Combined and Calibrated Features for Steganalysis of Motion Vector-Based Steganography in H.264/AVC

Liming Zhai
School of Computer,
Wuhan University
Wuhan 430072, China
limingzhai@whu.edu.cn

Lina Wang
School of Computer,
Wuhan University
Wuhan 430072, China
lnwang@whu.edu.cn

Yanzhen Ren
School of Computer,
Wuhan University
Wuhan 430072, China
renyz@whu.edu.cn

ABSTRACT

This paper presents a novel feature set for steganalysis of motion vector-based steganography in H.264/AVC. First, the influence of steganographic embedding on the sum of absolute difference (SAD) and the motion vector difference (MVD) is analyzed, and then the statistical characteristics of these two aspects are combined to design features. In terms of SAD, the macroblock partition modes are used to measure the quantization distortion, and by using the optimality of SAD in neighborhood, the partition based neighborhood optimal probability features are extracted. In terms of MVD, it has been proved that MVD is better in feature construction than neighboring motion vector difference (NMVD) which has been widely used by traditional steganalyzers, and thus the inter and intra co-occurrence features are constructed based on the distribution of two components of neighboring MVDs and the distribution of two components of the same MVD. Finally, the combined features are enhanced by window optimal calibration, which utilizes the optimality of both SAD and MVD in a local window area. Experiments on various conditions demonstrate that the proposed scheme generally achieves a more accurate detection than current methods especially for videos encoded in variable block size and high quantization parameter values, and exhibits strong universality in applications.

KEYWORDS

Video steganalysis; sum of absolute difference (SAD); partition modes; motion vector difference (MVD); calibration; combined and calibrated features (CCF)

1 INTRODUCTION

Steganalysis is the counter measure to steganography. Its main purpose is to determine if there is a secret message hidden in digital media such as image, video and audio. Most current steganalysis focuses on image steganography, while the steganalysis for digital video has received relatively limited attention. With the popularity of video capture devices and internet video applications, digital video has become a readily available information hiding carrier. Besides, the video volume is usually large and can provide enough space for secret messages. More seriously, the digital video, especially the compressed video, has rich and varied components, which are favorable to design various steganographic methods, such as motion vector (MV) [1, 3, 6, 10, 27, 29], transformed coefficients [14, 19], prediction modes [9, 28], partition modes [11, 30] and entropy encoded bitstream [13, 18]. As a result, steganography tools or algorithms based on digital video have been gradually increasing recently, and they pose a severe challenge to video steganalysis.

Among all video steganography, the MV based steganography in H.264/AVC is chosen as the target for steganalysis for two reasons: First, MV based steganography is prevailing owing to its security and high embedding capacity [29]. Second, H.264/AVC is currently the most widely used video coding standard, and is likely to be the carrier of video steganography in many practical applications [22].

The MV based steganography is usually accomplished by modifying the MVs and adjusting the corresponding prediction errors (PEs) simultaneously. There are many MV based steganography using different ideologies. Some early methods use predefined selection rules (SRs) to select candidate MVs for embedding. Jordan [10] firstly proposed a MV based embedding method by replacing the least significant bits (LSBs) of the horizontal and vertical components of all MVs with secret message bits. Xu [27] proposed to select the MVs that satisfy a certain threshold, and modify the MV components with larger magnitude to embed messages. In [6], Fang suggested hiding messages by the phase angle between two components of MVs. However, the fixed SRs above will bring some potential risks, so the steganography later introduces adaptive mechanisms or steganographic codes to improve the steganographic security. Aly [1] designed an adaptive PE threshold selection scheme, and embedded secret messages into both components of MVs associated with larger PEs. A steganographic method with perturbed motion estimation was presented in [3], where the embedding is implemented by incorporating wet paper code (WPC). Moreover, Zhang [29] proposed to modify MVs with preserved local optimality, and syndrome-trellis code (STC) is employed to minimize the embedding distortion.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
IH&MMSec '17, June 20-22, 2017, Philadelphia, PA, USA
© 2017 ACM. ISBN 978-1-4503-5061-7/17/06...$15.00
DOI: http://dx.doi.org/10.1145/3082031.3083237

To detect the MV based steganography, some feature-based steganalytic methods have been proposed in recent years, and they can be approximately divided into three types. The first type of steganalytic methods constructs features based on neighboring motion vector difference (NMVD). Su [20] proposed to use the features derived from the statistical characteristics of NMVDs in spatial and temporal domain. Wu [26] used the joint distribution of the NMVDs between one macroblock (MB) and the other two MBs as features. The NMVD, which is obtained by the subtraction between two neighboring MVs, is the key point for the construction of these features. However, in H.264/AVC and other advanced video coding standards, when two neighboring MBs have different partition modes or neighboring MV does not exist, it is difficult to calculate the NMVD.

The second type of steganalytic methods uses the statistics of sum of absolute differences (SADs) to design features. In Wang's work [24], a feature set called AoSO (Add-or-Subtract-One), which bases on the assumption that the local optimality of SAD will be changed if the corresponding MV is modified, was proposed for steganalysis. Ren [16] also proposed a SPOM (Subtractive Probability of Optimal Matching) feature by using the local optimal SAD. The ideologies of AoSO and SPOM are similar, but they all face the same problem: When the quantization parameter (QP) value is high or the distribution of PE is smooth, the stability of the local optimality of SAD will decrease due to the quantization distortion, thus leading to deteriorated detection performance. Also, for some steganography that considers preserving the statistics of SADs during embedding [3, 29], the performance of AoSO and SPOM are also not ideal.

The third type of steganalytic methods uses calibrations to enhance the features. Cao [2] presented a recompressed calibration method, and then the motion vector reversion-based (MVRB) features were drawn from the difference of MVs and SADs before and after calibration. However, the coding parameters for two compression processes need to be the same, otherwise the detection performance will deteriorate. Deng [5] proposed another calibration using local polynomial kernel regression model (LPKRM) to recover original MVs. The LPKRM depends on the correlation among neighboring MVs; however, like NMVD features [20, 26], the locations and existence of neighboring MVs were also not considered.

To sum up, the current steganalytic features are mainly designed from MVs/NMVDs and PEs/SADs, or supplemented by certain calibration techniques, but they all have their own limitations. In addition, MV based steganography simultaneously modifies the MVs and PEs, both statistics of which are disturbed to some extent. However, apart from [2], all the features above are derived either from MVs or from PEs separately. It is necessary that both statistics of MVs and PEs should be combined for steganalysis.

In this paper, we propose a feature set, named combined and calibrated features (CCF), to detect MV based steganography. The CCF is derived from the statistical characteristics of SADs and motion vector differences (MVDs) and further improved by calibration. The rationale are as follows: For SAD features, we found that the quantization distortion of SAD is often associated with partition modes, which can be used to measure the quantization distortion. Moreover, the steganographic embedding not only changes the local optimality of SAD, but also the neighborhood optimality. So the partition based neighborhood optimal probability (PB-NOP) features are extracted. For MVD features, we found that there exists an inter distribution of components of neighboring MVDs and an intra distribution of components of the same MVD. These two distributions of MVDs are more sensitive to embedding and more tractable to feature design than those of NMVDs. Then the inter and intra co-occurrence (IIC) features of MVD are constructed as a new statistical representation based on the above two distributions. For calibration, a new calibration method called window optimal calibration (WOC) is proposed by using the optimality of both SAD and MVD in a local window area.

The contribution of this paper is three fold: First, we propose a partition based quantization method, and the influence of quantization distortion on SAD features is greatly reduced. Second, we for the first time propose to construct features based on MVD, which is superior to the traditional NMVD. Third, we propose a highly universal calibration that can be applied to various coding conditions without any restriction.

This paper is organized as follows. Section 2 introduces the encoding process of H.264/AVC and then describes the effects of MV based steganography on video statistics. Section 3 elaborates the relation between quantization distortion and partition modes, followed by the construction of PB-NOP features. The subsequent Section 4 contrasts the statistical characteristics of MVDs and NMVDs, and then presents the IIC features. The description for WOC appears in Section 5. In Section 6, the experimental results are given. Finally, the paper is concluded in Section 7.

Throughout the paper, calligraphic font is reserved for sets. For a finite set $\mathcal{X}$, $|\mathcal{X}|$ denotes the cardinality. Macroblock (MB) specifically refers to the block of size 16×16. "Block" and "partition" have the same meaning and are used in different contexts. "Block" generally refers to the blocks of various sizes. "Partition" refers to the subblocks of MB, it is used to emphasize the division of a MB.

2 PRELIMINARIES

2.1 Basics of H.264/AVC

The H.264/AVC adopts a block-based predictive/transform coding model. For an inter block of size $M\times N$ in the current frame, it searches a similar and best matching region in the encoded reference frame through a block matching algorithm. The best matching region is called reference block, and the offset from the current block to the reference block is called motion vector (MV). This process is known as motion estimation (ME), which reduces temporal redundancy between neighboring frames. The reference block is subtracted from the current block to form a residual block (i.e., PE), and then the PE is subjected to transform and quantization to

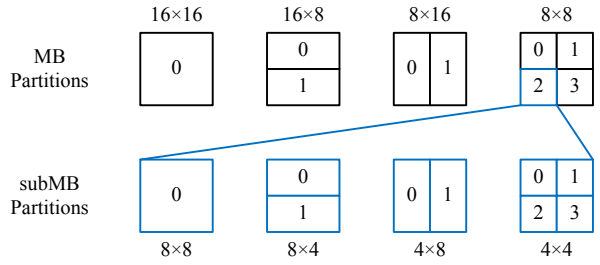

Figure 1: MB partitions and subMB partitions.

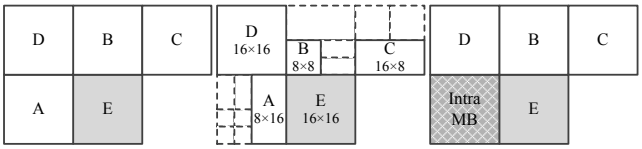

Figure 2: Current and neighboring partitions. Left: Same partition mode. Middle: Different partition modes. Right: Discontinuous inter-partitions.

further reduce spatial redundancy. Finally, the transformed and quantized PE together with motion vector difference (MVD) is entropy encoded to reduce statistical redundancy. To make this paper self-contained, the following concepts on inter coding need to be further introduced.

2.1.1 Partition Modes. The size of a MB is fixed at 16×16. In order to achieve a more accurate ME, the MB is often divided into some subblocks, and this is so called variable block size (VBS); while the opposite case is fixed block size (FBS). For the VBS, a MB can be divided into one 16×16 partition, two 16×8 partitions, two 8×16 partitions, or four 8×8 partitions; they are all called MB partitions. The 8×8 partition, also named subMB, can be further divided into one 8×8 partition, two 8×4 partitions, two 4×8 partitions, or four 4×4 partitions; they are all called subMB partitions. Figure 1 illustrates the partitioning of a MB.

2.1.2 Motion vector prediction and Motion vector difference. The motion of neighboring blocks is often similar, so the MV of neighboring block is served as a starting point for ME. This starting MV is called predicted motion vector (PMV), whose forming depends on the partition modes of neighboring partitions and the availability of neighboring MVs.

Let E be the current partition, and A, B, C, and D be the left, top, top-right and top-left partition immediately next to E respectively. Figure 2 shows examples of the positions of A, B, C, D and E in different cases. In general, the PMV of E is the median of the MVs of A, B and C. If the MV of C does not exist (the partition beyond the frame boundary or belongs to an intra MB), the MV of D is taken instead. If other neighboring MVs do not exist, the forming of PMV will change accordingly [21].

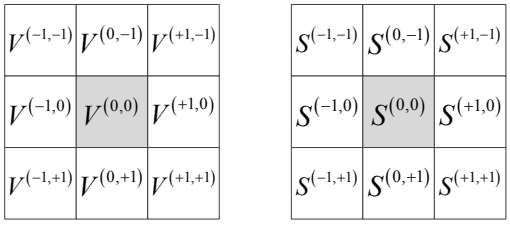

Figure 3: The structures of $\mathcal{V}_l$ (left) and the corresponding $\mathcal{S}_l$ (right) in 3×3 window.

For saving bitstream, the motion vector difference (MVD) instead of MV will be entropy encoded and transmitted. MVD is the difference between MV and PMV defined by

$$\bar{D} = V - P \tag{1}$$

where $\bar{D}$, V, and P denote the MVD, MV, and PMV respectively.

2.1.3 Block Matching Criterion. The ME locates the best matching block in the reference frame using a rate-distortion criterion [17]: minimizing the loss of video quality under the constraint of bit rate. This criterion is usually fulfilled by minimizing the following Lagrangian cost function:

$$J = S + \lambda \cdot B(\bar{D}) \tag{2}$$

where J is the Lagrangian cost; S denotes the SAD; $\bar{D}$ is the MVD calculated by (1); $B(\bar{D})$ refers to the bits required for coding $\bar{D}$; λ is the Lagrangian multiplier, whose recommended value is obtained by the following expression [25]:

$$\lambda = \sqrt{0.85 \times 2^{(Q-12)/3}} \tag{3}$$

where Q stands for quantization parameter (QP).

2.2 Effects of MV Based Steganography on Video Statistics

According to current MV based steganography [1, 3, 6, 10, 27, 29], one or two components of a MV can be modified during emmbeding. Let $V_l = \left(V_l^h, V_l^v\right)$ be the l-th MV in a cover video frame, V_l^h and V_l^v are the horizontal component and vertical component of V_l. For LSB embedding, the possible variations of V_l will form a 3×3 window area

$$\mathcal{W}_9 = \{(\Delta h, \Delta v)\,|\Delta h, \Delta v = -1, 0, +1\} \tag{4}$$

where Δh and Δv denote the modification amplitude of V_l^h and V_l^v respectively.

Focusing only on the actual modifications of V_l, namely $\Delta h, \Delta v = \pm 1$, a local 8-neighborhood will be defined as

$$\mathcal{N}_8 = \mathcal{W}_9 \backslash \{(0,0)\} \tag{5}$$

where $(0,0) \in \mathcal{W}_9$ means no modification of V_l.

The $\mathcal{W}_9$ corresponds to a set consisting of possible MVs as follows

$$\mathcal{V}_l = \left\{ V_l^{(\Delta h, \Delta v)} \middle| V_l^{(\Delta h, \Delta v)} = \left(V_l^h + \Delta h,\ V_l^v + \Delta v\right), \right. \\ \left. (\Delta h, \Delta v) \in \mathcal{W}_9 \right\} \tag{6}$$

among which $V_l^{(0,0)}$ denotes the original MV, and $V_l^{(\Delta h,\Delta v)}$, $(\Delta h, \Delta v) \in \mathcal{N}_8$ refers to the modified MVs. During embedding, not only the MVs but also the PEs are modified, so $\mathcal{W}_9$ corresponds to a set $\mathcal{S}_l = \left\{ S_l^{(\Delta h,\Delta v)} \right\}$ for possible SADs as well ($S_l^{(\Delta h,\Delta v)}$ denotes the SAD corresponding to $V_l^{(\Delta h,\Delta v)}$). The structures of $\mathcal{V}_l$ and $\mathcal{S}_l$ are shown in Figure 3.

According to Section 2.1, some statistical properties of PEs/SADs and MVs/MVDs can be easily observed. For example, most SADs of PEs are local optimal for the cover videos, i.e., $S_l^{(0,0)} \leq S_l^{(\Delta h,\Delta v)}, (\Delta h, \Delta v) \in \mathcal{N}_8$. The MVs of neighboring blocks in cover videos are highly correlated because of motion similarity. However, the above statistical properties are often disturbed by MV based steganography, e.g., the SADs of PEs will be changed from local optimal to suboptimal, and the strength of the correlations between neighboring MVs will be weakened and thus the distributions of MVDs will also be changed. All of these leave detectable traces for steganalysis.

In the following sections, we will construct steganalytic features and then calibrate the features by using the statistical characteristics of both SADs and MVDs.

3 PB-NOP: PARTITION BASED NEIGHBORHOOD OPTIMAL PROBABILITY FEATURES

3.1 Relation between Quantization Distortion and Partition Modes

The SAD of PE is modified along with MV during message embedding. Both [24] and [16] pointed out that the proportion of local optimal SADs will be changed before and after embedding, and this is the basis for AoSO and SPOM features. Since the local optimality of SAD is affected by quantization distortion [24], the performance of AoSO and SPOM will also be affected inevitably.

Reference [24] assumed that the 2D-DCT (discrete cosine transform) coefficients of PE follow a Laplace distribution. Furthermore, [24] also proved that the quantization distortion is determined by QP and α, where α is the parameter of the Laplace distribution.

QP represents the compression degree, and α refers to the shape of the Laplace distribution. However, for the video compressed in constant bit rate, the value of QP is dynamic. The α is related to the movement of video content, texture complexity, and ME method [24]. So both QP and α are not easy to determine and measure. This also means that the degree of quantization distortion for blocks exhibits variability and uncertainty, which are harmful to the steganalytic features based on the local optimality of SAD (experiments for AoSO in Section 6 prove this point). If the quantization distortion can be quantized to a definite and tractable form, it will help to improve the features' performance.

For the ME using variable block size (VBS), we found that the quantization distortion is often associated with partition modes, and this can be demonstrated experimentally. Two video sequences, akiyo (slow moving and flat texture) and soccer (fast moving and complex texture), are selected for this experiment. The quantization distortion[1], partition proportion, and local optimal probability[2] (LOP) for different partition modes and different QPs are shown in Figure 4. It is shown that the smaller the partition size, the larger the quantization distortion, and the smaller the LOP. Moreover, with the increase of QP value, the quantization distortion also increases.

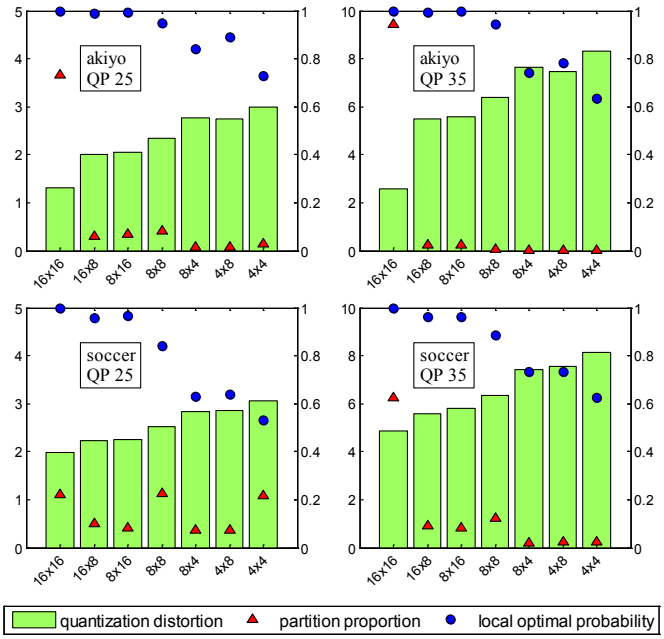

Figure 4: The quantization distortion, partition proportion, and local optimal probability for different partition modes and different QPs.

The relationship between quantization distortion and partition modes can be analyzed with QP and α. From (2) and (3), the λ (also the QP) controls the tradeoff between rate and distortion. For a high QP, the cost in (2) will be dominated by MVD, and a larger partition size will limit the number of MVs/MVDs for a MB and hence reduces the bitstream [17]. For the example shown in Figure 4, the proportion of 16×16 sized partitions is larger for a higher QP value. As for the α, it reflects the variance of the DCT coefficients of PE. In addition to ME method that is hard to interpret, the fast moving objects and textured areas (i.e. small α) are more in need of small partitions to fit the details [17]. As shown in Figure 4, the proportion of small partitions is larger for soccer than akiyo. Since the partition modes are related to QP and α, so we use partition modes instead of QP and α to quantize the quantization distortion.

[1]The quantization distortion is represented by the difference between current block and reconstructed block, and it is normalized by dividing by its block area for a fair comparison.

[2]The probability of occurrence of local optimal SADs is defined as local optimal probability (LOP), which is represented by the proportion of blocks with local optimal SADs in a video.

In the next subsection, we will extract features from separate partition modes to reduce the influence of quantization distortion, and we call it the partition based quantization method.

3.2 Neighborhood Optimal Probability of SAD

As described in Section 2.2, in cover videos, there always exists $S_l^{(0,0)} \leq S_l^{(\Delta h,\Delta v)}, (\Delta h, \Delta v) \in \mathcal{N}_8$ because of the local optimality of SAD; such a SAD is defined as local optimal SAD (LO-SAD). On the other hand, in stego videos, for the MV whose LSB has been modified, its corresponding original MV remains in 8-neighborhood, and thus may result in $S_l^{(\Delta h,\Delta v)} \leq S_l^{(0,0)}, (\Delta h, \Delta v) \in \mathcal{N}_8$; such a SAD is defined as neighborhood optimal SAD (NO-SAD). This is the case at the encoder side, and the situation will be maintained to some extent at the decoder side [24]. Therefore, it is obvious that the probability of a block that has at least one NO-SAD in stego videos is usually larger than that in cover videos.

Let the set of NO-SADs for a block be defined as

$$\mathcal{S}_l^{no} = \left\{ S_l^{(\Delta h,\Delta v)} \middle| S_l^{(\Delta h,\Delta v)} \leq S_l^{(0,0)},\ (\Delta h, \Delta v) \in \mathcal{N}_8 \right\} \quad (7)$$

The NO-SADs in cover videos are mainly caused by quantization distortion, while the NO-SADs in stego videos mostly result from steganographic embedding other than quantization distortion, so the number of NO-SADs for a block, i.e., $|\mathcal{S}_l^{no}|$, in stego videos is usually larger than that in cover videos. The probability of a block that has a specified number of NO-SADs is called neighborhood optimal probability (NOP) and is denoted as $\Pr\left(|\mathcal{S}_l^{no}| = i\right), i = 0, 1, \cdots, 8$, where $i = 0$ means that the block does not have NO-SADs, but has a LO-SAD instead.

The comparison of NOP $\Pr\left(|\mathcal{S}_l^{no}| = i\right)$ for different partition modes in cover and stego videos is shown in Figure 5. The stego videos are created by random LSB matching embedding on the larger component of all MVs, and the QP value is 25. It is evident that the $\Pr\left(|\mathcal{S}_l^{no}| = i\right)$ of stego is less than that of cover for $i = 0$, but larger for $i > 0$ (especially for $i = 1, 2, 3, 4$). In addition, no matter whether it is cover or stego, partitions of smaller size tend to have more NO-SADs, demonstrating the fact that the influence of quantization distortion on SAD varies with different partition modes (it is consistent with the LOP shown in Figure 4).

Following the steganalytic features are designed by using NOP incorporated with partition based quantization method.

Let $P_{16\times16}, P_{16\times8}, P_{8\times16}, P_{8\times8}, P_{8\times4}, P_{4\times8}, P_{4\times4}$ be the seven original partition modes mentioned in Section 2.1, then four sets containing new partition modes are defined as follows

$$\begin{aligned}
\mathcal{P}_1 &= \{P_{16\times16} \vee P_{16\times8} \vee P_{8\times16} \vee P_{8\times8} \vee P_{8\times4} \vee P_{4\times8} \vee P_{4\times4}\} \\
\mathcal{P}_2 &= \{P_{16\times16} \vee P_{16\times8} \vee P_{8\times16} \vee P_{8\times8}, P_{8\times4} \vee P_{4\times8} \vee P_{4\times4}\} \\
\mathcal{P}_5 &= \{P_{16\times16}, P_{16\times8} \vee P_{8\times16}, P_{8\times8}, P_{8\times4} \vee P_{4\times8}, P_{4\times4}\} \\
\mathcal{P}_7 &= \{P_{16\times16}, P_{16\times8}, P_{8\times16}, P_{8\times8}, P_{8\times4}, P_{4\times8}, P_{4\times4}\}
\end{aligned} \quad (8)$$

where the symbol '$\vee$' means merging original partition modes into a new partition mode, $\mathcal{P}_1$ has only one new partition mode without considering the sizes of the original partitions

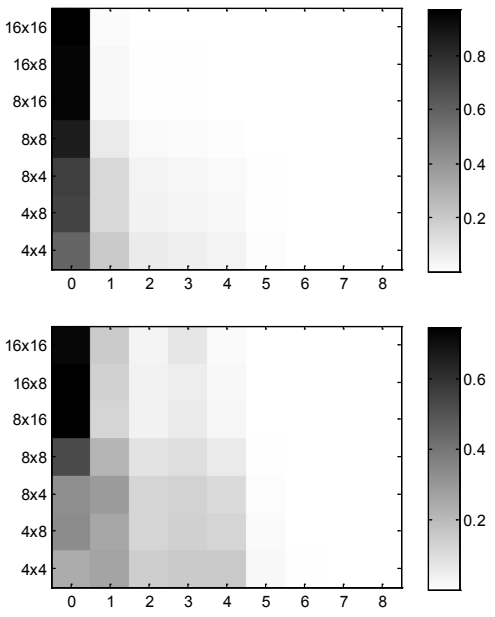

Figure 5: The NOP $\Pr\left(|\mathcal{S}_l^{no}| = i\right)$ for different partition modes in cover (top) and stego (bottom) videos.

(AoSO and SPOM belong to this type), $\mathcal{P}_2$ has two new partition modes based on MB partitions and subMB partitions, $\mathcal{P}_5$ forms five new partition modes according to the area of the original partitions, $\mathcal{P}_7$ contains all original partition modes without any merging.

Based on the new partition modes, the set of NO-SADs for a partition is redefined as

$$\mathcal{S}_{p,l}^{no} = \left\{ S_{p,l}^{(\Delta h,\Delta v)} \middle| S_{p,l}^{(\Delta h,\Delta v)} \leq S_{p,l}^{(0,0)}, (\Delta h, \Delta v) \in \mathcal{N}_8 \right\} \quad (9)$$

where $p = 1, \cdots, P$ is the index of the new partition modes in $\mathcal{P}_i$, $P = |\mathcal{P}_i|$, $i \in \{1, 2, 5, 7\}$. $l = 1, 2, \cdots, L_p$ is the index of partitions with new partition mode p in a frame. Then the partition based neighborhood optimal probability (PB-NOP) features are defined as

$$f_{p,i} = \Pr\left(\left|\mathcal{S}_{p,l}^{no}\right| = i\right) = \frac{1}{L_p} \sum_{l=1}^{L_p} \delta\left(\left|\mathcal{S}_{p,l}^{no}\right|, i\right),\ i = 0, 1, \cdots, 8 \quad (10)$$

where $\delta(x, y) = 1$ if $x = y$ and 0 otherwise.

As can be seen from Figure 4 and Figure 5, the partitions with the same area have a similar LOP or NOP, so $\mathcal{P}_5$ is adopted to measure the quantization distortion (see Section 6.2 for more discussions). It can also be seen from Figure 5 that when $\left|\mathcal{S}_{p,l}^{no}\right|$ has large values ($i > 5$), the $f_{p,i}$ is all very small, and their difference is negligible. To get a more compact and robust form, a threshold is used to merge the underpopulated features together. Then the final PB-NOP features are as follows

$$F_{p,i} = \begin{cases} f_{p,i}\,, & if\ i = 0, 1, \cdots, T_1 - 1 \\ \sum\limits_{j=T_1}^{8} f_{p,j}\,, & if\ i = T_1 \end{cases} \quad (11)$$

where T_1 is the threshold. In this paper we set $T_1 = 5$ (the discussion of T_1 is postponed to Section 6.2), so the dimensionality of PB-NOP features is $P \times (T_1 + 1) = 30$.

4 IIC: INTER AND INTRA CO-OCCURRENCE FEATURES

4.1 MVD vs. NMVD

As described in Section 2.2, the statistics of MVs are disturbed by MV based steganography. Like some classic steganalytic features for image [8, 15], MV steganalytic features [20, 26] are usually constructed by the subtraction of neighboring elements, i.e., NMVD, to reveal the statistical anomalies. The MVD, which can be viewed as a special case of NMVD, shows more superiority than NMVD on feature construction.

4.1.1 Compactness of Distribution. Owing to the correlation between neighboring MVs, the distribution of NMVDs exhibits zero-mean and symmetry. As mentioned in Section 2.1, the PMV is the median value of three neighboring MVs. So according to (1), MVD is also the median of three corresponding NMVDs. In order to better describe the characteristics of NMVDs and MVDs and compare their difference, the distributions of NMVDs and MVDs are analyzed experimentally.

Figure 6(a) shows the histogram of the horizontal components of MVDs and the histogram of the horizontal components of NMVDs calculated from neighboring MVs in horizontal direction. As shown, the two histograms are both Laplacian-like, but the histogram of MVDs is much steeper. So it can be concluded that the distribution of MVDs is more compact than that of NMVDs. As for the joint distributions of MVDs/NMVDs, i.e., the co-occurrence for two components of neighboring MVDs/NMVDs, and the co-occurrence for two components of the same MVD/NMVD, the argument is also tenable (this can also be demonstrated experimentally but is not shown here due to lack of space).

For steganalytic features based on residual signals in image [8, 15] or video [26], a thresholding technique is often used to reduce feature dimensionality. Since the threshold in practice usually takes a small value, it will lose some useful statistical information to an extent. Therefore, a compact distribution will be helpful to capture more statistical information with a small threshold. From this point of view, MVD is more favorable than NMVD.

4.1.2 Inter Distribution. The features in [20] and [26] are based on the first-order distributions (histograms) of NMVDs and the joint distributions of NMVDs. The NMVD features are under the assumption that all blocks are of the same size (i.e., FBS) and all neighboring blocks have their own MVs. However, this assumption is unsuitable for many advanced video coding standards, thus limiting the construction and application of NMVD features. In contrast, the MVD is automatically generated by the video encoder without considering the consistency of neighboring partitions and the continuity of MVs. So the features based on MVD can be easily applied to various coding conditions.

The joint distribution of horizontal or vertical components of neighboring MVDs is defined as inter distribution. The inter distribution can be denoted by $\Pr(x^{\mathrm{E}}, x^{\mathrm{N}})$, where x^{E} and x^{N} are the horizontal or vertical components of MVDs for current block and neighboring block respectively.

Like the joint distributions of NMVDs [26], the inter distributions will also be changed by MV based steganography. There are four inter distributions of MVDs can be used for steganalysis. Let A, B, C and D be four neighboring blocks next to current block E as shown in Figure 2, the inter distributions of current MVD and neighboring MVDs in location A (horizontal), B (vertical), C (minor diagonal) and D (main diagonal) are concerned, i.e., $\mathrm{N} \in \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}\}$ in x^{N}. Note that the nonexistent neighboring MVD pairs are not counted. Figure 6(b) shows the $\Pr(x^{\mathrm{E}}, x^{\mathrm{A}})$ of two neighboring horizontal components of MVDs, it is observed that most of the neighboring MVD components have similar values, and are located around the origin. The inter distributions $\Pr(x^{\mathrm{E}}, x^{\mathrm{B}})$, $\Pr(x^{\mathrm{E}}, x^{\mathrm{C}})$ and $\Pr(x^{\mathrm{E}}, x^{\mathrm{D}})$ are also similar.

4.1.3 Intra Distribution. The NMVD features are derived from the statistical characteristics of components of neighboring MVs [20, 26]. However, both NMVD and MVD have two components, and the statistical characteristics of components of the same NMVD or MVD have not yet been fully studied for steganalysis.

The joint distribution of two components of the same MVD is defined as intra distribution, which can also be denoted by $\Pr(x^h, x^v)$, where x^h and x^v are the horizontal and vertical components of the same MVD. The intra distribution can be interpreted from a coding point of view. The $B(\bar{D})$ in (2) indicates that the coding weights of two components of a MVD are equal, so the two components of a MVD will have a higher probability to obtain similar values. Figure 6(c) shows the $\Pr(x^h, x^v)$, it is observed that the distributional characteristics of $\Pr(x^h, x^v)$ are similar to that of $\Pr(x^{\mathrm{E}}, x^{\mathrm{A}})$. $\Pr(x^h, x^v)$ is also affected by steganographic embedding, and can be used for steganalysis.

4.1.4 Statistical Discriminability. Statistical discriminability is the key factor that has to be considered for feature design. To compare the statistical discriminability of MVD and NMVD, the K-L divergence between the cover and stego distributions is adopted as a benchmark. The K-L divergence between the histograms of MVDs and NMVDs and the K-L divergence between the joint distributions of MVDs and NMVDs are shown in Figure 6(d) and (e) respectively. The stego videos are created by random LSB matching embedding on the larger component of all MVs. To get the joint distributions of NMVDs, the videos for Figure 6(e) are only encoded in 16×16 partitions (i.e., FBS). The NMVD-A, NMVD-B, NMVD-C and NMVD-D are the NMVDs calculated from horizontal, vertical, minor diagonal and main diagonal directions respectively.

As shown in Figure 6(d) and (e), in all cases, the divergence value of MVD is greater than that of NMVD. Greater K-L divergence value represents larger statistical distortion in the

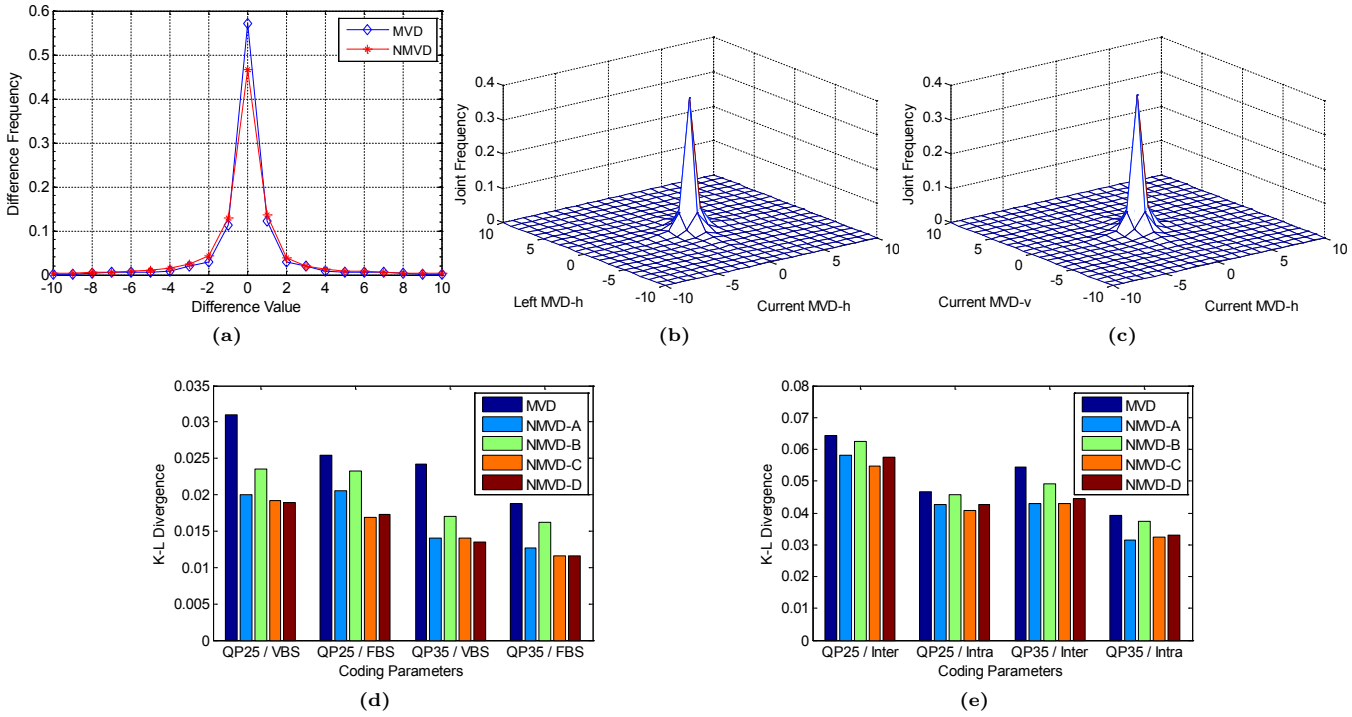

Figure 6: (a) Histograms of MVDs and NMVDs. (b) Inter distribution $\Pr(x^{\mathbf{E}}, x^{\mathbf{A}})$ of MVDs. (c) Intra distribution $\Pr(x^h, x^v)$ of MVDs. (d) K-L divergence between histograms of MVDs and NMVDs. (e) K-L divergence between joint distributions of MVDs and NMVDs. Histograms and inter distributions only use horizontal components.

stego, that is to say, the statistical representation of MVD is better at revealing the traces of steganographic embedding than that of NMVD.

4.2 Inter and Intra Co-occurrences of MVD

From the above comparative analysis, the MVD is exploited to construct features based on inter and intra distributions, which are represented by co-occurrence matrices.

4.2.1 Inter Co-occurrence. Let $D_l(h)$ and $D_l(v)$ denote the horizontal and vertical components of the l-th MVD in a frame. $D_l^*(x)$ is a neighboring MVD component to $D_l(x)$, where $* \in \{\leftarrow, \uparrow, \nearrow, \nwarrow\}$ stands for the neighboring relation in horizontal, vertical, minor diagonal, or main diagonal direction corresponding to location A, B, C, or D in Figure 2. $x \in \{h, v\}$ stands for the horizontal or vertical component of MVD. The inter co-occurrence matrix of $D_l(h)$ and $D_l^{\leftarrow}(h)$, for example, is defined as follows:

$$C_{m,n}^{\leftarrow}(h) = \frac{1}{Z_1} \left| \left\{ \left(D_l(h), D_l^{\leftarrow}(h) \right) | D_l(h) = m, D_l^{\leftarrow}(h) = n \right\} \right| \tag{12}$$

where Z_1 is a normalization factor ensuring that $\sum_{m,n} C_{m,n}^{\leftarrow}(h) = 1$.

By analogy, a total of 8 inter co-occurrence matrices $C_{m,n}^{*}(x)$ can be obtained.

4.2.2 Intra Co-occurrence. Similar to the $C_{m,n}^{*}(x)$, the intra co-occurrence matrix of $D_l(h)$ and $D_l(v)$ is defined as

$$C_{m,n}^{ia} = \frac{1}{Z_2} \left| \left\{ \left(D_l(h), D_l(v) \right) | D_l(h) = m, D_l(v) = n \right\} \right| \tag{13}$$

where the normalization factor Z_2 ensures that $\sum_{m,n} C_{m,n}^{ia} = 1$.

4.2.3 Co-occurrence Concentration. As shown in Figure 6(b) and (c), most of the MVDs are located around the bin zero. To make the co-occurrence bins well populated, a threshold is utilized to compactify the co-occurrences. For a predefined threshold T_2, only the bins of $C_{m,n}^{*}(x)$ and $C_{m,n}^{ia}$ belonging to $[-T_2, T_2]$ are selected for features.

The Figure 6(b) and (c) also show that the joint distributions of MVDs are symmetrical about the origin. For the sake of robustness and lower dimensionality, the co-occurrence matrices are symmetrized by both direction and sign inspired by [8]. The directional symmetry is under the assumption that the inter and intra distributions do not change after swapping two elements' positions. From a coding perspective, the coding cost of MVD is also independent of the sign. Then the symmetrizations by direction and sign for an inter co-occurrence matrix in the horizontal direction, for example, are defined as follows:

$$\overleftrightarrow{C}_{m,n}^{\leftarrow}(h) \Leftarrow C_{m,n}^{\leftarrow}(h) + C_{n,m}^{\leftarrow}(h) \tag{14}$$

$$\bar{C}^{\leftarrow}_{m,n}(h) \Leftarrow \overleftrightarrow{C}^{\leftarrow}_{m,n}(h) + \overleftrightarrow{C}^{\leftarrow}_{-m,-n}(h) \tag{15}$$

The symmetrizations of the other inter and intra co-occurrence matrices are defined analogically.

In our additional experiments, we observed that the statistical properties of 8 inter co-occurrences are quite similar despite their components or directions are different. To further decrease the feature dimensionality and improve the robustness, the inter co-occurrence matrices are averaged by components and directions as follows:

$$\bar{C}^{\leftarrow}_{m,n} = \frac{1}{2}\left[\bar{C}^{\leftarrow}_{m,n}(h) + \bar{C}^{\leftarrow}_{m,n}(v)\right] \tag{16}$$

$$\bar{C}^{ir}_{m,n} = \frac{1}{4}\left[\bar{C}^{\leftarrow}_{m,n} + \bar{C}^{\uparrow}_{m,n} + \bar{C}^{\nearrow}_{m,n} + \bar{C}^{\nwarrow}_{m,n}\right] \tag{17}$$

Note that the direction used for symmetrization and averaging are different. After a series of concentration (thresholding, symmetrization, and averaging), the final inter and intra co-occurrence (IIC) features will be obtained as

$$F_{1,\cdots,k} = unique\left(\bar{C}^{ir}_{\cdot}\right) \tag{18}$$

$$F_{k+1,\cdots,2k} = unique\left(\bar{C}^{ia}_{\cdot}\right) \tag{19}$$

where $unique(\cdot)$ represents eliminating the duplicates produced by symmetrization from a co-occurrence matrix, and $k = (T_2+1)^2$ is the feature dimensionality for a concentrated co-occurrence matrix. In this paper we use $T_2 = 3$ (see Section 6.2 for the selection of T_2), obtaining thus 16-dimensional features for inter co-occurrence and intra co-occurrence respectively.

5 WOC: WINDOW OPTIMAL CALIBRATION

Calibration, which enhances the feature sensitivity and thus improves the detection accuracy, has been frequently applied to image steganalysis [7, 12] and video steganalysis [2, 5].

Calibration of narrow sense is to estimate the statistical properties of a cover from a stego object [7]. Existing calibration methods in video steganalysis [2, 5] all attempt to recover the original MVs. Following this idea, assuming the MV based steganography only modifies the LSB of MV component, i.e., the modified MV is located in $\mathcal{W}_9$ of original MV, then the calibrated (original) MV should also be located in $\mathcal{W}_9$ of the modified MV. More specifically, for the MV that has not been modified, the calibrated MV is likely to be the $V_l^{(0,0)}$ in $\mathcal{W}_9$; while for the modified MV, the calibrated MV is probably the $V_l^{(\Delta h,\Delta v)}, (\Delta h,\Delta v) \in \mathcal{N}_8$. If the calibrated MV can be obtained, then the calibrated features for the proposed PB-NOP and IIC can also be extracted. Therefore, the key is how to get the calibrated MV.

The SAD with minimal value in $\mathcal{W}_9$ is defined as window optimal SAD (WO-SAD), which forms the following set

$$\mathcal{S}^{wo}_{p,l} = \Big\{ S^{(\Delta h,\Delta v)}_{p,l} \Big| S^{(\Delta h,\Delta v)}_{p,l} \le S^{\left(\Delta\bar{h},\Delta\bar{v}\right)}_{p,l}, (\Delta h,\Delta v) \in \mathcal{W}_9, \left(\Delta\bar{h},\Delta\bar{v}\right) \in \mathcal{W}_9 \Big\} \tag{20}$$

The MV that corresponds to a WO-SAD and has minimal bitstream size for coding MVD is defined as window optimal MV (WO-MV). The set of WO-MVs is as follows

$$\mathcal{V}^{wo}_{p,l} = \Big\{ V^{(\Delta h,\Delta v)}_{p,l} \Big| B\Big(D^{(\Delta h,\Delta v)}_{p,l}\Big) \le B\Big(D^{\left(\Delta\bar{h},\Delta\bar{v}\right)}_{p,l}\Big), S^{(\Delta h,\Delta v)}_{p,l} \in \mathcal{S}^{wo}_{p,l}, S^{\left(\Delta\bar{h},\Delta\bar{v}\right)}_{p,l} \in \mathcal{S}^{wo}_{p,l} \Big\} \tag{21}$$

where $D^{(\Delta h,\Delta v)}_{p,l}$ is the MVD corresponding to $S^{(\Delta h,\Delta v)}_{p,l}$ and $V^{(\Delta h,\Delta v)}_{p,l}$, $B(\cdot)$ refers to the bitstream size of MVD.

Let $\widehat{V}_{p,l}$ be the calibrated MV. In most cases, $\left|\mathcal{V}^{wo}_{p,l}\right| = 1$, then $\widehat{V}_{p,l} = V^{(\Delta h,\Delta v)}_{p,l}$, $V^{(\Delta h,\Delta v)}_{p,l} \in \mathcal{V}^{wo}_{p,l}$. If $\left|\mathcal{V}^{wo}_{p,l}\right| > 1$, the $\widehat{V}_{p,l}$ will be selected from $\mathcal{V}^{wo}_{p,l}$ randomly or in a certain order[3]. According to Section 2.2 and Section 3.2, the $\widehat{V}_{p,l}$ is most likely the original MV generated during ME owing to the optimality of SAD and MVD in a local window area. So this calibration is called window optimal calibration (WOC). The detailed steps of WOC for a video frame are as follows:

Step 1. For a MV $V_{p,l}$ in a video frame, get its calibrated MV $\widehat{V}_{p,l}$ using (20) and (21).

Step 2. Get the SAD corresponding to $\widehat{V}_{p,l}$ and take it as the calibrated SAD denoted as $\widehat{S}_{p,l}$.

Step 3. Repeat Step 1 and Step 2, and calibrate all the MVs in the frame in a coding order, thus forming a calibrated frame consisting of blocks with calibrated SADs and calibrated MVs.

Step 4. For each $\widehat{S}_{p,l}$, form a local window $\mathcal{W}_9$ centered at $\widehat{S}_{p,l}$, and then compute the $\widehat{\mathcal{S}}^{no}_{p,l}$ for $\widehat{S}_{p,l}$ based on $\mathcal{W}_9$ using (9).

Step 5. Repeat Step 4 until all $\widehat{\mathcal{S}}^{no}_{p,l}$ has been calculated, and then compute the PB-NOP features for the calibrated frame according to Section 3.2.

Step 6. For each $\widehat{V}_{p,l}$, update its PMV by the neighboring calibrated MVs in the calibrated frame, and then compute the calibrated MVD corresponding to $\widehat{V}_{p,l}$ using (1).

Step 7. Repeat Step 6 until all calibrated MVDs have been calculated, and then compute the IIC features for the calibrated frame according to Section 4.2.

Unlike the difference calibration [7] used in [2, 5], the calibrated features in this paper are processed as a Cartesian form [12]. The main reasons for that are as follows. The difference calibrated features can be completely derived from Cartesian calibrated features, but not vice versa. In other words, the Cartesian calibrated features contain more discriminative information than difference calibrated features.

Even though the Cartesian calibration doubles the feature dimensionality, it is not difficult for classifier training owing to the low dimensionality of original features. Table 1 shows the dimensionalities of all feature components. The total dimensionality of the combined and calibrated features (CCF) is $(30+32)\times 2 = 124$.

[3]In this paper, the $V^{(\Delta h,\Delta v)}_{p,l}$ in $\mathcal{W}_9$ is scanned in a top-to-bottom, and left-to-right order, and the first $V^{(\Delta h,\Delta v)}_{p,l}$, $V^{(\Delta h,\Delta v)}_{p,l} \in \mathcal{V}^{wo}_{p,l}$, is selected as $\widehat{V}_{p,l}$.

Table 1: Feature dimensionality of PB-NOP and IIC with and without WOC

Calibration	Feature	Threshold	Dimensionality
Non-WOC	PB-NOP	5	30
	IIC	3	32
WOC	PB-NOP	5	30
	IIC	3	32

Compared with the existing MV calibration methods, the window optimal calibration (WOC) is more universal. As mentioned before, the calibration methods in [2, 5] are used under some particular conditions. While the WOC has no additional constraints, its calibration process only depends on the MVs and SADs in a local window $\mathcal{W}_9$ which is independent of any coding parameters, so WOC can be applied to a variety of coding conditions without any restriction.

6 EXPERIMENTS

6.1 Experimental Setup

6.1.1 Video Sequences. A video database consisting of 36 standard test sequences downloaded from the internet is used for experiments. All video sequences are stored in 4:2:0 YUV format and have the size of CIF (352×288). The original video sequences have various scenes and various frames (mostly 300 frames), in order to uniformly disperse the video sequences, only the first 240 frames of each video sequence are utilized, thus forming trimmed sequences.

6.1.2 Steganographic Methods. To evaluate the detection performance of the steganalytic features, four typical MV based steganography, i.e., Xu's method [27], Aly's method [1], Cao's method [3] and Zhang's method [29], are included. These steganographic methods are implemented using a well-known H.264/AVC codec named x264 [23] to generate the stego videos, and the basic profile is adopted for simplicity.

The random bit stream is used for embedding, and the embedding rate or payload is denoted by bpnsmv (bits per non-skip MV), which represents the ratio of embedded bits' number to the total number of non-skip MVs in each frame (excluding the MV of a skip MB is due to the fact that zero-valued SAD and MVD are susceptible to steganographic embedding, which leads to deteriorated steganographic security and compression efficiency).

6.1.3 Training and Classification. For stability, the steganalytic features are extracted from the frames within a fixed size sliding window which scans each trimmed sequence without overlapping. The sliding window size is set to be 6 based on experimental experiences.

The soft-margin support vector machine (C-SVM) with Gaussian kernel [4] is used as classifier, and the penalty parameter C and kernel parameter γ of the C-SVM are optimized using five-fold cross-validation on the following grid space $(C,\gamma) \in \left\{\left(2^i, 2^j\right) | i=-5,-4,\cdots,15, j=-15,-14,\cdots,3\right\}$.

Table 2: Effect of the partition mode sets $\mathcal{P}_i$ on the detection accuracy rate of PB-AoSO

$\mathcal{P}_i$	$i=1$	$i=2$	$i=5$	$i=7$
AR	0.7570	0.7953	0.7952	0.7796

Table 3: Effect of the partition mode sets $\mathcal{P}_i$ on the detection accuracy rate of PB-NOP

$\mathcal{P}_i$	$i=1$	$i=2$	$i=5$	$i=7$
AR	0.6126	0.6441	0.6879	0.6869

Table 4: Effect of the threshold T_1 on the detection accuracy rate of PB-NOP

T_1	4	5	6	7	8
AR	0.6876	0.6879	0.6789	0.6834	0.6817

Table 5: Effect of the threshold T_2 on the detection accuracy rate of IIC

T_2	2	3	4	5
AR	0.7235	0.7256	0.7277	0.7132

A binary classifier is trained for each specific feature set, steganographic method, and embedding rate. For each binary classifier, half of cover and the corresponding half of stego are randomly selected for training, and the remaining half pairs of the cover and stego are used for testing. The detection performance is measured by accuracy rate (AR) computed as $\mathrm{AR} = (\mathrm{TPR} + \mathrm{TFR})/2$, where TPR and TNR represent the true positive rate and true negative rate respectively. The training and testing process is repeated 50 times, and the mean value of all results is calculated as the final AR.

6.2 Parameters Selection

To promote the effectiveness of the steganalytic features, some parameters should be tuned before feature extraction. In this subsection, a total of four parameters, the partition mode sets $\mathcal{P}_i$ for AoSO and PB-NOP, the threshold T_1 for PB-NOP, and the threshold T_2 for IIC, are discussed through experiments. The videos are encoded in VBS for QP 25. For simplicity, we only detect Xu's method at 0.1 bpnsmv.

6.2.1 Partition Mode Sets $\mathcal{P}_i$. The partition based quantization method uses partition mode sets $\mathcal{P}_i$, $i \in \{1,2,5,7\}$ to reduce the quantization distortion and thus enhances the SAD features, but different $\mathcal{P}_i$ have different effects on features' detection performance. Moreover, to evaluate the universality of partition based quantization method, the AoSO [24] is extended by the $\mathcal{P}_i$ and we call it partition based AoSO (PB-AoSO). In Table 2 and Table 3, the effects of the different $\mathcal{P}_i$ on the detection AR of PB-AoSO and PB-NOP are given. Note that the PB-AoSO with $\mathcal{P}_1$ is the original AoSO.

Table 6: Accuracy rate of AoSO, PB-AoSO and CCF for videos encoded in VBS and QP 25

Method	bpnsmv	AoSO	PB-AoSO	CCF
Xu's	0.05	0.6654	0.6947	0.6751
	0.1	0.7570	0.7953	0.7750
	0.2	0.8359	0.8691	0.8713
	0.3	0.8708	0.9090	0.9319
Aly's	0.05	0.8481	0.9085	0.9264
	0.1	0.9139	0.9477	0.9641
	0.2	0.9426	0.9691	0.9794
	0.3	0.9547	0.9731	0.9817
Cao's	0.05	0.5211	0.5181	0.5832
	0.1	0.5173	0.5171	0.6492
	0.2	0.5227	0.5106	0.7678
	0.3	0.5515	0.5344	0.8426
Zhang's	0.05	0.5182	0.5106	0.6095
	0.1	0.5219	0.5383	0.6796
	0.2	0.5531	0.5903	0.7596
	0.3	0.5739	0.6553	0.8301

Table 7: Accuracy rate of AoSO, PB-AoSO and CCF for videos encoded in VBS and QP 35

Method	bpnsmv	AoSO	PB-AoSO	CCF
Xu's	0.05	0.5127	0.5202	0.5570
	0.1	0.5279	0.5664	0.6179
	0.2	0.5722	0.6238	0.7268
	0.3	0.6027	0.6610	0.7998
Aly's	0.05	0.5967	0.6869	0.7307
	0.1	0.6607	0.7543	0.8457
	0.2	0.7148	0.8072	0.8968
	0.3	0.7149	0.8179	0.9054
Cao's	0.05	0.5072	0.5043	0.5436
	0.1	0.5116	0.5090	0.6031
	0.2	0.5149	0.5126	0.7009
	0.3	0.5150	0.5153	0.7602
Zhang's	0.05	0.5119	0.5103	0.5727
	0.1	0.5185	0.5196	0.6322
	0.2	0.5321	0.5375	0.7357
	0.3	0.5342	0.5605	0.7951

Table 8: Accuracy rate of Su's feature, MVRB, AoSO and CCF for videos encoded in FBS and QP 25

Method	bpnsmv	Su	MVRB	AoSO	CCF
Xu's	0.05	0.5125	0.7088	0.7772	0.6822
	0.1	0.5338	0.7611	0.8272	0.7194
	0.2	0.5929	0.8631	0.8877	0.8195
	0.3	0.6340	0.9137	0.8815	0.8603
Aly's	0.05	0.5490	0.8273	0.9442	0.8886
	0.1	0.5911	0.8994	0.9678	0.9539
	0.2	0.6714	0.9454	0.9797	0.9816
	0.3	0.6978	0.9715	0.9825	0.9854
Cao's	0.05	0.5141	0.5219	0.5236	0.6176
	0.1	0.5283	0.5615	0.5231	0.6356
	0.2	0.5643	0.6095	0.5412	0.6507
	0.3	0.5861	0.6668	0.5782	0.6823
Zhang's	0.05	0.5037	0.5047	0.5456	0.6337
	0.1	0.5126	0.5165	0.5465	0.6333
	0.2	0.5396	0.5519	0.5490	0.6635
	0.3	0.5673	0.5671	0.5577	0.6994

Table 9: Accuracy rate of Su's feature, MVRB, AoSO and CCF for videos encoded in FBS and QP 35

Method	bpnsmv	Su	MVRB	AoSO	CCF
Xu's	0.05	0.5106	0.5087	0.5333	0.5160
	0.1	0.5289	0.5332	0.5726	0.5786
	0.2	0.5805	0.5766	0.6344	0.6644
	0.3	0.6314	0.6131	0.6501	0.7334
Aly's	0.05	0.5173	0.5880	0.7079	0.6742
	0.1	0.5538	0.6609	0.7757	0.7583
	0.2	0.6144	0.7379	0.8213	0.8418
	0.3	0.6410	0.7817	0.8467	0.8641
Cao's	0.05	0.5142	0.5025	0.5129	0.5323
	0.1	0.5349	0.5091	0.5061	0.5758
	0.2	0.5597	0.5216	0.5192	0.6235
	0.3	0.5690	0.5322	0.5090	0.6684
Zhang's	0.05	0.5044	0.5048	0.5133	0.5658
	0.1	0.5133	0.5145	0.5110	0.6143
	0.2	0.5322	0.5300	0.5237	0.6766
	0.3	0.5683	0.5332	0.5208	0.7284

From Table 2 and Table 3, it can be seen that the $\mathcal{P}_2$ is more appropriate to PB-AoSO while the $\mathcal{P}_5$ is best suited for PB-NOP. So we select $\mathcal{P}_2$ for PB-AoSO and $\mathcal{P}_5$ for PB-NOP respectively in this paper.

6.2.2 Threshold T_1. The threshold T_1 is another parameter for PB-NOP. In Table 4, the effects of the threshold T_1 on the detection AR of PB-NOP are shown.

From Table 4, it can be seen that the detection performance of PB-NOP is insensitive to threshold T_1. When T_1 is larger than 5, the detection AR is no longer increasing. So the T_1 is set to be 5 in this paper.

6.2.3 Threshold T_2. The threshold T_2 is used to curb the dynamic range of MVD. A larger T_2 keeps more statistical information, whereas the dimension of IIC will increases sharply. In Table 5, the effects of the threshold T_2 on the detection AR of IIC are given.

As shown in Table 5, when the T_2 increases, the detection AR of IIC does not improve obviously. So we set $T_2 = 3$ for a balance between feature dimensionality and detection performance.

6.3 Comparison With Prior Art

6.3.1 Comparison on VBS and QP. To evaluate various steganalytic features on the videos encoded in VBS and different QPs, the MBs are allowed to be divided into small partition sizes as shown in Figure 1, and the QP values are set to be 25 (high bit rate) and 35 (low bit rate). The reason why

we use constant QP instead of constant bit rate is that the constant QP is easier to measure the influence of compression degree on steganalytic features than dynamic QP. The current steganalytic features that can be fully applicable to the VBS videos are AoSO [24] and SPOM [16], and the AoSO has the best performance, so AoSO is chosen for comparison. Moreover, the PB-AoSO with $\mathcal{P}_2$ is also included, and its feature dimensionality is 36. The AR of AoSO, PB-AoSO and CCF on VBS videos with QP 25 and 35 are reported in Table 6 and Table 7 respectively.

It is observed that for low QP, the proposed CCF generally performs better than AoSO and PB-AoSO (the AR of CCF is slightly lower than that of PB-AoSO on detecting Xu's method at low payloads). For high QP, CCF delivers the best performance across all tested steganographic methods and all embedding rates.

By comparing AoSO and PB-AoSO, the PB-AoSO outperforms AoSO especially under the condition of high QP. When detecting Aly's method for QP 35, the PB-AoSO can even increase AR by about 10%. This demonstrates that the partition based quantization method can effectively deal with the issue of quantization distortion, and can also be applied to other steganalytic features that are based on LO-SAD or NO-SAD.

As for the Cao's method and Zhang's method, the AoSO is basically invalid, and the PB-AoSO does not work either. Cao's method replaces original optimal MV with suboptimal MV. Both SADs corresponding to optimal MV and suboptimal MV are quite close, and the local optimality of these two SADs are likely to be consistent with each other at the decoder side. In other words, their differences are more easily obscured by quantization distortion. Zhang's method preserves the local optimality of SADs at the decoder side during embedding. So both mechanisms lead to the failure of AoSO that is based on LO-SAD. The PB-NOP also faces the same problem against these two methods (see Table 10), but owing to the robustness of IIC, the CCF can still detect Cao's method and Zhang's method.

It can also be seen from Table 6 and Table 7 that the AR of three feature sets decreases with the increase of QP. This is mainly caused by the aggravation of quantization distortion. In addition, for higher QP, more MBs tend to choose larger partition size (see the partition proportion in Figure 4), which can be viewed as a transition from VBS to FBS (see experiments below), thus weakening the effectiveness of partition based quantization method used in PB-AoSO and PB-NOP.

6.3.2 Comparison on FBS and QP. To evaluate various steganalytic features on the videos encoded in FBS and different QPs, all MBs are of size 16×16, and the QP values are also set to be 25 and 35. The feature sets for comparison are Su's feature [20] (features are derived only from MVs), MVRB [2] (features are derived from MVs and SADs), and AoSO [24] (features are derived only from SADs). The [5, 16, 26] are omitted, because their ideologies are similar to [2, 20, 24] and the latter are more representative. All coding parameters are kept the same for two compressions of MVRB. The dimensionality of PB-NOP is only 6 due to the FBS, so the total dimensionality of CCF is $(6+32)\times2=76$. The AR of Su, MVRB, AoSO and CCF on FBS videos with QP 25 and 35 are reported in Table 8 and Table 9 respectively.

Table 10: Accuracy rate of CCF and its components for four methods with 0.1 bpnsmv on VBS videos

QP	Method	Calibrated PB-NOP	Calibrated IIC	PB-NOP and IIC	CCF
25	Xu's	0.6879	0.7256	0.6929	0.7750
	Aly's	0.9349	0.9275	0.9181	0.9641
	Cao's	0.5076	0.6453	0.5806	0.6492
	Zhang's	0.5662	0.6784	0.6209	0.6796
35	Xu's	0.5609	0.5992	0.5972	0.6179
	Aly's	0.7948	0.7534	0.7969	0.8457
	Cao's	0.5031	0.6017	0.5959	0.6031
	Zhang's	0.5203	0.6313	0.6215	0.6322

As seen from Table 8, for the steganography with high security (Cao's method and Zhang's method), CCF achieves the best detection performance. For other steganography, AoSO provides the highest AR, CCF and MVRB have comparable performance (CCF is better in detecting steganography that modifies two components of MV, while MVRB can better detect steganography that modifies one component of MV), and Su performs worst in most tested cases.

The Table 9 shows that for the VBS videos with high QP, CCF is the best performer in most cases. Of particular note is the Su's feature, which achieves a higher AR for Cao's method and Zhang's method than AoSO and MVRB owing to the robustness of NMVD features.

By the comprehensive comparison from Table 6 – Table 9, it can be concluded that CCF is better suited to detect videos encoded in VBS and high QP values. This is because for the VBS video, the various partition modes can be fully utilized to quantize the quantization distortion, and it can also be considered to increase the feature diversity. As for the QP, the SAD features are more sensitive to QP due to quantization distortion, so MVRB and AoSO perform worse for the higher QP. While CCF is combined by the SAD features and MVD features, and the latter is robust to QP, so CCF is still in effect for high QP owing to IIC.

6.4 Feature Component Analysis

The CCF consists of three components: PB-NOP, IIC, and WOC. To evaluate the performance of different components and to validate the importance of combining all components together, calibrated PB-NOP features (60-D)[4], calibrated IIC features (64-D), combined PB-NOP and IIC features (62-D), and CCF (124-D) are subject to test. The videos are encoded in VBS for QP 25 and 35, and only embedding rate of 0.1 bpnsmv is considered for simplicity. The detection AR of CCF and its components are shown in Table 10.

[4]The number in bracket denotes a feature dimensionality, see Table 1 for details. The same below.

As expected, the detection performance of CCF is superior to any single component for all tested steganographic methods. The comparison among CCF, calibrated PB-NOP and calibrated IIC validates the viewpoint that the statistical characteristics of different aspects should be combined to improve the detection capability. Besides, the performance of combined features with and without calibration also proves the effectiveness of WOC.

7 CONCLUSION

The steganographic embedding in MVs changes the statistical characteristics of both SADs and MVDs. According to this phenomenon, the combined and calibrated features (CCF) for steganalysis of MV based steganography in H.264/AVC is introduced in this paper.

The CCF consists of three components: partition based neighborhood optimal probability (PB-NOP) features, inter and intra co-occurrence (IIC) features, and window optimal calibration (WOC). The performance of CCF was carefully examined by various experiments. The experimental results show that CCF achieves in general a higher accuracy than current steganalytic methods especially for videos encoded in VBS and high QP values. Although it has been emphasized that CCF is applied to H.264/AVC videos, owing to the universality of PB-NOP, IIC and WOC, CCF can be easily extended to other video coding standards, such as MPEG-4 and MPEG-2.

For future work, the CCF will be further optimized. For instance, the new partition modes in partition based quantization method will be formed in an adaptive way. In addition, testing of more MV based steganography and comparison of CCF with more steganalytic methods are also on our agenda of future research.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Nos. U1536204, U1536114), and the National Key Technologies R&D Program of the Ministry of Science and Technology of China (No. 2014BAH41B00).

REFERENCES

[1] Hussein A Aly. 2011. Data hiding in motion vectors of compressed video based on their associated prediction error. *IEEE Trans. Inf. Forensics Security* 6, 1 (Mar. 2011), 14–18.

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

[3] Yun Cao, Xianfeng Zhao, Dengguo Feng, and Rennong Sheng. 2011. Video steganography with perturbed motion estimation. In *Proc. 13th Int. Conf. IH*, Vol. 6958. 193–207.

[4] Chih-Chung Chang and Chih-Jen Lin. 2015. LIBSVM: A Library for Support Vector Machines. (Feb. 2015). http://www.csie.ntu.edu.tw/~cjlin/libsvm

[5] Yu Deng, Yunjie Wu, and Linna Zhou. 2012. Digital video steganalysis using motion vector recovery-based features. *Appl. Opt.* 51, 20 (Jul. 2012), 4667–4677.

[6] Ding-Yu Fang and Long-Wen Chang. 2006. Data hiding for digital video with phase of motion vector. In *Proc. IEEE Int. Symp. Circuits Syst.* 1422–1425.

[7] Jessica Fridrich. 2004. Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes. In *Proc. 6th Int. Conf. IH*. 67–81.

[8] Jessica Fridrich and Jan Kodovskỳ. 2012. Rich models for steganalysis of digital images. *IEEE Trans. Inf. Forensics Security* 7, 3 (Jun. 2012), 868–882.

[9] Yang Hu, Chuntian Zhang, and Yuting Su. 2007. Information hiding based on intra prediction modes for H.264/AVC. In *Proc. IEEE Int. Conf. Multimedia Expo*. 1231–1234.

[10] Fred Jordan, Martin Kutter, and Touradj Ebrahimi. 1997. Proposal of a watermarking technique for hiding/retrieving data in compressed and decompressed video. *ISO/IEC Doc. JTC1/SC29/WG11 MPEG97/M2281* (Jul. 1997).

[11] Spyridon K Kapotas and Athanassios N Skodras. 2008. A new data hiding scheme for scene change detection in H.264 encoded video sequences. In *Proc. IEEE Int. Conf. Multimedia Expo*. 277–280.

[12] Jan Kodovskỳ and Jessica Fridrich. 2009. Calibration revisited. In *Proc. 11th ACM Multimedia Security Workshop*. 63–74.

[13] Ke Liao, Shiguo Lian, Zhichuan Guo, and Jinlin Wang. 2012. Efficient information hiding in H.264/AVC video coding. *Telecomm. Syst.* 49, 2 (2012), 261–269.

[14] Xiaojing Ma, Zhitang Li, Hao Tu, and Bochao Zhang. 2010. A data hiding algorithm for H.264/AVC video streams without intra-frame distortion drift. *IEEE Trans. Circuits Syst. Video Technol.* 20, 10 (Oct. 2010), 1320–1330.

[15] Tomáš Pevnỳ, Patrick Bas, and Jessica Fridrich. 2010. Steganalysis by subtractive pixel adjacency matrix. *IEEE Trans. Inf. Forensics Security* 5, 2 (Jun. 2010), 215–224.

[16] Yanzhen Ren, Liming Zhai, Lina Wang, and Tingting Zhu. 2014. Video steganalysis based on subtractive probability of optimal matching feature. In *Proc. 2nd ACM Workshop Inf. Hiding Multimedia Security*. 83–90.

[17] Iain E Richardson. 2011. *The H.264 advanced video compression standard*. John Wiley & Sons.

[18] Young-Ho Seo, Hyun-Jun Choi, Chang-Yeul Lee, and Dong-Wook Kim. 2008. Low-complexity watermarking based on entropy coding in H.264/AVC. *IEICE Trans. Fundamentals Electron. Commun. Comput. Sci.* E91.A, 8 (Aug. 2008), 2130–2137.

[19] Zafar Shahid, Marc Chaumont, and William Puech. 2013. Considering the reconstruction loop for data hiding of intra-and inter-frames of H.264/AVC. *Signal Image Video Process.* 7, 1 (Jan. 2013), 75–93.

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

[21] Joint Video Team. 2003. Advanced video coding for generic audiovisual services. *ITU-T Rec. H.264 and ISO/IEC 14496-10 AVC* (May 2003).

[22] Yiqi Tew and KokSheik Wong. 2014. An overview of information hiding in H.264/AVC compressed video. *IEEE Trans. Circuits Syst. Video Technol.* 24, 2 (Feb. 2014), 305–319.

[23] VideoLAN. 2015. x264. (Feb. 2015). http://www.videolan.org/developers/x264.html

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

[25] Thomas Wiegand, Heiko Schwarz, Anthony Joch, Faouzi Kossentini, and Gary J Sullivan. 2003. Rate-constrained coder control and comparison of video coding standards. *IEEE Trans. Circuits Syst. Video Technol.* 13, 7 (Jul. 2003), 688–703.

[26] Hao-Tian Wu, Yuan Liu, Jiwu Huang, and Xin-Yu Yang. 2014. Improved steganalysis algorithm against motion vector based video steganography. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*. 5512–5516.

[27] Changyong Xu, Xijian Ping, and Tao Zhang. 2006. Steganography in compressed video stream. In *Proc. 1st Int. Conf. Innov. Comput., Inf. Control*, Vol. 1. 269–272.

[28] Gaobo Yang, Junjie Li, Yingliang He, and Zhiwei Kang. 2011. An information hiding algorithm based on intra-prediction modes and matrix coding for H.264/AVC video stream. *AEU Int. J. Electron. Commun.* 65, 4 (Apr. 2011), 331–337.

[29] Hong Zhang, Yun Cao, and Xianfeng Zhao. 2016. Motion vector-based video steganography with preserved local optimality. *Multimedia Tools and Applications* 75, 21 (2016), 13503–13519.

[30] Hong Zhang, Yun Cao, Xianfeng Zhao, Weiming Zhang, and Nenghai Yu. 2014. Video steganography with perturbed macroblock partition. In *Proc. 2nd ACM Workshop Inf. Hiding Multimedia Security*. 115–122.