UNIVERSAL SPATIAL FEATURE SET FOR VIDEO STEGANALYSIS

Xikai Xu, Jing Dong and Tieniu Tan

National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences
E-mail: xikaixu@gmail.com, { jdong, tnt }@nlpr.ia.ac.cn

ABSTRACT

In this paper, we propose a universal spatial feature set for video steganalysis. This feature set comprehensively exploits the correlation of adjacent pixels and can be viewed as a generalized extension of most correlation based features. We also develop a new approach to extract inter-frame features for video steganalysis. Our method can be universally applied to detect different video steganographic algorithms regardless of video format. The experimental results show that it outperforms current correlation based methods.

Index Terms**—**Video steganalysis, stegnography, data hiding, inter-frame feature, video slice

1. INTRODUCTION

Steganography is the art of hiding the very presence of a secrete message in innocuous-looking cover medium, such as text, audio, image and video [1]. The main purpose of steganography is for secret communication via public channel without drawing any suspicion. As the opposite of steganography, steganalysis is developed to detect the hidden messages transmitted through the cover media. With the development of network and multimedia technologies, various videos can be acquired from the Internet easily. Thus, video becomes a very promising cover candidate that has a very high payload capacity for data hiding. Nowadays, although many steganographic algorithms for video have been proposed [2-7], there are very few video steganalysis algorithms [8-13], since video steganalysis is very complex especially for compressed videos.

The basic assumption for steganalysis is that the embedding of a message changes some statistical properties of the cover-object. Hence, the goal of the steganalyzer is to find and measure these distortions. Usually pattern classification technique is employed, in which discriminative features are extracted from cover and stego objects and then a classifier as detector is trained using machine learning methods. For image steganalysis, most schemes try to extract effective features which are insensitive to image content but discriminative between cover and stego images. Currently, most effective features are based on the correlation of adjacent elements under an intuitive assumption that natural image has strong dependence in local region and such dependence will be weakened by the data hiding operation. Therefore, how to use the correlation of adjacent elements becomes a key problem. A feasible way is using Markov chain to model the inter-pixel dependence and then taking the elements of the empirical transition matrices as features, such as Zou et al.'s 324-D feature [14], SPAM feature [15] and NIP feature [16]. These feature sets can be applied directly to detect video frames, but they only consider the correlation of adjacent pixels in few directions, thus they are insufficient in describing the relationship of adjacent pixels and unable to describe inter-frame correlation. Hence it is not enough to apply them in detecting steganographic algorithms in video, especially in compressed format. To solve this problem, in this paper, we focus on exploiting and describing the inter-dependence of elements in all kinds of adjacency structure in video frame to get a universal spatial feature set for video steganalysis. To improve the detection performance, we also develop a new approach to utilize the inter-frame correlation. Experimental results show that our method can effectively detect stego-videos embedded by four popular video steganographic methods.

The remainder of the paper is organized as follows: Section 2 is the motivation of our method. In Section 3, we derive a generalized description of the correlation of adjacent pixels and propose our universal spatial feature set. Then the experimental results are presented in Section 4. Finally, the conclusion is drawn in Section 5.

2. MOTIVATION

Videos are often stored and transmitted in compressed format. Many international standards are set for video compression, such as MPEG-2, MPEG-4, ITU-T H.263 and H.264, etc. and most of them adopt hybrid coding. These schemes are based on the principle of reducing the redundancies in spatial domain and temporal domain by using block-based transform coding and motion compensated prediction. Video steganography may carry out in the process of compression and encoding, which means modification for embedding can occur in every stage of video compression. Even so, most data hiding operations may leave their traces in the video after decompression, which allows us to extract features from decompressed video for steganalysis. As a universal steganalysis method for video, the advantage of extracting features from uncompressed domain is obvious: we do not need to deal with the specialties of compression and encoding process for certain video. The key point is that such features should be able to capture the changes introduced by data hiding after video decompression.

Inspired by image steganalysis, we believe that extracting features based on the correlation of intra-frame and inter-frame

978-1-4673-2533-2/12/$26.00 ©2012 IEEE

pixels is a promising way for video steganalysis. Currently, many Markov based methods designed for image steganalysis only consider the correlation of adjacent pixels along certain directions. They are not sufficient for video steganalysis, thus, we want to derive a generalized description that can exploit comprehensively the correlation of pixels in any adjacency structures and involve all correlation based features to improve the steganalysis performance. Similar strategy was previously developed for image steganalysis and named Rich Model proposed by Fridrich et al [17] which obtained very good result. Our work can also be viewed as an extension of a spatial domain based Rich Model feature for video steganalysis.

3. PROPOSED METHOD

Since video data can be viewed as a set of continuous frames, in this section we describe our feature respectively in two aspects: intra frame and inter frame.

Usually, the frames of video will be split into Y, Cb, Cr channels and resampled in compression process. The Cb and Cr channels are down sampled. Thus, most video steganographic algorithms embed messages in the Y channel instead of Cb or Cr channel for its capacity. Therefore, in this paper, we just focus on the Y channel (8-bit gray value).

As most data hiding operations are very slight, the description of pixels' correlation should be very precise and sensitive for steganalysis. It is appropriate to use the joint probability mass function which can be estimated from the co-occurrence matrix of adjacent pixels.

3.1. Intra frame

First, let us consider the case of intra frame. In fact, our feature set is composed of several subsets and each subset corresponds to a specific type of adjacency structure of pixels. Although these structures are different, the processing steps of them are similar. Without loss of generality, we first describe the simplest case: adjacency structures of three pixels. Other subsets corresponding more complex cases can be obtained by similar manner.

Adjacency of three pixels can be selected in 20 different types of arrangement as shown in Fig.1, where I_i is the gray value of pixel.

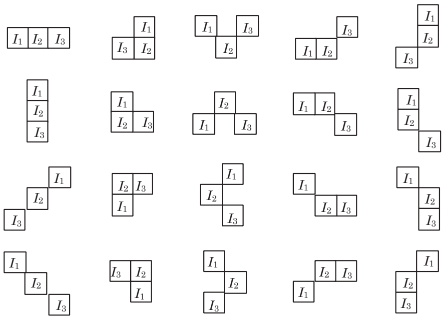

Fig.1: Adjacency of three pixels in a video frame

Each type of arrangement can be considered as an adjacency structure. In a frame, for each of these structures, we can calculate the co-occurrence matrix of its three pixels and get the joint probability matrix: $Pr(I_1, I_2, I_3)$, which has 256^3 elements. However, this matrix is too large to compute and useless for steganlaysis, because most of its elements reflect the image content rather than the stego information. Actually, we just care about the interdependence of adjacent pixels, not their absolute values for steganalysis. So we can choose I_2 as reference, and calculate the differences between I_2 and its adjacent pixels as following (d_i denote the difference):

$$d_1 = I_1 - I_2, \qquad d_2 = I_3 - I_2$$

Then $Pr(I_1, I_2, I_3)$ can be transformed to $Pr(d_1, 0, d_2)$ which is equivalent to $Pr(d_1, d_2)$. That means the correlation of three adjacent pixels is reflected by dependence of d_1 and d_2.

It is believed that the adjacent pixels with small differences have higher correlation. Compared to the irregular sharp edges, they are more proper for steganalysis. For this reason, we define a threshold T, and the difference d_i is truncated according to the following rule:

$$Threshold(d_i)\begin{cases} d_i, & -T < d_i < T \\ -T, & d_i \le -T \\ T, & d_i \ge T \end{cases}$$

For each type of arrangement, we estimate the joint probability distribution of d_1 and d_2 by calculating the frequency of d_1 with value i and d_2 with value j occurs within a frame, as following:

$$M^k(i,\ j) = Pr(d_1 = i, d_2 = j)\ ,$$
$$\text{where } i, j \in \{-T, \ldots, T\},\ \ k \in \{1,2,\ldots,20\}.$$

We can get 20 joint probability matrices in total, and their elements can be cascaded as a feature vector for steganalysis.

Then, we take more adjacent pixels into account, such as pixels in a 3×3 region, and we can find a generalized description of the correlation.

First, we take the central pixel of the 3×3 region as reference, and then calculate the differences between the center and its eight adjacent pixels (as shown in Fig. 2) and we can get a set of differences: D: $\{d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8\}$.

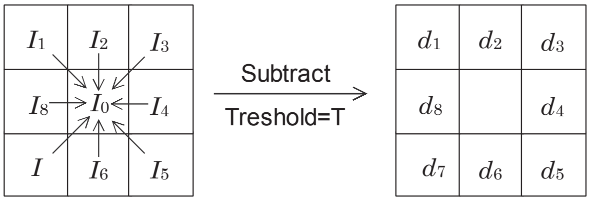

Fig. 2: Differences between the center and its 8 adjacent pixels in a 3×3 region

Take the 3×3 block and the differential operation as a template and let it go over the entire frame, for each d_i, we count the occurrence of every value from –T to T and then we can get 8 histograms. We estimate the joint probability of differences by calculating the co-occurrence of their histograms.

For each type of arrangement of three adjacent pixels, we can find the corresponding joint probability of two differences from D. In other words, to describe the interdependence of three adjacent pixels, we can calculate the joint probability of any two differences of D as following:

$M^{u,v}(i,j) = Pr(d_u = i, d_v = j)$, where $i,j \in \{-T,\dots,T\}$, $u,v \in \{1,2,\dots,8\}, u \neq v$.

As we all know, there are 28 different choices to choose any 2 from 8 differences, but some will be double counted when the 3×3 template goes over the entire frame. Therefore, there are 20 different choices correspond to those 20 different types of arrangement of three adjacent pixels.

Similarly, to describe the inter-dependence of four adjacent pixels, we can calculate the joint probability of any 3 differences of D as following:

$M^{u,v,w}(i,j,k) = Pr(d_u = i, d_v = j, d_w = k)$, where $i,j,k \in \{-T,\dots,T\}$, $u,v,w \in \{1,2,\dots,8\}, u \neq v \neq w$.

And there are 40 different choices correspond to 40 different types of arrangement of four adjacent pixels.

Therefore, this is a generalized description of pixels' correlation, and in this way, we can describe the interdependence of five, six, up to nine pixels in a 3×3 region. By an extension of this logic, we can also describe the interdependence of pixels in a larger local region such as 4×4 or 5×5 region.

In summary, we can calculate the joint probability of $N-1$ differences to describe the interdependence of N adjacent pixels. Each choice of the differences corresponds to a specific type of adjacency structure and yields a joint probability matrix. All these joint probability matrices compose a universal spatial feature set for steganalysis. This feature set includes all current correlation based features so that the dimensionality of the feature space is very high. In practice, we can use some subsets of it or employ some dimension reduction techniques to alleviate over-learning problem. Each subset can be described using three parameters(R, N, T), R – size of the local region, N – number of adjacent pixels involved, and T – threshold of the difference.

3.2. Inter frame

Besides intra frame, we also extract features from inter frame to improve the detection performance. For slow-moving sequences the frames in one scene are quite similar and the static or slow moving parts (background) are even the same. Even for high-motion movies, there are often some parts that are very similar in adjacent frames. Such temporal correlation among adjacent frames will be disturbed by the data hiding operation, which can be exploited for video steganalysis. The generalized description above can be also used to describe such inter-frame correlation. We extract features in a new way by stacking the video sequences as a 3D signal like a cube (as shown in Fig. 3, *xoy* denotes the frame plane, *t* denotes the index of frames). We section the cube along two different directions to get x-t and y-t slices. Each slice can be treated as an image. The inter-dependence of adjacent pixels in such image reflects the inter-frame correlation. Features are extracted from these slices using the same method as described in Section 3.1.

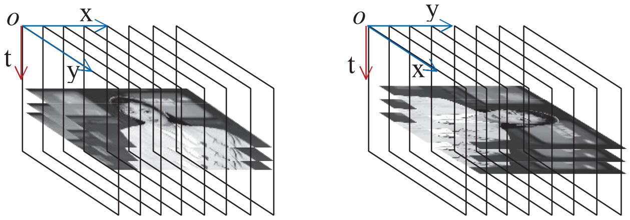

Fig. 3: Vertical section of the video "cube"

The greatest advantage of extracting features from the slices is that it is robust to the complex texture in video frames. No matter how complex the background in a video is, it changes smoothly between frames if the camera is not moving all the time. Hence, the hiding operation may be easily detected.

4. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed method, several experiments have been carried out. The video database is composed of 50 videos downloaded from the internet, the frame size is 288×352 and the total number of frames is about 15000. We hide data in these videos by using four different steganographic algorithms: two spread-spectrum watermarking algorithms in spatial domain: Hartung's SS [2] and JAWS [3], they all embed data in raw format video sequences. The other two are MSU StegoVideo [18] and Liu et al.'s 'compressed video secure steganography' (CVSS) [5]. MSU StegoVideo is a public tool for hiding information in video and can resist to video compression, but its algorithm details are not known yet. CVSS is a typical video steganographic method in the compressed domain which embeds data by modifying DCT (AC) coefficients selected using a security estimation strategy. For these two methods, in our experiments, the cover and stego videos are all MPEG-4 encoded using XVID codec with a bit rate of 500kbps.

Our experiments consist of two parts, and in both parts, we choose to use a feature subset with such parameters: $(R = 3, N = 3, T = 4)$, considering the trade-off between exploiting neighbor correlation and low dimensionality of feature space. We use SVM with RBF kernel as classifier and randomly select 75% samples as training set and the rest as testing set.

In the first part, the features are extracted individually from the frames of the videos. We compare our feature set with SPAM feature for two reasons: First, SPAM is one of the most effective correlation based spatial feature for image steganalysis and we can compare our feature set with it to verify the advantages of our method in utilizing the correlation of adjacent pixels. Second, there are very few video steganalysis algorithms and none of them has public source code, especially for universal video steganalysis.

In the other part, the features are extracted from x-t slices and y-t slices respectively. Here we set t=300 (frames), thus the x-t slice size is 352×300 and the y-t slice is 288×300.

The experimental results of the above two parts are respectively shown in Table 1 and Table 2. We use true positive (TP), true negative (TN), and average rate (AR) to compare the detection performance. True positive rate stands for proportion of stego samples be correctly classified, and vice versa the true negative. Average rate is the average value of TP and TN.

Table 1. Experimental results of frame steganalysis

	SPAM (%)			Our method (%)		
	AR	TN	TP	AR	TN	TP
Hartung's SS	95.6	95.3	95.9	98.1	97.7	98.5
JAWS	97.3	98.1	96.5	99.7	99.4	99.9
MSU StegVideo	93.4	92.4	94.3	96.3	96.2	96.4
CVSS	68.7	67.6	69.8	74.1	72.6	75.5

Table 2. Experimental results of slice steganalysis

	Our method (%)					
	x-t slice			y-t slice		
	AR	TN	TP	AR	TN	TP
Hartung's SS	98.5	98.3	98.7	98.2	97.9	98.4
JAWS	99.5	99.2	99.7	99.4	99.3	99.5
MSU StegVideo	95.1	94.2	96.0	94.4	96.2	92.6
CVSS	85.2	83.5	86.9	83.6	82.7	84.5

From the experimental results, it is clear that our feature set is very effective. We can also find that, for CVSS, the detection rate of slices is much higher than that of frames, because CVSS changes the correlation of pixels inter frame more than intra frame. This indicates the advantage of using slices for video steganalysis.

The "inter-frame consistency" is a basic assumption for inter-frame feature in our method. In practice, a video may include many different scenes, and we can first segment the video into pieces according to its content. For each piece, we can respectively extract features from frames, x-t slices and y-t slices and fuse their detection results by average, medium, or max value strategy.

5. CONCLUSION

This paper has introduced a universal spatial feature set for video steganalysis. The experimental results show that a subset of our feature set is more effective than SPAM, the state-of-the-art correlation based spatial feature. It can be expected that if we use more subsets, the accuracy will be higher. Also, the new way that extracts features from the slices of the video is proved to be useful. Although the validity of the proposed feature set should be verified in detecting more video steganography, our method provided a generalized and effective way of constructing feature set for video steganalysis. However, our method may not be very suitable for Motion Vector (MV) based video steganography since the modification of motion vector changes the correlation of adjacent pixels little.

In the future, we will try to use the correlation of more than three adjacent pixels to detect some new video steganographic algorithms including motion vector based ones. And we will also employ some dimensionality reduction techniques or ensemble classifier to cope with the high dimensionality problem of the feature space.

6. REFERENCES

[1] J. Fridrich and M. Goljan, "Practical steganalysis of digital images: state of the art," *Proc. of SPIE*, Vol. 4675, pp. 1-13, 2002.

[2] F. Hartung and B. Girod, "Watermarking of Uncompressed and Compressed Video," *Signal Processing*, Vol. 66, Issue 3, pp. 283-301, 1998.

[3] T. Kalker, G. Depovere, J. Haitsma, "A Video Watermarking System for Broadcast Monitoring," *Proc. of SPIE*, Vol. 3657, pp.103-112, 1999.

[4] Hua Cao, Jingli Zhou, and ShengshengYu, "An Implement of Fast Hiding Data into H.264 Bitstream based on Intra-Prediction Coding," *Proc. of SPIE*, Vol. 6043, pp. 123-130, 2005.

[5] Bin Liu, Fenlin Liu, Chunfang Yang, and Yifeng Sun. "Secure Steganography in Compressed Video Bitstreams," *Proceedings of the 3rd International Conference on Availability, Reliability and Security*, pp. 1382-1387, 2008.

[6] Bo Wang and Jiuchao Feng, "A chaos-based steganography algorithm for H.264 standard video sequences," *Proc. of ICCCAS*, pp. 750-753, May, 2008.

[7] KS Wong, K Tanaka, K Takagi, and Y Nakajima, "Complete Video Quality-Preserving Data Hiding," *IEEE Trans on Circuits and Systems for Video Technology*, Vol. 19, pp. 1499-1512, 2009.

[8] U. Budhia, D. Kundur, T. Zourntos, "Digital Video Steganalysis Exploiting Statistical Visibility in the Temporal," *IEEE Trans on Information Forensics and Security*, pp. 502-516, 2006.

[9] Vinod P. and A.T.S.Ho., "Improving video steganalysis using temporal Correlation," *Intelligent Information Hiding and Multimedia Signal Processing*, *Third International Conference on*, pp. 287-290, 2007.

[10] Bin Liu , Fenlin Liu, Ping Wang, "Inter-frame Correlation Based Compressed Video Steganalysis," *CISP*, *Congress on*, Vol. 3, pp. 42-46, 2008.

[11] Julien S. Jainsky, Deepa Kundur, Don R. Halverson, "Towards digital video steganalysis using asymptotic memoryless detection," *Proc. of Multimedia & security*, pp. 161-167, 2007.

[12] C. Zhang, Y. Su, and C. Zhang, "A New Video Steganalysis Algorithm against Motion Vector Steganography," *WiCOM*, *4th International Conference on*, pp. 1-4, 2008.

[13] C. Zhang and Y. Su, "Video SteganalysisBased on Aliasing Detection," *Electronics Letters*, Vol.44, no. 13, pp. 801-803, 2008.

[14] D. Zou, Y.Q. Shi, W. Su, and G. Xuan, "Steganalysis based on Markov Model of Thresholded Prediction-Error Image," *Multimedia and Expo, IEEE International Conference on*, pp. 1365-1368, 2006.

[15] T. Pevny, P. Bas, J. Fridrich, "Steganalysis by Subtractive Pixel Adjacency Matrix," *Information Forensics and Security*, *IEEE Transactions on*, Vol. 5, Issue: 2, pp. 215-224, 2010.

[16] Qingxiao Guan, Jing Dong, Tieniu Tan, "An effective image steganalysis method based on neighborhood information of pixels," *Proc. of ICIP*, pp. 2721-2724, IEEE, 2011.

[17] J. Fridrich, J. Kodovský, V. Holub and M.Goljan, "Steganalysis of Content-Adaptive Steganography in Spatial Domain", *Proc. of the Information Hiding Conference*, pp. 102-117, 2011.

[18] http://www.compression.ru/video/stego_video/index_en.html