optimality based steganalysis.

3.3 Statistical Distribution of Motion Vector

Because MVs in clean videos are obtained through ME process and represent the motion information of video content, there exists strong spatial correlation among MVs in each frame and strong temporal correlation between consecutive frames. The basic hypothesis for steganalysis is that some statistical characteristics of cover object can be changed by embedding process. The modifications of the MVs will introduce the changes of spatial-temporal correlation, which are signatures showing the existence of the embedded message [22]. Therefore, in order to reduce the evidence which can be utilized by steganalyzers, MVs' statistical distribution should be considered in designing the distortion function.

Due to motion's relevance and continuity, adjacent MBs have similar motion trends in spatial domain and the MBs of same locations also have same motion trajectories between adjacent frames. Considering both the spatial and temporal consistency, the MVs' changes in MBs with similar motion trends are often more possible to be detected, and thus should be set as high distortion values. Contrariwise modifications tend to be made in MBs with different motion trends, which should be set as low distortion values.

In order to model the MVs' statistical distribution, the horizontal and vertical components of MVs in one frame are separated to construct two $H\times W$ components matrix $\boldsymbol{MVX}_t$ and $\boldsymbol{MVY}_t$, where H and W denotes the height and width of the frame in the unit of block. In spatial domain, we use the second-order difference to calculate the components' statistical distribution in four directions respectively. As shown in Figure 5, denoted by a differential operation ∇, the second-order differences of $\boldsymbol{MVX}_t$ in horizontal, vertical, 45° and −45° direction can be computed by

$$\nabla x^{\rightarrow}_{i,j,t}(\boldsymbol{MVX}_t)=2mvx_{i,j,t}-mvx_{i,j+1,t}-mvx_{i,j-1,t} \tag{15}$$

$$\nabla x^{\uparrow}_{i,j,t}(\boldsymbol{MVX}_t)=2mvx_{i,j,t}-mvx_{i+1,j,t}-mvx_{i-1,j,t} \tag{16}$$

$$\nabla x^{\nearrow}_{i,j,t}(\boldsymbol{MVX}_t)=2mvx_{i,j,t}-mvx_{i+1,j-1,t}-mvx_{i-1,j+1,t} \tag{17}$$

$$\nabla x^{\nwarrow}_{i,j,t}(\boldsymbol{MVX}_t)=2mvx_{i,j,t}-mvx_{i+1,j+1,t}-mvx_{i-1,j-1,t} \tag{18}$$

where $i=2,\ldots,H-1$, $j=2,\ldots,W-1$. Then we use histograms to count the difference of the neighboring triples. The horizontal components' statistical distribution at four different directions could be obtained by

$$H^{\rightarrow}_d(\boldsymbol{MVX}_t)=\sum_{i=2}^{H-1}\sum_{j=2}^{W-1}||\nabla x^{\rightarrow}_{i,j,t}(\boldsymbol{MVX}_t)=d|| \tag{19}$$

$$H^{\uparrow}_d(\boldsymbol{MVX}_t)=\sum_{i=2}^{H-1}\sum_{j=2}^{W-1}||\nabla x^{\uparrow}_{i,j,t}(\boldsymbol{MVX}_t)=d|| \tag{20}$$

$$H^{\nearrow}_d(\boldsymbol{MVX}_t)=\sum_{i=2}^{H-1}\sum_{j=2}^{W-1}||\nabla x^{\nearrow}_{i,j,t}(\boldsymbol{MVX}_t)=d|| \tag{21}$$

$$H^{\nwarrow}_d(\boldsymbol{MVX}_t)=\sum_{i=2}^{H-1}\sum_{j=2}^{W-1}||\nabla x^{\nwarrow}_{i,j,t}(\boldsymbol{MVX}_t)=d|| \tag{22}$$

where $||I||$ equals 1 if I is true and 0 otherwise.

With regard to temporal correlation, the similar operations are applied to MVs of adjacent frames. As illustrated

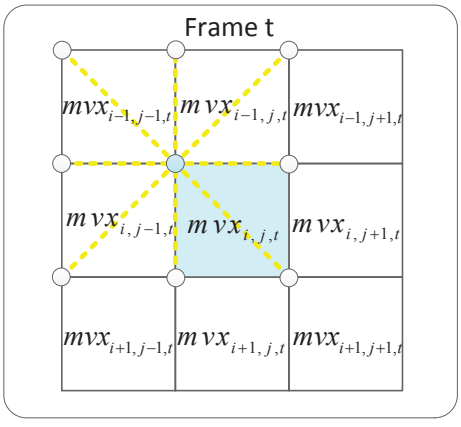


Figure 5: Schematic diagram of computing statistical distribution for spatial correlation.

in Figure 6, the $\boldsymbol{MVX}_t$'s second-order difference in temporal direction and its distribution histogram can be calculated by following formulas.

$$\nabla x^{\dagger}_{i,j,t}(\boldsymbol{MVX}_t) = 2mvx_{i,j,t} - mvx_{i,j,t+1} - mvx_{i,j,t-1} \tag{23}$$

$$H^{\dagger}_{d}(\boldsymbol{MVX}_t) = \sum_{i=2}^{H-1}\sum_{j=2}^{W-1} ||\nabla x^{\dagger}_{i,j,t}(\boldsymbol{MVX}_t) = d|| \tag{24}$$

By analogy, the second-order differences $\nabla x^{\rightarrow}_{i,j,t}(\boldsymbol{MVY}_t)$, $\nabla x^{\uparrow}_{i,j,t}(\boldsymbol{MVY}_t)$,$\nabla x^{\nearrow}_{i,j,t}(\mathbf{MVY}_t)$,$\nabla x^{\nwarrow}_{i,j,t}(\boldsymbol{MVY}_t)$,$\nabla x^{\dagger}_{i,j,t}$ $(\boldsymbol{MVY}_t)$ as well as their distribution histograms $H^{\rightarrow}_{d}(\boldsymbol{MVY}_t)$, $H^{\uparrow}_{d}(\boldsymbol{MVY}_t)$,$H^{\nearrow}_{d}(\boldsymbol{MVY}_t)$, $H^{\nwarrow}_{d}(\boldsymbol{MVY}_t)$,$H^{\dagger}_{d}(\boldsymbol{MVY}_t)$ can also be computed for vertical components $\boldsymbol{MVY}_t$. Based on above analysis, we can define the statistical distribution based distortion as follows.

Definition 3. (**Statistical Distribution Based Distortion, SDD**). The statistical distribution based distortion associated with $mv_{i,j,t}$ can be defined by

$$SDD_{i,j,t} = \sum_{\substack{d\in\{-128,\ldots,128\}\\ f\in\{\rightarrow,\uparrow,\nearrow,\nwarrow,\dagger\}}} \frac{1}{|d|+1}\cdot H^{f}_{t}(\boldsymbol{MV}_t) \tag{25}$$

$$H^{f}_{t}(\boldsymbol{MV}_t) = max(|H^{f}_{t}(\boldsymbol{MVX}_t) - H^{f}_{t}(\boldsymbol{MVX}^{'}_t)|, \\ |H^{f}_{t}(\mathbf{MVY}_t) - H^{f}_{t}(\mathbf{MVY}^{'}_t)|) \tag{26}$$

where $\mathbf{MV}^{'}_t$ denotes the MV field obtained from the tth frame of stego video, the original MV $mv_{i,j,t}$ has been replaced by the modified MV $mv^{'}_{i,j,t}$ in $\mathbf{MV}^{'}_t$ and $mv^{'}_{i,j,t} \in \mathbf{SMV}_{mv_{i,j,t}}$, is one of the SMVs defined in subsection 3.2.

Because the information is randomly embedded in MVs' horizontal or vertical component, the larger value of these two components' statistical differences will be used to define the statistical distribution based distortion for $mv_{i,j,t}$. By defining the weight by d, larger distortion will be obtained for MVs with less differences, which means that the MBs with similar motion trends are not applicable for embedding. And as a result, modifications are made in MVs with various motion trends and the security performance against statistical steganalysis is hoped to be enhanced.

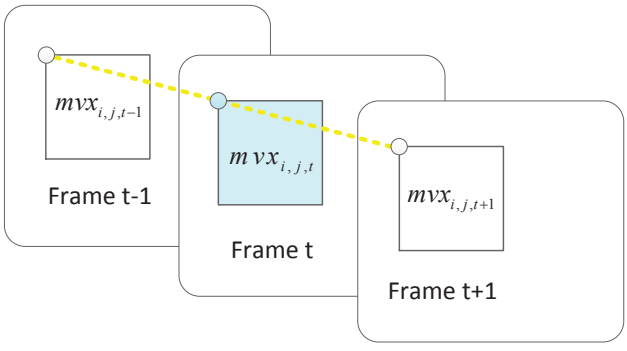


Figure 6: Schematic diagram of computing statistical distribution for temporal correlation.

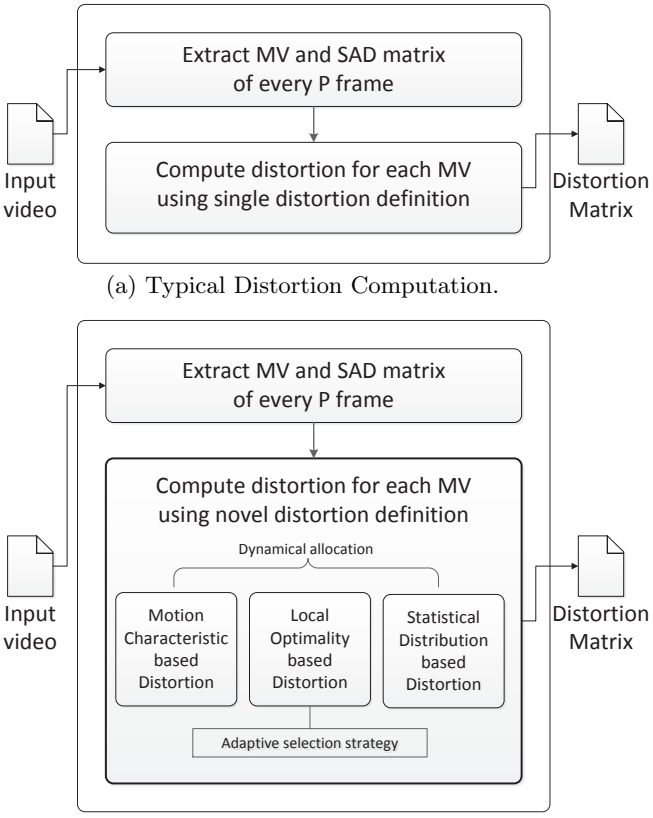


Figure 7: A comparison between typical and novel distortion computation.

3.4 The Definition of Distortion Function

As introduced in Section 2, the embedding impact can be measured by the non-negative additive distortions introduced by independent modifications. And the overall distortion should be minimized by steganographic codes. Thus the overall distortion of all the $mv_{i,j,t}$ in tth frame can be calculated by

$$D(\boldsymbol{MV}_t, \boldsymbol{MV}^{'}_t) = \sum_{i=1}^{H}\sum_{j=1}^{W} \Phi_{i,j,t}(mv_{i,j,t}, mv^{'}_{i,j,t}) \tag{27}$$

where H and W represents the number of block in frame's height and width respectively, $\boldsymbol{MV}_t = \{mv_{i,j,t}|i \in \{1,\ldots,H\}, j \in \{1,\ldots,W\}\}$, which is the original MV set, $\boldsymbol{MV}^{'}_t = \{mv^{'}_{i,j,t}|i \in \{1,\ldots,H\}, j \in \{1,\ldots,W\}\}$, which is the modified MV set. The variable $\Phi_{i,j,t}$ denotes the distortion function of $mv_{i,j,t}$, which can fully represent the modification's impact by considering the $MCD_{i,j,t}$, $LOD_{i,j,t}$ and $SDD_{i,j,t}$.

Definition 4. (**The Distortion Function**). The distortion of $mv_{i,j,t}$ can be defined by

Table 2: Coding performance, including achieved bit-rate (achieved BR, *kbps*) and PSNR (*dB*) under given compressed bit-rate (compressed BR, *kbps*) and embedding rate (ER, *bpmv*).

Sequence	Method	Compressed BR	500		1000		3000		10000	
		ER	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.5
Bus.yuv	STC	Achieved BR	489.6		983.52		2948.40		9960.25	
		PSNR	30.64		34.17		40.66		50.79	
	Ours	Achieved BR	489.6	490.32	983.76	983.28	2951.52	2956.08	9959.53	9964.09
		PSNR	30.51	30.3	34.05	33.81	40.58	40.39	50.73	50.64
	Cao's	Achieved BR	488.88	488.88	982.56	982.32	2948.4	2949.36	9959.29	9955.45
		PSNR	30.24	30.07	33.35	33.17	40.28	40.1	50.42	50.33
	Yao's	Achieved BR	489.12	490.32	982.8	984	2950.8	2951.04	9959.53	9973.45
		PSNR	29.48	29.37	33.16	32.95	39.63	39.65	50.23	50.17
Mobile.yuv	STC	Achieved BR	495.6		974.16		2873.76		9858.73	
		PSNR	28.15		30.95		36.73		48.3	
	Ours	Achieved BR	497.04	498.72	977.76	981.60	2882.4	2897.76	9880.90	9910.09
		PSNR	27.99	27.71	30.82	30.56	36.62	36.39	48.22	48.06
	Cao's	Achieved BR	495.84	497.04	974.64	978.24	2876.16	2882.16	9870.49	9892.33
		PSNR	27.4	27.23	30.09	29.81	36.02	35.86	47.74	47.33
	Yao's	Achieved BR	497.52	499.92	979.44	987.84	2887.2	2914.32	9888.25	9957.37
		PSNR	26.39	26.25	29.24	29.12	35.19	35.08	47.14	47.03
Stefan.yuv	STC	Achieved BR	486		991.92		3046.08		10302.01	
		PSNR	31.24		34.58		40.71		50.43	
	Ours	Achieved BR	486	487.68	994.8	996.24	3047.52	3042.24	10322.41	10322.97
		PSNR	31.04	30.72	34.4	34.1	40.63	40.44	50.41	50.34
	Cao's	Achieved BR	486.24	486.24	993.12	992.4	3043.92	3048.72	10305.61	10311.13
		PSNR	29.95	29.72	34.12	33.98	40.18	40.11	50.2	49.99
	Yao's	Achieved BR	487.92	489.36	997.44	997.68	3049.92	3048.96	10329.61	10349.77
		PSNR	29.65	29.59	33.28	33.09	39.67	39.51	50.09	50.02

$$\Phi_{i,j,t}(mv_{i,j,t}, mv^{'}_{i,j,t}) = WMCD_{i,j,t} \cdot WLOD_{i,j,t} \cdot WSDD_{i,j,t} \tag{28}$$

$$WMCD_{i,j,t} = MCD_{i,j,t}^{\beta_{MCD_{i,j,t}}} \tag{29}$$

$$WLOD_{i,j,t} = (LOD_{i,j,t} + \alpha_{LOD_{i,j,t}})^{\beta_{LOD_{i,j,t}}} \tag{30}$$

$$WSDD_{i,j,t} = (SDD_{i,j,t} + \alpha_{SDD_{i,j,t}})^{\beta_{SDD_{i,j,t}}} \tag{31}$$

where $WMCD_{i,j,t}$, $WLOD_{i,j,t}$ and $WSDD_{i,j,t}$ denotes the weighted distortion of $MCD_{i,j,t}$, $LOD_{i,j,t}$, and $SDD_{i,j,t}$ respectively. The parameter $\alpha_{LOD_{i,j,t}}$ and $\alpha_{SDD_{i,j,t}}$ can be selected as relatively small positive constants to ensure the embedding distortion keeps positive. Every element in $\{\beta_{MCD_{i,j,t}}, \beta_{LOD_{i,j,t}}, \beta_{SDD_{i,j,t}}\}$ is used to allocate the contributions of these three sub-distortions, which are dynamically set as follows.

$$\beta_{MCD_{i,j,t}} = \frac{1}{|\boldsymbol{MV}_{i,j,t}| + 1} \tag{32}$$

$$\beta_{LOD_{i,j,t}} = \frac{1}{K} \tag{33}$$

$$\beta_{SDD_{i,j,t}} = \sum H_t^f(\boldsymbol{MV}_t) \tag{34}$$

where $|\boldsymbol{MV}_{i,j,t}|$ is $mv_{i,j,t}$'s magnitude, K is the number of SMVs and $H_t^f(\boldsymbol{MV}_t)$ is the statistical distribution difference. This method adaptively adjusts the tradeoff between the sub-distortions and focuses on the embedding impact on fragile characteristic. Therefore, the distortion fully considering multi-level embedding influence can be adaptively defined in our proposed approach.

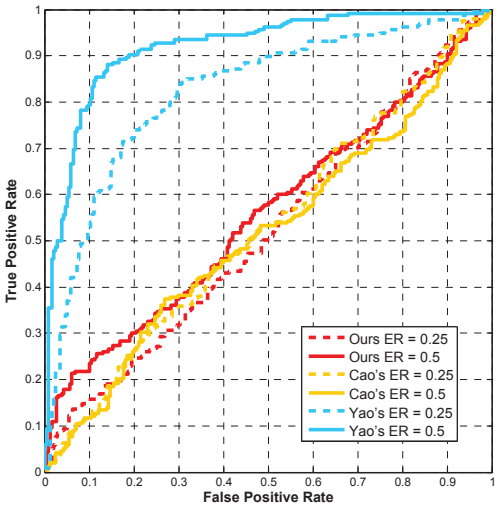
The comparison between typical and novel distortion computation is shown in Figure 7. In order to take advantage of the motion information for modification, the impact of motion characteristic is introduced in designing the distortion. By adopting the adaptive selection strategy, the local optimality of modified MV can be better preserved in our proposed method. With the hope of resisting statistical analysis, we also consider the change of statistical distribution before and after embedding. Thus comparing with traditional distortion defined only considering single embedding impact, the performance of our approach is hoped to be improved by using the novel distortion definition.

4. PRACTICAL IMPLEMENTATION

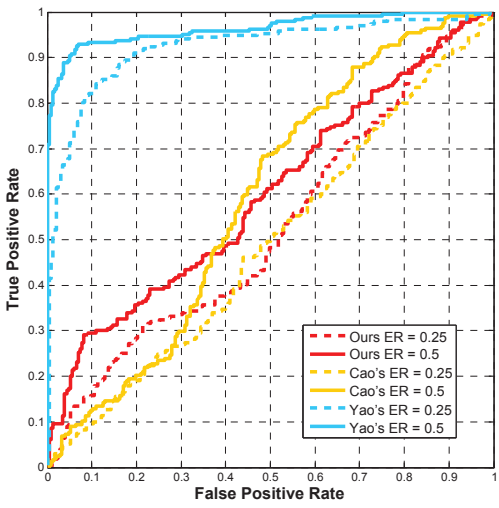
Based on the given distortion function, the practical implementation of proposed steganographic method using two-layered STCs is to be introduced in this section. In our proposed steganographic method, there are three phases for completing data embedding and extraction scheme.

Embedding Distortion Definition: For every frame $\boldsymbol{F}_t$ in the N frames, if it is not I frame, obtain its MV matrix $\boldsymbol{MV}_t$ and the corresponding prediction error matrix $\boldsymbol{E}_t$ through ME process. Then for each $MB_{i,j}$ in this frame, define the distortions using $\Phi_{i,j,t}(mv_{i,j,t}, mv^{'}_{i,j,t})$ in Eq.28 and store all the defined distortion in $\boldsymbol{D}_t = \{\Phi_{i,j,t}(mv_{i,j,t}, mv^{'}_{i,j,t})\}$.

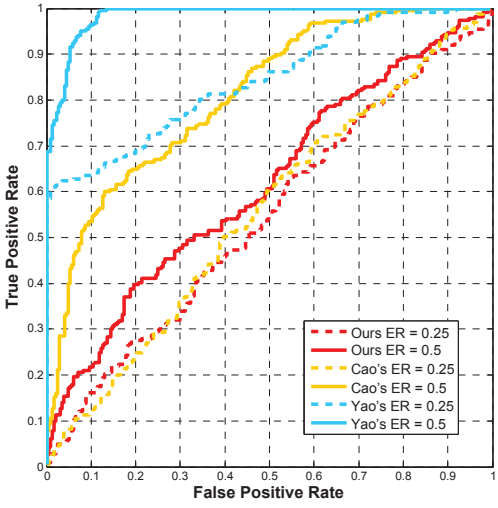
Data Embedding: If the length of the binary message sequence is l and the number of P frames is N_p, the cover length can be calculated by $n = H \times W \times N_p \times 2$. The embedding rate is $\gamma = l/n$, which is measured by the average embedded bits per motion vector (*bpmv*). The first and second embedding channels are constructed using the LSB layer and second LSB layer of $(mvx_{i,j,t} + mvy_{i,j,t})$. With the binary message sequence $\boldsymbol{m}$, the $\boldsymbol{MV}_t$ and the distortions $\boldsymbol{D}_t$, embed data in $LSB((mvx_{i,j,t} + mvy_{i,j,t}))$ and $LSB(\lfloor (mvx_{i,j,t} + mvy_{i,j,t})/2 \rfloor)$ using ± 1 two-layered STCs



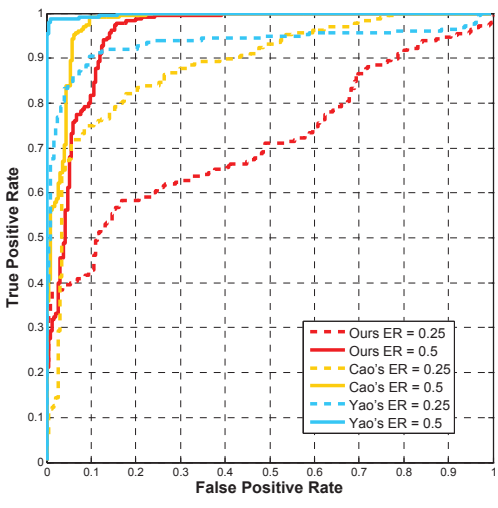
(a)BR = 500$kbps$



(b)BR = 1000$kbps$



(c)BR = 3000$kbps$



(d)BR = 10000$kbps$

Figure 8: ROC curves against AoSO.

with embedding rate $\gamma/2$. After embedding data for every $mv_{i,j,t}$ in N frames, encode frame $\boldsymbol{F}_t$ using the modified MV matrix $\boldsymbol{MV}_t^{'}$. As the result, the coded bit stream can be generated.

Data Extraction: Decode every received frame to obtain the MV matrix $\boldsymbol{MV}_t^{'}$. Then use the STCs decoding to extract the binary message sequence.

5. EXPERIMENTS

5.1 Experimental Setup

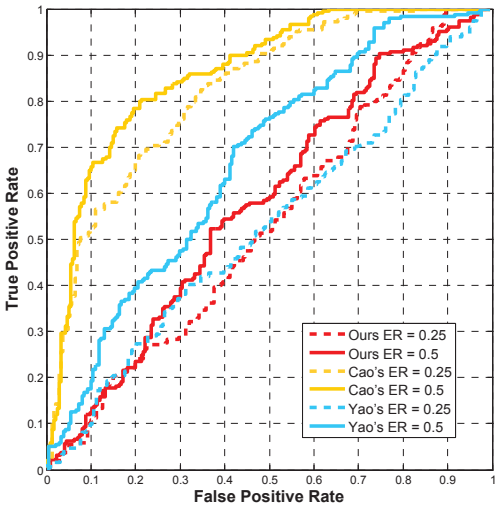
Our proposed embedding scheme is implemented on a well-known H.264/AVC codec named x264 [23]. The video database is composed of 30 standard 4:2:0 YUV sequences in CIF format. The raw sequences vary from 150 to 300 frames in length and are coded with 30fps frame rate.

In order to evaluate the security of our approach under different cases, various bit-rate (BR) including 500$kbps$, 1000$kbps$, 3000$kbps$ and 10000$kbps$ are considered with the achieved embedding rate (ER) 0.25$bpmv$ and 0.5$bpmv$ respectively. Beside, Cao's [12] and Yao's [8] methods are also implemented for performance comparisons. And the LibSVM toolbox [24] with the Gaussian kernel is used as classifier.

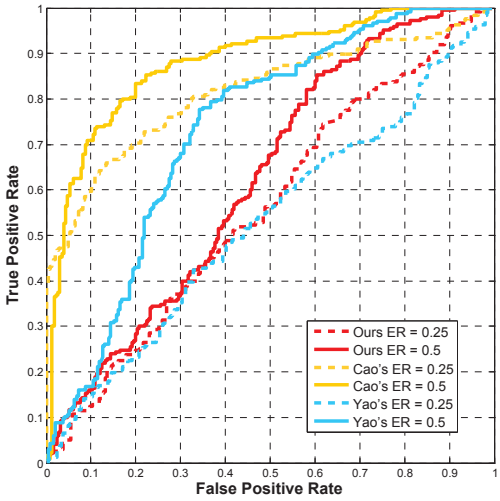
In our tests, stego and clean videos are generated by com-

Table 3: Average detection accuracies (%) against AoSO

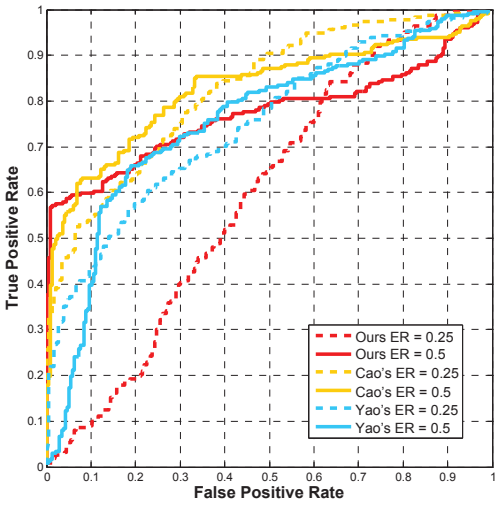
BR (*kbps*)	**500**		**1000**		**3000**		**10000**	
ER (*bpmv*)	**0.25**	**0.5**	**0.25**	**0.5**	**0.25**	**0.5**	**0.25**	**0.5**
Ours	50.78	54.20	50.71	56.43	52.23	56.40	63.05	88.81
Cao's	52.30	52.66	50.87	57.85	54.60	70.89	82.43	94.39
Yao's	76.27	87.44	85.71	92.00	73.78	93.03	90.26	98.07



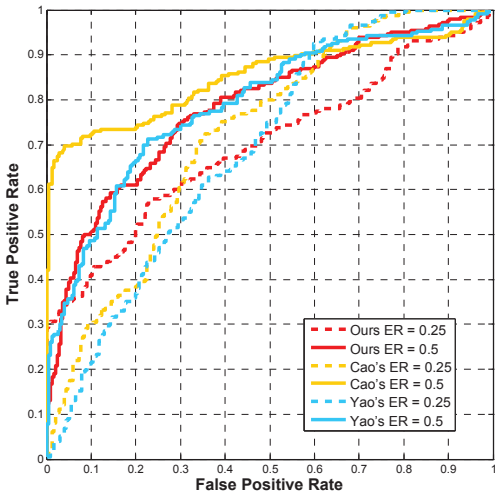
(a)BR = 500$kbps$



(b)BR = 1000$kbps$



(c)BR = 3000$kbps$



(d)BR = 10000$kbps$

Figure 9: ROC curves against MVRBR.

pressing 30 sequences at various bit-rates with and without embedding. Subsequently, AoSO [10] and MVRBR [14] features are extracted out from every 12 frames. Then features are fed into classifier, 60 percent sequence pairs (stego and cover) are randomly selected for training and the remaining ones for testing. Each training and testing is repeated 50 times and average results are used to evaluate the final performance. The true negative (TN) rate and the true positive (TP) rate are computed by counting the number of detections in the test sets. By averaging TN and TP, averaged detection accuracies are obtained.

Table 4: Average detection accuracies (%) against MVRBR

BR (*kbps*)	**500**		**1000**		**3000**		**10000**	
ER (*bpmv*)	**0.25**	**0.5**	**0.25**	**0.5**	**0.25**	**0.5**	**0.25**	**0.5**
Ours	51.75	56.79	53.50	57.04	56.95	71.51	64.13	71.94
Cao's	72.79	79.78	73.91	80.21	72.32	76.18	67.73	77.82
Yao's	52.21	61.39	53.30	69.52	68.43	70.93	63.00	72.22

5.2 Results and Discussion

5.2.1 Coding Efficiency

One of the conspicuous advantage of MV-based steganographic methods is that they do not affect the coding efficiency (PSNR and bit-rate) much. The error caused by modifying MV will be handled by the mechanism of ME and MC. For example, comparison coding results of three sequences, i.e., "Bus.yuv", "Mobile.yuv" and "Stefan.yuv" are listed in Table 2, where STC denotes the standard compression without embedding. It can be seen that both the video quality and the compression ratio are affected sightly for

all methods, and the embedding impact of our approach is smallest.

5.2.2 Steganalysis

For security evaluation, the recent proposed AoSO [10] and MVRBR [14] feature are leveraged against our proposed, Cao's and Yao's steganographic schemes.

The detection accuracies against AoSO feature are recorded in Table 3 and the corresponding receiver operation characteristic (ROC) curves are depicted in Figure 8. In comparison with Ours and Cao's methods, Yao's method performs worst in all tested cases due to local optimality unconsidered. Cao's method performs well at low bit-rate i.e, $500kbps$ in Figure 8(a) and $1000kbps$ in Figure 8(b). However, the performance deteriorates when bit-rate increases, which is shown by Figure 8(c)(d). It is because that less loss is induced by compression in higher quality videos. As the result, the number of SMVs obtained from lossy-compression reduces and the modifications have to be made on MVs without local optimality. The performance of our method is quite close to the average performance of Cao's at low bit-rate. With bit-rate increasing, our approach works better than Cao's method. Despite deterioration in performance under high bit-rate cases, our approach still enhances the security against AoSO obviously, which benefits from the adaptive selection strategy. Thus our approach generally performs best at different bit-rates, especially in the cases of high bit-rate.

When detected by MVRBR feature, the average detection accuracies are presented in Table 4. It can be seen from Figure 9 that Cao's method performs worst because of statistical distribution not included. The performance of Yao's is better than Cao's but inferior to ours. By adopting dynamical adjustment between three factors in defining distortion scale, our approach works best in these three methods. Both of Figure 8 and Figure 9 indicate that the steganalytic performances against all above steganographic methods are related to video quality. This is mainly because that AoSO and MVRBR features are extracted using MVs and SADs. The SADs are calculated from prediction residuals which are compressed in lossy way. Since the distortion of prediction residuals is less if video has higher quality, AoSO and MVRBR features perform better when bit-rate is higher.

Overall, in contrast with Cao's and Yao's methods, the detection accuracies and the ROC curves indicate that our proposed method reduces the probability of detection. This implies that the steganographic security of MV-based steganography can be enhanced by utilizing the novel distortion function, which dynamically adjusts the contributions of sub-distortions considering three different factors.

6. CONCLUSIONS

In this paper, a novel embedding distortion for MV-based steganography is proposed. Three important factors are simultaneously considered in distortion definition, which are motion characteristic of video content, MV's local optimality and statistical distribution. And dynamical allocation is implementation between these three distortion components. The practical embedding algorithm is implemented using two-layered STCs. In experiments against current effective steganalytic methods, the proposed approach outperforms other existing MV-based steganographic methods in security, especially for high quality videos.

As part of our future work, more impact of MVs' modification and optimization of distortion definition is to be further exploited. Moreover, method to reduce the heavily computation of distortion is also attempted in practical implementation.

7. ACKNOWLEDGMENTS

This work was supported by the NSFC under 61303259 and U1536105, the Strategic Priority Research Program of Chinese Academy of Sciences (CAS) under XDA06030600, and the Key Project of Institute of Information Engineering, CAS, under Y5Z0131201.

8. REFERENCES

[1] F. JORDAN. Proposal of a watermarking technique for hiding/retrieving data in compressed and decompressed video. *ISO/IEC Doc. JTC1/SC 29/QWG 11 MPEG 97/M 2281*, 1997.

[2] Changyong Xu, Xijian Ping, and Tao Zhang. Steganography in compressed video stream. In *Innovative Computing, Information and Control, 2006. ICICIC '06. First International Conference on*, volume 1, pages 269–272, Aug 2006.

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

[4] J. Fridrich, M. Goljan, P. Lisonek, and D. Soukal. Writing on wet paper. *Signal Processing, IEEE Transactions on*, 53(10):3923–3935, Oct 2005.

[5] TomÃąÅą Filler, Jan Judas, and Jessica Fridrich. Minimizing embedding impact in steganography using trellis-coded quantization, 2010.

[6] T. Filler, J. Judas, and J. Fridrich. Minimizing additive distortion in steganography using syndrome-trellis codes. *Information Forensics and Security, IEEE Transactions on*, 6(3):920–935, Sept 2011.

[7] Yun Cao, Xianfeng Zhao, Dengguo Feng, and Rennong Sheng. *Information Hiding: 13th International Conference, IH 2011, Prague, Czech Republic, May 18-20, 2011, Revised Selected Papers*, chapter Video Steganography with Perturbed Motion Estimation, pages 193–207. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[8] Yuanzhi Yao, Weiming Zhang, Nenghai Yu, and Xianfeng Zhao. Defining embedding distortion for motion vector-based video steganography. *Multimedia Tools and Applications*, 74(24):11163–11186, 2014.

[9] Yanzhen Ren, Liming Zhai, Lina Wang, and Tingting Zhu. Video steganalysis based on subtractive probability of optimal matching feature. In *Proceedings of the 2Nd ACM Workshop on Information Hiding and Multimedia Security*, IH&MMSec '14, pages 83–90, New York, NY, USA, 2014. ACM.

[10] Keren Wang, Hong Zhao, and Hongxia Wang. Video steganalysis against motion vector-based steganography by adding or subtracting one motion vector value. *Information Forensics and Security, IEEE Transactions on*, 9(5):741–751, May 2014.

[11] Hong Zhang, Yun Cao, and Xianfeng Zhao. Motion vector-based video steganography with preserved local

optimality. *Multimedia Tools and Applications*, pages 1–17, 2015.

[12] Yun Cao, Hong Zhang, Xianfeng Zhao, and Haibo Yu. Video steganography based on optimized motion estimation perturbation. In *Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security*, IH&MMSec '15, pages 25–31, New York, NY, USA, 2015. ACM.

[13] Yun Cao, Xianfeng Zhao, and Dengguo Feng. Video steganalysis exploiting motion vector reversion-based features. *Signal Processing Letters, IEEE*, 19(1):35–38, Jan 2012.

[14] Peipei Wang, Yun Cao, Xianfeng Zhao, and Bin Wu. Motion vector reversion-based steganalysis revisited. In *Signal and Information Processing (ChinaSIP), 2015 IEEE China Summit and International Conference on*, pages 463–467, July 2015.

[15] Tomáš Pevný, Tomáš Filler, and Patrick Bas. *Information Hiding: 12th International Conference, IH 2010, Calgary, AB, Canada, June 28-30, 2010, Revised Selected Papers*, chapter Using High-Dimensional Image Models to Perform Highly Undetectable Steganography, pages 161–177. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[16] V. Holub and J. Fridrich. Designing steganographic distortion using directional filters. In *Information Forensics and Security (WIFS), 2012 IEEE International Workshop on*, pages 234–239, Dec 2012.

[17] Danfeng Xie, Zhiwei Huang, Shizheng Wang, and Heguang Liu. Moving objects segmentation from compressed surveillance video based on motion estimation. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 3132–3135, Nov 2012.

[18] Wei Zeng, Jun Du, Wen Gao, and Qingming Huang. Robust moving object segmentation on h.264/avc compressed video using the block-based {MRF} model. *Real-Time Imaging*, 11(4):290 – 299, 2005.

[19] R. Venkatesh Babu and K.R. Ramakrishnan. Recognition of human actions using motion history information extracted from the compressed video. *Image and Vision Computing*, 22(8):597 – 607, 2004.

[20] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra. Overview of the h.264/avc video coding standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(7):560–576, July 2003.

[21] M. Tom and R.V. Babu. Fast moving-object detection in h.264/avc compressed domain for video surveillance. In *Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2013 Fourth National Conference on*, pages 1–4, Dec 2013.

[22] Yuting Su, Chengqian Zhang, and Chuntian Zhang. A video steganalytic algorithm against motion-vector-based steganography. *Signal Processing*, 91(8):1901 – 1909, 2011.

[23] VideoLan. x264. Available: http://www.videolan.org/developers/x264.html.

[24] C.Chang and C.Lin. LIBSVM: A Library for Support Vector Machines, 2001 [online]. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm.