SPATIO-TEMPORAL RICH MODEL FOR MOTION VECTOR STEGANALYSIS

Kasim Tasdemir *Fatih Kurugollu* *Sakir Sezer*

The Institute of Electronics, Communications and Information Technology (ECIT)
Queen's University Belfast, Belfast/UK
Email: {ktasdemir01, f.kurugollu, s.sezer}@qub.ac.uk

ABSTRACT

We propose a spatio-temporal rich model of motion vector planes as a part of a full steganalytic system against motion vector based steganography. Superior detection accuracy of the rich model over the previous methods has been lately demonstrated for digital images in both spatial and DCT domain. It has not been heretofore used for detection of motion vector steganography. We also introduced a transformation so as to extend the feature set with temporal residuals. We carried out the tests along with most recent motion vector steganalysis and steganography methods. Test results show that the proposed model delivers an outstanding performance compared to the previous methods.

Index Terms— video, steganalysis, motion vector, rich model

1. INTRODUCTION

As a traditional steganalysis approach, motion vector steganalysis methods start with adapting a video model within which steganalyzers are built using machine learning tools. Most of the motion vector (MV) steganalysis algorithms can be considered as a targeted steganalysis, exploiting aberrations introduced by a specific MV steganographic method. This was a common approach in the image side. However, there has been remarkable developments in image steganalysis, yet, MV steganalysis studies appear indifferent to the advancements[1]. There is an paradigm shift in image steganalysis towards employing many weak features rather than attempting to model low dimensional representation of the image. Detection success rate and flexibility of the framework of recently published steganalysis algorithm [4, 5] commanded attentions of researchers in the field of digital forensic. As the authors stated in [4, 5] that the idea can be applied to wide range of applications such as sound or video steganalysis, yet, it has not been tested in MV steganalysis studies. In Spatial Rich Model (SRM) method, noise is modelled as union of many diverse submodels formed by joint distributions of noise residuals obtained from many linear and non-linear high-pass filters. It employs ensemble classifier as final steganalyzer because of its low computational complexity. This is achieved by merging many smaller submodels and training them with a fast learning algorithm.

In this paper, we adapt the rich model to MV steganalysis and introduce a novel transformation so as to apply rich model filters to videos in temporal domain as well. We doubled the number of features by appending temporal features to spatial features. Test results illustrate that the proposed method outperforms the prominent motion based steganalysis methods (e.g. AoSO [6]) against a diverse set of stego videos. In addition, the test section of this paper has the most comprehensive comparison in the field of MV steganalysis, according to our best knowledge.

The paper is structured as follows. In Section 2 we give some background information about previous methods. The methodology to adapt the rich model to MV steganalysis and explanation of temporal to spatial transformation appears in Section 3. The experimental settings and discussions are detailed in Section 4. Lastly, in Section 5 we summarize the proposed work and stress the advantage of it.

2. RELATION TO PRIOR WORK

The contribution of the paper over previous works [4,5,7] is at least three-folds : 1) Work by Kodovsky and Fridrich [4, 5, 7] considers only spatial or DCT domain, which are both in 2D. Unlike images, video has temporal data as well. We take temporal correlation into account as well as spatial correlation which provides a better detection accuracy in video; 2) Previous MV steganalysis methods have proposed ad hoc methods [1, 2, 6, 8–12]. We show that actually image steganalysis methods could give a better performance with help of some modifications. 3) The paper presents the most comprehensive cross comparison among MV steganalysis studies. It gives a clear view of superiorities and drawbacks of both MV steganography and steganalysis methods.

Currently, there are limited number of MV steganography [13–20] and MV steganalysis algorithms [1, 2, 6, 8–12].

The very first MV steganalysis algorithm [1] and the following [2, 8] investigate the first order statistics of corrup-

This work was supported by the EPSRC under CSIT Project (EP/H049606/1)

[1]The only exception to this is that [1, 2] are inspired by [3]

tions caused by LSB embedding and models steganography as a noise added on MV magnitudes. Cao *et al.* states that expected values of MVs of recompressed video are equal to that of the cover video [9]. By using this fact, their method utilises distance between MVs of videos before and after recompression. Motion estimation is unknown to steganalyzer in a realistic scenario. It is stated in [6] that Cao's method suffers if the motion estimation method of recompression is different than that of the first compression. In [10], first a lost MV recovery algorithm using *polynomial kernel regression* on neighbouring eight MVs is proposed then the algorithm is employed for estimation of the cover MVs. A targeted MV steganalysis for LSB steganography is proposed by Tasdemir *et al.* [11]. It is stated that flat areas, which are common in MV patterns, are highly corrupted by LSB embedding and supported by a theoretical proof. The most recent MV steganalysis algorithm AoSO [6] perform a local motion estimation with search window width one pixel. It is reported that if the new MV is same as the received MV, then it is more likely to be a clean video. It is more likely to be a stego if the opposite is the case. It is not stated if half pel or full pel resolution used in the paper. We took the half pel resolution in our AoSO implementation since it is the most common option in real scenarios. AoSO has several pitfalls. It can not be used against phase modifying MV stego methods because they search MV in a different region. Hence, the new MV is always the locally optimal MV. Another problem is with bi-directionally estimated MVs. Their aggregate residual error is minimum rather than individual residual errors. Hence, they might not be locally optimal MVs individually.

3. RICH MODEL IN MOTION VECTOR DOMAIN

Each macroblock has MVs with x and y components. If the macroblock is of type B, it has four components, i.e., x and y for both backward and forward predicted MVs. In our method, we exploit the relations of MV components of the same type as it is a common way which was also used before [2, 8]. One reason for this is that the reference frame distance affects MV magnitudes [21, 22]. We are going to group MVs of same prediction type, direction and extract features from the each group separately. Thus, it would be useful to introduce a new term here. We name each MV matrices of same prediction type, same component and same frame as *MV plane*. For example, a typical B frame would have four MV planes (forward predicted x and y components, backward predicted x and y components). Let $V \in \{-W, \dots, W\}$ represent a MV component where W is the search window width in ME. Then, we denote $\mathbf{MV} = (MV_{i,j,k}) \subseteq \{\cup_{c\in\{x,y\},d\in\{f,b\}} V^{(c,d)}\}^{M\times N}$ where x and y are cartesian components of the vector, f and b are abbreviations for *forward prediction* and *backward prediction*, which are indicating the type of motion estimation direction. M and N represents row and column size of 2D macroblock array in a frame. MV plane is 2D matrix slice of $\mathbf{MV}$ at a specific frame, type and coordinate, $\mathbf{V}^{(\mathbf{c_1},\mathbf{d_1})}_{\mathbf{k_1}} = (V^{c=c_1,d=d_1}_{i,j,k=k_1}) \in \{-W, \dots, W\}^{M\times N}$.

We presently modify SRM algorithm, which is applicable to 2D data, to meet 3D MV patterns. For the detailed descriptions of the following definitions the reader is referred to the original paper [4].

The filters of rich models are only applicable in 2D. 2D SRM high-pass filters are not suitable for multidimensional MV pattern of a frame. A typical B frame would have four MV planes. There are two ways to alleviate this problem. First, one can reduce the dimension by a transformation. Second way is to apply SRM on MV planes, $\mathbf{V}^{(c_1,d_1)}_{k_1}$ individually. If the latter approach is considered, the features will correspond to a MV plane rather than a whole frame. That means a frame could possess half cover half stego MV pattern, e.g., abscissas of MVs might be stego where ordinates of the MVs are cover. Latter approach is more compelling because most of the MV steganography algorithms embed into only one component of a MV and leave the other untouched. Therefore, the image $X_{ij} \in \mathbb{R}^{n_1\times n_2}$ in SRM ([4]) is replaced with MV plane $\mathbf{V}^{(c_1,d_1)}_{i,j,k_1} \in \{-W, \dots, W\}^{M\times N}$ as follows:

$$R_{ij} = \hat{V}_{ij}(\mathcal{N}_{ij}) - cV^{(c_1,d_1)}_{i,j,k_1} \tag{1}$$

where $\hat{V}_{ij}(.)$ is a predictor of $cV^{(c_1,d_1)}_{i,j,k_1}$ using neighborhood of $V^{(c_1,d_1)}_{i,j,k_1}$. Rest of SRM algorithm does not require any modification for spatial feature extraction (q and T are taken as $\{1, 1.5, 2\}$ and 2 as in [4]). However, in order to exploit the temporal correlation of MVs, a new transformation from a 3D temporally cascaded MV planes matrices to a 2D MV planes matrix is introduced. As shown in Fig. 1a, first three MV planes corresponding to three consequent frames of the same type are grouped together. The current frame is the one in the middle because the previous and the following frames are the most related ones to the current frame. We considered only the previous and the next frame but one can extend the temporal group size. First, we take first column of the previous MV plane, $V^{(c_1,d_1)}_{i,j,k_1-1}, i = 1, \dots, M\, j = 1$, and append the first column of current MV plane, $V^{(c_1,d_1)}_{i,j,k_1}, i = 1, \dots, M\ j = 1$, and append the first and second column of next MV plane, $V^{(c_1,d_1)}_{i,j,k_1+1}, i = 1, \dots, M\ j = 1, 2$, as in Fig. 1b. We pick up the columns as we temporally go back and forth. This process resembles unfolding of an accordion so that it is named as *accordion folding*. The benefit of this transformation is rich model filters can exploit the correlation of not only temporally collocated MVs but also neighboring collocated MVs in the next or previous frames. For example, as demonstrated in Fig. 1b, a 4^{th} degree filter can examine the relation between $V^{(c_1,d_1)}_{i,j,k_1}, i = 3\ j = 1$ and $V^{(c_1,d_1)}_{i,j,k_1+1}, i = 7\ j = 2$. This simple transformation gives a surprising amount of extra accuracy in test results.

In Section4, we test both spatial only and spatio-temporal SRM steganalysis results.

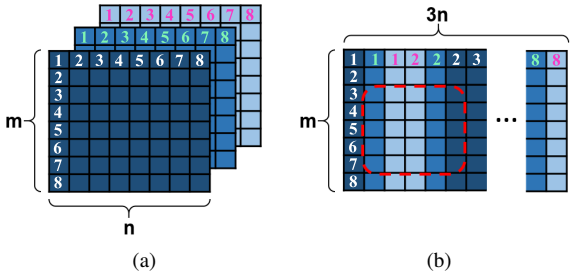

Fig. 1. Accordion unfolding. a) three consequent MV planes are concatenated to form a 3D matrix b)3D to 2D transformation by accordion unfolding.

After unfolding the MV plane group, SRM method is employed. Note that SRM is employed to both abscissas ($V_{i,j,k}^{(x,d)}$) and to the ordinates ($V_{i,j,k}^{(y,d)}$) of the MVs separately since one of them might be clean where the other is carrying a message. Our final feature set has $(34671 + 34671) = 69342$ as it includes spatial and temporal features.

4. EXPERIMENTS

The most recent MV steganalysis method AoSO [6] is implemented as well as other three methods, namely, DengCom [2], DengRec [10], Su [8]. Xu [15], Aly [19], He [17], Pan [18] and Fang [16] are implemented for message embedding by the help of an open source library [23]. Some steganalysis methods above are only applicable to P frames only. If such methods were being tested, IPPP GOP structure was used. IBPBI GOP structure is used for the rest. 100 unique CIF sized videos each contains 100 frames are encoded with the same setting (i.e. One Mb/s bitrate, full MV search, no field frames, progressive, 4:2:2 chroma subsampling ...) apart from GOP type. The final data set was comprised of total 700 stego and cover videos which makes 70000 frames[2] on total. All MV steganography algorithms allow user to choose threshold rather than payload. Thus, in a realistic case a data set with a predefined payload can not be built. Tests are carried out for a realistic scenario and algorithms are not modified to meet our test set. They are tested for ten different ranges of payloads. Nevertheless, we slightly abuse the conventional meaning of the term *payload* here because of the reasons we presently elaborate on.

A *unit* is group of consequent frames of the same type. The feature arrays of some steganalysis algorithms are extracted from a unit where each frame in it has different payload. This makes it impossible to carry out a pairwise comparison because of two reasons:
Firstly, some steganalysis algorithms give decision results in frame or MV plane resolution whereas some others give the results in unit resolution.
Secondly, it is unlikely to have a unit with full embedding rate, where every MV component is carrying a message bit. There would be many gaps in the comparison due to lack of samples with all payloads. To overcome this problem, we considered the maximum of total bits embedded by a steganography method in a unit, which is defined by each steganalysis differently, in our data set as the maximum payload of the steganography method for that steganalysis. Then the test set is divided into ten payload ranges from [0,0.1] to [0.9,1].

Aly's algorithm requires setting values for lower and upper limits of Prediction Error Frame (PEF). In our tests, they are set to 15db and 60db respectively. Thresholds of other embedding methods are set to 5 and number of regions are set to 16, 16 and 8 for Pan, He and Fang respectively.

Test videos are generated using raw image sequences [24]. Each video has five stego and two cover versions, one with IPPP and the other with IBPB GOP sequence. A randomly generated secret message bit sequence is embedded to each stego video. Half of the stego videos are used for training the steganalytic systems and the other half is used for tests. If the stego set of a steganalysis method has IPPP GOP sequence, we trained and test it with the cover set which has IPPP GOP sequence as well. IBPB is used otherwise. All settings of training and testing were applied as explained in the steganalysis papers.

All test results are shown in Fig. 2. Each data point in this figure is the average of 10 train-test results. Each graph shows the accuracy vs payload plots of all steganalysis algorithms against a steganography algorithm given in the title of the graph. Accuracy is calculated as:

$$P_A = 1 - (P_F + P_M) \tag{2}$$

where P_A is accuracy, P_F and P_M are probability of false detection and miss detection. We tested our proposed Accordion Unfolding SRM (ASRM) with spatial only features (SRM spatial) and spatial + temporal features together (ASRM spatial+temporal) in order to observe the effects of temporal and spatial features individually.

Fig. 2 shows that proposed ASRM method have better accuracy against all steganography methods tested. It has an outstanding detection accuracy especially against Xu's and Aly's steganography methods. Temporal features give extra around 5% accuracy in mid and high payloads, 20% in low payloads against Aly and Xu. Still, temporal features improve detection accuracy for other steganography methods as well.

Deng Rec, Su, AoSO and Deng Com steganalysis algorithms follow ASRM spatio-temporal and SRM spatial methods in this order. However, when the payload is greater than 0.5, Su's method has better accuracy than AoSO against Xu.

[2]This big data set does not cause a *curse of dimensionality* because every payload range is trained-tested separately.

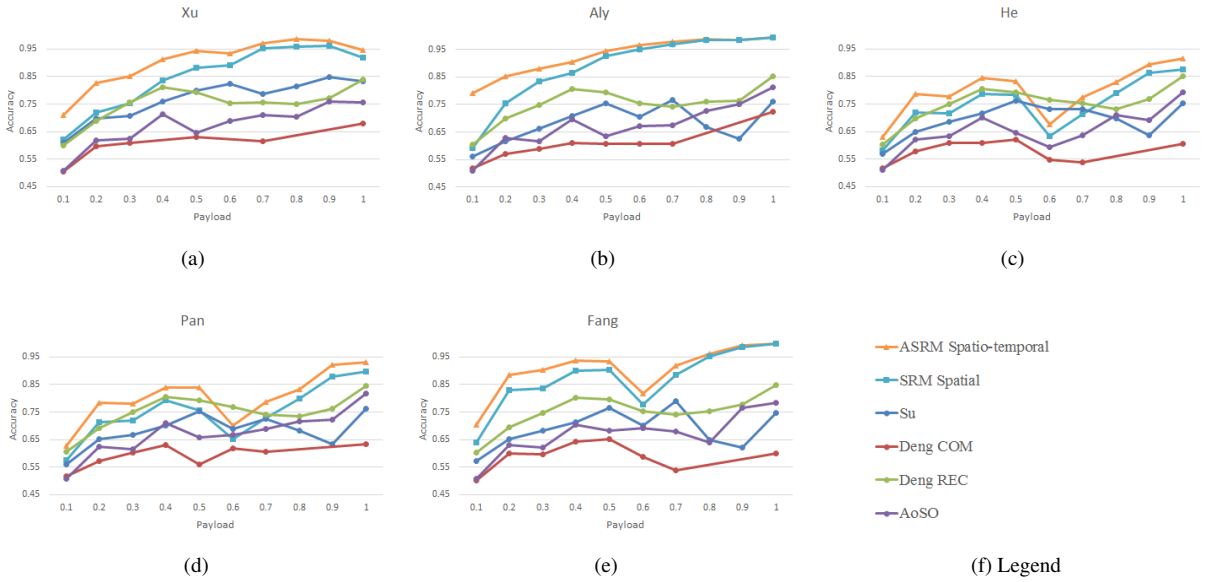

Fig. 2. Test results are given as comparison of all steganalytic systems against a selected stego algorithm. a) Xu, b) Aly, c) He, d) Pan, e) Fang, f) legend for steganlaysis methods.

Cartesian coordinate was divided into 16 regions in Pan and He, eight regions in Fang. Fig. 2e shows that extra eight regions give more security against ASRM algorithm but coset syndrome coding used in Pan Fig. 2d does not improve the security of it.

The most degradation in MV patterns is caused by Aly's algorithm because it does not alter only the nonzero MVs but also the zero MVs. Typically, MV patterns have wide regions with zero MVs. These regions are readily detectable and unavoidable traps for Aly's steganography. Xu's method is coming after Aly in terms of detectability. Test results also show that phase based MV modification is safer against current adversary methods than magnitude base steganographies.

Since we do not have control over payload directly, our test and training set have different sizes at different payloads. There are less samples in between 0.5 and 0.8 payload than the rest. This is the reason why there is a drop in detection accuracy of the steganalysis methods in that range.

AoSO did not perform better than previous steganalysis methods. The reason is half pel resolution, which is more usual, is used in motion estimation stage of our test set. Locally optimal MV search is bounded to a 3×3 half pel sized search box. Another reason is that it is only supposed to work if the MV has not been changed more than one pel. Nevertheless, Pan, He and Fang methods rotate the MV much further than one pel distance. Even the extensive modification makes it conducive to be easily detected. This precludes AoSO from discerning the abnormality. Hence, AoSO can not be used against phase based MV steganography algorithms.

5. CONCLUSION

The purpose of the paper is to present a supervised universal MV steganalysis, which has the best detection accuracy among previous methods.

Of late, there has been remarkable improvements in digital image steganalysis. Especially the rich model has drawn researchers' attention by virtue of its generic and flexible framework. In this study, rich model is adapted for MV steganalysis. To the our best knowledge, this study is the first to import an image steganalysis to the video side. Furthermore, we introduced a novel temporal transformation so that rich filters enjoy the benefits of temporal correlation of video. The introduced simple 3D to 2D transformation to exploit both spatial and temporal correlation in MV patterns gives an extra 5-20% accuracy.

The test section has the most comprehensive comparison in the field of MV forensics. Five steganography algorithms are tested against six adversary methods including the proposed ASRM spatio-temporal method. The tests show that overall detection accuracy of the proposed method is above all other adversary methods. It has an outstanding detection rate especially against Xu's and Aly's method.

6. REFERENCES

[1] Chengqian Zhang, Yuting Su, and Chuntian Zhang, "A new video steganalysis algorithm against motion vector steganography," in *Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on*, Oct 2008, pp. 1–4.

[2] Yu Deng, Yunjie Wu, Haibin Duan, and Linna Zhou, "Digital video steganalysis based on motion vector statistical characteristics," *Optik - International Journal for Light and Electron Optics*, vol. 124, no. 14, pp. 1705 – 1710, 2013.

[3] AD. Ker, "Steganalysis of lsb matching in grayscale images," *Signal Processing Letters, IEEE*, vol. 12, no. 6, pp. 441–444, June 2005.

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

[5] Jan Kodovsk and Jessica Fridrich, "Steganalysis of jpeg images using rich models," *Proc. SPIE*, vol. 8303, pp. 83030A–83030A–13, 2012.

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

[7] J. Kodovsky, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 2, pp. 432–444, April 2012.

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

[9] Yun Cao, Xianfeng Zhao, and Dengguo Feng, "Video steganalysis exploiting motion vector reversion-based features," *Signal Processing Letters, IEEE*, vol. 19, no. 1, pp. 35–38, Jan 2012.

[10] Yu Deng, Yunjie Wu, and Linna Zhou, "Digital video steganalysis using motion vector recovery-based features," *Appl. Opt.*, vol. 51, no. 20, pp. 4667–4677, Jul 2012.

[11] K. Tasdemir, F. Kurugollu, and S. Sezer, "Video steganalysis of lsb based motion vector steganography," in *EUVIP, 2013 4th European Workshop on*, June 2013, pp. 260–264.

[12] Hui Ye, Weiming Zhang, Yuanzhi Yao, Cong Kong, Hao Huang, and Yu Nenghai, "Motion vector-based video steganalysis using spatial-temporal correlation," in *Image and Signal Processing (CISP), 2013 6th International Congress on*, Dec 2013, vol. 01, pp. 148–153.

[13] Jun Zhang, Jiegu Li, and Ling Zhang, "Video watermark technique in motion vector," in *Computer Graphics and Image Processing, 2001 Proceedings of XIV Brazilian Symposium on*, Oct 2001, pp. 179–182.

[14] Bodo Yann, Laurent Nathalie, and Dugelay Jean-Luc, "A scrambling method based on disturbance of motion vector," in *Proceedings of the Tenth ACM International Conference on Multimedia*, 2002, pp. 89–90.

[15] Changyong Xu, Xijian Ping, and Tao Zhang, "Steganography in compressed video stream," in *ICICIC '06. First International Conference on*, Aug 2006, vol. 1, pp. 269–272.

[16] Ding-Yu Fang and Long-Wen Chang, "Data hiding for digital video with phase of motion vector," in *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*, May 2006, pp. 4 pp.–.

[17] Xuansen He and Zhun Luo, "A novel steganographic algorithm based on the motion vector phase," in *Computer Science and Software Engineering, 2008 International Conference on*, Dec 2008, vol. 3, pp. 822–825.

[18] Feng Pan, Li Xiang, Xiao-Yuan Yang, and Yao Guo, "Video steganography using motion vector and linear block codes," in *ICSESS, 2010 IEEE International Conference on*, July 2010, pp. 592–595.

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

[20] Yun Cao, Xianfeng Zhao, Dengguo Feng, and Rennong Sheng, "Video steganography with perturbed motion estimation," in *Information Hiding*, vol. 6958 of *Lecture Notes in Computer Science*, pp. 193–207. Springer Berlin Heidelberg, 2011.

[21] Kasim Tasdemir and A.Enis Cetin, "Content-based video copy detection based on motion vectors estimated using a lower frame rate," *Signal, Image and Video Processing*, vol. 8, no. 6, pp. 1049–1057, 2014.

[22] Kasim Tasdemir and A.Enis Cetin, "Motion vector based features for content based video copy detection," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, Aug 2010, pp. 3134–3137.

[23] "Mpeg1/2 - mpeg software simulation group," 2014.

[24] "Video test media [derf's collection]," 2014.