# *IET Image Processing*

## Special issue

## Call for Papers

**Be Seen. Be Cited.**
**Submit your work to a new IET special issue**

**"Advancements in Fine Art Pattern Extraction and Recognition"**

**Guest Editors: Fabio Bellavia, Gennaro Vessio, Giovanna Castellano and Sinem Aslan**

**Read more**

The Institution of Engineering and Technology

Received: 10 October 2021 | Revised: 23 May 2022 | Accepted: 14 June 2022

IET Image Processing

DOI: 10.1049/ipr2.12573

The Institution of Engineering and Technology WILEY

REVIEW

# A comprehensive review of video steganalysis

**Mourad Bouzegza**[1,2] | **Ammar Belatreche**[2] | **Ahmed Bouridane**[3] | **Mohamed Tounsi**[1]

[1]College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia

[2]Department of Computer Science and Digital Technologies, Engineering and Environment, Northumbria University at Newcastle, UK

[3]College of Computing and Informatics, Sharjah University, Sharjah, United Arab Emirates

**Correspondence**
Mourad Bouzegza, College of Computer and Information Sciences, Prince Sultan University, P.O.Box No. 66833, Rafha Street, Riyadh 11586, Saudi Arabia.
Email: mbouzegza@psu.edu.sa

**Funding information**
Prince Sultan University

**Abstract**

Steganography is the art of secret communication and steganalysis is the art of detecting the hidden messages embedded in digital media covers. One of the covers that is gaining interest in the field is video. Presently, the global IP video traffic forms the major part of all consumer Internet traffic. It is also gaining attention in the field of digital forensics and homeland security in which threats of covert communications hold serious consequences. Thus, steganography technicians will prefer video to other types of covers like audio files, still images, or texts. Moreover, video steganography will be of more interest because it provides more concealing capacity. Contrariwise, investigation in video steganalysis methods does not seem to follow the momentum even if law enforcement agencies and governments around the world support and encourage investigation in this field. In this paper, the authors review the most important methods used so far in video steganalysis and sketch the future trends. To the best of the authors' knowledge this is the most comprehensive review of video steganalysis produced so far.

## 1 | INTRODUCTION

As is well known in the field, steganography is the art of secret communication. It is the art and science of hiding information by embedding messages within a digital cover, which could be a text, an image, or a video. Many steganography methods exist in order to let the loaded messages remain seemingly harmless. As an obvious and a priori condition, the best carrier for secret messages must possess two features. Firstly, the carrier should be popular. Secondly, the steganogram insertion-related modifications of the carrier should not be 'visible' to the third party not aware of the steganographic procedure [1], whereas steganalysis is the art of detecting the hidden messages embedded in digital media covers by the means of steganography. The steganalysts are usually something of forensic statisticians, and must start by reducing the suspect set of data files to the subset most likely to have been altered [2]. In addition to the forensics and homeland security use [3], the steganalysis strategies are also beneficial in civilian applications by detecting all kinds of malware meant to harm the computers [4].

So far, images are probably the cover that received the biggest part of investigators effort in steganalysis. Nevertheless and as the digital video is overwhelming the Internet traffic [5], researchers start working on it in order to develop new methods because the ones developed for image steganalysis are not always easily adaptable to videos [6, 7]. Therefore, investigators should make an extra effort in order to adapt image algorithms or develop new ones that take into account the video characteristics like the temporal domain. This has to become a strong trend in the coming years, especially when we know that the capacity for hidden messages in video streams is enormous when compared with other steganographic mediums, and countless algorithms can be used to embed information in various domains of the video [8]. Unfortunately and despite the recent interest in the video cover by steganalysts, research in this field is still in its first steps when compared with image cover-related investigation. It is obvious that due to the huge amount of data to analyze in videos, the algorithm designers should put more effort in producing heuristic-based solutions in order to reduce the computation costs. We suggest, in Figure 1, a general framework for the video steganalysis approaches.

This paper contains a comprehensive review of steganalysis methods applied to videos. We are using a new taxonomy principally based on distinguishing whether the algorithms are built upon spatial, temporal, or spatial-temporal steganalysis. We then look at specific considerations for each of the above classes.

Video steganalysis strategies use the spatial and temporal features obtained from intra- and inter-frames levels. The human

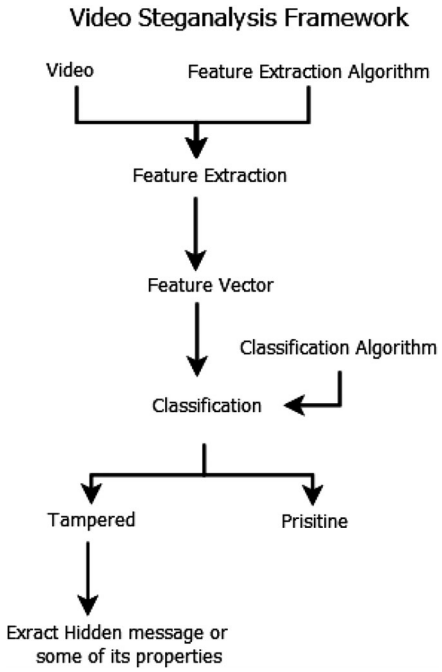


**FIGURE 1** A suggested video steganalysis framework

visual system (HVS), naturally containing certain specific characteristics, can be deceived by covert data if a given embedding threshold is respected. This makes the temporal domain more privileged to fit larger amounts of hidden messages, through creating higher redundancy that gives steganalysts greater attack opportunities [6].

In this paper, the terms 'algorithm', 'approach', or 'technique' are used interchangeably. They mean the same. Also, the term 'cover' is used to refer to a media devoid of any hidden secret information and the term 'stego' is used to refer to a media that has hidden secret information.

In the following, we will start, in Section 2, by comparing image to video when used as covers for hidden data. The objective is to determine how possible it is to extend the existing image steganalysis methods to video. Then, in Section 3, we will give general steganalysis rules and lessons extracted from the literature to the benefit of the new investigators or sometimes the experienced ones. In Section 4, we explain, in an acceptable detail, the differences between active and passive steganalysis. Further detail about the motion vectors-based methods is given. Another section addresses active video steganalysis principles and techniques. The paper then continues, in Section 5, by enumerating the chronological evolution of the video steganalysis features. The taxonomy we used in this paper in order to make it easily readable appears in Section 6. In Section 7, we gather, in two tables, the different characteristics of the surveyed video steganalysis methods. As datasets are essential for any experience, we decided to include a chapter about the most used and useful ones in the topic. In Section 9, we discuss the general findings of this research to conclude the paper, in Section 10, with the research directions and open problems that appear in the recent investigations to help the researchers push forward research in this very complex field.

**TABLE 1** Advantages (+) and disadvantages (−) of image versus video as covers

| | Image | Video |
|---|---|---|
| Ease of implementation | + | − |
| Popularity | + | + |
| Steganography capacity | − | + |
| Options to embed message | − | + |
| Maturity of research in steganalysis | + | − |

## 2 | IMAGE VERSUS VIDEO STEGANALYSIS

From the steganalysts perspective, using videos as covers is interesting since the embedding capacity is high when compared to the other mediums such as images. Nevertheless, and as asserted in [9], there is a greater chance of message detection because of the statistical redundancy existence in the temporal domain. So far, images are probably the cover that received the biggest part of investigators effort in steganalysis. However, as the digital video is overwhelming the Internet traffic [5], researchers start working on it in order to develop new methods because the ones developed for image steganalysis are not always adaptable to videos [7]. Therefore, investigators should make an extra effort in order to adapt image algorithms or develop new ones that take into account the video characteristics like the temporal domain. This has to become a strong trend in the coming years, especially when we know that the capacity for hidden messages in video streams is enormous when compared to other steganographic mediums, and countless algorithms can be used to embed information in various domains of the video [8]. A quick comparison between image and video based on different criteria is given in Table 1. The advantages are represented by the sign (+) while the disadvantages are represented by the sign (−).

## 3 | GENERAL STEGANALYSIS RULES AND PRINCIPLES

When we went through the video steganalysis literature, we found many lessons and rules worth extracting for the benefit of the new investigators or even more experienced ones who will need to complete their knowledge in order to extend the scope of their research. For example, slow-motion videos contain more exploitable information than high-motion ones. The reason for that is that slow-motion video successive frames present more common statistical and perceptual characteristics [10]. That is why the messages hidden in non-moving or having only translational motion in the video are easily detectable. Here is a list of those similar useful lessons in this very restrained topic of research

- The objective of steganalysis is to detect hidden messages with high probability and low complexity [9].

- There exists a tradeoff between the visibility of the concealed message and its embedding rate in video sequences. In fact, below a given embedding rate, the message becomes statistically invisible leading to the used steganalysis method to fail. However, above that threshold, the message becomes robust to destruction caused by compression or other means while disturbing statistics that make the message visible to some steganalysis approaches [6, 11, 12]. Precisely, and in the temporal domain, the statistic visibility is tightly bound to the temporal correlation between the successive frame correlations.
- Farid et al. in [13] and [14] show that the embedding of a message disrupts the higher order statistical regularity within an image. The fact applies to the video spatial domain steganography too.
- So far, the majority of video steganalysis research work is oriented towards H.264-based videos. However, a few researchers start putting forward some investigation on the H.265 (HEVC) videos. In that perspective, we can cite, for example, Huang et al. in [15] and Shi et al. in [16]. A lot is yet to be done with this specific type of videos.
- In theory, it is argued that for a steganographic algorithm to be defeated, it is necessary for the steganography to be detected with more success than a random guess, that is, with a true positive rate greater than 50% and a false positive rate less than 50%. However, it is commonly agreed that the real performance of an efficient steganalysis should be far better.
- Based on the message embedding location in the processing chain of video compression, digital video steganography can be classified into two main categories [17].
- Pre-compression steganography: The pre-compression steganography, the secret message is embedded into the cover video before compression [18, 19]. The primary advantage of pre-compression steganography is that the embedding technique is independent of the video compression standard. Robustness of the embedded message cannot be guaranteed in this case.
- In-compression steganography: In case of in-compression steganography, the message is embedded into the cover video during the compression process, for example, into the motion vectors [20] or quantized DCT coefficients [21] during video compression. These schemes, however, are sensitive to distortion introduced during transmission and are dependent on the video compression standard.
- Research on video steganalysis is developing slowly because video steganography and steganalysis require good backgrounds in video compression [22].
- Noise level is proportional to embedding strength and different noise levels correspond to different detecting thresholds [22].
- Compression has influence on mode detection results. The higher the bitrate, the less the embedding modes are destroyed and the easier the detection will be [22].
- Because of the strong temporal correlation between adjacent frames, the frame difference signal of the cover video can be commonly approximated by a Laplacian distribution [22].
- Active steganalysis is barely addressed by investigators because it is very greedy in processing time and often severely affects the cover media [8].
- The video steganalysis algorithms that utilize the temporal redundancies at the frame level and inter-frame level have been observed to be more effective than algorithms based on spatial redundancies only. Nevertheless, some video steganalysis algorithms that simultaneously exploit both the temporal and spatial redundancies have also been proposed and shown to be effective [23].
- Because video stream is first offered in a compressed format, steganography algorithms that are not applicable in compressed bit-stream would require complete or at least partial decompression, which is an unnecessary burden best avoided when the steganalysis method used is capable of attacking compressed videos. In the recent years, a good number of video steganography algorithms were designed for the compressed domain [24].
- Motion vector (MV)-based steganography has been used for the last decade because the statistical characteristics of the spatial/frequency coefficients are indirectly affected and because the motion compensation technique is adopted by most advanced compression standards and the MVs are lossless coded, little degradation of the reconstructed visual quality would be introduced [25].
- MV-based steganography embed messages during motion estimation within the compression process.
- The MV-based steganography is less detectable than those utilizing spatial/frequency coefficients directly [26].
- Because the secret messages are embedded into MVs, the statistical characteristics of the spatial/frequency coefficients of video frames will not be changed aggressively [27].
- The steganographic algorithms based on motion vectors may reduce the embedding rates, but produce little quality degradation of the reconstructed frames and little influence on the statistical characteristics of coefficients [28].
- Data can be embedded in raw video [29, 30] or in compressed ones [20, 25, 31–35]. The techniques to do so range from extending the image steganography methods, which embeds data in DCT coefficients or intra-frame, as done in JPEG images, to hiding data in motions vectors during the encoding process. It is important to continue developing steganalysis techniques in order to deter the MV-based steganography as the approaches used in the image are not optimal [36].
- The rich model steganography descriptors have been tested sometimes and showed their efficiency as done by Tasdemir in [37] and [7].
- Automatic detection of object-based video forgery is still in its infancy [38].
- When the descriptor (feature vector) is too large, it cannot be trained using an SVM classifier. Instead, an ensemble classifier should be used [7].
- It is relatively difficult to distinguish between the double compressed video frames from the tampered ones [38].

- The video sequences with fast and moderate movements exhibit stronger MV correlations than the slow moving video sequences, and this phenomenon has been explained in [7].
- Low bitrate videos usually have shorter motion vectors and greater prediction residuals than their high bitrate versions. On the contrary, high bitrate videos usually have smaller prediction residuals than their low bitrate versions [39].

## 4 | TYPES OF VIDEO STEGANALYSIS

In this section of the paper, we gather the different types of steganalysis along with the related used terminology found in the literature. The steganalysis approach can be passive, active, or even the combination of both. The researcher could also look at steganalysis from the spatial and temporal points of view. Thus, different assumptions could be done before introducing a new steganalysis approach.

### 4.1 | Passive video steganalysis

Passive steganalysis is based on a-priori knowledge of the steganography tool or embedding algorithm used to embed the secret message. It is also called specific or targeted steganalysis [8]. Kumar in [40] states that if the steganalysts know the embedding algorithm and its statistical signature then this type of steganalysis is very effective. This approach is similar to how anti-virus software works by looking for specific signatures that could be, for example, a statistical signature. As with anti-virus software, targeted steganalysis does not perform well with new or unknown embedding algorithms. Passive attacks are often very efficient at detecting stego messages as they are built against a specific subset of steganographic algorithms [41]. Thus, they are generally sufficient to defeat them. Otherwise, these types of steganalysis attacks do not need to alter the cover media to detect the presence of a hidden load [42]. However, despite these positive aspects of these technics, they present a set of disadvantages. We may point out that they are extremely difficult to generalize and can only be applied to a very specific type of steganographic algorithms. In addition, such strategies can easily be overcome by even a slight alteration of the steganographic algorithms. Consequently, we can state that these steganalysis passive attacks are very useful in a given limited interval of steganography methods. Beyond that, they present a large set of disadvantages like being rigid and inflexible.

### 4.2 | Active video steganalysis

The active steganalysis could be defined as the detection or extraction of a hidden message with little or no a priori information [10]. That is the blind attempt to detect the presence of covert data without knowing the particular steganographic algorithm used. For the first glance, this approach seems interesting. It would be impeccable if we reduce the complexity and do not loose in efficiency. Unfortunately, a very few active steganalysis attacks have been proposed [43] and [44] due to the broad range of existing steganography approaches and methods. Moreover, they are much fewer when the cover is a video. Thus, we do not consider this as an independent category in our taxonomy. Finally, each of these two approaches has its advantages and disadvantages. When enough a priori information is available, the passive attack is privileged and is more efficient but difficult to generalize to other steganography methods. Whereas, with no or little a priori information, it is better to use an active approach which needs a large dataset for training and test. In this case, the performance produced to attack a specific embedding algorithm is suboptimal.

### 4.3 | Other definitions of active and passive steganalysis

In [10, 45], the authors consider the role of passive steganalysis is only to detect the presence or absence of a hidden message, whereas the active steganalysis objective is to extract, destroy or alter it. The same definition is also adopted by Wu in [46]. The estimation of the hidden message 'length' can be sometimes part of the passive approach like in [45], or part of the active approach as stated by Trivedi in [47]. In addition to this, the above authors consider 'the location of the hidden load' to be part of the active approach. We may add to those pieces of information, 'the secret key used in the embedding process', some other parameters of the steganography algorithm used and eventually 'the message extraction' which is considered to be the ultimate objective when possible. It is also important to point out the difference between the active steganalysis as defined here and the active warden case where the steganalysts will destroy the message instead of making the effort to extract it. In some cases, the steganalysts may alter messages even if no trace of a hidden message is detected [48]. The amount of alteration made is subject to the model and the cover used. In this paper, we will consider the active and passive steganalysis as defined in Section 4.2.

### 4.4 | Spatial, temporal and combined spatial-temporal video steganalysis

In this paper, we define the spatial video steganalysis as the steganalysis method applied to video sequences based on features extracted from the spatial domain. Similarly, the temporal video steganalysis is the steganalysis method applied to video sequences based on features extracted from the temporal domain. Consequently, the combined spatial-temporal steganalysis is the steganalysis method applied to video sequences based on features extracted from both the spatial and temporal domains. It is important to point out that this is completely independent of whether the steganography method used to load the video is spatial, temporal or combines both the domains.

### 4.5 | Motion vector steganalysis

The temporal domain is very characteristic of videos. Thus, it should be privileged by steganalysts to hide messages especially that it gives more space for that. Surprisingly, the techniques that have been proposed so far are in majority utilizing the spatial domain, which is common to other traditional covers like image or text. For example, the techniques described in [49] and [50] encode information within the frames of the video using image steganography. Thus, they look to the video as a set of images without considering the temporal video specific domain. Some other techniques, in contrary, described for example in [51], [24], [52], and [53] among others suggest algorithms which use motion vectors. Most of these techniques involve encoding information within motion vectors as calculated by motion estimation blocks of various compression algorithms [8]. Investigation in this matter goes on and more descriptors tend to be built using the inter-frame domain. More recently, a few works taking into account both the spatial and the temporal domains were presented. They are more complex but seem to be robust like in [7] or [54]. The work of Sadat et al. in [55] is original and proposes a new method for motion vector steganalysis. They use the statistical entropy value and combine it with the optimized motion vector features.

In another high level view of steganalysis, it can be divided into different ways based on different criteria as follows:

- Original domain steganalysis and transform domain steganalysis according to its scope.
- Steganalysis based on the identification and steganalysis based on statistics.
- Specific steganalysis and universal steganalysis.

The present studies of video steganalysis mainly aim at specific steganographic algorithm, but there are still no mature methods for universal steganalysis [56].

## 5 | CHRONOLOGICAL EVOLUTION OF VIDEO STEGANALYSIS FEATURES

The video steganalysis features used in the surveyed papers can be split into three categories: spatial, temporal and the combination of both spatial and temporal (spatial-temporal) domains. In the following, they are dispatched into those classes with respect to their publication chronology since 2003, when, for the first time, a formal framework for video steganalysis was proposed by Chandramouli et al. [10] and 2004, when the first video steganalysis algorithm was proposed by Budhia et al. [9].

### 5.1 | Temporal

In 2004, Budhia et al. [10] first suggested using inherent temporal redundant information in the form of Gaussian watermarks. They used the statistical values: kurtosis, entropy and the 25th percentile as a feature vector. In 2006, the same authors used the same features above but worked in [6] on generalizing, extending and providing thorough analytic justification and simulations to their previous method. In 2007, Jainsky et al. [45] exploited the high redundancy correlation between consecutive frames and used motion interpolation and non-classical asymptotic relative (ARE) memoryless detection to be implemented on networks involving multimedia sensor systems. In the same year, Pankajakhshan et al. [57] proposed a method that forms a 39-dimensional feature vector by concatenating the three moments of the histogram characteristic functions of the corresponding wavelet sub-bands as in [58]. In 2008, Su et al. introduced in [22] a new steganalysis algorithm, which uses correlation between adjacent frames and detects a special distribution mode across the frames to distinguish videos tampered by StegoVideo [59]. In the same year, Zhang et al. [60] were, to the best of our knowledge, the first to introduce a steganalysis technique to detect messages hidden in the motion vectors of compressed videos. They consider the motion vector first-order statistics alterations caused by the least significant bit (LSB) embedding. Almost simultaneously, Zhang and Su [61] used the aliasing effects (distortion) between adjacent frames that are generated by the embedded data in raw videos. More specifically, they used the probability mass function of the stego-video frame difference signal. It seems like the researchers, from this date on, preferred forming their descriptors from a combination of both spatial and temporal features.

### 5.2 | Spatial

In 2006, Pankajakshan et al. [62] suggested capturing the statistical changes introduced in the motion trajectories of the video. The features used are the local variance and the local mean taken from the motion-compensated prediction error frames (PEFs) of the video. In 2008, Rana et al. proposed a frame-by-frame blind video steganalysis method based on spatial average filtering of frames. They used the same features as in Budhia et al. [6]. Pankajakshan et al. [63] tried to detect, in 2009, whether a video stream contains any motion-incoherent component. Their method relies on some features extracted from the error frames after motion compensation, like the local variance. In 2013, Tasdemir et al. [64] introduced a steganalysis method that targets LSB-based motion vector steganography. They used a flatness measure that indicates the corruption in flat areas in order to detect the existence of a message embedded into MVs in the LSB fashion. In 2014, Chen et al. [65] proposed a passive forensics steganalysis approach to attack object-based forgery. They looked at some statistical features around the object boundaries. The same year, Wang et al. [66] presented another video steganalysis method called AoSO (add or subtract one) against MV-based steganography. Their method considers the influence of the stego message on the sum of absolute difference (SAD). In 2015, Wu [46] investigated the detection of hidden messages in H265-encoded videos using an LSB matching algorithm. Fan et al. [67] proposed in 2016 an approach targeting hash-based LSB steganography. They compared the correlation between two consecutive frames in the stego video

and the correlation between the same frames in the reconstructive video. The same year, Zarmehi and Akhaee [68] presented an interesting method that estimates both the concealed message and the used spread spectrum gain factor. They calculated a residual matrix from which they extracted their feature vector. Again in 2016, Chen et al. [38] developed a way that localizes the forged video segment. Their approach first detects the frame manipulation using motion residuals and then locates the forged segment within the suspicious video. They enhanced a method that was originally built for still image steganalysis to extract forensic features from the motion residuals. Again, in 2016, Hong Zhang et al. were among the rare researchers to present a blind approach even if it is meant to attack motion vector-based steganography only. Their approach checks the local optimality of motion vectors in a rate-distortion sense. In 2017, Wang et al. [69] divided the video into detection intervals (DI) with fixed-length and then extracted the NPELO features [70] from every DI. Sadat et al. [55] proposed a motion vector-based method that extracts intrinsic and statistical features obtained by local optimization of the cost function. Specifically, they used entropy, combined with features from the optimized motion vector. Lately, in 2020, Peng Liu et al. [71] proposed a method that extracted its features from the noise residuals. It was among the first works to have tested deep learning for video steganalysis.

### 5.3 | Spatial-temporal

In 2007, Pankajakshan et al. [57], to the best of our knowledge, were the first to propose a combination of spatial and temporal aspects in video steganalysis. The features for the steganalysis are extracted from the residual frames after spatio-temporal prediction as described in [58]. In 2009, Kancherla and Mukkamala [72] used the average intensities of adjacent frames and especially the Discrete Cosine Transform and Markov features. Su et al. [73] proposed in 2011 to use the aliasing degrees between neighbouring frames as the spatial features and the frequency features of the centre of mass (COM) as the temporal ones. Otherwise, 1 year later, in 2012, Htet and Mya proposed to use images higher order statistics in order to observe a particular behaviour that would help separate corrupted videos from uncorrupted ones via the use of a Bayes classifier. These statistics include the entropy, the contrast, the angular second moment and the inverse difference moment. Cao et al. [26] targeted hidden messages by considering the disturbance that the steganography process creates in the motion vectors. They considered the probabilities of MV shift distances and the given shift distances. Another contribution related to steganalysis attacking forgery in motion vectors is the work of Deng et al. [27] in 2012. They presented a feature vector based on the 'local polynomial kernel regression model' and the calibration distances between the original MVs and the corresponding predicted one. In the same year, Zhao et al. [17] used YouTube videos and exploited the spatial-temporal correlation by extracting statistical features from the 3D DCT domain. Later, in 2013, Deng et al. [28] used statistical quantified correlations between motion vectors that relied on adjacent motion vector difference. Hui Ye et al. [36] presented in the same year an approach against motion vector steganography. They extracted the similar trends between neighbouring macroblocks from the spatial domain, and from the temporal domain, they considered the similar motion trend of macroblocks at the same position between successive frames. Keren Wang et al. [74], in 2014, built a model called SPEAM (subtractive prediction error adjacency model) which is an enhancement of the SPAM (subtractive pixel adjacency model) model that was originally designed for image steganalysis. Dependencies between adjacent samples in a PEF (prediction error frame) are modelled by a first-order Markov chain, and subsets of the empirical matrices are employed to build the feature vector. In 2015, Arijit Sur et al. [75] proposed a method that attacks motion vector steganography by exploring its effects on motion vectors, prediction error, flicker, and on motion vector histogram. The same year, Ting Da et al. [56] obtained a co-occurrence matrix of feature vector by Grey-level Co-occurrence Matrix as the correlation between macroblocks, and then established Markov model of inter-frames. Tasdemir and Kurugollu [7] proposed in 2016 rich model derived from both temporal and spatial correlations of motion vectors. Different filters have been used in order to capture different types of dependencies among MVs in a wide spatio-temporal range. In 2017, Yuting Su et al. [54] presented a method based on a spatial and temporal detector called ST_D that considers the impact of local motion intensity and texture complexity combining the histogram distributions of the well-known three descriptors RM, SPAM and SPEAM. Ye Yao et al. [76] proposed, in 2018, a deep convolutional neural network learning-based approach to detect object-based forgery in H.264-encoded videos. In 2019, Zhai et al. [77] used a 12-dimensional descriptor extracted from the motion vector consistency (MVC). The last works, as per the time of publication of this paper, namely 2021, Liu et al. [78] extracted a one-dimensional descriptor that computes the difference between the coding cost of a video before and after recoding. The same year, Ghamsarian et al. in [79] extracted 54 features exploiting the dependencies among neighbouring blocks, and the modification of the statistical Lagrangian multiplier value.

## 6 | TAXONOMY

Investigators, through different attempts, have shown that some video steganalysis algorithms have been well adapted from standard image steganalysis algorithms and worked as efficiently as for images [7, 37]. However, the adaptation stage is a sine qua non condition for such a move because of the clear differences that exist between images and videos, principally and namely the motion aspects, which are temporally dependent. However and generally speaking, the frame-by-frame use of image steganalysis methods to video has yet to be investigated and dug down as it has shown clear low performance results so far [23].

Unlike image steganalysis, only a handful of video steganalysis methods exist in the literature. Thus, a very few attempts have been made to survey and classify them. We

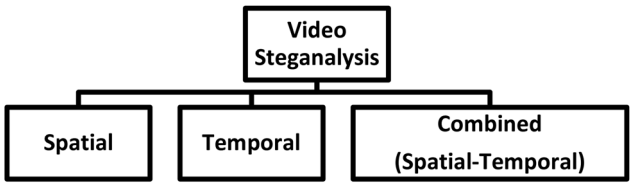


**FIGURE 2** Video steganalysis feature-based taxonomy adopted in this paper

will classify the methods based on the use or not of the temporal information used to extract the features used in classification. Other parameters will be considered in this survey, namely the mode detection, and the spatial and temporal prediction.

In this paper, the taxonomy of the video steganalysis methods we present consists of dividing these methods based on the extracted features' domains. The videos are categorized to either spatial, temporal or combined (spatial and temporal together) domains. We, then, cite the most relevant papers in each of the categories. The active steganalysis approaches are not considered in here because this category contains a very publications and this topic is yet to be investigated due to the lack of interesting algorithms and due to its high processing costs. The following Figure 2 illustrates the taxonomy used in this paper.

In Section 7, Table 2 gathers those methods and classifies them in the way explained above, whereas Table 3 discusses the strengths and weaknesses of those methods.

## 7 | CHARACTERISTICS OF THE SURVEYED VIDEO STEGANALYSIS METHODS

We gather, in the following two tables, the characteristics related to the surveyed methods. Table 2 contains the 'layout' of the proposal like the year of publication, the domain from which the features were chosen, the classifier used for discrimination. For more strength, we also added whether the suggested method was compared to the previous one and which steganography method was targeted if any. Table 3 contains deeper notions like the features extracted, their strengths and weaknesses. We also added whether a future work was proposed, which will help the investigators pick up their papers.

## 8 | DATASETS

One of the biggest challenges in video steganalysis is the absence of a canonical steganalysis video dataset. Nevertheless, a few attempts have been made to let the investigators test their works even if this cannot be used for all the works. In fact, many researchers do develop their own datasets based on their specific parameters and conditions. The problem becomes quickly crucial when the approach is deep learning based.

### 8.1 | The surrey university library for forensic analysis: SULFA [117]

This is composed of 150 videos shot by different cameras in different positions so that the temporal and spatial domains are considered. They mimic the real-life conditions as much as possible. All the video lengths are around 10 s with a resolution of 320 × 240 at the rate of 30 frames per second.

### 8.2 | Sysu-Objforg

SYSU-OBJFORG is one of the most interesting video datasets according to the report in [121]. It contains 100 pristine video sequences and 100 tampered video sequences. All video sequences are of 11 s, 1280×720 H.264/MPEG-4 encoded sequences with a bitrate of 3 Mbit/s and a frame rate of 25. In order to extract positive samples and negative samples from this dataset for the training of deep learning, an annotation algorithm should be presented to help to mark the forged areas in the forged video frames as there is no annotation information provided with the dataset.

### 8.3 | YUV

A set of test video test in the 4:2:0 YUV format [119]. All the sequences are compressed in the 7-Zip format

### 8.4 | NIST 2018 media forensics challenge

The Development datasets are provided by the NIST [123] Media Forensics Challenge5 for the Video Manipulation Detection task. There are two separate development datasets, namely, Dev1 and Dev2. The first one consists of 30 video pairs (i.e. 30 tampered videos and their 30 untampered sources), and the second of 86 video pairs, containing approximately 44K and 134K frames respectively. The task also includes a large number of distractor videos. These two datasets, Dev1 and Dev2, are treated as independent sets, but since they originate from the same source, they likely exhibit similar features.

### 8.5 | The InVID Fake Video Corpus

The InVID Fake Video Corpus [124] was developed over the course of the InVID project. The Fake Video Corpus (FVC) contains 110 real and 117 fake newsworthy videos from social media sources, which include not only videos that have been tampered but also videos that are contextually false (e.g. whose description on YouTube contains misinformation about what is shown). This dataset contains 163K frames, equally split between tampered and untampered videos. The temporal annotation of the datasets is incomplete, that is, one does not always know where and when the tampering takes place, only that a video contains a tampered part.

**TABLE 2** General characteristics of the surveyed video steganalysis methods

| Year | Author | Domain (spatial, temporal, or both) | Classifier | Comparisons (with previous methods) | Video format (compressed/uncompressed) | Embedding method (steganography) |
|---|---|---|---|---|---|---|
| 2004 | Budhia et al. [9] | Temporal | kNN | No | Uncompressed | Spread-spectrum-based Gaussian watermarks |
| 2006 | Pankajakshan et al. [62] | Spatial | kNN | No | MPEG compressed | Spread-spectrum |
| 2006 | Budhia et al. [6] | Temporal | kNN | No | Uncompressed | Spread-spectrum |
| 2007 | Pankajakshan et al. [57] | Spatial-temporal | k-NN with single neighbourhood | No | Uncompressed | Frame-by-frame additive Gaussian spread-spectrum watermarking scheme |
| 2007 | Jainsky et al. [45] | Temporal | kNN | No | Uncompressed | Spread-spectrum steganography |
| 2008 | Rana et al. [80] | Spatial | kNN | With 'StegWall' as in [81]. | Uncompressed | Spread-spectrum steganography |
| 2008 | Su et al. [22] | Temporal | No classifier. discrimination is threshold-based. | No | Uncompressed | Spread spectrum using StegoVideo Vatolin [59] |
| 2008 | Liu et al. [24] | Spatial-temporal | FFNN | No | MPEG-2 compressed | Message embedded in the Y component of the I-frames |
| 2008 | Zhang et al. [60] | Temporal | SVM | No | TM5 compressed | LSB simulation within motion vectors |
| 2008 | Zhang et al. [61] | Temporal | The majority-takes-all strategy | No | Uncompressed | Spread spectrum techniques in each frame like in Hartung [21]. |
| 2009 | Kancherla et al. [72] | Spatial-temporal | SVM, NN, kNN, and RF | Budhia et al. [6] and Jainsky et al. [45] | Uncompressed (avi) | Spread spectrum using StegoVideo Vatolin [59] and another tool developed locally. |
| 2009 | Pankajakshan et al. [63] | Spatial | kNN with single nearest neighbour | Budhia et al. [9] and to Budhia et al. [6]. | Tried both compressed and uncompressed | Spread spectrum and motion coherent (MCoh) described in Vinod et al. [62] |
| 2011 | Su et al. [73] | Spatial-temporal | SVM | No | Compressed | Motion vector based algorithms: Kutter [82], Dai [83], and Zhang [20]. |
| 2012 | Yun Cao et al. [26] | Spatial-temporal | Chang's LIBSVM [84] | Zhang et al. [60] | Compressed | Aly [25], Xu [33], Fang [34], and Cao [35]. |
| 2012 | Deng et al. [27] | Spatial-temporal | LIBSVM | [61] and [26]. | Uncompressed (CIF) | Cao [35] |
| 2012 | Zhao et al. [17] | Spatial-temporal | Unsupervised K-means clustering and SVM | No | Uncompressed | Adaptive spread spectrum |
| 2013 | Tasdemir et al. [64] | Spatial | Threshold-based | No | Uncompressed (CIF) | Xu et al. [33] and Aly's [25] |
| 2013 | Yu Deng et al. [28] | Spatial-temporal | SVM (RBF) | Su et al. [73] and Cao [26] | Uncompressed (CIF) | Fang [34] and Aly [25] |
| 2013 | Hui Ye et al. [36] | Spatial-temporal | Ensemble classifier [85] | Cao et al. [26] | Uncompressed (CIF and QCIF) | Xu [33], Aly [25], and Cao [35] |
| 2014 | Wang et al. [74] | Spatial-temporal | SVM | Budhia [6] and Pankajakshan [57] | Uncompressed (CIF) | Spread spectrum |
| 2014 | Chen et al. [65] | Spatial | LIBSVM | No | Raw (AVI) and WMV (windows media video). | Nothing cited |
| 2014 | Wang et al. [66] | Spatial | Linear SVM | [73] and [26] | Uncompressed (CIF) | Xu [33] and [25] |
| 2015 | Wu [46] | Spatial | SVM | No | Uncompressed (CIF and QCIF) | Spread spectrum within LSB |

(Continues)

**TABLE 2** (Continued)

| Year | Author | Domain (spatial, temporal, or both) | Classifier | Comparisons (with previous methods) | Video format (compressed/uncompressed) | Embedding method (steganography) |
|---|---|---|---|---|---|---|
| 2015 | Sur et al. [75] | Spatial-temporal | Linear discriminant analysis (LDA) | [61] and [35] | Uncompressed (CIF and QCIF) | Aly [25], Xu [33], and [86] |
| 2015 | Ting Da et al. [56] | Spatial-temporal | SVM (RBF kernel function) | [80] | Compressed (H.264). | LSB (1 bit per pixel). |
| 2015 | Y. Zhao et al. [87] | Spatial | LibSVM | PMC [88] and TPMC | Uncompressed (CIF) | Yang's [89] and Bouchama's [90] |
| 2016 | Fan et al. [67] | Spatial | Threshold-based | No | Uncompressed (AVI) | Hash based LSB substitution |
| 2016 | Zarmehi et al. [68] | Spatial | SVM [91]. | SPEAM method [74] | Uncompressed | Spread spectrum |
| 2016 | Tasdemir et al. [7] | Spatial-temporal | An ensemble classifier was used as commonly done with rich model vectors. | AoSO [74], Deng et al. [27], Deng et al. [28], and Su et al. [73] | Uncompressed (CIF) | Xu et al. [33], Fang and Chang [34], He and Luo[92], Pan et al.[93], and Aly [25] |
| 2016 | Zhang et al. [70] | Spatial | LibSVM | AoSO [66] and MVRB [94]. | Uncompressed (CIF) then compressed in the experimental phase | Aly et al. [25], Yao et al. [95], Cao et al. [96], Zhang et al. [97], and He et al. [92]. |
| 2016 | Chen et al. [38] | Spatial | The ensemble classifier described in [85] | Chen et al. in [65] | Compressed (H.264) | Videos taken from the SYSU-OBJFORG database |
| 2017 | Y. Su [54] | Spatial-temporal | SVM | Fridrich [98], SPAM in [99], and (SPEAM) | Uncompressed (CIF and QCIF) and (lossy) compressed | Spread spectrum technique described in Hartung [100] |
| 2017 | Wang et al. [69] | Spatial | SVM | Zhang et al. NPELO method [70] | Uncompressed (CIF) | Cao's [26] and Wang's [101] |
| 2018 | Yao et al. [76] | Spatial-temporal | CNN (DL) | Chen [38]. | Compressed (MPEG-4) | Object-based |
| 2018 | Sadat et al. [55] | Spatial | SVM [102]. | AoSO [66], IMVRB [94], NPE [70], and Arijit [75] | Compressed | Aly [25], Cao [96], and Xuansen [92] |
| 2019 | Zhai et al. [77] | Spatial-temporal | Gaussian-kernel SVM | [103], AoSO [66], NPE [70], and CCF [104]. | Uncompressed | Partition Mode Kapotas et al. [105], [89] and motion vector [35, 96, 97]. |
| 2020 | Peng Liu et al. [71] | Spatial | CNN (deep learning). | IS-Net [76] | Uncompressed | Hu et al. [106] and Hong Zhang et al. [97] |
| 2020 | Huang et al. [39] | Temporal | Deep CNN | No | Compressed using HEVC standard | Aly [25] and Xu [33] |
| 2020 | Wu et al. [107] | Spatial and temporal | Deep CNN | SRM [108]and Bayar [109] | Uncompressed | GAN-based approach [110] |
| 2021 | Yun Cao et al. [111] | Prediction error domain | Ensemble v2.0 [85] | COMMM [56] and VDCTR [69]. | Uncompressed (CIF) | Y. Cao et al. [111], Ma et al. [112], Lin et al. [113], and Chen et al. [114] |
| 2021 | Liu et al. [78] | Spatial-temporal | SVM | AoSO [66], IMVRB [94] and SPOM [115]. | Uncompressed (CIF) | Xu [33], Aly [25], and Cao [96] |

**TABLE 3** Video steganalysis methods extracted features, strengths, and weaknesses

| Year | Author | Features | Strengths | Weaknesses | Mention of future work (yes/no and how?) |
|---|---|---|---|---|---|
| 2004 | Budhia et al. [9] | Statistical values: kurtosis, entropy, and the 25th percentile. | Started the new era of the temporal video steganalysis and showed its potential. | In comparison to spatial methods of image steganalysis, this temporal method gives slightly poorer performance for lower embedding strengths. | Yes.<br>Use other classifiers.<br>Use both spatial and temporal domains.<br>Target the statistics of the video. |
| 2006 | Pankajakshan et al. [62] | The local variance and the local mean extracted from the PEFs (prediction error frame). | Reduces the computation cost while preserving a good result. | Tied to watermark Gaussianity and assumes steganography modifies all the pixels. | Yes.<br>Obtain PEFs from MPEG streams.<br>Assess the impact of the watermark on the human perception. |
| 2006 | Budhia et al. [6] | Statistical values: kurtosis, entropy, and the 25th percentile. | Low in complexity. Highlighted the tradeoff between the load size and the detection efficiency. | Inefficient with fast motion videos. Tied to watermark Gaussianity and assumes steganography modifies all the pixels. | Yes.<br>Decrease the unit of interest to only part of the frame instead of the whole frame. |
| 2007 | Pankajakshan et al. [57] | 39-dimensional feature vector extracted from the residual frames after spatio-temporal prediction as in [58] | Partial decoding of the compressed sequences which reduces computation. | Poor performance with fast motion and non-translational videos. Used with MPEG coded sequences only. | No |
| 2007 | Jainsky et al. [45] | Asymptotic relative efficiency (ARE) approach (MoViSteg algorithm). Used non-classical detection theory. | Works even only a subset of the video frames are watermarked. More proactive than the previous methods. | Assumes the video respects the Gauss–Markov correlation model from frame-to-frame and that the watermark is zero-mean. | Yes.<br>test MoViSteg on other steganography techniques other than SSS methods |
| 2008 | Rana et al. [80] | The same features used by Budhia et al. in [6]. | Invulnerable to temporal domain attacks like frame dropping etc. The scheme is robust against spatial attacks like frame cropping etc. | Suboptimal with fast moving objects. | Yes.<br>Check the proposed method on contrasting neighbourhoods and fast-moving object scenes. |
| 2008 | Su et al. [22] | A special distribution mode across the frames is compared to a threshold. | The model seems to be efficient when the stego video is obtained by using the StegoVideo tool. | The model ignores the spatial domain. Also, it is uncertain how the threshold is computed. | Yes.<br>Design a more precise algorithm by investigating intra-frame and inter-frame in the MSU stegoVideo |
| 2008 | Liu et al. [24] | 14-dimensional of AC and DC coefficients are extracted from the DCT domain of compressed videos using the inter-frame correlation (IFCS). | The proposed algorithm is executed in the compressed domain, which reduces the computing cost. | The method is suboptimal when the embedding rate is less than 30% | No |
| 2008 | Zhang et al. [60] | 12-dimensional statistical feature vector (aliasing degrees and COM). | Interesting detection rate. | Not very effective with low embedding strengths [26]. | No |

(Continues)

**TABLE 3** (Continued)

| Year | Author | Features | Strengths | Weaknesses | Mention of future work (yes/no and how?) |
|---|---|---|---|---|---|
| 2008 | Zhang et al. [61] | Uses the aliasing effect in the PMF (probability mass function) of the frame difference signal caused by embedding the hidden data. | Works for videos compressed with different bitrates. | Considers the hidden messages to be independent of the cover video. Also, the embedding is done in all the frames. | No |
| 2009 | Kancherla et al. [72] | The feature vector contains 274 features with 193 DCT features and 81 Markov features. | A rich vector of features that showed its efficiency compared to previous methods. | The watermark being embedded at the same location for all the frames. | No |
| 2009 | Pankajakshan et al. [63] | Some features extracted from the prediction error frames (PEF) after motion compensation. | Less computation as features directly extracted from compressed videos, capable of dealing with hybrid (static and dynamic areas) watermarking systems. | Less efficient with low embedding strengths. | Yes |
| 2011 | Su et al. [73] | 12-dimensional feature vector using aliasing degrees and centre of mass (COM). | Acceptable detection rate. | Low detection rate with low embedding strengths and abrupt changing in the scenes. | No |
| 2012 | Yun Cao et al. [26] | 15-dimensional feature vector based on the probabilities of MV shift distances. | Detect some typical MV-based steganography even with a low embedding strength. | Detection performance is likely to drop especially if the videos are recompressed under different settings. | Yes.<br>Use higher-order features and adopt certain feature selection/fusion techniques. |
| 2012 | Deng et al. [27] | Statistical: distance between MVs and absolute moments. | Improvement in detection accuracy. | Detection will drop if the message carrier is the MVs themselves. | Yes.<br>Improve the model by using a large-scale temporal model and adopting the self-adaptation sliding window. |
| 2012 | Zhao et al. [17] | Four statistical features extracted from the 3D DCT domain: (1) absolute central moments, (2) Skewness, (3) Kurtosis, and (4) Markov features. | Effective for most of the tested videos. | Ineffective for high texture or fast moving objects videos. | Yes.<br>Exploit more sensitive features to data hiding and use a more powerful unsupervised classifier. |
| 2013 | Tasdemir et al. [64] | (1) Flatness measure and (2) distance between reference and current frame. | Effective for most of the tested videos. | Inefficient with abrupt scene or random motions of a crowd. | No |
| 2013 | Yu Deng et al. [28] | 12-dimensional relying on adjacent MV difference. | The proposed difference operator is more sensitive to the statistical characteristics' changes of MV than the first-order one. | More samples needed for better training. The reference frame distance not considered in motion prediction. | No |

(Continues)

**TABLE 3** (Continued)

| Year | Author | Features | Strengths | Weaknesses | Mention of future work (yes/no and how?) |
|---|---|---|---|---|---|
| 2013 | Hui Ye et al. [36] | 324-dimensional feature vector based on the difference between neighbouring MVs in four directions. | More efficient than previous methods in detecting the MV-based steganography. | If the MVs and MV reversion tendency are modified, the detection is likely to drop. | Yes.<br>Apply a reach model of features. |
| 2014 | Wang et al. [74] | SPEAM (subtractive prediction error adjacency model). | Calculation of features is of low complexity and is suitable for real-time applications. | Not enough efficient for fast-moving compressed videos. | Yes.<br>Investigate more the inter- and intra-frame dependencies. |
| 2014 | Chen et al. [65] | The moment features of detailed wavelet coefficients and the average gradient intensity of each colour channel. | Difficult to assess as the dataset is tiny. | The scene background is static. The dataset used is far from being sufficient. | Yes.<br>Develop more robust features such as the motion trajectory and create more complete and close to reality datasets. |
| 2014 | Wang et al. [66] | AoSO 'for add or subtract one'. It calculates the difference between the actual SAD and locally optimized SAD after adding or subtracting one on the motion-value. | Applicable for various codecs, various LSB on the MV and various frame types | Fails to detect phase modifying stego algorithms. Efficiency deteriorates when the video quality is low. | Yes.<br>Consider the correlation between neighbouring MVs, extract features from those MVs, and combine those features with AoSO to achieve more favourable detection accuracy. |
| 2015 | Wu [46] | Based on LSB matching. | This is one of the rare works done on H265-encoded videos so far | Decrease of detection when the embedding rate decreases. | Yes.<br>Expand the use of the proposed algorithm by using other video steganalysis algorithms. |
| 2015 | Sur et al. [75] | 15-dimentional vector based on MV, PE, flicker, and absolute MV histogram. | Features are little or no correlated which makes the descriptor discrimination interesting | Lack of performance in high embedding rates. | No |
| 2015 | Ting Da et al. [56] | Special correlation change between video frames | Considers both spatial and temporal correlations. | Performs less in low bitrate videos. | Yes.<br>Improve this method to low bitrate videos |
| 2015 | Y. Zhao et al. [87] | 13-dimensional vector called (IPMC) composed of IPM and SATD | Sensitive to IPM-based steganography even at low embedding rates. | Detection drops when rate-distortion (RD) cost function is adopted to reselect IPM during H.264 encoding. | Yes.<br>Blend RD and SATD during the calibration process to improve the performance of IPMC features. |
| 2016 | Fan et al. [67] | Cross correlation | Not greedy in computation. | Not suitable to fast-moving videos. | Yes.<br>Extend cross correlation based steganalysis to generic video signals. |

(Continues)

**TABLE 3** (Continued)

| Year | Author | Features | Strengths | Weaknesses | Mention of future work (yes/no and how?) |
|---|---|---|---|---|---|
| 2016 | Zarmehi et al. [68] | Six-dimensional vector from the frames and the residual matrix. | It estimates the gain factor and original frame. | The algorithm shows less accuracy when $\alpha = 1\%$. | No |
| 2016 | Tasdemir et al. [7] | A rich model obtained from many diverse high-pass filters. | The proposed algorithm surpasses the previous methods in terms of classification accuracy in almost any payload. | The feature vector size was too large for SVM. The detection accuracy in high payload videos is lower than in low payload ones. | No |
| 2016 | Zhang et al. [70] | 36-dimensional feature set by computing the SAD-based or SATD-based Lagrangian cost function. | One of the rare steganalysis methods that have a blind approach. | Dependent on whether rate-distortion optimized has been conducted for the compressed videos. | No |
| 2016 | Chen et al. [38] | Motion residuals | Tested on the largest object-based forged video database in the literature. | The performance declines with high-resolution and high bitrate videos. | Yes.<br>Focus on more precise localization algorithms that can detect the actual location of forged objects in the video scene. |
| 2017 | Su [54] | Local motion intensity and texture complexity. | More efficient with compressed videos. | The best results are obtained with compressed videos. | No |
| 2017 | Wang et al. [69] | NPELO [70] and MVRBR [94] | Seems to be efficient in low bitrate and low embedding rates. | Dataset not enough large. | Yes.<br>Use a bigger dataset with more motion diversity. |
| 2018 | Yao et al. [76] | High dimensional features | Good results | Efficiency declines with low resolution and low bitrate videos. | Yes.<br>Apply trained CNN-based model to lower bitrate or lower resolution videos. |
| 2018 | Sadat et al. [55] | Intrinsic and statistical features obtained by local optimization of the cost function | It generally preserves its detection accuracy through different bitrates and different resolutions. | Questions on dataset size. | Yes.<br>More attention can be devoted to the H.265 video standard. |

(Continues)

**TABLE 3** (Continued)

| Year | Author | Features | Strengths | Weaknesses | Mention of future work (yes/no and how?) |
|---|---|---|---|---|---|
| 2019 | Zhai et al. [77] | 12-dimensional from the motion vector consistency (MVC). | Low computation. Can detect steganography from two different domains. | MVC features can only be applied to the variable block size videos. | No |
| 2020 | Peng Liu et al. [71] [71] | Extracted from noise residuals. | Among the first works to have tested deep learning for video steganalysis. | Features taken only from the spatial domain. Method compared to only one other method. | No |
| 2020 | Huang et al. [39] | (Deep CNN) deep learning-based quantitative steganalyzer for video | Effective and robust on HEVC videos | Not tested on other compression-strategies-based videos. | No |
| 2020 | Wu et al. [107] | The spatial features are extracted using a deep CNN for and temporal features are extracted by a single LSTM. | Outperforms the other steganalysis methods on the FF++ dataset [116]. | Not tested extensively. | No |
| 2021 | Yun Cao et al. [111] | Derived from SPEBs and IPM transition probabilities. | Shows good results even with extensive testing. | Not compared to a number of well-known steganalyzers like MVRBF [26], AoSO [66], and NPELO [70]. | Yes. Apply the same approach to the H.265 videos. |
| 2021 | Liu et al. [78] | One feature: the difference between the coding cost of a video before and after recoding. | Simplicity of feature extraction. | Limited when it comes to low encoding rate videos. | No |
| 2021 | Ghamsarian et al. [79] | 54-dimensional exploiting the dependencies among neighbouring blocks, and the modification of the statistical Lagrangian multiplier value. | Can be adjusted to various settings of the state-of-the-art video codec standards. Less vulnerable to overfitting compared to some rival methods. | Less efficient when the embedding is done exclusively in the MVs of deformable objects. | Yes. Detect hidden messages in deformable objects using object detection and tracking techniques. |

## 9 | DISCUSSION

In steganalysis, detecting concealed messages in video covers is by far more complex than detecting them in other covers like still images or text. The reason is that the message could be hidden, not only in the spatial domain, but it can also be concealed in the temporal domain like Motion Vectors for example. The temporal domain, in itself, presents many other features that can be used for this same purpose. Researchers used these features to form descriptors in order to build off-the-shelf systems based on active steganalysis. Only a few algorithms based on Motion Vector were developed so far. For example, Tasdemir et al., in [64], proposed a flatness measure for video steganalysis targeting LSB-based motion vector steganography. It considers the anchor frame and current frame distances and directions, which affect the correlation strength of adjacent motion vectors. Their proposed algorithm successfully classifies cover and stego videos. Similarly, Su et al. in [73] proposed a steganalysis method to detect information hidden in the motion vectors of video bit-streams. Another approach suggests using the motion vector recovery-based features like in [27]. Generally, the features extracted rely on properties related to motion vector (MV) like adjacent MV difference. It could be also the distance between neighbouring MV in four directions. Some other researchers used the distance between MVs and other MVs estimated from eight neighbours. Others were interested by the noise extracted from many high-pass filters and showed in many works they reached good results. The number of filters can be increased and there is a big chance that better results could be reached [37]. Of course, the number of features extracted can range from a few to thousands and containing a fewer number of unique features. Otherwise, the classifiers used by the investigators are in majority the support vector machine or sometimes ensemble of classifiers. Surprisingly, the reading of many works showed that a big number of published papers did not compare their results with other results in the field.

The different strategies used are either active or passive. The passive methods have to remain and follow the different strategies introduced by the steganography specialists. This should remain exactly as the antivirus algorithms are written based on the new created viruses. However, the active strategies have to acquire more strength even if the computational costs are globally huge. Thus, researchers should focus and introduce new heuristics and release constraints in order to alleviate the computation weight. The spatio-temporal and the temporal methods are the most effective but the most expensive at the same time. Also, the spatial methods are cost-effective but generally inefficient when the cover is video as the stenographers use the temporal domain more extensively than the spatial alone. This could be of course also combined with the spatial domain in order to achieve better results.

In the literature, the readers can easily find out how statistical redundancy in the cover video can help the steganalysis specialists in detecting hidden messages. More and more inter-frame correlation will, for sure, improve performance. Furthermore, the block-based scheme demonstrates how slow-moving video sequences are not an ideal choice for steganography [7].

A changing bit rates and variation in textures and movements are real challenges in video steganalysis. Old algorithms that analyze motion vectors are not sufficiently efficient in maintaining intrinsic features of the video [55]. Consequently, investigators should propose better strategies in order to reach an optimal property of the video based on the weight of each block with more attention to be given to the H.265 video standard. This is said based on the fact that the H.265 encoded videos use half the bandwidth used by the H.264 encoded ones for the same videos.

It has been largely proven that, in videos, motion vectors (MVs) have strong temporal and spatial dependency. In order to cover distant neighbouring MVs, the filter order is sometimes increased to the fourth or fifth order as in [37]. Of course, increasing the filter order gives better detection accuracy but generates more computational cost. The other idea consists of considering a stack of consequent MV planes of multiple frames. One of the strategies that researchers should investigate is trying to prove that these two approaches applied together will improve detection performance. Sequentially, each of the methods should be applied alone and then apply them together on two different layers on the same stack of filters. The results of the three experiments should be then compared to the already published ones.

Based on the above-mentioned reasons, the least that we can say is that video steganalysis is not largely investigated and is yet to be visited by the researchers. Thus, many open problems and suggestions are found in some research works.

## 10 | RESEARCH DIRECTIONS THAT APPEAR IN THE RECENT INVESTIGATION PAPERS

After nearly two decades of video steganalysis, we still feel that more work is yet to be done in order to deter messages concealed in real-life videos. Almost all the investigation efforts propose attacks aimed to specific types of steganography and with video sequences containing slow motion action rather than the fast-moving ones. In fact, slow-moving video sequences are not an ideal choice for steganography [10]. Thus, the steganographers would privilege fast-moving ones. Unfortunately, the majority of steganalysis approaches proposed so far perform more in slow- than in fast-moving videos. Consequently, researchers should research the possibility of designing a more generalized video steganalysis methodology, regardless of the types of steganography, formats, or compressions used in order to provide more general applications in media [125]. Another important concern that researchers still point out is the lack of rich and robust databases containing sequences close to reality like suggested by Chen et al. in [65]. Even if respectable efforts have been made to develop interesting datasets for video steganalysis, it is recurrent in the recent literature that more work is still to be done in this topic. As a big share of the videos in different media are low quality, some authors like Wang, in [69], consider developing solutions for low-quality video especially that these solutions will need only a limited and light

computation. As the H.265 (HEVC) video encoding format is making good progress on the Internet and other means, it starts attracting the attention of steganographers. Consequently, another trend in video steganalysis is designing algorithms that will attack embedded messages in H.265-encoded videos like suggested by Sadat in [55] and Liu et al. in [78]. Even if we present some detail about the future directions, they will all congregate in the description above. For example, Hui et al. in [36] suggested investigating how to apply image high-dimension features called Rich Model introduced in [98] to video steganalysis. It is to be noted that in 2016, Tasdemir et al. proposed a rich model called spatio-temporal rich model (STRM) which they stated ensured better results than the previous models based on smaller descriptors. More rich models should be proposed in the future, especially those that gather features from different domains. Wang et al. pointed out in [74] that more effort should be done to investigate advanced measures to merge the utilization of temporal and spatial redundancies to apply for content fast moving compressed videos with irregular trails or of high texture complexity. They also encouraged to check the effectiveness of the steganalysis features on videos of various codec, and research on dependencies between intra-frame MVs and correlation within inter-frame MVs, and derive favourable features for steganalysis. From another perspective, Wang et al. advise, in [66], considering the correlation between neighbouring MVs, extract features from those MVs, and combine those features with AoSO feature [66] to achieve more favourable detection accuracy. Otherwise, and as adaptive steganalysis is becoming a trend, Zhao et al. [87] suggest improving the adaptability of their intra prediction mode calibration (IPMC) features. Possible schemes include higher-order features and adaptive feature extraction/selection techniques.

In the recent few years, some researchers such as Yao et al. [76] suggested using deep learning in video steganalysis. They proposed, for example, focusing on the localization for forged regions in each of the tampered video frames. Also see how to apply the trained CNN-based model to detect object forgery for lower bitrate video sequence or lower resolution video sequence, which is named as transfer learning in deep learning research. This approach can even be associated with the suggestion of Ghamsarian et al. in [79] to detect hidden messages in deformable objects using object detection and tracking techniques like in [114, 126].

## CONFLICT OF INTEREST

To whom it may concern, I hereby declare that I have no pecuniary or other personal interest, direct or indirect, in any matter that raises or may raise a conflict with the publication of the article I am submitting and that is entitled 'A comprehensive review of video steganalysis'.

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

## ORCID

*Mourad Bouzegza* https://orcid.org/0000-0003-2451-5319
*Ammar Belatreche* https://orcid.org/0000-0003-1927-9366
*Ahmed Bouridane* https://orcid.org/0000-0002-1474-2772

## REFERENCES

1. Zielińska, E., Mazurczyk W., Szczypiorski K.: Trends in steganography. Commun. ACM. 57(3), 86–95 (2014)
2. Nissar, A., Mir A.: Classification of steganalysis techniques: A study. Digit. Signal Process. 20(6), 1758–1770 (2010)
3. Kessler, G.C.: Steganography: Implications for the Prosecutor and Computer Forensics Examiner. American Prosecution Research Institute, Alexandria (2004)
4. Burgess, C., et al.: Detecting packed executables using steganalysis. In: 2014 5th European Workshop on Visual Information Processing (EUVIP). IEEE (2014)
5. papers, C.W.: Cisco VNI Forecast and Methodology, 2015–2020. CISCO, (2016)
6. Budhia, U., Kundur D., Zourntos T.: Digital video steganalysis exploiting statistical visibility in the temporal domain. IEEE Trans. Inform. Forensics Secur. 1(4), 502–516 (2006)
7. Tasdemir, K., Kurugollu F., Sezer S.: Spatio-temporal rich model-based video steganalysis on cross sections of motion vector planes. IEEE Trans. Image Process. 25(7), 3316–3328 (2016)
8. Sharp, A., et al.: A novel active warden steganographic attack for next-generation steganography. In: 2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC). IEEE (2013)
9. Budhia, U., Kundur D.: Digital video steganalysis exploiting collusion sensitivity. In: Defense and Security. International Society for Optics and Photonics (2004)
10. Chandramouli, R.: A mathematical framework for active steganalysis. Multim. Syst. 9(3), 303–311 (2003)
11. Su, K., Kundur D., Hatzinakos D.: Statistical invisibility for collusion-resistant digital video watermarking. IEEE Trans. Multimedia. 7(1), 43–51 (2005)
12. Su, K., Kundur D., Hatzinakos D.: Spatially localized image-dependent watermarking for statistical invisibility and collusion resistance. IEEE Trans. Multim. 7(1), 52–66 (2005)
13. Farid, H.: Detecting hidden messages using higher-order statistical models. In: Proceedings of the International Conference on Image Processing. IEEE (2002)
14. Lyu, S., Farid H.: Detecting hidden messages using higher-order statistics and support vector machines. In: International Workshop on Information Hiding. Springer (2002)
15. Huang, K., et al.: Combined features for steganalysis against PU partition mode-based steganography in HEVC. Multimedia Tools and Applications, 79, 31147–31164 (2020)
16. Shi, H., et al.: A HEVC video steganalysis against DCT/DST-based steganography. Int. J. Digit. Crime Forensics (IJDCF). 13(3), 19–33 (2021)
17. Zhao, H., Wang H., Malik H.: Steganalysis of youtube compressed video using high-order statistics in 3d DCT domain. In: 2012 Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP). IEEE
18. Huang, H.-Y., Yang C.-H., Hsu W.-H.: A video watermarking technique based on pseudo-3-D DCT and quantization index modulation. IEEE Trans. Inf. Forensics Secur. 5(4), 625–637 (2010)
19. Esen, E., Alatan A.A.: Robust video data hiding using forbidden zone data hiding and selective embedding. IEEE Trans. Circuits Syst. Video Technol. 21(8), 1130–1138 (2011)
20. Zhang, J., Li J., Zhang L.: Video watermark technique in motion vector. In: Proceedings of XIV Brazilian Symposium on Computer Graphics and Image Processing. IEEE (2001)
21. Hartung, F., Girod B.: Watermarking of uncompressed and compressed video. Signal Process. 66(3), 283–301 (1998)
22. Su, Y., et al.: A new video steganalysis based on mode detection. In: 2008 International Conference on Audio, Language and Image Processing. IEEE (2008)

23. Meghanathan, N., Nayak L.: Steganalysis algorithms for detecting the hidden information in image, audio and video cover media. Int. J Netw Secur. Appl. (IJNSA). 2(1), 43–55 (2010)
24. Liu, B., Liu F., Wang P.: Inter-frame correlation based compressed video steganalysis. In: Congress on Image and Signal Processing, 2008. CISP'08. IEEE (2008)
25. Aly, H.A.: Data hiding in motion vectors of compressed video based on their associated prediction error. IEEE Trans. Inf Forensics Secur. 6(1), 14–18 (2011)
26. Cao, Y., Zhao X., Feng D.: Video steganalysis exploiting motion vector reversion-based features. IEEE Signal Process. Lett. 19(1), 35–38 (2012)
27. Deng, Y., Wu Y., Zhou L.: Digital video steganalysis using motion vector recovery-based features. Appl. Opt. 51(20), 4667–4677 (2012)
28. Deng, Y., et al.: Digital video steganalysis based on motion vector statistical characteristics. Optik-Int. J. Light Electron Optics. 124(14), 1705–1710 (2013)
29. Chae, J.J., Manjunath B.: Data hiding in video. In: Proceedings of the 1999 International Conference on Image Processing (Cat. 99CH36348). IEEE (1999)
30. Pazarci, M., Dipcin V.: Data embedding in scrambled digital video. In: Proceedings of the Eighth IEEE Symposium on Computers and Communications. ISCC 2003. IEEE (2003)
31. Giannoula, A., Hatzinakos D.: Compressive data hiding for video signals. In: Proceedings of the 2003 International Conference on Image Processing (Cat. No. 03CH37429). IEEE (2003)
32. Caccia, G., Lancini R.: Data hiding in mpeg-2 bit stream domain. In: EUROCON'2001. International Conference on Trends in Communications. Technical Program, Proceedings (Cat. No. 01EX439). IEEE (2001)
33. Xu, C., Ping X., Zhang T.: Steganography in compressed video stream. In: First International Conference on Innovative Computing, Information and Control, 2006. ICICIC'06. (2006)
34. Fang, D.-Y., Chang L.-W.: Data hiding for digital video with phase of motion vector. In: Proceedings of the 2006 IEEE International Symposium on Circuits and Systems (ISCAS 2006). IEEE (2006)
35. Cao, Y., et al.: Video steganography with perturbed motion estimation. In: International Workshop on Information Hiding. Springer (2011)
36. Ye, H., et al.: Motion vector-based video steganalysis using spatial-temporal correlation. In: 2013 6th International Congress on Image and Signal Processing (CISP). IEEE (2013)
37. Tasdemir, K.: Moving from image steganalysis to motion vector based video steganalysis, Queen's University Belfast (2015)
38. Chen, S., et al.: Automatic detection of object-based forgery in advanced video. IEEE Trans. Circuits Syst. Video Technol. 26(11), 2138–2151 (2016)
39. Huang, X., et al.: Deep learning-based quantitative steganalysis to detect motion vector embedding of HEVC videos. In: 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC). IEEE (2020)
40. Kumar, M., Steganography and Steganalysis of JPEG Images: A Statistical Approach to Information Hiding and Detection. LAP Lambert Academic Pub (2011)
41. Li, B., et al.: A survey on image steganography and steganalysis. J. Inf. Hiding Multim. Signal Process. 2(2), 142–172 (2011)
42. Kharrazi, M., Sencar H.T., Memon N.: Image Steganography and Steganalysis: Concepts and Practice, in Mathematics And Computation In Imaging Science And Information Processing, pp. 177–207. World Scientific, Singapore (2007)
43. Zheng, D., et al.: A survey of RST invariant image watermarking algorithms. ACM Comput. Surv. (CSUR). 39(2), 5 (2007)
44. Dalal, M., Juneja M.: Steganography and steganalysis (in digital forensics): A cybersecurity guide. Multim. Tools Appl. 80(4), 5723–5771 (2021)
45. Jainsky, J.S., Kundur D., Halverson D.R.: Towards digital video steganalysis using asymptotic memoryless detection. In: Proceedings of the 9th Workshop on Multimedia & Security. ACM (2007)
46. Wu, K.: Research of video steganalysis algorithm based on H265 protocol. In: MATEC Web of Conferences. EDP Sciences (2015)
47. Trivedi, S., Chandramouli R.: Active steganalysis of sequential steganography. In: Security and Watermarking of Multimedia Contents V. 2003. International Society for Optics and Photonics
48. Avcibas, I., Memon N., Sankur B.: Steganalysis using image quality metrics. IEEE Trans. Image Process. 12(2), 221–229 (2003)
49. Balaji, R., Naveen G.: Secure data transmission using video Steganography. In: 2011 IEEE International Conference on Electro/Information Technology (EIT). IEEE (2011)
50. Hu, S.D.: A novel video steganography based on non-uniform rectangular partition. In: 2011 IEEE 14th International Conference on Computational Science and Engineering (CSE), IEEE (2011)
51. Sharp, A.T., Devaney J., Steiner A.E.: Digital video authentication with motion vector watermarking. In: 2010 4th International Conference on Signal Processing and Communication Systems (ICSPCS). IEEE (2010)
52. Mohaghegh, N., Fatemi O.: 264 copyright protection with motion vector watermarking. In: International Conference on Audio, Language and Image Processing, 2008. ICALIP 2008. IEEE (2008)
53. Ceddillo-Hernandez, A., et al.: Robust video watermarking using perceptual information and motion vector. In: IEEE Northeast Workshop on Circuits and Systems. 2007. NEWCAS 2007. IEEE (2007)
54. Su, Y., Yu F., Zhang C.: Digital video steganalysis based on a spatial temporal detector. KSII Trans. Internet Information Syst. (TIIS) 11(1), 360–373 (2017)
55. Sadat, E., Faez K., Saffari Pour M.: Entropy-based video steganalysis of motion vectors. Entropy. 20(4), 244 (2018)
56. Da, T., Li Z., Feng B.: A video steganalysis algorithm for H. 264/AVC based on the Markov features. In: International Conference on Intelligent Computing. Springer (2015)
57. Pankajakshan, V., Ho A.T.: Improving video steganalysis using temporal correlation. In: Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2007). IEEE (2007)
58. Xuan, G., et al.: Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions. In: International Workshop on Information Hiding. Springer (2005)
59. Vatolin, D., Petrov O.: StegoVideo steganalysis tool. 2001–2020; Available from: http://www.compression.ru/video/stego_video/index.html
60. Zhang, C., Su Y., Zhang C.: A new video steganalysis algorithm against motion vector steganography. In: 4th International Conference on Wireless Communications, Networking and Mobile Computing, 2008. WiCOM'08. IEEE (2008)
61. Zhang, C., Su Y.: Video steganalysis based on aliasing detection. Electron. Lett. 44(13), 801–803 (2008)
62. Vinod, P., Doerr G., Bora P.: Assessing motion-coherency in video watermarking. In: International Multimedia Conference: Proceeding of the 8th Workshop on Multimedia and Security. (2006)
63. Pankajakshan, V., Doërr G., Bora P.K.: Detection of motion-incoherent components in video streams. IEEE Trans. Inf. Forensics Secur. 4(1), 49–58 (2009)
64. Tasdemir, K., Kurugollu F., Sezer S.: Video steganalysis of LSB based motion vector steganography. In: European Workshop on Visual Information Processing (EUVIP). (2013). IEEE
65. Richao, C., Gaobo Y., Ningbo Z.: Detection of object-based manipulation by the statistical features of object contour. Forensic Sci. Int. 236, 164–169 (2014)
66. Wang, K., Zhao H., Wang H.: Video steganalysis against motion vector-based steganography by adding or subtracting one motion vector value. IEEE Trans. Inf. Forensics Secur. 9(5), 741–751 (2014)
67. Fan, M., et al.: Cross correlation feature mining for steganalysis of hash based least significant bit substitution video steganography. Telecommun. Syst. 63(4), 523–529 (2016)
68. Zarmehi, N., Akhaee M.A.: Digital video steganalysis toward spread spectrum data hiding. IET Image Process. 10(1), 1–8 (2016)
69. Wang, P., Cao Y., Zhao X.: Segmentation based video Steganalysis to detect motion vector modification. Security Commun. Netw. 2017, 8051389 (2017)

70. Zhang, H., Cao Y., Zhao X.: A steganalytic approach to detect motion vector modification using near-perfect estimation for local optimality. IEEE Trans. Inf. Forensics Secur. 12(2), 465–478 (2016)
71. Liu, P., Li S.: Steganalysis of intra prediction mode and motion vector-based steganography by noise residual convolutional neural network. In: IOP Conference Series: Materials Science and Engineering. IOP Publishing (2020)
72. Kancherla, K., Mukkamala S.: Video steganalysis using spatial and temporal redundancies. In: International Conference on High Performance Computing & Simulation, 2009. HPCS'09. IEEE (2009)
73. Su, Y., Zhang C., Zhang C.: A video steganalytic algorithm against motion-vector-based steganography. Signal Process. 91(8), 1901–1909 (2011)
74. Wang, K., Han J., Wang H.: Digital video steganalysis by subtractive prediction error adjacency matrix. Multim. Tools Appl. 72(1), 313–330 (2014)
75. Sur, A., et al.: Detection of motion vector based video steganography. Multimed. Tools Appl. 74(23), 10479–10494 (2015)
76. Yao, Y., et al.: Deep learning for detection of object-based forgery in advanced video. Symmetry. 10(1), 3 (2018)
77. Zhai, L., Wang L., Ren Y.: Universal detection of video steganography in multiple domains based on the consistency of motion vectors. IEEE Trans. Inf. Forensics Secur. 15, 1762–1777 (2019)
78. Liu, J., et al.: A video steganalysis method based on coding cost variation. Int. J. Distrib. Sens. Netw. 17(2), 1550147721992730 (2021)
79. Ghamsarian, N., Schoeffmann K., Khademi M.: Blind MV-based video steganalysis based on joint inter-frame and intra-frame statistics. Multimed. Tools Appl. 80(6), 9137–9159 (2021)
80. Rana, V., et al.: Novel scheme of video steganalysis for detecting antipodal watermarks. In: TENCON 2008-2008 IEEE Region 10 Conference, IEEE (2008)
81. Voloshynovskiy, S.V., et al.: Stegowall: Blind statistical detection of hidden data. In: Security and Watermarking of Multimedia Contents IV. 2002. International Society for Optics and Photonics
82. Mkutter, F.J., Ebrahimi T.: Proposal of a watermarking technique for hiding/retrieving data in compressed and decompressed video. JTCI/SC29/WG11. (1997)
83. Dai, Y., Zhang L., Yang Y.: A new method of MPEG video watermarking technology. In: Proceedings of the International Conference on Communication Technology 2003. ICCT 2003. IEEE(2003)
84. Chang, C.-C., Lin C.-J.: LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol. (TIST). 2(3), 1–27 (2011)
85. Kodovsky, J., Fridrich J., Holub V.: Ensemble classifiers for steganalysis of digital media. IEEE Trans. Inf. Forensics Secur. 7(2), 432–444 (2012)
86. Jue, W., Min-qing Z., Juan-li S.: Video steganography using motion vector components. In: 2011 IEEE 3rd International Conference on Communication Software and Networks. IEEE (2011)
87. Zhao, Y., et al.: Video steganalysis based on intra prediction mode calibration. In: International Workshop on Digital Watermarking. Springer (2015)
88. Li, S., et al.: Steganalysis of prediction mode modulated data-hiding algorithms in H. 264/AVC video stream. Ann. Telecommun. Annales des Télécommunications 69(7), 461–473 (2014)
89. Yang, G., et al.: An information hiding algorithm based on intra-prediction modes and matrix coding for H. 264/AVC video stream. AEU-Int. J. Electron. Commun. 65(4), 331–337 (2011)
90. Bouchama, S., Hamami L., Aliane H.: 264/AVC data hiding based on intra prediction modes for real-time applications. In: Proceedings of the World Congress on Engineering and Computer Science. (2012)
91. Theodoridis, S.: Pattern Recognition. Pattern recognition. (2003)
92. He, X., Luo Z.: A novel steganographic algorithm based on the motion vector phase. In: 2008 International Conference on Computer Science and Software Engineering. IEEE (2008)
93. Pan, F., et al.: Video steganography using motion vector and linear block codes. In: 2010 IEEE International Conference on Software Engineering and Service Sciences. IEEE (2010)
94. Wang, P., et al.: Motion vector reversion-based steganalysis revisited. In: 2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP). IEEE (2015)
95. Yao, Y., et al.: Defining embedding distortion for motion vector-based video steganography. Multim. Tools Appl. 74(24), 11163–11186 (2015)
96. Cao, Y., et al.: Video steganography based on optimized motion estimation perturbation. In: Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security. ACM (2015)
97. Wang, P., et al.: 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. (2016)
98. Fridrich, J., Kodovsky J.: Rich models for steganalysis of digital images. IEEE Trans. Inf. Forensics Secur. 7(3), 868–882 (2012)
99. Pevny, T., Bas P., Fridrich J.: Steganalysis by subtractive pixel adjacency matrix. IEEE Trans. Inf. Forensics Secur. 5(2), 215–224, (2010)
100. Hartung, F.H., Girod B.: Digital watermarking of raw and compressed video. In: Proceedings Digital Compression Technologies and Systems for Video Communications. Int. Soc. Optics Photonics (1996)
101. Wang, H., et al.: Background modeling and foreground extraction method based on depth image. Google Patents. (2016)
102. Lin, C.-C.C.a.C.-J.: LIBSVM – A Library for Support Vector Machines. (2016); Available from: https://www.csie.ntu.edu.tw/~cjlin/libsvm/
103. Zhang, H., et al.: Video steganography with perturbed macroblock partition. In: Proceedings of the 2nd ACM Workshop on Information Hiding and Multimedia Security. (2014)
104. Zhai, L., Wang L., Ren Y.: 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. (2017)
105. Kapotas, S.K., Skodras A.N.: A new data hiding scheme for scene change detection in H. 264 encoded video sequences. In: 2008 IEEE International Conference on Multimedia and Expo. IEEE (2008)
106. Hu, Y., Zhang C., Su Y.: Information hiding based on intra prediction modes for H. 264/AVC. In: 2007 IEEE International Conference on Multimedia and Expo. IEEE (2007)
107. Wu, X., et al.: Sstnet: Detecting manipulated faces through spatial, steganalysis and temporal features. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE (2020)
108. Zhou, P., et al.: Learning rich features for image manipulation detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2018)
109. Bayar, B., Stamm M.C.: A deep learning approach to universal image manipulation detection using a new convolutional layer. In: Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security. (2016)
110. Shen, T., et al.: "deep fakes" using generative adversarial networks (gan). (2018)
111. Cao, Y., et al.: Steganalysis of H. 264/AVC videos exploiting subtractive prediction error blocks. IEEE Trans. Inf. Forensics Secur. 16, 3326–3338, (2021)
112. Ma, X., et al.: A data hiding algorithm for H. 264/AVC video streams without intra-frame distortion drift. IEEE Trans. Circuits Syst. Video Technol. 20(10), 1320–1330 (2010)
113. Lin, T.-J., et al.: An improved DCT-based perturbation scheme for high capacity data hiding in H. 264/AVC intra frames. J. Syst. Software. 86(3), 604–614 (2013)
114. Chen, Y., et al.: Adaptive video data hiding through cost assignment and STCs. IEEE Trans. Dependable Secure Comput. 18(3), 1320–1335 (2019)
115. Ren, Y., et al.: Video steganalysis based on subtractive probability of optimal matching feature. In: Proceedings of the 2nd ACM Workshop on Information Hiding and Multimedia Security. (2014)

116. Cao, Y., et al.: Covert communication by compressed videos exploiting the uncertainty of motion estimation. IEEE Commun. Lett. 19(2), 203–206 (2014)
117. Zhang, H., Cao Y., Zhao X.: Motion vector-based video steganography with preserved local optimality. Multim. Tools Appl. 75(21), 13503–13519 (2016)
118. Rossler, A., et al.: Faceforensics++: Learning to detect manipulated facial images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. (2019)
119. Xiph.org, *Xiph.org Video Test Media*. [derf's collection]. Xiph.org
120. Qadir, G., Yahaya S., Ho A.T.: Surrey university library for forensic analysis (SULFA) of video content. (2012)
121. University, S.: The SYSU-OBJFORG dataset. (2019); Available from: http://media-sec.szu.edu.cn/sysu-objforg/index.html
122. University, A.S.: YUV Video Sequences. (2017); Available from: http://trace.eas.asu.edu/yuv/index.html
123. Technology, T.N.I.o.S.a. Media Forensics Challenge 2018. (2018); dataset for video steganalysis]. Available from: https://www.nist.gov/itl/iad/mig/media-forensics-challenge-2018
124. Papadopoulou, O., et al.: A corpus of debunked and verified user-generated videos. Online Inf. Rev. 43(1), 72–88 (2018)
125. Tabares-Soto, R., et al.: Digital media steganalysis. In: Digital Media Steganography, pp. 259–293. Elsevier (2020)
126. Chen, Y., et al.: Research of improving semantic image segmentation based on a feature fusion model. J. Ambient Intell. Humanized Comput. 1–13 (2020)

**How to cite this article:** Bouzegza, M., Belatreche, A., Bouridane, A., Tounsi, M.: A comprehensive review of video steganalysis. IET Image Process. 16, 3407–3425 (2022). https://doi.org/10.1049/ipr2.12573