VIDEO STEGANALYSIS IN THE TRANSFORM DOMAIN BASED ON MORPHOLOGICAL STRUCTURE OF THE MOTION VECTOR MAPS

Ismahane Cheheb,Abdellatif Zouak, Ahmed Bouridane

Department of Computer and Information Sciences
Northumbria University
Newcastle-upon-Tyne, UK

Yves Michels, Salah Bourennane

Institut Fresnel
Ecole Centrale Marseille
Marseille, France

ABSTRACT

Steganography is the art of transmitting hidden messages through a cover object without raising any suspicion. In contrast, steganalysis is the science of detecting the presence of hidden information and a significant amount of research has been focused on multimedia steganalysis. In this paper, a video steganalysis method is proposed to detect the presence of hidden data by analysing the structure of the motion vectors in the compressed video data. The proposed method is based on the classification of features extracted from the morphology of the motion vector map. The proposed method has been evaluated on a large dataset of short videos with variable resolution and quality and the results suggest the effectiveness of the proposed modelling scheme.

Index Terms— Video steganalysis, motion vector, MPEG-4, H.264, SVM.

1. INTRODUCTION

In contrast to cryptography, which aims is to avoid an outsider to read the information, steganography is the science and the art of covert communications. The objective of steganography is to transmit the secret message without drawing any suspicion. Fundamentally, the steganographic goal is not to prevent outsiders from decoding the hidden message, but to prevent them from suspecting the existence/presence of the secret message. In terms of modern steganography, any digital objects, such as images, sounds, text document or video can be used as a message carrier. Courtesy of the recent developments of informatics and the internet, steganography has become a large field of research. For example, several steganalysis methods have been developed to detect the embedded messages in images [1, 2]. However, video steganalysis research and development has received much less interest compared to image detection. In addition, the availability of cheap and user friendly video sharing platforms has made sharing of video data on untrusted networks such as the internet an easy task resulting in a widespread of illegal transmission of covert communications for various purposes including espionage, terrorism and fake news to name a few [3, 4].

Owing to the fact that the majority of videos are transmitted and shared in the compressed domain, steganographic methods are usually developed to operate in the transform domain so that they are robust against compression attacks. This means that the embedding is carried out at the motion vectors since they resist compression distortions by design [5, 6, 7]. In addition, practical systems used for covert communications are blind systems, and as such they do not use known video clips as carriers rather they use their own video. Therefore, our proposed method is based on the analysis of the motion vectors (MVs) used in the H.264/MPEG-4 Advanced Video Coding (AVC) standards and is based on a structural analysis of the motion.

This paper is organised as follows: section 2 introduces the main characteristics of H.264/MPEG-4 AVC standards. In section3 we analyse and discuss the general morphology of the MV map. In section 4, the proposed algorithm for the detection of hidden messages is explained. Finally, experiments and obtained results are depicted in section 5.

2. H.264/MPEG-4 ADVANCED VIDEO CODING STANDARDS

The H.264/MPEG-4 AVC was developed in order to provide better compression of video compared to previous standards (H.261, H.263).[8, 9].

While H.264 performs a more efficient compression, this comes at the cost of increased complexity. It kept similar elements to the previous standards such as quantization for bitrate control, motion compensated prediction for reduction of temporal correlation and entropy encoding for reduction of statistical correlation, while bringing new functions such as intra-picture prediction, the use of multiple reference pictures and different block sizes. It also aims to reduce both spatial and temporal redundancy. [9].

H.264/MPEG-4 AVC also uses group of pictures (GOP) structure for encoding. A GOP is constituted of a number of frames that can form different sequences. Frames can be of 3 types:

- I-frame: intra frame or key frame.

- P-frame: forward predicted frame from one previously decoded frame.
- B-frame: bi-directional predicted frame from one or two previously coded frames.

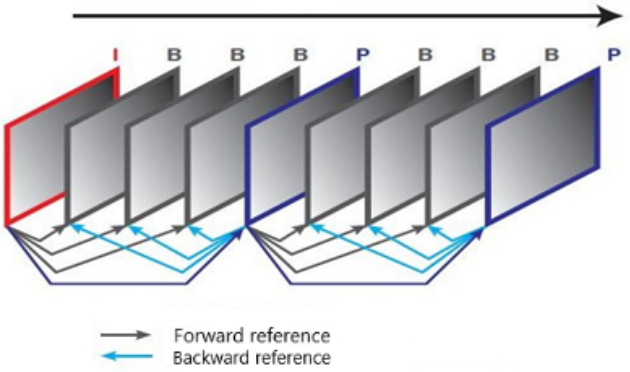

Fig. 1. Frame Prediction in a GOP of sequence: I-B-B-B-P-B-B-B-P

Moreover, each frame is divided into a set of macroblocks (MB) of sizes 16x16, 16x8, 8x16 or 8x8. Each MB can be divided further into sub-MBs of sizes as small as 4x4. The MBs are organised in slices, each slice represents a region in a frame that can decoded separately. [10]

H.264 encoding has 2 main prediction modes: intra-prediction which starts by performing spatial prediction on either a 16x16 or 4x4 macroblocks. inter-prediction where macroblocks coding is done using motion compensation in order to determine block prediction error.

Data hiding in the MPEG4 or H.264 coding standard is ideally performed by directly modifying the code word in the stream and embed information. This ensures bit rate stability and the preserving of video quality all while maintaining efficiency, as the information is hidden during the encoding process, thus avoiding decoding and re-encoding. However, this may still impact the motion vectors which has been investigated in the next sections.

3. MOTION VECTOR PATTERNS

Natural video data is generally modelled as a background subjected to simple transformations as translations, rotations and scaling caused by a shift or a zoom of the camera including independent moving objects subject to the same types of transformations/distortions. This induces high correlations between the neighbour pixels in space and time domains. This correlation is used in H.264/MPEG-4 AVC standards to compress the video file by modelling the frames as a grid of 16x16 pixels macro-blocks (MB) subjected to translation distortions. The compressed information is only a sample of the entire frame, MV maps, which contain the information of the MB's translations and residual frames to complete the loss of information. Even with this compression method, there remains spatial and temporal correlations. In this paper, only spatial correlation of the MV map is investigated and employed.

The MV map can be seen as image frames with continuous areas for the background and the moving objects. In the H.264/MPEG-4 AVC standards, the MVs are estimated with half a pixel or a quarter pixel precision. This sampling introduces homogeneous areas (Fig.2).

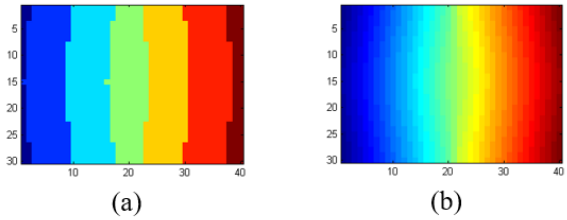

Fig. 2. simulated vertical component of the MV map with (a) and without (b) sampling.

In Fig.2 the simulated motion introduces 7 different areas. An area is defined by the same value here, MV components and a spatial continuity regarding the 4 neighbours.

The majority of the steganographic schemes are based on the Least Significant Bit (LSB) such as the method proposed by Kutter and al. [11]. It consists of defining a rule to choose the MV carriers hence introducing information in it by adding a noise corresponding to the information bit. These techniques are modelled by C.Zhang and al. [12] as an independent noise added to the vertical and horizontal components of the motion vectors. See Equation.1.

$$X_i = S_i + \eta_i \tag{1}$$

where S_i and X_i are the original and stego component of the MV corresponding to the ith MB, and η is the added steganographic noise. For the LSB method $\eta \in -w, 0, w$ where w is the range of value which was set to either 0.5 or 0.25, according to the H.264/MPEG-4 AVC standards.

The random variables represented by the variable $r\eta$ is independent and identically distributed and follow the probability mass function as given by equation.2.

$$\begin{aligned} P(\eta = 0) &= 1 - \frac{1}{2}.p \\ P(\eta = -w) &= \frac{1}{4}.p \\ P(\eta = w) &= \frac{1}{4}.p \end{aligned} \tag{2}$$

where p is the embedding rate, varying between 0, which means the video has not been modified, and 1, meaning that all the MVs are carrying the message.

The modifications of the MV value are almost negligible, but the morphology of the MV maps is subject to significant changes, even at low embedding rates. Except during the changes of scenes or chaotic scenes like smoke or crowd, the MV map represents sampled continuous curves (See Fig.2). Thus, the MV map is constituted of homogeneous areas. When a message is embedded, these areas are modified. The proposed steganalysis method is based on features extracted from these MV map areas.

4. DESCRIPTION OF THE PROPOSED METHOD

The proposed method is based on the number of continuous areas per MV map, with the areas defined in section.3, where the number of isolated MVs surface is 1 Macro Block (MB) while the 4 neighbours MVs do not have closer values; this corresponds to flat peaks. In contrast to statistical steganalysis based on the values of the MVs, the proposed method does not take into account the MV values but only the regularity/structure of the MV map. If one of the MBs is changed using the model proposed by C. Zhang and al.[12], the probability of adding isolated MBs is defined in equation.3.

$$\begin{aligned} P(addisolated = 1) &= 1 - Pb + Pb.\frac{1}{4} \\ &= 1 - \frac{3}{4}.Pb \end{aligned} \tag{3}$$

where Pb is the proportion of MVs which have at least one neighboured MV with a difference of w. For non-chaotic scenes, $Pb << 1$.

At low embedding rates, if changes occur to isolated MBs, the average number of added surfaces is defined in equation 4.

$$< N >= (1 - Pb.\frac{3}{4}.Pb).p.Nb \tag{4}$$

where p is the embedding rate and Nb the number of MB per frames.

At high embedding rates the average number of generated surfaces is inferior to the one given by the equation, but the number of added surfaces does still increase when p is increased. This increase of isolated areas has a direct impact on the number of areas present in the MV map. For a given resolution of the video sequence, the number of areas can vary significantly; for example, between 100 and 1000 for 640x480p videos. But for natural videos, the number of areas per MV map is almost proportional to the standard deviation (See Fig.3). This observation was inferred by empirical data from 1059 N9M2 Groups Of Pictures (GOP) extracted from numerous videos with different resolutions.

The slope of this relation varies with the resolution of the video. In order to separate original videos from steganographic (stego) videos, we have used four features: the number of isolated MVs per frame, the number of areas per frame,

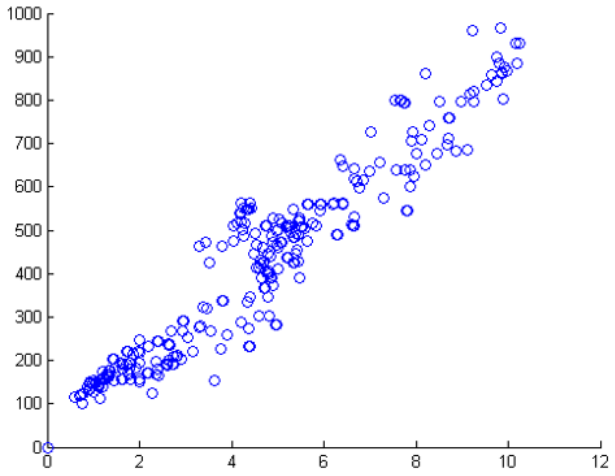

Fig. 3. Number of areas in function of the standard deviation of the MV for 640x480p video.

the variance and the number of MBs per frame (Fig.4). As videos have different scenes with different global motions, each video is separated in GOP. For each video, the decision is taken by comparing the number of positive and negative decisions in all the video GOPs.

To separate the original videos and stego videos, a Support Vector Machine (SVM) classifier is employed [13] and the quadratic kernel used. This kernel steers an optimisation problem in a 19 dimensions space. The chosen optimisation method uses a golden section search [14] applied sequentially to the 19 coefficients. To avoid the convergence in a local optimum, a Gaussian noise, decreasing with the iterations is added to the parameters.

5. EXPERIMENTS AND RESULTS

Experiments have been carried out using 110 videos with a duration of 10 to 100 seconds containing a large set of scenes: television coverage, home videos, extracts from films. Some sequences with few motions, other with complex motions or scene changes. These samples have allowed the generation of 3896 GOP N9M2 from the original videos. The same number was also obtained from the modified videos. The results are summarised using Receiver Operating Characteristic (ROC) curves as shown in Fig.5. Each curve has been obtained with one training step by varying the number of videos, and one testing step with the rest of the videos. The videos used for training were randomly chosen to have representative sets of videos. Each curve is the average of the 11 ROC curves obtained using a training sample of 100 videos and 10 test samples. Each test video is tested once only by varying the rate of insertion and the experiments carried out are independent with respect to the insertion rates.

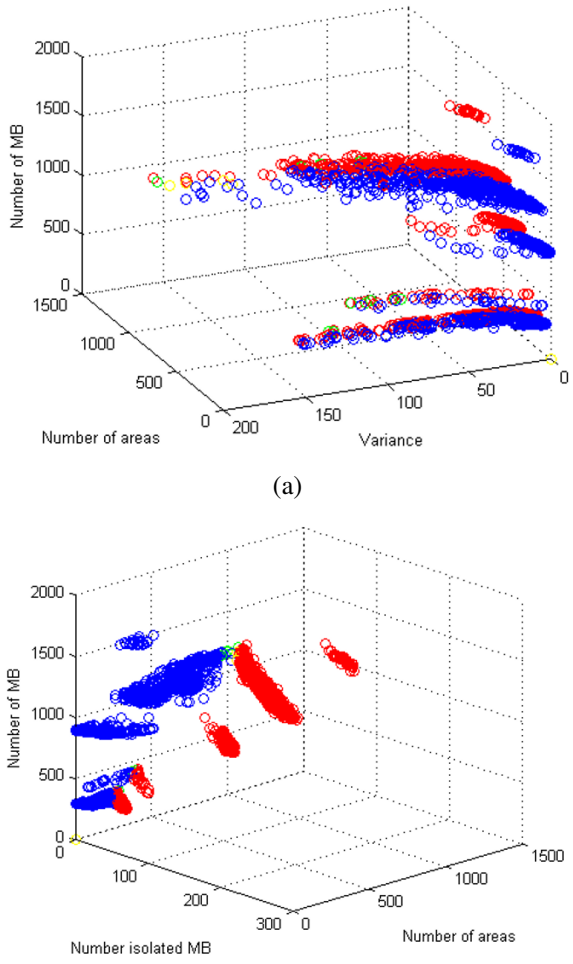

Fig. 4. Scatterplots of the four defined feature for 1440 GOP extracted from 35 original (blue) and stego (red) videos with 100% embedding rate.

The obtained results show that the proposed method is capable of separating the original 110 videos from the 110 modified ones when insertion rates are over 40%; i.e., when 20% of MVs are modified with ± 0.5 pixel. The proposed method is also capable of detecting modifications with lower rates of 10% achieving a detection rate of 85% with, however, 20% of false positives.

6. CONCLUSION

This paper presented a video steganalysis method based on the analysis of the structure of the motion vectors in the compressed video data. More precisely the investigation of features extracted from the morphology of the motion vector maps. Tests have been conducted on a large number of videos, of varying quality and length, and have produced satisfactory results suggesting the effectiveness of the proposed scheme.

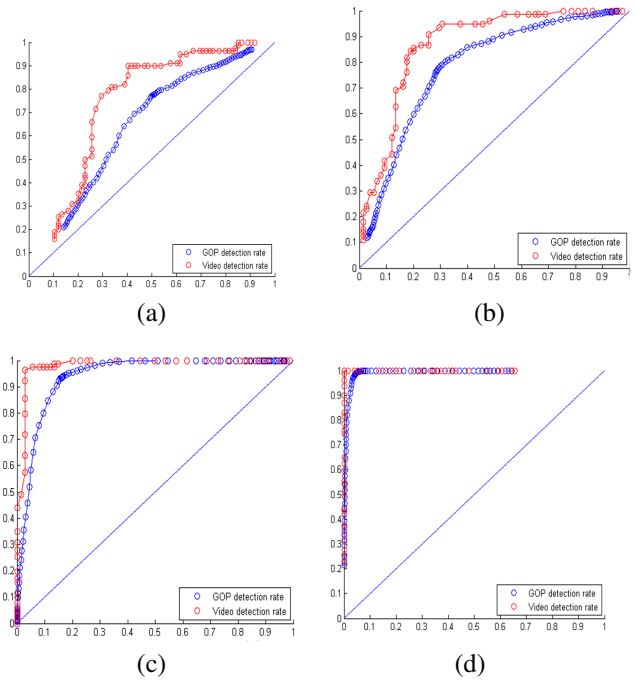

Fig. 5. ROC curves for insertion rate of 5% (a), 10% (b), 20% (c), 40% (d).

Acknowledgment

This publication was made by NPRP grant # NPRP11S-0113-180276 from the Qatar National Research Fund (a member of the Qatar Foundation). The statements made herein are solely the responsibility of the authors.

7. REFERENCES

[1] Chunhua Chen and Yun Q Shi, "Jpeg image steganalysis utilizing both intrablock and interblock correlations," in *2008 IEEE International Symposium on Circuits and Systems*. IEEE, 2008, pp. 3029–3032.

[2] Konstantinos Karampidis, Ergina Kavallieratou, and Giorgos Papadourakis, "A review of image steganalysis techniques for digital forensics," *Journal of information security and applications*, vol. 40, pp. 217–235, 2018.

[3] Tom Kellen, "Hiding in plain view: Could steganography be a terrorist tool," *SANS institute infosec reading room*, 2001.

[4] Ian Murphy, "Steganography used in attack on industrial enterprises," *enterprise times*, 2020.

[5] Hussein 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, 2010.

[6] Xuansen He and Zhun Luo, "A novel steganographic algorithm based on the motion vector phase," in *2008 international conference on computer science and software engineering*. IEEE, 2008, vol. 3, pp. 822–825.

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

[8] H ITU-T RECOMMENDATION, "264 "advanced video coding for generic audiovisual services"," 2003.

[9] Mohammed Ghanbari, *Standard codecs: Image compression to advanced video coding*, Number 49. Iet, 2003.

[10] Detlev Marpe, Thomas Wiegand, and Gary J Sullivan, "The h. 264/mpeg4 advanced video coding standard and its applications," *IEEE communications magazine*, vol. 44, no. 8, pp. 134–143, 2006.

[11] M Kutter, F Jordan, and T Ebrahimi, "Proposal of a watermarking technique for hiding/retrieving data in compressed and decompressed video," *Technical reportM2881, ISO/IEC document, JTC1/Sc29/WG11*, 1997.

[12] C. Zhang, Y. Su, and C. Zhang, "A new video steganalysis algorithm against motion vector steganography," in *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*, 2008, pp. 1–4.

[13] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144–152.

[14] Floyd Hanson, "Mcs 471 class optimization notes: Method of golden section search," .