Hindawi
Security and Communication Networks
Volume 2024, Article ID 9824673, 1 page
https://doi.org/10.1155/2024/9824673

Retraction

Retracted: A Review of Motion Vector-Based Video Steganography

Security and Communication Networks

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Li, M. Zhang, K. Niu, and X. Yang, "A Review of Motion Vector-Based Video Steganography," *Security and Communication Networks*, vol. 2022, Article ID 2946812, 19 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 2946812, 19 pages
https://doi.org/10.1155/2022/2946812

Review Article

A Review of Motion Vector-Based Video Steganography

Jun Li,[1,2] **Minqing Zhang**,[1,2] **Ke Niu**,[1,2] **and Xiaoyuan Yang**[1,2]

[1]*Key Laboratory of Network and Information Security Under the Chinese People's Armed Police Force (PAP), Xi'an 710086, China*
[2]*College of Cryptography Engineering in Engineering University of PAP, Xi'an 710086, China*

Correspondence should be addressed to Minqing Zhang; api_zmq@126.com

Received 11 June 2022; Accepted 16 August 2022; Published 26 September 2022

Academic Editor: Yuchuan Luo

Steganography is a popular research direction in the field of information security. Due to the widespread use of video media, video steganography has received much attention from the research community. Among video steganography, motion vector (MV)-based video steganography has become one of the critical concerns of researchers due to its large embedding capacity and high visual quality. In this article, we focus on the research of MV-based video steganography. Firstly, the basic principles and evaluation criteria for MV-based steganography are discussed. Secondly, according to the different technical characteristics, the MV-based steganography is divided into three categories: the traditional MV domain steganography, the code-based MV domain steganography, and the adaptive MV domain steganography based on the framework of minimizing embedding distortion. The advantages and possible improvement directions of the above representative methods are illustrated. And then, the MV-based video steganalysis is outlined according to different perspectives of feature extraction, which is conducive to the design of better steganography algorithms. Finally, five future research directions are presented, such as designing distortion functions based on multiple factors, embedding methods based on new video coding standards, deep learning-based MV steganography, multi-domain embedding strategies, and moving the MV-based steganography from the laboratory into the real world.

1. Introduction

The development of modern information technology has dramatically changed how people obtain and transmit information. People protect information security by encrypting secret information into unintelligible ciphertext through cryptography. However, cryptography cannot conceal the existence of communication behavior, and information hiding technology can make up for this defect. The generalized information hiding technology achieves the purpose of information security by hiding the secret message into common carriers, which has two main branches: digital watermarking and digital steganography [1]. Digital watermarking is to embed authentication information into multimedia to achieve the purpose of copyright protection, and it is mainly concerned with embedding capacity and robustness. On the other hand, digital steganography embeds secret information in a common carrier to conceal the fact that communication is taking place without attracting the attention of third parties, thus achieving the purpose of covert communication, which is mainly concerned with embedding capacity and security. However, steganography also has the possibility of being misused by terrorists [2, 3], so the corresponding steganalysis technique was born. The main goal of steganalysis is to determine whether the detected object has steganographic traces or not. It achieves the purpose of discriminative classification with the help of knowledge from fields such as pattern recognition and machine learning. Steganography and steganalysis are two sides of a game, which confront and promote each other, and have made significant progress in the last two decades.

The typical carriers used for steganographic are text, image, audio, video, etc. Since images are very widely used, the research on image steganography is the earliest and most profound. Hence, the development of image steganography has some guidance to the development of audio, video, and text steganography. The development of image steganography has gone through four main stages: the first stage is

traditional steganography, typically the LSB (least significant bit) replacement algorithm [4], the PVD (pixel value differencing) algorithm [5], and the quantization index modulation algorithm [6]. Their primary purpose is to embed the information into the image without causing noticeable visual distortion. The second stage is based on steganography code [7–9], and the primary goal of this class of methods is to improve the embedding efficiency and to be able to resist the attacks of steganalysis with low-order statistical features. The third stage is adaptive steganography based on the framework of minimizing embedding distortion [10]. The core task of these methods is to assign a reasonable cost value to each carrier element in the image and then encode it using STCs (syndrome trellis codes) [10], polar codes [11], etc. These methods are convenient and have high statistical security, and the typical algorithms are HUGO (Highly Undetectable steGO) [12], UNIWARD (UNIversal WAvelet Relative Distortion) [13], etc. The fourth stage is the implementation of embedding by drawing on the research results in the field of deep learning, which has a promising future. There are two main research works in this stage. On the one hand, deep learning is used to learn the distortion function [14] automatically. On the other hand, the method is called generative-based steganography [15, 16], in which the stego images are usually generated automatically by neural networks without the modification process.

With the improvement of network bandwidth and the development of video coding standards, video-on-demand and live streaming services have rapidly gained popularity, and video has gradually replaced images as the most popular and adopted information transmission medium. Globally, the amount of time people spend viewing short videos per week is climbing [17], with data showing that 14.9% of millennials aged 26 to 35 watch 10 to 20 hours of online video per week as of August 2020. Moreover, according to the 48th China Internet Development Report [18], the size of Chinese online video users reached 944 million by June 2021, accounting for 93.4% of Internet users as a whole. Therefore, video media is considered an ideal vehicle for steganography compared to images and text. However, compared with image steganography, video steganography started late and developed slowly, and there are still many urgent scientific problems to be solved. The basic component unit of video is the image, so video steganography has many similarities with image steganography. However, due to its complex coding rules, video has more embedding domains suitable for steganography than images, so video steganography has many features that distinguish it from image steganography.

Video steganography can be classified into the spatial domain and the compressed domain according to whether it is compressed by an encoder or not. Uncompressed spatial domain video is similar to the image spatial domain, so the relevant algorithms are mainly based on image steganography. The application of spatial video steganography is limited because it is difficult to preserve the embedded information after compression, while compression domain video steganography mainly refers to the combination of the embedding process and the video compression process, which can be classified according to the video coding standards they use, mainly MPEG series [19], H.264/AVC [20], and H.265/HEVC [21]. In the past ten years, H.264/AVC has been the most widely used standard, while H.265 is expected to be promoted in the future. Although different video coding standards have different performances, they all use a hybrid coding framework, which usually contains techniques such as prediction, variation, quantization, entropy coding, intraframe prediction, interframe prediction, and loop filtering. Therefore, video steganography can be classified into intraframe prediction modes [22, 23], interframe prediction modes [24, 25], MVs [26, 27], transform coefficients [28, 29], quantization parameters [30], and entropy coding coefficients [31]. Table 1 lists the advantages and disadvantages of various embedding domain steganography in the video. Among these classifications, MV-based steganography has a larger embedding capacity because the compressed domain video has a large amount of MVs. Moreover, MV-based steganography is usually closely related to the encoding process. The embedding perturbation to the MV is handled automatically by the subsequent encoding process so that the MV-based steganography method can obtain better visual quality and coding efficiency. In short, the MV-based steganography has long received wide attention from researchers, so we focus on the MV-based steganography in this article.

Some literature has been reviewed on video steganography or steganalysis in recent years. Sadek et al. [32] summarized the early video steganography techniques. Zhang et al. [33] mainly summarized video steganalysis's research status and development direction for different embedding domains. Dalal et al. [34] reviewed the video steganography based on the spatial domain. Liu et al. [35] classified video steganography into intra-embedding, pre-embedding, and postembedding. They also summarized the reversible steganography and robust steganography. Dalal et al. [36] conducted a qualitative and quantitative analysis of video steganography and steganalysis. The experimental analysis of some prominent techniques using different quality metrics has also been performed. Patel et al. [37] provide a systematic overview of video steganography in the compressed and uncompressed domains. However, there is no published literature specifically summarizing MV-based video steganography techniques to the best of our knowledge. Moreover, the above review articles on MV-based steganography are not comprehensive enough. Therefore, to promote the development of MV-based video steganography, it is necessary to summarize and sort out the current status of research on MV-based steganography and steganalysis in recent years and discuss the possible future research directions, which would provide a reference for researchers in related fields. The remainder of this article is organized as follows: Section 2 presents the process of interframe predictive coding and the basic principles of MV-based steganography. Section 3 analyzes the development stages of MV-based video steganography and various types of steganographic embedding methods. Section 4 reviews the current research status of MV-based video steganalysis. The existing problems and possible future research

Table 1: Advantages and disadvantages of video steganography with different embedding domains.

Embedding domain	Cover	Embedding technology	Advantages	Disadvantages
Spatial domain	Pixels, transform domain coefficients	Modifies pixel or transform domain coefficients.	Independent of coding, can use research results from image steganography.	Distortion exists after compression.
Compressed domain	MV	Modify the horizontal or vertical component of the MV.	High embedding capacity, high visual quality, and no distortion drift.	Higher complexity.
	Intraframe prediction modes	Establishing the mapping relationship between intraframe prediction modes and secret message.	Lower complexity.	Usually used only in I-frames, lower embedding capacity, distortion drift.
	Interframe prediction modes	Establishing the mapping relationship between interframe prediction modes and secret message.	Embedding message in P-frames or B-frames.	Distortion drift.
	Transform domain coefficients	Modify the quantized DCT (discrete cosine transform) coefficients.	Can use research results from JPEG image steganography.	Low embedding capacity, distortion drift.
	Entropy coding coefficients	Modify the entropy coding coefficients.	Lower complexity, lower bit rate increment.	Low embedding capacity.

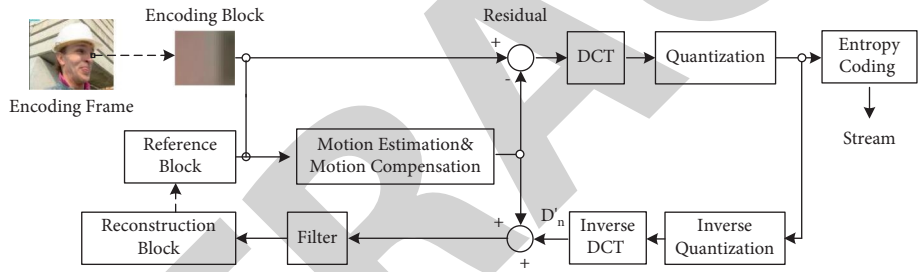

Figure 1: Interframe prediction for video coding.

directions are discussed in Section 5. Section 6 concludes the article.

2. Relevant Knowledge

Since MVs are generated in video interframe coding, this section first introduces the basic process of video interframe prediction coding, then describes the general principle of MV-based video steganography, and finally composes the common evaluation metrics of MV-based steganography.

2.1. Interframe Prediction for Video Coding. Most of the current video coding standards adopt a hybrid video coding framework, mainly consisting of intraframe prediction, interframe prediction, and entropy coding processes for compressing spatial redundancy, temporal redundancy, and statistical redundancy. Among them, temporal redundancy is the most extensive redundancy in the video because natural video consists of consecutive frames, and adjacent frames usually contain the same content between them, especially in scenarios such as surveillance and conferences. Interframe prediction coding can reduce these temporal redundancies, whose framework is shown in Figure 1. Since H.264 and HEVC videos are mainly used for steganography, this section focuses on these two standards' interframe prediction coding process.

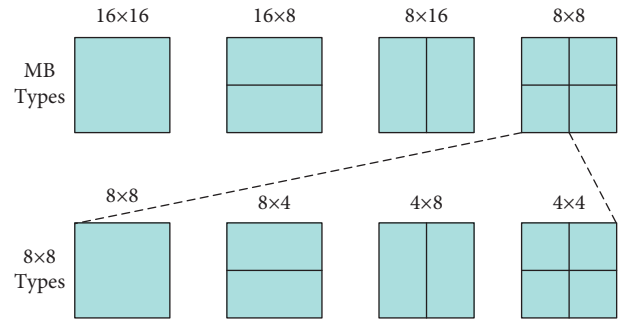

Figure 2: Segmentations of the macroblock for motion compensation (MC) in H.264/AVC. Top: segmentation of MB, bottom: segmentation of 8×8 partitions.

In the H.264 coding standard, the currently encoding frame is divided into 16×16 pixel-sized nonoverlapping MBs (MacroBlocks), and the luminance MBs are divided into various sizes such as 16×16, 16×8, 8×16, and 8×8, and the 8×8 MBs can continue to be divided into 8×8, 8×4, 4×8, or 4×4 sub-blocks, as shown in Figure 2. For encoding block B, the interframe prediction algorithm finds the most suitable

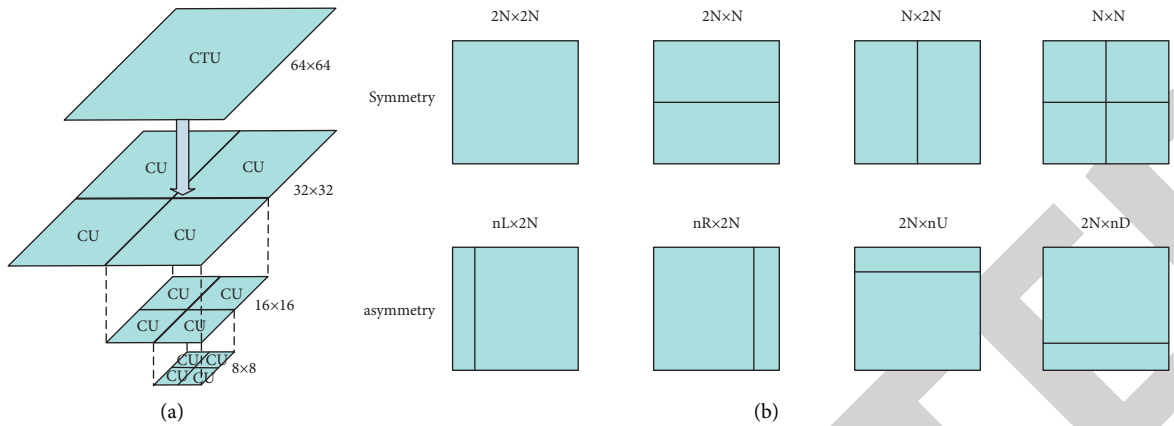

Figure 3: Partition modes of blocks in H.265/HEVC. (a) Subdivision of a CTU into CUs. (b) Modes for splitting a CU into PUs.

reference block T in the reference frame based on the Lagrangian rate distortion optimization model using motion estimation (ME):

$$J_{\text{motion}}(B, mv(h,v)) = D(B,T) + \lambda R(mv(h,v)), \quad (1)$$

where J_{motion} is the Lagrangian distortion of the interframe prediction and $D(B,T)$ is the pixel distortion between the encoding block B and the corresponding prediction block T. λ is the Lagrangian parameter to control the balance between the code rate and distortion. The relative distance of B and T is the MV, which contains the horizontal and the vertical component. $R(mv(h,v))$ is the number of bits needed to transmit the MV. On the one hand, the residual block outputs a video compressed stream after DCT transformation, quantization, and entropy coding. On the other hand, the quantified coefficients should be carried out by reverse quantified, and inverse DCT transformed to reconstruct the residual. Finally, the reconstructed residual was added with the prediction block to obtain the reconstructed block as the reference block for the subsequent encoded block.

Compared with H.264's macroblock model that fixed size, the HEVC coding standard adopts a more flexible way of dividing coding blocks. For each encoding frame, HEVC divides it into nonoverlapping coding tree units (CTUs), which are similar in concept to the MBs in H.264, and the size is specified by the encoder (usually 64×64). According to the quadratic tree division principle, each CTU can be further divided into smaller coding units (CUs) of 64×64, 32×32, 16×16, or 8×8, which are shown in Figure 3(a). In the interframe prediction mode, a CU with the size of $2N \times 2N$ can be divided into eight different sizes of prediction unit (PU) for MV prediction according to the symmetric and asymmetric approaches, as shown in Figure 3(b). It can be seen that HEVC has a more flexible interframe prediction mode than H.264. In addition, HEVC adopts new technologies such as AMVP (advanced MV prediction) to predict MVs, so there are richer ways to perform steganographic embedding on MVs.

2.2. The MV-Based Steganography. The process of MV-based video steganography is closely combined with the process of video compression coding, and its basic block diagram is shown in Figure 4. First, motion estimation and motion compensation are carried out according to the normal coding process to obtain the original MV, coding parameters, quantized DCT coefficients, and other information. Then, the original MV is modified according to the embedding algorithm to get the stego MV. Since the modified MV will affect the corresponding reference and reconstruction blocks, it is necessary to update the QDCT coefficients, coding parameters, and other information. Finally, the updated information is encoded to obtain the video code stream.

Specifically, MV-based video steganography takes the $mv(h,v)$ obtained according to motion estimation as the original cover, and then, the cover is modified by the embedding algorithm E:

$$\begin{aligned} mv(h',v') &= E(mv(h,v)) \\ &= mv(h \pm \Delta h, v \pm \Delta v), \end{aligned} \quad (2)$$

where Δh and Δv are 0 or positive integers, indicating the magnitude of modification, which usually should not be too large. Figure 5 shows the case after the MV is modified from $mv(h,v)$ to $mv(h',v')$. Obviously, the reference block T will change to T'. At the same time, the reconstruction block used for the subsequent reference will also be changed, thus affecting the whole subsequent encoding process. Therefore, the steganography embedding will inevitably affect the various original statistical properties of the MVs, etc., leaving space for possible attacks. The main goal of the MV-based steganographic algorithm is to ensure that the stego video is as close as possible to the original video in terms of visual quality, bit rate, and various statistical features. Steganalysis aims at destroying the covert communication by mining the statistical differences between the cover video and the stego video.

2.3. Performance Assessment Metrics for MV-Based Steganography. The core purpose of steganography is covert communication, so the most important metric to evaluate a video steganographic algorithm is statistical security, represented by the ability to resist steganalysis. The performance assessment metrics for MV-based steganography should also

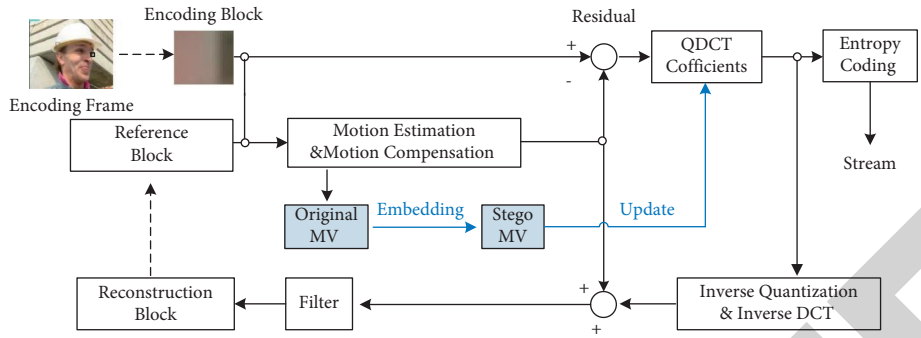

Figure 4: The block diagram of the MV-based steganography.

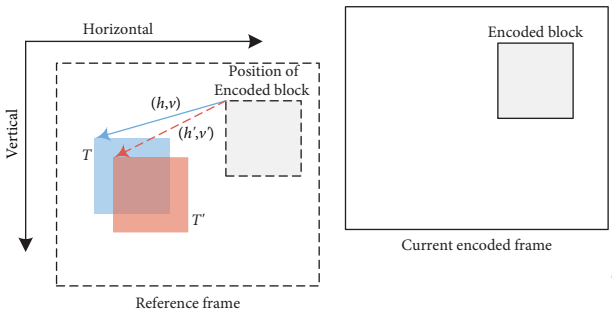

Figure 5: Modifying the MV in interframe prediction.

include embedding capacity, visual quality, coding efficiency, and computational complexity.

2.3.1. Embedding Capacity. The embedding capacity can be classified into absolute embedding capacity and relative embedding capacity. The primary absolute embedding capacity metric is bpf (bits per frame); that is, the number of bits that can be embedded into one frame. The relative embedding capacity is commonly used as cmvr (corrupted MV ration), representing the ratio of corrupted MVs' number to the total number of MVs in each frame. It is also possible to use bpmv (bits per MV), the average number of bits that can be embedded per MV. In addition, there is no MV in the situation when the macroblock is in skip mode, so the bpnsmv (bits per non-skip MV) is used as an evaluation metric. Since adaptive steganographic algorithms based on minimizing embedding distortion usually extract the lowest bit of the MV as the cover first and then combine it with STC coding for embedding, the relative capacity in STC coding can also be used to describe the overall embedding capacity.

Different MV-based steganography methods usually embed messages with different types of MVs as covers. For example, some methods select MVs with values larger than a certain threshold as covers, while some select only MVs with macroblocks divided into 16×16 or 8×8 as covers, and some embed for all MVs. The number of MVs varies widely for different encoding parameters, so in this case, we believe that it is unfair to conduct a cross-sectional comparison of different methods with relative embedding capacity. Therefore, an absolute embedding capacity (e.g., bpf) should be used for comparison.

2.3.2. Visual Quality. The visual quality of the compressed video is an important metric to judge an encoder, and the most widely used evaluation criteria are peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). Embedding operation inevitably impacts the visual quality, so MV-based steganography also uses these metrics to evaluate the algorithm. The PSNR represents the difference between the original video and the compressed reconstructed video in the pixel domain, which is defined as

$$\begin{aligned} \mathrm{PSNR} &= 10 \cdot \log \frac{\mathrm{MAX}^2}{\mathrm{MSE}}, \\ \mathrm{MSE} &= \frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} (f(i,j) - d(i,j))^2, \end{aligned} \tag{3}$$

where f and d are the encoded frame and the reconstructed decoded frame, respectively, with dimensions of $W \times X$. MSE is the mean square error of the encoded and decoded frames, and MAX is the maximum number of colors. Larger PSNR values indicate better visual quality, and this metric is simple to implement but does not reflect the visual properties of the human eye.

The other metric, SSIM, models distortion as three different factors: luminance, contrast, and structure. It is defined as

$$\mathrm{SSIM} = \frac{\left(2\mu_f\mu_d + c_f\right)\left(2\sigma_{fd} + c_d\right)}{\left(\mu_f^2 + \mu_d^2 + c_f\right)\left(\sigma_f^2 + \sigma_d^2 + c_d\right)}, \tag{4}$$

where μ_f and μ_d are the mean values of the original and decoded frames, $\sigma_f^2, \sigma_d^2, \sigma_{fd}$ are the variances of the original and decoded frames and their covariances. $c_f = (0.01 * \mathrm{MAX})^2$ and $c_d = (0.03 * \mathrm{MAX})^2$ are hyperparameters, and MAX is defined as above. The value range of SSIM is $[-1, 1]$, and the larger value means the better visual quality, which can reflect the visual characteristics of human eyes.

2.3.3. Video Bit Rate and Computational Complexity. The bit rate after video compression is a very important metric, and the core task of every new coding standard is to reduce the bit rate while maintaining the visual quality. Steganographic algorithms usually lead to an increase in the

video bit rate. To not attract the attention of attackers, steganographic algorithms should minimize the variation of bit rate. The impact of steganography on coding efficiency can be measured by bit rate increment (BRI), which is defined as follows:

$$\mathrm{BRI} = \frac{\mathrm{BR}_s - \mathrm{BR}_c}{\mathrm{BR}_c} \times 100, \tag{5}$$

where BR_c is the bit rate for normal encoding and BR_s is the bit rate for encoding with the steganographic embedding. The computational complexity of the steganographic algorithm is another important metric. Since video compression coding is a process of finding the optimal parameters among a large number of coding parameters, the encoder usually possesses a high complexity. For the steganographic algorithm to be applied to practical scenarios, the computational complexity of the steganographic algorithm should not be too high. And the computational complexity of steganography is usually described using the embedding speed.

3. Review of MV-Based Video Steganography

Referring to the development history of image steganography, we divide the development process of MV-based steganography into three stages. The first stage is traditional steganography, which is mainly implemented by modifying the magnitude of MV, MVD (MV difference), and phase of MV. The second stage is based on matrix coding [7], wet paper coding [8], and other steganographic codes to achieve the primary goal of improving embedding efficiency. The third stage is adaptive steganography based on minimizing embedding distortion [10], and these steganography methods have high embedding capacity and high security, which are still the current research hotspots.

3.1. Traditional MV-Based Steganography. Before the emergence of steganography code, the primary goal of MV-based steganography was to find suitable MVs for message embedding by setting up certain rules, which examined objects such as the magnitude of the MV, the magnitude of the prediction error of the encoding block, and the phase of the MV.

Xu et al. [38] designed a steganography method based on the magnitude of MVs in P-frames and B-frames with the MPEG coding standard. They concluded that the larger the magnitude value of the modified MVs, the less impact it brings to the cover in the coding process. First, the MVs whose magnitude values (the sum of the squares of the horizontal and vertical components) are greater than a certain threshold are selected as the embedding cover. Then, the decision of which component to modify is based on the magnitude of phase between horizontal and vertical components. Finally, the corresponding MVs' component is modified using LSB replacement. However, Aly et al. [39] argue that the magnitude of the MVs does not accurately reflect the effect of the steganographic perturbation, but the magnitude of the corresponding prediction residuals of the coding block reflects the steganographic perturbation. Those coding blocks with prediction residuals larger than a threshold are first selected. The MVs in these blocks are used as embedding covers, and steganographic embedding is performed using LSB substitution on both the horizontal and vertical components of the MVs. This method is equivalent to selecting those MVs corresponding to texturally complex regions as embedding covers, which can effectively improve the visual quality of stego videos.

Fang et al. [40] embed secret information by establishing the correspondence between the magnitude of the phase between two MVs' components and the secret messages. First, the MV with an amplitude greater than a threshold value is selected as the candidate. Then, the phase between the horizontal and vertical components in each MV is calculated. Finally, judging whether the phase angle difference between two adjacent sets of MVs satisfies the preset message mapping rule. If the mapping rule is satisfied, the corresponding two sets of MVs remain unchanged; otherwise, a new search for a new MV is required to realize the message embedding. This method does not use the LSBs of the MV as the embedding covers but adjusts the MV corresponding to the encoded block with the preset message mapping rule. Rana et al. [41] embed the secret message mainly in homogeneous regions. Since homogeneous or smooth regions contain macroblocks with similar prediction errors, it helps to reduce the detected by masking the embedding noise in adjacent macroblocks with similar prediction errors possibility. Van et al. [42] embed secret information in QDCT coefficients and MVs based on the HEVC standard, and the reduction in the PSNR is controlled within 1 db with better visual quality.

From the above literature, it can be seen that the early traditional MV-based steganography methods mainly focus on the improvement of embedding capacity and visual quality and pay less attention to statistical security. Although these algorithms cannot resist steganalysis attacks based on statistical properties, these MV candidate rules provide a good basis for the development of later steganography techniques.

3.2. Embedding Methods Based on Steganographic Codes. With the emergence of matrix codes [7], wet paper codes [8], and ZZW coding construction [9] in image steganography, MV-based video steganography has been developed rapidly. The primary goal of algorithms at this stage is to reduce the modification of covers by combining steganographic codes, thus improving the embedding efficiency and achieving the goal of higher security and visual quality.

In order to improve the embedding efficiency, Pan et al. [43] constructed (n, k) linear block codes based on the LSBs of MVs to embed secret messages. Hao et al. [44] proposed a low modification rate steganography method using matrix codes. Firstly, the candidate MVs were selected as embedding covers by the MVs' magnitude and phase, and the LSBs of $2^k - 1$ MVs are as a set of covers named A. After calculating $X = A \cdot H$, where H is the matrix with $(2^k - 1) \times k$, they take k secret messages $C = \{c_1 \dots c_k\}$ and then calculated $R = C \oplus X$. Finally, they modified 1 bit in A by judging the

value of R. This algorithm can achieve the purpose of embedding k bits of secret information with only modifying 1 bit in A, which effectively improves the embedding efficiency.

Cao et al. [45] aim at improving the algorithm's security by constructing MVs' suboptimal alternatives with wet paper codes. In the first step, they consider the motion estimation as a process of outputting the optimal prediction block for the current coding block. In order to achieve information embedding, a suboptimal prediction block can be filtered out according to the prediction residuals less than a certain threshold value. Then, the MVs corresponding to these coding blocks with suboptimal prediction blocks are "dry" covers, which can be used to embed secret messages, while the MVs corresponding to other blocks are "wet" covers, which cannot be used to embed messages. In the second step, the sender and the receiver share a key used to construct the random matrix. In the third step, the sender determines whether the MVs corresponding to the encoded blocks needs to be modified by computing a linear system of equations. If they need to be modified, the MVs are replaced by the MVs corresponding to the suboptimal prediction blocks. The advantage of this algorithm is that it achieves adaptive embedding, which allows secret embedding information in "dry" covers that result in better statistical security and visual quality. Based on the literature [45], Cao et al. [46] proposed a steganographic algorithm with better performance by obtaining more suitable suboptimal alternative MVs through multipath motion estimation and ZZW construction to achieve higher embedding efficiency.

Duan et al. [47] argued that steganographic algorithms that maintain the statistical features of MV residuals can maintain the spatiotemporal correlation of MVs. They combined variable-length matrix codes to embed secret messages in MVD, which is more secure in resisting attacks on spatiotemporal correlation features, but the limited embedding capacity. Yang et al. [48] proposed a space coding steganography method based on the HEVC standard. They gave the construction and encoding method of MV space. They defined the mapping relationship between the set of MVs and the points in this space, which can achieve the effect of embedding a $2N + 1$ binary number by changing at most one component among N MV components, with high embedding capacity.

The performance comparison of the above typical code-based steganographic algorithms is shown in Table 2. From these papers, it can be seen that these steganographic algorithms usually first select certain candidate MVs from all MVs as the covers to be embedded. And then, they combine with one of the codes to improve the embedding efficiency, achieving a certain degree of adaptive steganography embedding. Overall, these methods' visual quality and security are better than the traditional MV steganography.

3.3. MV-Based Adaptive Steganography Using Framework of Minimizing Embedding Distortion. Also inspired by the framework of minimizing embedding distortion in image steganography, the third stage of MV-based steganography is mainly based on the adaptive steganography method of this framework, which is the mainstream framework of the whole multimedia steganography direction at present. The basic idea is to minimize the overall distortion by assigning a cost value to each MV cover and then encoding it with STC. This class of methods dramatically improves the security of steganographic algorithms, and it facilitates the steganographic algorithms to move from the laboratory to the practical application scenarios. According to the design perspective of the distortion function, it can be divided into methods based on complexity, methods based on the MV's local optimality, and methods based on multiple factors.

3.3.1. Designing of Distortion Function Based on Complexity. In image steganography, it is a fundamental principle to prioritize embedding messages in those texture complexity regions. There is more high-frequency redundant information in texture regions than smooth regions, and steganalysis features are challenging to model in these regions. Therefore most adaptive image steganographic algorithms aim at embedding secret messages in texture complexity regions, such as HUGO [12], WOW [52], and S-UNIWARD [13] in the spatial domain, and J-UNIWARD [13] and UERD [53] in the compressed domain. From the perspective of information hiding, any steganographic algorithm adds a certain amount of noise to a specific digital cover, so that the higher the texture complexity (statistical complexity) of the original cover, the more complex the added noise will be detected and the less impact it will have on the statistical properties of the original cover. Similarly, in MV-based video steganography, treating MVs as ordinary digital covers, they have their unique statistical properties, both spatial and temporal. Thus, secret messages should be embedded in those statistical complexity regions.

Yao [54] et al. considered that modifying MVs would bring perturbations to their temporal and spatial correlations and leave spaces for steganalysis. They designed a distortion function based on the covariance matrix of MV residuals and interframe prediction errors and combined it with STC coding to achieve message embedding. In the first step, for the t-th video frame containing $H \times W$ inter-frame coding blocks, the horizontal and vertical components of the MV are constructed into matrices MVXt and MVYt with the dimension of $H \times W$, respectively. In the second step, for any element in MVXt and MVYt, the second-order difference arrays in four directions (horizontal, vertical, diagonal, and antidiagonal) are calculated to obtain the spatial statistics distortion. Similarly, the second-order difference arrays of horizontal and vertical components are computed in the time direction of adjacent frames to obtain the temporal statistical distortion. The MV's statistical distribution change (SDC) before and after modification is obtained based on the temporal and spatial distortions. The third step is to calculate the prediction error change (PEC) of the corresponding coding block before and after the MV's modification. In the fourth step, the final embedding distortion of the MV is calculated based on SDC and PEC, and the message embedding is performed with STCs. This scheme introduces the framework of minimizing embedding distortion to MV-

Table 2: The performance comparison of typical code-based steganographic algorithms.

Literature	Codes	Security	Visual quality (PSNR)	Embedding capacity
Pan et al. [43]	Linear block codes	Without test	Minimum 39.38 db for foreman sequence	Maximum 0.67 bpmv when using (6, 4) linear block codes
Hao et al. [44]	Matrix codes	Without test	Minimum 36.08 db for foreman sequence	Depends on k
Cao et al. [45]	Wet paper codes	High (resist [49, 50])	Average increase 0.49 db for foreman sequence	Average 33.2 bpf for foreman sequence
Cao et al. [46]	ZZW construction	High (resist [50, 51])	Average increase 0.61 db	Average 40 bpf
Yang et al. [48]	Space codes	High (resist [51])	High	Embedding a $2N+1$ binary number in N MV components

based steganography for the first time, which changes the traditional embedding model and improves the statistical safety and visual quality of stego videos. However, when calculating the statistical complexity of MVs, only the fixed size mode of macroblocks in the H.264 standard is considered, and the variable block size mode is not considered, which has a limited application. In addition, the algorithm has high computational complexity.

Wang et al. [55] gave a formal description related to MVs in the process of video coding and analyzed two factors of MVs, including local optimality and adjacency correlations. They pointed out that modifying the MV corresponding to a simple macroblock of texture does not easily cause significant local optimality anomalies. They also discussed the distribution law of the MV components, arguing that modifications that make the components closer to the mean of the distribution can better maintain correlation. They proposed a distortion function designing method (adaptive macroblock complex, AMC) based on the coding block's complexity. In addition, this scheme can also directly use the complexity of the coding block as the threshold value for whether to embed or not, without using coding methods such as STC, and thus has the characteristics of flexible usage and large embedding capacity. It is worth noting that the concept of complexity in this algorithm contains two aspects: one refers to the complexity of the pixel content, that is, the image texture complexity; and the other refers to the correlation of the MV itself. However takes MVs that correspond to smooth coding blocks for embedding, which does not apply to the case of variable size macroblock partition. Because regions with complex textures will have a finer division of coding blocks in H.264 and HEVC standards mean that there are more different MVs available for embedding in that coding block, thus having higher security under the same conditions. In addition, this algorithm chooses smooth regions for embedding to ensure that the local optimality of the MVs is not subject to large perturbations. However, with the appearance of other different types of steganalysis features [56, 57], the security of this algorithm will be reduced.

3.3.2. Designing of Distortion Function Based on Local Optimality. Macroscopically, video coding is an output process of optimal coding parameters. In a normal interframe coding process, from the coding side, the rate distortion $J_{\text{motion}}(B, mv(h, v))$ should be minimal after the motion estimation process determines the $mv(h, v)$; that is, the MV at the coding side is locally optimal. But after the steganography operation, this local optimality is likely to be disturbed. Therefore, some scholars have designed steganalysis features, such as AoSO (adding or subtracting one) feature sets [58], NPELO (near-perfect estimation for local optimality) feature sets [59], and generalized local optimality (GLO) feature sets [60]. These features are still effective for detecting steganography in the MV domain. Therefore, it is essential to consider whether the modified MVs can maintain local optimality when embedding messages from the perspective of designing steganographic algorithms. Table 3 lists the primary comparisons of typical algorithms.

Cao et al. [61] explored the possibility that the stego MV is still determined to be locally optimal based on the uncertainty of the surrounding SAD (Sum of Absolute Differences) matrix caused by video compression. They first defined a number of MVs with a "1-distance optimal neighbor" to measure the magnitude of distortion. And then proposed an adaptive video steganography method based on motion estimation perturbation optimization. This method tries to modify only those MVs that are still judged to be locally optimal after modification and constructs a double-layer embedding channel by combining wet paper codes (WPCs) [8] and STC [10], thus improving the security under AoSO's attacks. However, this method considers local optimality only from the perspective of SAD, not from the perspective of rate distortion, and thus cannot resist attacks with rate distortion local optimality features such as NPELO. In addition, the method may not have enough alternative MVs as carriers for embedding under high bit rate compression and has limited application in practical scenarios [64].

Cao et al. [62] argued that the output MVs conform to the local optimality of the surrounding SAD matrix from the encoder side but not necessarily from the decoder side. This is because different motion estimation algorithms may lead to searching for different MVs. Generally speaking, the information of the motion estimation algorithm used on the encoder side is not available on the decoder side, so there is uncertainty in the local optimality of the MVs. Therefore, they proposed that the embedding behavior of the secret

Table 3: Comparisons of typical algorithms based on local optimality.

Algorithms	Motivation	Distortion calculation	Codes	Against features
Cao et al. [61]	Uncertainty of the surrounding SAD matrix at the decoder	1-distance optimal neighbor	STC + WPC	AoSO
Cao et al. [62]	ME's uncertainty	The degree of ME's uncertainty	STC	AoSO
Zhang et al. [26]	Uncertainty of the surrounding SAD matrix at the decoder	The difference of Lagrangian rate distortion between original and modified MVs	STC	AoSO, SPOM [63], MVRB [51]

information should be confused with the different motion estimation behaviors. In this case, the steganalysis detector cannot distinguish whether the local optimality perturbation is caused by motion estimation or steganographic embedding.

Zhang et al. [26] proposed an MV-based steganography method called MVMPLO (Motion vector Modification with Preserved Local Optimality), which can guarantee the local optimality of modified MVs. This algorithm uses the difference Lagrangian rate distortion between the original MV and the alternative MV as the embedding cost value, which is more reasonable and effective. However, this method needs to search for candidate optimal MVs in a large range, which may lead to excessive modifications and thus bring about large perturbations in the spatiotemporal correlation of MVs. That is to say, although local optimality is guaranteed, it leads to the risk of being attacked by other statistical features.

3.3.3. Designing of Distortion Function considering Multiple Factors. Due to the complexity of interframe prediction coding, it is difficult to resist different types of steganalysis attacks when designing distortion functions using only complexity features or local optimality features. More and more MV-based steganographic algorithms consider multiple factors, such as complexity, local optimality, consistency within block groups, coding block prediction errors, etc., to obtain higher statistical security.

Wang et al. [64] designed a distortion function considering three factors, such as motion characteristics of video content, local optimality of MVs, and statistical distribution of MVs, to resist attacks from different steganalysis features. First, they claimed that embedding information in regions with rich motion characteristics is more beneficial to maintaining concealment. They measure motion characteristics based on the magnitude of MVs and the QP difference of neighboring macroblocks. Second, by combining the advantages in the literature [26, 61], they design a strategy to select candidate MVs adaptively, thus being able to resist the attack of AoSO features. Then, the second-order residuals of the MVs are constructed in the spatial and temporal domains, and the cost value representing the statistical distribution properties of the MVs was proposed. Finally, an adaptive integrated distortion function is designed by considering the three aspects. Experiments show that the security can be effectively improved against AoSO and MVRBR (MV reversion-based steganalysis revisited) [65]. However, the algorithm does not consider the case of variable macroblock size and thus has limited application in the real world.

Zhu et al. [66] considered the steganography system as a multiobjective optimization problem and designed the distortion function by considering the MV distribution correlation, local optimality, and reconstructed frame distortion. They used the CV (Coefficient of Variance) of the MV residuals to measure the statistical distribution properties of the MVs. The number of 0 values in the QDCT coefficients and the SATD (Sum of Absolute Transformed Difference) values of the coding block are used for calculating the distortion of local optimality. The SAD difference before and after the MV modification is used to calculate the reconstruction frame distortion. In addition, the algorithm treats the horizontal and vertical components of the MV as separate covers. They calculated the cost value of the horizontal component and performed steganography embedding first, then calculated the distortion of the vertical component and performed steganography embedding, which can effectively improve the security. However, this algorithm only aims to maintain Gaussian distribution when calculating the distortion of the statistical distribution of MVs. It does not give the calculation method of distortion when macroblock using the model of variable block size. Ghamsarian et al. [67], investigated the effect of the modified MVs and the cover order on the statistical properties of intraframe and interframe coding. The algorithm first finds the optimal alternative MV from all candidate MVs based on the principle of minimal change in Lagrangian rate distortion. It then calculates its spatial and temporal statistical distortion as steganographic coding distortion. The algorithm considers the case of variable size in macroblocks, which is beneficial for application in practical scenarios.

The MVC (motion vector consistency) feature sets proposed in the literature [57] point out that the MVs within the same block group are weakly correlated. And the MV values are often different, indicating that these MVs have weak consistency. However, the common ś1 operation in the embedding process will significantly change this MV consistency. In order to resist the attacks of MVC, Liu et al. [27] proposed the first algorithm that can resist MVC attacks. They considered MV consistency within a block group, statistical complexity, and local optimality. First, the degree of consistency of MV is described based on the number of identical MVs in the block group. Then, the complexity is measured using the difference between the MV and other MVs in the same block group. The local optimality of the MV before and after embedding is kept constant. Finally, a comprehensive distortion function is designed based on these three factors. The embedding is combined with a two-stage embedding strategy, i.e., the embedding of the horizontal and vertical components of the MVs is performed

Table 4: Comparison of distortion function considering multiple factors.

Algorithms	Factors to consider	Whether to consider variable size macroblock partition	Against features
Wang et al. [64]	Motion characteristics of video content, local optimality, statistical distribution	No	AoSO, MVRBR
Zhu et al. [66]	MV distribution correlation, local optimality, reconstructed frame distortion	No	NPELO
Ghamsarian et al. [67]	Statistical properties of intraframe, statistical properties of interframe, cover order, local optimality	Yes	AoSO, MVRB, NPELO
Liu et al. [27]	Consistency, statistical complexity, local optimality	Yes	NPELO, MVC
Li et al. [68]	Statistical complexity, local optimality, consistency	Yes	NPELO, MVC, CCF [56]

separately. The experimental results show that the algorithm significantly improves both security and coding efficiency.

From the above literature, it can be seen that the design of the distortion function must consider various factors due to the emergence of different types of steganalysis features. It can be summarized that the main factors that should consider are the statistical complexity, the local optimality, and the consistency. Based on the above observations, in our previous work [68], we proposed a method based on distortion design principles, which summarized three principles of distortion assignment: local optimality, consistency, and complexity priority. We designed three new distortion function assignment methods, respectively, and finally defined them as a joint distortion. The experimental results show that not only the three independent distortion assignment methods can effectively resist the corresponding steganalysis attacks, but also the final joint distortion can resist the attacks of the three steganalysis features simultaneously, in addition to obtaining good visual quality and coding efficiency.

Table 4 lists the comparison of distortion functions considering multiple factors. Although different algorithms consider roughly the same factors, the specific design details differ greatly, which indicates that different algorithms have some disagreement in defining distortion, and there is still space for research.

3.3.4. Other Methods for Designing Distortion Function. In addition to the above methods, scholars have also proposed different distortion function design methods from other perspectives. There are mainly methods based on multi-embedded domain strategy, methods based on HEVC coding characteristics, methods based on capacity allocation, and methods based on nonadditive distortion.

To make full use of the multiple embedding domain covers provided in video coding, Zhai et al. [69] proposed a video steganography method based on MVs and interframe prediction modes. They presented two embedding strategies: sequential embedding and simultaneous embedding. The multiple domain steganography methods can effectively improve capacity and security through reasonable capacity allocation and distortion function definition. However, there is a correlation between each embedding domain, and their mutual influence needs to be further studied and explored.

Guo et al. [70], for the HEVC coding standard, first counted the motion trend of each frame and established an MTB (motion trend-based) mapping strategy between the MV and the secret message. Then, they used the SATD difference before and after the MV modification as the embedding distortion. This algorithm only uses the SATD value without considering the overall rate distortion has some limitations. In the base of SAMVP (steganography by advanced MV prediction) [71], Liu et al. [72] proposed the Adaptive-SAMVP algorithm based on the HEVC standard by defining the distortion function and combining it with STC coding. Since AMVP encodes MVs by index numbers and MV residuals, they embed the information in the index values of the candidate list and use the bite rate difference between two candidate MVs to define the distortion function. Unlike the general information embedding based on the interframe motion estimation and compensation process, this algorithm's embedding and extraction process is only implemented in the entropy decoding process of the video stream of HEVC. Thus, the complexity is lower, and the MV is not directly modified, so there is no degradation in visual quality. However, this algorithm theoretically implements information embedding by selecting different candidate MVs, and thus it will destroy the local optimality of the candidate MVs, and there is a security risk.

Yao et al. [73] asserted that when the MVs are modified, it causes residual offset propagation for subsequent frames. Based on the residual offset propagation analysis results, they designed a capacity allocation strategy to try to allocate capacity to frames that cause less offset propagation, which helps to maintain the overall safety. It is worth noting that the algorithm does not design a new distortion function but improves security through a capacity adjustment strategy, which is theoretically applicable to all distortion functions. However, the algorithm essentially concentrates the embedding capacity on specific frames, which will have security crises if the attacker adopts an adaptive steganalysis strategy [74, 75].

Usually, embedding distortion is nonadditive since the impact of embedding in individual cover elements on the overall distortion is not independent. However, the optimal solution under nonadditive conditions is difficult to solve, and the optimal problem under additive conditions is well solved by codes such as STC. Therefore, in the initial research stage on steganographic algorithms, most scholars

design steganographic algorithms by assuming that distortion is additive, which is obviously out of touch with the actual situation. Many nonadditive steganographic algorithms appeared in the spatial domain [76–78] and JPEG domain [79, 80] in image steganography, which greatly promoted the development of steganography in a more efficient and practical direction. However, in video steganography, nonadditive research is still in its infancy [28, 81]. In the MV domain, Li et al. [82] designed a joint distortion function reflecting the influence of embedding for MVs based on the joint distortion in the image spatial domain [78]. The algorithm first transforms the joint distortion into a joint modification probability. It decomposes the joint modification probability into an edge modification probability for the horizontal component and a conditional modification probability for the vertical component. Then the two modification probabilities are transformed into the corresponding distortion values, and finally, the secret message is embedded into the two components of the MV using STCs. However, the algorithm only considers the interaction between MVs' horizontal and vertical components, but not the interaction between the elements in each component.

4. Review of MV-Based Video Steganalysis

As a rival of steganography, steganalysis aims at detecting whether the multimedia contains secret messages. The block diagram of MV-based steganalysis is shown in Figure 6. The basic process of MV-based video steganalysis is first to decode the video compressed stream and extract the statistical features related to MV modification from the decoding parameters. Then, train using a classifier and finally classify the detected objects and obtain the discriminative results

In image steganalysis, since the steganography operation mainly destroys the correlation between pixels or DCT coefficients, the designed steganalysis features are mainly used to reflect the correlation anomalies before and after steganography. And due to the complexity of video coding, the MV-based video steganography leads to the perturbation of different types of coding parameters, and the angles of extracting steganalysis features are more diverse and complex. Therefore, according to the starting point of feature extraction, MV-based video steganalysis can be divided into five categories: the first category is based on the spatiotemporal statistical properties of MVs [83, 84], and the motivation of this category is the existence of a correlation between MVs. The second category is based on MV calibration methods [56] because the MVs tend to recover to the original MVs after recompression of the stego video. The third category is based on the local optimality of the MV [58, 59]. Since the MV is a locally optimal output process in the sense of rate distortion, the steganography operation is likely to destroy it. The fourth category is steganalysis algorithms [57] designed based on the fact that MVs of subblocks in a macroblock are usually different. In addition, there are also steganalysis methods based on convolutional neural networks [85, 86], which is the fifth category.

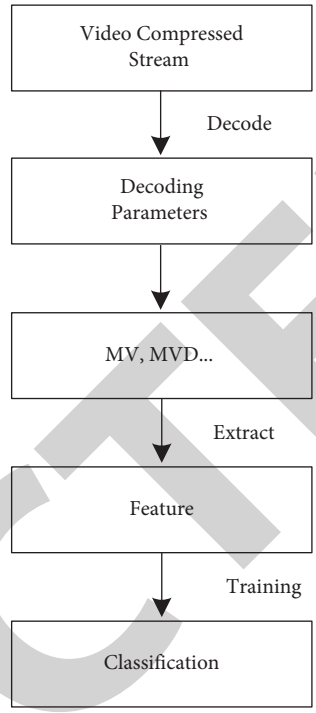

Figure 6: The block diagram of MV-based steganalysis.

4.1. Steganalysis Based on the SpatioTemporal Statistical Properties of MVs. Because of the strong correlation between adjacent coding blocks within a frame and between coding blocks at the same position between adjacent frames, the MVs have strong spatial and temporal correlations. Steganography will inevitably destroy this spatiotemporal correlation, so constructing statistical features of the difference in spatiotemporal correlation before and after steganography will effectively distinguish the cover video from the stego video.

For the fixed macroblock division, such as the MPEG-2 standard, Zhang et al. [50] and Su et al. [83] considered that the steganography operation is equivalent to adding additive noise to the horizontal and vertical components of the MV, respectively, and designed steganalysis feature sets based on the aliasing effect. The method extracts 3-dimensional features (including probability mass function and center of mass) from the temporal and spatial domains of the horizontal and vertical components, respectively. It can detect earlier conventional video watermarking algorithms that mainly modify the amplitude of MVs [87]. Based on the literature [83], Deng et al. [88] extended the MV first-order difference to second-order difference with improved performance.

Inspired by the rich model feature sets in image steganalysis [89] and combined with preprocessing techniques such as high-pass filtering, quantization, stage, and dimensionality reduction, Tasdemir et al. [84] proposed a 44785-dimensional video spatiotemporal rich model (STRM) feature sets, with higher correct detection accuracy. The literature [83, 84, 88] and others consider the temporal and spatial statistical properties of MVs for fixed macroblock

sizes. They cannot extract features for variable size macroblock divisions with limited applications. Li et al. [90] averaged all MVs within a macroblock to obtain an MV and then combined with the correlation network model to design steganalysis features, which could not consider the actual situation of variable sizes. In contrast, Wang et al. [91] proposed a four-way scanning method applicable to variable block size and designed 392-dimensional features based on the correlation anomaly of MVs, which can effectively detect methods such as those in the literature [39, 45]. In addition, Ghamsarian et al. [92] designed a method by considering intraframe statistical features, interframe statistical features, and local optimality of MVs, which further improved the performance of steganalysis.

From the above literature, it can be seen that steganalysis algorithms based on spatiotemporal statistical properties of MVs are mainly designed for early traditional MV-based steganography and fixed macroblock size division. Although high-dimensional features based on rich models have been widely used in image steganalysis, the current mainstream video coding standards usually use variable block size division, leading to difficulties for feature extraction. In addition, since video coding and decoding have high complexity, it is difficult to promote steganalysis algorithms with too high feature dimensions in practical applications. Therefore, how to design features with low dimensionality that can fully reflect spatiotemporal statistical features for variable block size division is a problem worthy of study.

4.2. Steganalysis Based on Calibration. The idea of calibration [93] is derived from JPEG image steganalysis, which refers to the fact that the coding parameters of JPEG images can be returned to the original state to some extent after recompression. For videos after MV-based embedding, it is possible to recover the original MVs by calibration techniques, which provides a basis for determining whether they have secret messages or not.

Cao et al. [51] found that after the calibration of the stego video with the same parameters as the first time, the MVs will show the nature of returning to the original values. Therefore, they designed the 15-dimensional feature sets of MVRB (MV reversion-based) based on the difference between the MV and prediction error before and after calibration. However, its performance is greatly affected by the encoding parameters. The second coding cannot achieve the detection purpose when it uses a different motion estimation algorithm and macroblock partition mode. Deng et al. [94] proposed calibration-based steganalysis features from the perspective of adjacent MV prediction, but there is still the problem of poor applicability. Therefore, Wang et al. [65], in order to solve the problem of encoding parameter mismatch, first collected various types of invariant encoding parameters (e.g., size and bit rate) and then obtained the best motion estimation algorithm by a search method, which has a certain performance improvement, but has high computational complexity.

To construct steganalysis features with rich statistical properties and applicable to various types of coding standards from multiple perspectives, Zhai et al. [56] constructed a joint calibration feature with a dimension of 124 from three aspects: neighborhood optimality of MVs, MVs' residual distribution, and MV calibration. The features contain optimality probability features based on segmentation neighborhood, inter- and intra-co-occurrence features based on the MV residuals, and window optimal MV calibration features. Since the algorithm considers several factors and macroblock's variable size, its steganalysis performance and applicability are strong and can be applied to mainstream coding standards. However, this algorithm did not consider the interaction between the locally optimal features and the statistical features of the residual distribution.

4.3. Steganalysis Based on Local Optimality. Video coding maintains visual quality and reduces the bit rate through a search process of optimal parameters, and is an output process of optimal coding parameters. The mainstream compression standards, such as H.264/AVC and H.265/HEVC, use a rate distortion optimization model based on the Lagrangian optimization algorithm to achieve interframe coding control. The goal of the encoder is to find the MV that minimizes $J_{\text{motion}} = D + R$, where D is the coding block distortion, and R is the bit rate required to transmit the coded information. Therefore, in the normal interframe coding process, from the encoder side, the rate distortion J_{motion} should be locally minimal after the motion estimation process determines the MV. However, this local optimality is likely to be disturbed after the embedding operation, so the attacker can design the steganalysis features based on this.

Wang et al. [58] performed the adding or subtracting one operation on the decoded MVs to obtain the candidate MVs. They proposed the 18-dimensional AoSO (adding or subtracting one) features based on the MV matrix and the reconstructed SAD matrix. AoSO can effectively detect the traditional MV-based and code-based algorithms. However, AoSO only uses the surrounding SAD matrix of the reconstructed block to determine whether the MV is locally optimal without considering other factors such as rate distortion. Therefore, in the literature [59], the local optimality of the MV is considered in the sense of rate distortion. Both SAD and SATD distortion measures are used to evaluate the value of distortion, and the 36-dimensional feature NPELO (near-perfect estimation for local optimality) is proposed. This feature can describe local optimality more accurately and effectively detect earlier steganographic algorithms designed based on local optimality [26, 62]. Ren et al. [63] used calibration technique to counting the variation of the local optimality of MVs, but the application is limited due to the mismatch of encoding parameters in the calibration technique itself.

Zhai et al. [60] proposed a steganalysis feature for generalized local optimality of H.264/AVC. The so-called generalized local optimality has two aspects. First, the local optimality measured in a rate distortion sense is jointly determined by MV and predicted motion vector (PMV). The variability of PMV will affect the estimation for local

optimality. Hence, they generalize the local optimality from a static estimation to a dynamic one. Second, they generalize the local optimality from the MV domain to the PMV domain. The proposed features effectively improve the MV-based steganographic algorithm's detection performance for H.264/AVC coding and is one of the best algorithms at present.

For the HEVC coding standard, MV-based steganography can only modify the MV index but not the MV itself [72] to embed messages. Therefore, the traditional MV-based steganalysis features are ineffective for this steganographic algorithm. However, if the MV index value is modified, the local optimality of the MVs in the index list will also be destroyed. Based on this observation, Liu et al. [95] constructed their steganalysis features based on local optimality on both the MVs and the MV candidate list, which effectively improved the detection performance in HEVC videos.

The above literature review shows that MV local optimality-based features are effective and can be applied to all coding standards. However, the starting point of such features is the assumption of an optimal parameter output process at the encoder side. However, in practical applications, due to lossy compression, the attacker at the decoder side cannot obtain accurate information at the encoder side. Therefore, the steganalysis features based on local optimality may have inevitable errors. It is a worthwhile direction to explore how to predict the original unknown information from the decoder side.

4.4. Steganalysis Based on MVs' Consistency. The current mainstream video coding standards usually adopt a variable block size for dividing macroblocks or coding tree units. In order to examine the relationship between individual MVs within a macroblock, Zhai et al. [57] defined the concepts of "big-block" and "small-block": if a coding block can be divided into multiple smaller blocks, the block is called a big-block; and if the sub-block that makes up a big-block is not further divided, such a block is called a small-block. All the small-blocks corresponding to the same big-block compose a small-block group. In addition, a block is said to have MV consistency if at least two horizontally or vertically adjacent small-blocks within the same group have identical values. They pointed out that in the cover H.264 video, the MVs within the same group are weakly correlated, and the MV values are often different, indicating that the MVs in the same group in the cover video have low MV consistency. The common ś1 operation in the embedding process will cause a significant change in this MV consistency. They proposed a 12-dimensional universal steganalysis features MVC (MV consistency) based on this phenomenon, which can detect video steganography in both the interframe prediction mode domain and the MV domain, achieving the best current detection accuracy.

Shanableh et al. [96] extended the MV consistency feature from the H.264 to HEVC standard. They redefined the concept of block group based on the coding depth according to the characteristics of the HEVC standard. They proposed steganalysis features based on MV consistency and coding unit residuals, which can effectively detect MV-based steganography in the HEVC standard.

4.5. Steganalysis Based on Convolutional Neural Network. Deep learning-based steganalysis has made significant progress in image steganalysis. Huang et al. [85] introduced convolutional neural networks to the quantitative steganalysis of MV video based on the HEVC standard. They proposed the VSRNet (Video Steganalysis Residual Network) network structure, whose input data contain the MV matrix and the prediction residual matrix. Independent VSRNet subnetworks are constructed for different embedding rates, and finally, all subnetworks are connected to form a quantitative steganalysis convolutional neural network capable of capacity estimation. They performed experimental validation for the traditional MV steganographic algorithm [38, 39] in the HEVC standard and obtained better results. Based on this, Huang et al. [86] further introduced the selection-channel-aware mechanism to improve the performance of steganalysis. The literature [85, 86] has made useful explorations in deep learning-based steganalysis of MV domains. However, it is still a challenge to propose more effective convolutional neural networks for minimizing embedding distortion in adaptive MV-based steganography.

5. Future Research Directions and Recommendations

According to the above literature review, the research on MV-based video steganography has made significant progress despite the late start. However, video steganography still has many issues that deserve further exploration due to the complexity of video coding and the emergence of new coding standards, which are shown in Figure 7.

5.1. Designing of Distortion Function considering Multiple Factors. Section 4 summarizes that the current mainstream MV-based steganalysis features are based on spatiotemporal statistical complexity, calibration differences, local optimality, consistency, etc. Therefore, the distortion functions summarized in Section 3.3 are also designed based on one or more of these factors. Since there are big differences in these steganalysis features, the design of distortion functions must consider all existing factors. The current main approach is to view the design of the distortion function as a multiobjective optimization problem. However, the interplay between factors has not been fully studied: how to optimize the distortion function, whether there is a conflict between factors, and how to deal with it if there is a conflict? For example, from the perspective of image content texture complexity, the more complex the texture, the finer the partition of macroblocks or coding tree units. It means that there are more MVs available as embedding covers, and the difference in MV magnitude/phase is also larger, so it is more favorable to maintain the spatiotemporal statistical properties of MVs. However, existing studies have shown

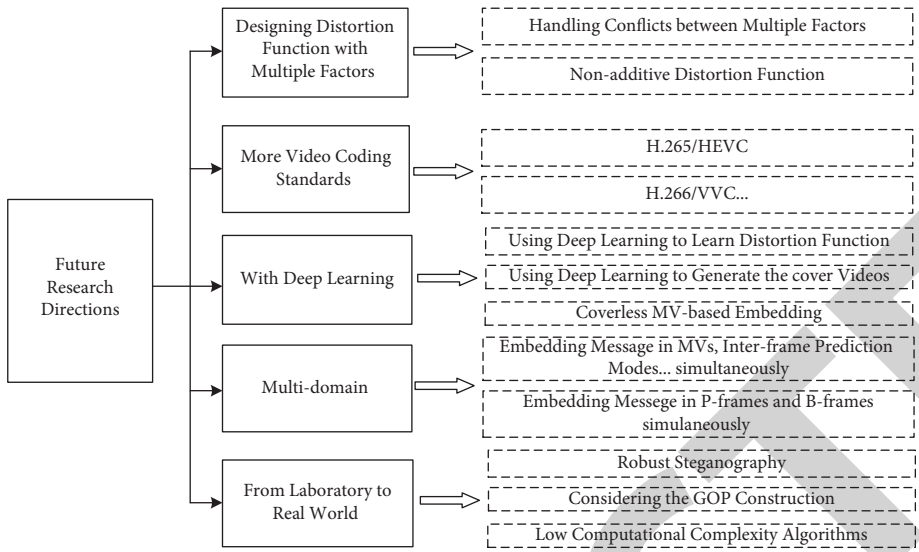

Figure 7: Future research directions.

that embedding in complex texture regions tends to lead to the destruction of local optimality of MVs [55], and steganography in smooth regions should be chosen, meaning that there exists conflict between the statistical properties of MVs and local optimality. We believe that similar conflicts exist between other factors as well. Therefore, it is an important research topic to propose algorithms that can maintain all kinds of statistical features simultaneously by fully considering the interplay between factors.

Secondly, most of the current distortion functions are based on additive assumptions, obviously out of touch with the actual situation. Compared with image nonadditive steganography, video MV-based nonadditive steganography has more challenges. Since the video coding adopts sequential coding, when the MV of a block in a frame is modified, the corresponding reconstruction block will be changed. This reconstruction block will provide a reference for the subsequent blocks, thus affecting the motion estimation and motion compensation of the subsequent blocks. Thus, it can be seen that the interplay between MVs is more significant, and how to fully exploit their interplay to design a nonadditive distortion function is a crucial issue worth exploring.

In addition, the design of all current distortion functions is based on the MVs of one frame; that is, STC coding can only achieve the overall optimum in one frame's range and cannot operate for the whole video sequence. That is to say, the current adaptive embedding only targets frames and does not consider the differences between different frames in the same video sequence. There are differences in complexity and motion speed between frames, so how to design a uniform distortion function for all MVs in all frames and use once STC coding for embedding is another crucial issue.

5.2. MV-Based Steganography Using Other Video Coding Standards. MV-based video steganography is still mainly focused on MPEG-2, H.264/AVC, while algorithms using H.265/HEVC and H.266/VVC [97] are less studied. At present, it is not easy to be replaced in a short time because of the broad application of H.264/AVC standards, but the promotion and application of new standards is an inevitable trend. In addition, the current research based on the new standard algorithm mainly focuses on DCT coefficients, intraframe prediction mode and interframe prediction mode [98], while the research based on MV is still rare.

New coding standards always require a huge improvement in compression efficiency over the previous standard; for example, both VVC and HEVC aim at doubling the coding efficiency over the previous generation, and thus, information redundancy will become less and less under the new standards. Essentially, information hiding is the embedding of secret information in data redundancy. On the one hand, as compression standards iterate, data redundancy becomes less and less, so theoretically, there will be fewer and harder "covers" for steganographic algorithms to embed information. Therefore, algorithms under the old standard may not be directly portable to the new standard. On the other hand, new coding standards are bound to introduce more coding techniques and more complex coding details, providing new entry points for steganographic algorithms. For example, the HEVC adopts advanced MV prediction (AMVP), so MV-based steganography can modify the MV and the index value of the corresponding MV candidate list. The interframe prediction in VVC is more elaborate, thus providing new opportunities and challenges for the design of steganographic algorithms.

5.3. MV-Based Steganography Using Deep Learning. Deep learning techniques have made breakthroughs in image steganography and steganalysis, and their performance has caught up with or even surpassed that of traditional algorithms. However, the application of deep learning technology in MV-based video steganography is still in its initial stage. We believe that the main reasons may be the

following: firstly, the complexity of video coding itself is high, and adding deep learning technology to the video coding process for message embedding will greatly improve the overall complexity of the algorithm. Secondly, steganography based on deep learning usually requires a large number of samples for model training. Currently, MV-based steganography usually takes the MVs in a frame as the basic embedded cover, and the number of MVs is limited.

Deep learning-based MV domain video steganography can be studied from the following aspects. One is the design of distortion functions based on deep learning. The current manual design of the distortion function usually needs to consider various factors. This way is subject to human experience interference, and it is difficult to achieve optimally. The use of deep neural networks (such as generative adversarial networks) to automatically learn distortion can effectively reduce the interference of human factors. Secondly, the deep learning technology directly generates the cover video sequence, based on which the existing adaptive steganography technology is used to embed, which can enrich the application scope of the existing MV domain steganography technology. Thirdly, using deep learning techniques to carry out coverless MV-based embedding (generative steganography). We can establish the mapping relationship between MVs and secret information without modifying original MVs, which can resist the current steganalysis attacks based on the statistical differences between the cover video and the stego video.

5.4. Multi-domain Video Steganography. Various coding elements in video coding can all be used as embedding covers. However, most current algorithms are usually based on single-domain embedding and do not fully utilize all embedding covers. More importantly, single-domain-based embedding algorithms ignore the fact that embedding in one domain also causes anomalies in the statistical properties of other domains. For example, the literature [57] demonstrated that embedding in the interframe prediction model leads to statistical anomalies in the MV domain. Therefore, it is important to make comprehensive use of the different characteristics of each embedding domain, to spread the steganographic capacity over different embedding covers.

Specifically, the MV domain is closely related to the interframe prediction mode domain. The division of interframe prediction patterns directly determines the number and distribution of MVs in macroblocks or coding tree cells. Their mutual constraint relationship can be fully utilized to design algorithms conducive to masking steganographic perturbation signals. In addition, except for interframe prediction coding blocks, there can also exist intraframe prediction coding blocks in P-frames or B-frames. The coding parameters of these intraframe prediction coding blocks can be fully utilized to effectively improve the steganographic capacity and security.

5.5. Moving Video Steganography from Laboratory Environment to Practical Application. In digital steganography and steganalysis, it has been an important concern for researchers to be able to apply the research results in the laboratory environment to the real-world environment [99]. Similarly, in MV-based steganography, there are still many "laboratory conditions" in the current research, and it is necessary to investigate algorithms that can be applied to real-world scenarios under more complex conditions.

The first one is about robust steganography. The videos that people upload to various platforms are usually compressed twice. Ensuring that the secret information can still be extracted normally after the secondary compression is a crucial issue. Although the research of video robust watermarking has been developed relatively mature, most of the current steganographic algorithms based on the MV domain do not consider robustness, limiting its application in practical scenarios. Therefore, how guaranteeing the embedding capacity, security, and robustness of MV-based steganographic algorithms is an important research direction.

The second one is about the GOP (group of pictures) structure in the video. Most current MV-based steganographic algorithms are usually studied in terms of GOP structure as IPPP. . ., and the size of GOP is usually within 15. In practice, there are usually tens or even hundreds of frames between two I-frames, which means that the residual perturbation generated by the MV modification in the previous frames of the GOP will propagate to the subsequent frames in the GOP. This perturbation propagation becomes more severe as the number of frames in the GOP increases. In addition, there are usually many B-frames in a GOP, and since the MVs in B-frames are obtained from the MVs of the two reference frames before and after, modifying the MVs in B-frames will affect more coding blocks. Therefore, it is also essential to fully consider the real-world GOP structure in the design of the steganographic algorithm.

The third one is about the computational complexity of the steganographic algorithm. Video coding, especially in real-time communication, is a technology with very high requirements for real-time performance. With the iteration of new video coding standards, the complexity is getting higher. The computational complexity of steganographic algorithms is also increasing due to the constant pursuit of security and its consideration of more and more factors. Suppose the complexity of the steganographic algorithm is too high and affects the normal video coding. In that case, it will not only be detrimental to the normal video coding process but also cause suspicion of attackers. Therefore, designing a lightweight MV-based video steganographic algorithm is also an important issue.

6. Conclusion

Although video steganography has received less attention than image steganography, the increasing proportion of video media on the Internet has contributed significantly to the development of video steganography, and many significant results have been achieved in recent years. This article presents a comprehensive overview of the basic principles and development process of MV-based video steganography, focusing on the research status and problems

of adaptive MV domain steganography based on minimizing embedding distortion. The typical MV-based steganalysis techniques were also reviewed from the perspective of feature extraction. Finally, because of the development status of video coding and the problems of existing algorithms, possible future research directions are elaborated, hoping to provide some reference for readers.

Data Availability

The data supporting this review are from previously reported studies and datasets, which have been cited.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant no. 61872384 and Basic Research Foundation of Engineering University of PAP under Grant no. WJY202140.

References

[1] K. T. Cox I, M. Miller, J. Bloom, and J. Fridrich, *Digital Watermarking and Steganography*, Morgan Kaufmann Publishers Inc, San Francisco, 2nd ed edition, 2008.

[2] B. Schneier, *Terrorists and Steganography*, ZDNet, San Francisco, CA, USA, 2001, http://www.zdnet.com/article/terrorists-and-steganography.

[3] The State of Security, "Hackers Exfiltrating Data with Video Steganography via Cloud Video Services," 2014, https://www.tripwire.com/state-of-security/incident-detection/hackers-exfiltrating-data-with-video-steganography-via-cloud-video-services.

[4] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding-a survey," in *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1062–1078, 1999.

[5] D.-C. Wu and W.-H. Tsai, "A steganographic method for images by pixel-value differencing," *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1613–1626, 2003.

[6] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[7] J. Fridrich and D. Soukal, "Matrix embedding for large payloads," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 3, pp. 390–395, 2006.

[8] J. Fridrich, M. Goljan, P. Lisoněk, and D. Soukal, "Writing on wet paper," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3923–3935, 2005.

[9] W. Zhang, J. Liu, X. Wang, and N. Yu, "Generalization and analysis of the paper folding method for steganography," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 694–704, Dec. 2010.

[10] T. Filler, J. Judas, and J. Fridrich, "Minimizing additive distortion in steganography using syndrome-trellis codes," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 920–935, 2011.

[11] W. Li, W. Zhang, L. Li, H. Zhou, and N. Yu, "Designing near-optimal steganographic codes in practice based on polar codes," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 3948–3962, 2020.

[12] T. Pevný, T. Filler, and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *Proceedings of the 2010 International Workshop on Information Hiding*, pp. 161–177, Calgary, Canada, June 2010.

[13] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP Journal on Information Security*, vol. 2014, no. 1, pp. 1–24, 2014.

[14] W. Tang, S. Tan, B. Li, and J. Huang, "Automatic steganographic distortion learning using a generative adversarial network," *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1547–1551, 2017.

[15] D. Volkhonskiy, I. Nazarov, and E. Burnaev, "Steganographic generative adversarial networks," in *Proceedings of the 12th International Conference on Machine Vision (ICMV 2019)*, p. 97, Amsterdam, The Netherlands, January 2020.

[16] D. Hu, L. Wang, W. Jiang, S. Zheng, and B. Li, "A novel image steganography method via deep convolutional generative adversarial networks," *IEEE Access*, vol. 6, pp. 38303–38314, 2018.

[17] Statista, "Weekly Time Spent with Online Video According to Internet Users Worldwide as of August 2020, by Age Group," 2020, https://www.statista.com/statistics/611750/millennial-time-spent-with-online-video/.

[18] China Internet Network Information Center, "The 48th Statistical Report on China's Internet Development," 2021, http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/202108/P020210827326243065642.pdf.

[19] T. Sikora, "MPEG digital video-coding standards," *IEEE Signal Processing Magazine*, vol. 14, no. 5, pp. 82–100, 1997.

[20] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[21] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[22] Y. Hu, C. Zhang, and Y. Su, "Information hiding based on intra prediction modes for H.264/AVC," in *Proceedings of the Multimedia and Expo, 2007 IEEE International Conference on*, pp. 1231–1234, Beijing, China, July 2007.

[23] Y. Wang, Y. Cao, X. Zhao, Z. Xu, and M. Zhu, "Maintaining rate-distortion optimization for IPM-based video steganography by constructing isolated channels in HEVC," in *Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security*, pp. 97–107, Innsbruck Austria, June 2018.

[24] S. K. Kapotas and A. N. Skodras, "A New Data Hiding Scheme for Scene Change Detection in H.264 Encoded Video Sequences," in *Proceedings of the 2008 IEEE International Conference on Multimedia and Expo*, pp. 277–280, Hannover, Germany, June 2008.

[25] H. Zhang, Y. Cao, X. Zhao, W. Zhang, and N. Yu, "Video steganography with perturbed macroblock partition," in *Proceedings of the 2nd ACM Workshop on Information Hiding and Multimedia Security–IH&MMSec'14*, pp. 115–122, Salzburg Austria, June 2014.

[26] H. Zhang, Y. Cao, and X. Zhao, "Motion vector-based video steganography with preserved local optimality," *Multimedia Tools and Applications*, vol. 75, no. 21, pp. 13503–13519, 2016.

[27] Y. Liu, J. Ni, W. Zhang, and J. Huang, "A Novel Video Steganographic Scheme Incorporating the Consistency Degree of Motion Vectors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 7, pp. 4905–4910, 2021.

[28] Y. Wang, Y. Cao, and X. Zhao, "CEC: cluster embedding coding for H.264 steganography," *IEEE Signal Processing Letters*, vol. 27, no. c, pp. 955–959, 2020.

[29] Y. Chen, H. Wang, K. K. R. Choo et al., "DDCA: a distortion drift-based cost assignment method for adaptive video steganography in the transform domain," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 4, pp. 2405–2420, 2022.

[30] T. Shanableh, "Data hiding in MPEG video files using multivariate regression and flexible macroblock ordering," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 455–464, Apr. 2012.

[31] C. Di Laura, D. Pajuelo, and G. Kemper, "A novel steganography technique for SDTV-H.264/AVC encoded video," *International Journal of Data Mining and Bioinformatics*, vol. 2016, Article ID 6950592, 9 pages, 2016.

[32] M. M. Sadek, A. S. Khalifa, and M. G. M. Mostafa, "Video steganography: a comprehensive review," *Multimedia Tools and Applications*, vol. 74, no. 17, pp. 7063–7094, 2015.

[33] H. Zhang, W. You, and X. Zhao, "A survey of video steganalysis," *Journal of Cyber Security*, vol. 3, no. 6, pp. 13–27, 2018.

[34] M. Dalal and M. Juneja, "Video steganography techniques in spatial domain—a survey," in *Proceedings of the International Conference on Computing and Communication Systems*, vol. 24, pp. 705–711, 2018.

[35] Y. Liu, S. Liu, Y. Wang, H. Zhao, and S. Liu, "Video steganography: a review," *Neurocomputing*, vol. 335, pp. 238–250, 2019.

[36] M. Dalal and M. Juneja, "A Survey on Information Hiding Using Video Steganography," *Artificial Intelligence Review*, vol. 54, no. 8, 2021.

[37] R. Patel, K. Lad, and M. Patel, "Study and investigation of video steganography over uncompressed and compressed domain: a comprehensive review," *Multimedia Systems*, vol. 27, no. 5, pp. 985–1024, 2021.

[38] C. Xu, X. Ping, and T. Zhang, "Steganography in compressed video stream," in *Proceedings of the 2006 1st International Conference on Innovative Computing, Information and Control–Volume I (ICICIC'06)*, pp. 269–272, Beijing, China, September 2006.

[39] H. A. Aly, "Data hiding in motion vectors of compressed video based on their associated prediction error," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pp. 14–18, 2011.

[40] D. Fang and L. Chang, "Data hiding for digital video with phase of motion vector," in *Proceedings of the 2006 IEEE International Symposium on Circuits and Systems (ISCAS)*, p. 4, Kos, Greece, May 2006.

[41] S. Rana, R. Kamra, and A. Sur, "Motion vector based video steganography using homogeneous block selection," *Multimedia Tools and Applications*, vol. 79, no. 9-10, pp. 5881–5896, 2020.

[42] L. P. Van, J. De Praeter, G. Van Wallendael, J. De Cock, and R. Van de Walle, "Out-of-the-loop information hiding for HEVC video," in *Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP)*, pp. 3610–3614, Quebec, Canada, September 2015.

[43] F. Pan, L. Xiang, X.-Y. Yang, and Y. Guo, "Video steganography using motion vector and linear block codes," in *Proceedings of the 2010 IEEE International Conference on Software Engineering and Service Sciences*, pp. 592–595, Beijing, China, July 2010.

[44] B. Hao, L. Zhao, and W. Zhong, "A novel steganography algorithm based on motion vector and matrix encoding," in *Proceedings of the 2011 IEEE 3rd International Conference on Communication Software and Networks, ICCSN*, pp. 406–409, Xi'an, China, May 2011.

[45] Y. Cao, X. Zhao, D. Feng, and R. Sheng, "Video steganography with perturbed motion estimation," *Lecture Notes in Computer Science*, vol. 6958, pp. 193–207, 2011.

[46] Y. Cao, X. Zhao, F. Li, and N. Yu, "Video steganography with multi-path motion estimation," *SPIE Proceedings*, vol. 8665, 2013.

[47] D. Ran and C. Dan, "Video steganography algorithm uses motion vector difference as carrier," *Journal of Image and Graphics*, vol. 23, no. 2, pp. 163–173, 2018.

[48] J. Yang and S. Li, "An efficient information hiding method based on motion vector space encoding for HEVC," *Multimedia Tools and Applications*, vol. 77, no. 10, pp. 11979–12001, 2018.

[49] G. Xuan, Y. Q. Shi, J. Gao et al., "Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions," in *Science and Technology*, M. Barni, J. Herrera-Joancomartí, S. Katzenbeisser, and F. Pérez-González, Eds., pp. 262–277, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[50] C. Zhang, Y. Su, and C. Zhang, "A new video steganalysis algorithm against motion vector steganography," in *Proceedings of the 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 4–7, Dalian, China, October 2008.

[51] Y. Cao, X. Zhao, and D. Feng, "Video steganalysis exploiting motion vector reversion-based features," *IEEE Signal Processing Letters*, vol. 19, no. 1, pp. 35–38, 2012.

[52] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," in *Proceedings of the 2012 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 234–239, Tenerife, Spain, December 2012.

[53] L. Guo, J. Ni, W. Su, C. Tang, and Y. Q. Shi, "Using statistical image model for JPEG steganography: uniform embedding revisited," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 12, pp. 2669–2680, 2015.

[54] Y. Yao, W. Zhang, N. Yu, and X. Zhao, "Defining embedding distortion for motion vector-based video steganography," *Multimedia Tools and Applications*, vol. 74, no. 24, pp. 11163–11186, 2015.

[55] L. Wang, Y. Xu, L. Zhai, and Y. Ren, "An adaptive video motion vector steganography based on macroblock complexity," *Chinese Journal of Computers*, vol. 40, no. 5, pp. 1044–1056, 2017.

[56] L. Zhai, L. Wang, and Y. Ren, "Combined and calibrated features for steganalysis of motion vector-based steganography in H.264/AVC," in *Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security*, pp. 135–146, Philadelphia, PA, USA, June 2017.

[57] L. Zhai, L. Wang, and Y. Ren, "Universal detection of video steganography in multiple domains based on the consistency

of motion vectors," *IEEE Transactions on Information Forensics and Security*, vol. 15, no. c, pp. 1762–1777, 2020.

[58] K. Wang, H. Zhao, and H. Wang, "Video steganalysis against motion vector-based steganography by adding or subtracting one motion vector value," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 741–751, 2014.

[59] H. Zhang, Y. Cao, and X. Zhao, "A steganalytic approach to detect motion vector modification using near-perfect estimation for local optimality," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 2, pp. 465–478, 2017.

[60] L. Zhai, L. Wang, Y. Ren, and Y. Liu, "Generalized Local Optimality for Video Steganalysis in Motion Vector Domain," pp. 1–13, 2021, https://arxiv.org/abs/2112.11729.

[61] Y. Cao, H. Zhang, X. Zhao, and H. Yu, "Video steganography based on optimized motion estimation perturbation," in *Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security*, pp. 25–31, Portland, OR, USA, June 2015.

[62] Y. Cao, H. Zhang, X. Zhao, and H. Yu, "Covert communication by compressed videos exploiting the uncertainty of motion estimation," *IEEE Communications Letters*, vol. 19, no. 2, pp. 203–206, 2015.

[63] Y. Ren, L. Zhai, L. Wang, and T. 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*, pp. 83–90, Salzburg, Austria, June 2014.

[64] P. Wang, H. Zhang, Y. Cao, and X. Zhao, "A novel embedding distortion for motion vector-based steganography considering motion characteristic, local optimality and statistical distribution," in *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security*, pp. 127–137, Vigo, Spain, June 2016.

[65] P. Wang, Y. Cao, X. Zhao, and B. Wu, "Motion vector reversion-based steganalysis revisited," in *Proceedings of the 2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, pp. 463–467, Chengdu, China, July 2015.

[66] B. Zhu and J. Ni, "Uniform embedding for efficient steganography of H.264 video," in *Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 1678–1682, Athens, Greece, October 2018.

[67] N. Ghamsarian and M. Khademi, "Undetectable video steganography by considering spatio-temporal steganalytic features in the embedding cost function," *Multimedia Tools and Applications*, vol. 79, no. 27-28, pp. 18909–18939, 2020.

[68] J. Li, M. Zhang, K. Niu, and X. Yang, "Investigation on Principles for Cost Assignment in Motion Vector-Based Video Steganography," pp. 1–16, 2022, https://arxiv.org/abs/2209.01744.

[69] L. Zhai, L. Wang, and Y. Ren, "Multi-domain embedding strategies for video steganography by combining partition modes and motion vectors," in *Proceedings of the 2019 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1402–1407, Shanghai, China, July 2019.

[70] M. Guo, T. Sun, X. Jiang, Y. Dong, and K. Xu, "A motion vector-based steganographic algorithm for HEVC with MTB mapping strategy," in *Proceedings of the 2019 International Workshop on Digital Watermarking*, pp. 293–306, Chengdu, China, November 2019.

[71] Y. Hu, W. Gong, F. Liu, L. Liu, and M. Zhu, "Large-capacity lossless HEVC information hiding based on index parameter modification," *Journal of South China University of Technology*, vol. 46, no. 5, pp. 1–8, 2018.

[72] S. Liu, B. Liu, Y. Hu, and X. Zhao, "Non-degraded adaptive HEVC steganography by advanced motion vector prediction," *IEEE Signal Processing Letters*, vol. 28, pp. 1843–1847, 2021.

[73] Y. Yao and N. Yu, "Motion vector modification distortion analysis-based payload allocation for video steganography," *Journal of Visual Communication and Image Representation*, vol. 74, Article ID 102986, 2021.

[74] P. Wang, Y. Cao, X. Zhao, and H. Yu, "An adaptive detecting strategy against motion vector-based steganography," in *Proceedings of the 2015 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, Torino, Italy, June 2015.

[75] P. Wang, Y. Cao, and X. Zhao, "Segmentation Based Video Steganalysis to Detect Motion Vector Modification," *Security and Communication Networks*, vol. 2017, Article ID 8051389, pp. 1–12, 2017.

[76] B. Li, M. Wang, X. Li, S. Tan, and J. Huang, "A strategy of clustering modification directions in spatial image steganography," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 9, pp. 1905–1917, 2015.

[77] T. Denemark and J. Fridrich, "Improving steganographic security by synchronizing the selection channel," in *Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security*, pp. 5–14, Portland, OR, USA, June 2015.

[78] W. Zhang, Z. Zhang, L. Zhang, H. Li, and N. Yu, "Decomposing joint distortion for adaptive steganography," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 10, pp. 2274–2280, 2017.

[79] Y. Wang, W. Li, W. Zhang, X. Yu, K. Liu, and N. Yu, "BBC++: enhanced block boundary continuity on defining non-additive distortion for JPEG steganography," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 5, pp. 2082–2088, May 2021.

[80] Y. Wang, W. Zhang, W. Li, and N. Yu, "Non-additive cost functions for JPEG steganography based on block boundary maintenance," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1117–1130, 2021.

[81] Y. Wang, Y. Cao, and X. Zhao, "Minimizing embedding impact for H.264 steganography by progressive trellis coding," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 333–345, 2021.

[82] L. Li, Y. Yao, X. Zhang, W. Zhang, and N. Yu, "Video steganography based on modification probability transformation and non-additive embedding distortion," *Journal of Electronics & Information Technology*, vol. 42, no. 10, pp. 2357–2364, 2020.

[83] Y. Su, C. Zhang, and C. Zhang, "A video steganalytic algorithm against motion-vector-based steganography," *Signal Processing*, vol. 91, no. 8, pp. 1901–1909, 2011.

[84] K. Tasdemir, F. Kurugollu, and S. Sezer, "Spatio-temporal rich model-based video steganalysis on cross sections of motion vector planes," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3316–3328, 2016.

[85] X. Huang, Y. Hu, Y. Wang, B. Liu, and S. Liu, "Deep learning-based quantitative steganalysis to detect motion vector embedding of HEVC videos," in *Proceedings of the 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC)*, pp. 150–155, Hong Kong, China, July 2020.

[86] X. Huang, Y. Hu, Y. Wang, B. Liu, and S. Liu, "selection-channel-aware deep neural network to detect motion vector embedding of HEVC videos," in *Proceedings of the 2020 IEEE International Conference on Signal Processing,*

Communications and Computing (ICSPCC), pp. 1–6, Xi'an, China, August 2020.

[87] J. Zhang, H. Maitre, J. Li, and L. Zhang, "Embedding watermark in MPEG video sequence," in *Proceedings of the 2001 IEEE Fourth Workshop on Multimedia Signal Processing (Cat. No.01TH8564)*, pp. 535–540, Cannes, France, October 2001.

[88] Y. Deng, Y. Wu, H. Duan, and L. Zhou, "Digital video steganalysis based on motion vector statistical characteristics," *Optik*, vol. 124, no. 14, pp. 1705–1710, 2013.

[89] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 868–882, 2012.

[90] S. Li, J. Yang, P. Liu, and L. Wang, "Steganalysis of motion vector-based steganography in H.264/AVC by correlation network model," *Journal of Applied Sciences—Electronics and Information Engineering*, vol. 37, no. 5, pp. 663–672, 2019.

[91] L. Wang, M. J. Wang, L. M. Zhai, and Y. Z. Ren, "H.264/AVC video steganalysis algorithm based on motion vector abnormal correlation," *Acta Electronica Sinica*, vol. 42, no. 8, pp. 1457–1464, 2014.

[92] N. Ghamsarian, K. Schoeffmann, and M. Khademi, "Blind MV-based video steganalysis based on joint inter-frame and intra-frame statistics," *Multimedia Tools and Applications*, vol. 80, no. 6, pp. 9137–9159, 2021.

[93] J. Fridrich, "Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes," *Lecture Notes in Computer Science*, vol. 3200, pp. 67–81, 2004.

[94] Y. Deng, Y. Wu, and L. Zhou, "Digital video steganalysis using motion vector recovery-based features," *Applied Optics*, vol. 51, no. 20, pp. 4667–4677, 2012.

[95] S. Liu, Y. Hu, B. Liu, and C. T. Li, "An HEVC steganalytic approach against motion vector modification using local optimality in candidate list," *Pattern Recognition Letters*, vol. 146, pp. 23–30, 2021.

[96] T. Shanableh, "Feature extraction and machine learning solutions for detecting motion vector data embedding in HEVC videos," *Multimedia Tools and Applications*, vol. 80, no. 18, pp. 27047–27066, 2021.

[97] M. Wien and B. Bross, "Versatile video coding - algorithms and specification," in *Proceedings of the 2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pp. 1–3, Macau, China, December 2020.

[98] J. Liu, Z. Li, X. Jiang, and Z. Zhang, "A high-performance CNN-applied HEVC steganography based on diamond-coded PU partition modes," *IEEE Transactions on Multimedia*, vol. 24, pp. 2084–2097, 2022.

[99] A. D. Ker, P. Bas, R. Bohme et al., "Moving steganography and steganalysis from the laboratory into the real world," in *Proceedings of the first ACM workshop on Information hiding and multimedia security–IH&MMSec'14*, p. 45, Montpellier, France, June 2013.