Improving Video Steganalysis Using Temporal Correlation

Vinod Pankajakshan and A. T. S. Ho
University of Surrey
Department of Computing
Guildford - GU2 7XH, UK
{v.pankajakshan, a.ho}@surrey.ac.uk

Abstract

Steganalysis, the method to detect steganographically embedded hidden messages in digital data, has received an increasing interest in recent years. Although significant research efforts have been devoted to develop steganalysis techniques for still-images, video steganalysis remains largely an explored area. In this paper, we investigate how the temporal correlation in the neighbouring frames of a video sequence can be exploited to improve the video steganalysis performance. The reported experimental results indicate that significant improvement in the detection rate, as high as 30%*, can be achieved by exploiting the temporal correlation.*

1. Introduction

In [2], Budhia *et al.* proposed a steganalysis method to detect the presence of additive Gauassian spread-spectrum watermarks in a video sequence. The steganalysis method is based on the sensitivity of watermarking algorithms to *temporal frame averaging* (TFA) attack. The TFA is an effective inter-frame collusion attack in which the frames in a temporal window are averaged to remove uncorrelated watermarks. The difference between a test sequence and the corresponding TFA attacked sequence will exhibit different statistics depending on the sensitivity of the watermark carried by the sequence to the TFA attack. If the watermark embedded in the successive frames are uncorrelated with each other, then the difference signal will be Gaussian distributed. On the other hand, if the sequence does not carry any watermark, the difference signal will be non-Gaussian. Thus, based on the level of Gaussianity in the difference frames between the test and its TFA attacked version, one can decide on the presence/absence of a watermark. A pattern classifier, trained using some features that measure the level of Gaussianity in the difference frames, can be used to make this decision. The authors of [2] proposed to use the *kurtosis*, *entropy* and the 25^{th} *percentile* as the feature vectors.

In a related work, V.Pankajakshan *et al.* [4] proposed an *oracle* to detect the presence of uncorrelated watermarks embedded along the motion-trajectories of a given video sequence. The basic idea behind the oracle is to capture the statistical changes introduced in the motion-compensated prediction error frames (PEFs) of a video sequence due to the presence of Gaussian watermarks. The feature vectors extracted from the PEFs are used to train a pattern classifier which detects the presence/absence uncorrelated watermarks along the motion-trajectories. It has been shown that there is a clear difference in the histograms of the local variances estimated from the PEFs corresponding to a sequence without any watermark and those carrying uncorrelated watermarks. The histogram of the local variances estimated from each PEF is approximated as a 2-parameter Gamma distribution and the shape parameter of the Gamma approximation is considered as the feature vector. One advantage of this scheme is that the PEFs can be obtained directly from a compressed sequence, thereby greatly reducing the computational requirements of the oracle.

The above mentioned video steganalysis methods depend on the statistics of the embedded watermark, i.e., Gaussianity, and hence not suitable for detecting non-Gaussian watermarks. Another drawback of these steganalysis techniques is that it is assumed that each pixel in the frames are modified by the watermark embedding process. However, this assumption may not valid since the steganographer may modify only a portion of the pixels to avoid detection by steganalysis techniques. It could be possible to develop a video steganalysis technique by adapting any of the existing blind-steganalysis techniques already proposed for still images [1, 3, 7], in a frame-by-frame manner. However such an approach is sub-optimal in the sense that the temporal correlation in the video sequences is not taken into account.

This paper proposes a new blind video steganalysis scheme. In this scheme, a well-designed blind-steganalysis scheme for still images is modified by exploiting the temporal correlation in the video sequences. The rest of the paper is organized as follows: Section 2 describes the features for blind-steganalysis. Section 3 presents the proposed video steganalysis scheme. The experimental results are reported in Section 4 and finally, Section 5 concludes the paper.

2 Moments of wavelet characteristic functions

In [7], Xuan *et al.* proposed an image steganalysis method using the statistical moments of the wavelet characteristics function (CF) as the feature vectors. The image is first decomposed using 3-level spatial discrete wavelet transform (DWT) and features are extracted from the sub-band images. Due to the decorrelating property of the wavelet transform, the features extracted from each of the sub-bands are independent. The characteristic function of a random variable is essentially the Fourier transform (with the sign in the exponential reversed) of its probability density function (PDF). The nth moment of the CF in a wavelet sub-band can be expressed as:

$$M_n = \frac{\sum_{k=0}^{N/2} f_k^n |H(f_k)|}{\sum_{k=0}^{N/2} |H(f_k)|} \qquad (1)$$

where $H(f_k), \quad k = 0, 1, \cdots N-1$ is the discrete Fourier transform of the coefficient histogram and N is the number of bins in the histogram.

In a subsequent paper [5], Shi *et al.* showed that the interference of the cover image in the estimated feature vectors can be reduced by using spatial prediction. Each pixel in the image is first predicted from its neighbouring pixels as:

$$\hat{x} = \begin{cases} max(a,b) & c \leq min(a,b) \\ min(a,b) & c \geq max(a,b) \\ a+b-c & otherwise \end{cases} \qquad (2)$$

where $\hat{x}$ is the predicted value of pixel x and a, b, c are the neighbouring pixels as shown in Fig. 1. The prediction-error image is then decomposed using DWT and the moments of the CFs are then computed for each sub-band using Eq. 1. Further, they showed that by excluding the DC-component of the CF $\big(H(f_0)\big)$ from the computation of the moments in Eq. 1 will increase the moment's sensitivity to data embedding.

3 Proposed method

The neighbouring frames in a video sequence contain high degree of spatio-temporal correlation, which is effectively exploited in the design of state-of-the-art video coding standards. To reduce this temporal redundancy, the MPEG coding schemes employ a motion-compensated prediction step in which a given frame is predicted from its neighbouring *reference frames* using the motion-compensation [6]. In the MPEG coding scheme, there are two types of predicted frames: the P-frames and the B-frames. The P-frames uses a single past frame as the reference frame where as the B-frames uses two reference frames: one past frame and one future frame. The prediction-error frames (PEFs) corresponding to the P- and B-frames are then coded using the transform coding techniques.

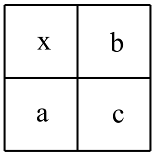

Figure 1. Pixels used in spatial prediction

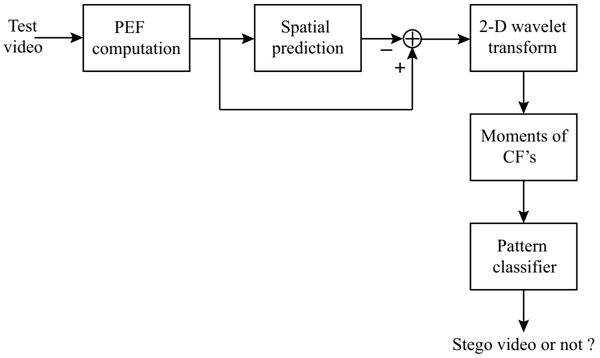

Figure 2. Proposed steganalysis scheme

As mentioned earlier, the interference of the host image on the feature vectors in the steganalysis can be reduced by exploiting the spatial correlation. Similarly, in the case of video sequences, both the temporal and spatial prediction may be exploited to reduce the interference of the host sequence on the feature vectors and thereby improving the steganalysis performance. We propose such a video steganalysis scheme.

The block-diagram of the proposed steganalysis scheme is given in Fig. 2. The first step in this scheme is the motion-compensated prediction to obtain the PEFs. It should be noted that since the video sequences are generally stored and transmitted in the compressed format, the PEFs can be obtained by partial decoding of the compressed stream. On the other hand, if the sequence is in the uncompressed format, the motion-vectors for the prediction has to be estimated. As shown in Fig. 3, the PEFs still exhibits some amount of spatial correlation. To reduce this spatial cor-

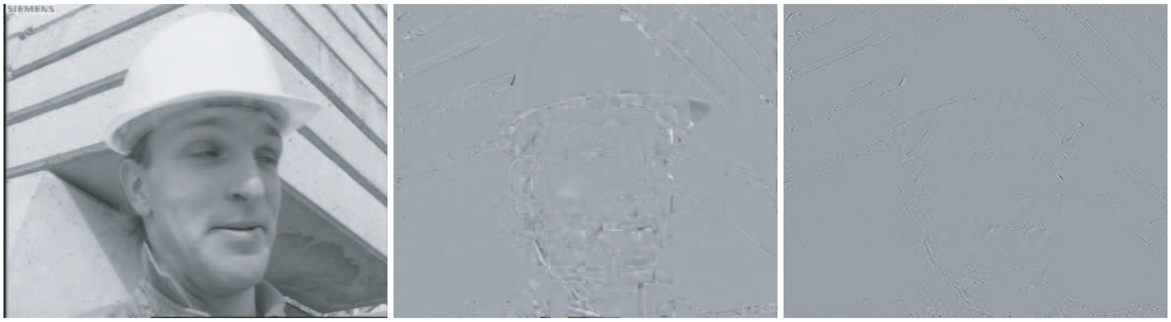

Figure 3. (a) Original frame (b) Temporal prediction error frame and (c) Spatio-temporal prediction error frame

relation, the spatial prediction step given in Eq. 2 is applied to each PEF and the difference between the PEFs and their spatially-predicted frames are computed. The resulting frames, which are the residues after spatio-temporal prediction, are then decomposed using the 3-level DWT and the first three moments of the CFs in each of the sub-bands are computed. The resulting 39-dimensional feature vector is used to train a pattern classifier which discriminates between the stego and non-stego videos.

4. Experimental results

Detailed experimental studies have been conducted to evaluate the performance of the proposed steganalysis scheme. The experiments are performed on 10 grayscale videos in the uncompressed format. The sequences are of CIF resolution and 124 frames from each of the sequences have been considered. Each sequence is watermarked using a frame-by-frame additive Gaussian spread-spectrum watermarking scheme, given as

$$\mathbf{y}_k = \mathbf{x}_k + \mathbf{w}_k \tag{3}$$

where $\mathbf{x}_k$ is the kth host frame, $\mathbf{w}_k \sim N(0,1)$ is the zero-mean, unit variance watermark added to the kth host frame and $\mathbf{y}_k$ is the kth watermarked frame. The watermark for each of the host frame is chosen pseudorandomly such that uncorrelated watermarks are added to different frames.

Each sequence is first divided into a standard MPEG-2 group of pictures (GOP) structure $IBBPBBPBBPBBI\cdots$ with a GOP length of 12 frames. The motion-vectors for the P and B frames are then estimated using a simple block-based motion estimation with a fixed block size of 16×16 pixels and integer-pixel accuracy for the motion-vectors. The PEFs corresponding to the P- and the B-frames are computed and from each of the PEFs, the feature vectors are extracted.

In order to evaluate the performance improvement in the proposed steganalysis method, it is compared with two other steganalysis schemes. In the first scheme [7], the features are extracted from each frame in the test sequence without spatial or temporal prediction. In the second scheme [5], feature vectors are obtained from each of the frames after spatial prediction. In other words, each of the steganalysis scheme uses a 39-dimensional feature vector, which are the first three moments of the histogram characteristic functions of the corresponding wavelet sub-bands.

In all the three steganalysis methods, a k-NN classifier with single neighbourhood citebudhia:ifs2006 is used. The classifiers are trained with feature vectors extracted from 40% of the frames in each sequence and the remaining frames are used for testing the classifiers. For cross-validation, the frames for training and testing the classifiers are randomly chosen and the reported results are the average of 100 experiments, each with a different training set. The *false positive rate* (FP), *true positive rate* (TP), *true negative rate* (TN) and the *average detection rate* (AD), defined as:

$$\begin{aligned} \mathrm{FP} &= \%\text{ of host frames detected as carrying watermark} \\ \mathrm{TP} &= \%\text{ of frames correctly detected as carrying watermark} \\ \mathrm{TN} &= \%\text{ of frames correctly detected as host frames} \\ \mathrm{AD} &= \frac{\mathrm{TP}+\mathrm{TN}}{2} \end{aligned}$$

are used as the performance measures.

Table 1 reports the comparative performance of the steganalysis schemes for a watermark embedding rate of 0.1 bits/pixel (bpp). It can be observed that the spatial prediction improves the steganalysis performance (reduced FP and increased TP) as compared to that without prediction.

As expected, the proposed method using spatio-temporal prediction outperforms both the other methods. Note that the performance improvement of the proposed scheme depends on the motion in the sequences. For slow-moving sequences (*Container*, *News* and *Akiyo*) and the sequence containing only translational motion (*Antibes*), the block-based motion estimation accurately estimates the motion and hence the host interference in the prediction error frame is significantly reduced. On the other hand, for other sequences which contain fast and non-translational motion, the motion-estimation technique fails to capture the motion and results in significant host interference in the extracted features. This is the reason behind the poor performance of the proposed steganalysis scheme for these sequences. However, its performance for these sequences is still better than the other two schemes. The average detection rate of the steganalysis schemes for different embedding rates are plotted in Fig. 4. It can be observed that the proposed steganalysis scheme performs better than the other two schemes for all the embedding rates. However, the performance improvement diminishes with an increase in the watermark embedding rate.

Sequence	**No prediction**		**Spatial Prediction**		**Spatio-temporal prediction**	
	FP	TP	FP	TP	FP	TP
Container	58.00	47.46	50.38	89.84	5.94	99.53
News	53.01	70.34	29.00	87.84	0	99.66
Akiyo	55.58	38.88	44.79	76.66	0	92.93
Mobile	60.82	42.19	49.31	69.15	23.90	73.87
Coastguard	30.87	79.28	36.21	99.03	16.16	99.91
Tempete	64.10	49.84	49.94	72.34	21.85	90.53
Foreman	31.71	62.97	6.71	99.62	8.16	99.31
Antibes	3.66	84.62	0.44	95.26	0	100.00
Stefan	61.56	46.47	41.35	87.16	41.02	92.53
Bike	50.31	53.54	48.09	84.75	32.66	88.94
Avg.	**46.96**	**57.56**	**35.62**	**86.17**	**14.97**	**93.72**

Table 1. Comparative performance of the proposed steganalysis method

5. Conclusions

This paper proposed a new blind steganalysis scheme for video sequences. The features for the steganalysis are extracted from the residual frames after spatio-temporal prediction. The experimental results show that the temporal prediction step significantly improves the performance of the proposed steganalysis scheme, particularly at the low embedding rates. One advantage of this scheme is that the temporal prediction error frames can be directly obtained by partial decoding of the compressed sequences, thereby greatly reducing the computational requirements.

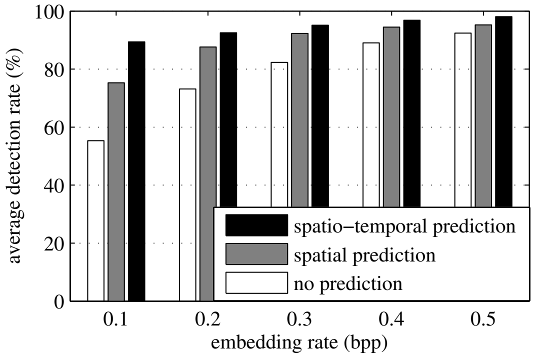

Figure 4. Average detection performance of the steganalysis schemes

References

[1] I. Avcibas, N. Memon, and B. Sankur. Steganalysis using image quality metrics. *IEEE Transactions on Image Processing*, 12(2):221–229, February 2003.

[2] U. Budhia, D. Kundur, and T. Zourntos. Digital video steganalysis exploiting statistical visibility in the temporal domain. *IEEE Transactions on Information Forensics and Security*, 1(4):502–516, December 2006.

[3] S. Lyu and H. Farid. Steganalysis using higher-order image statistics. *IEEE Transactions on Information Forensics and Security*, 1(1):111–119, March 2006.

[4] V. Pankajakshan, G. Doërr, and P. K. Bora. Assessing motion-coherency in video watermarking. In *Proceedings of the ACM Multimedia and Security Workshop*, pages 114–119, September 2006.

[5] Y. Q. Shi, G. Xuan, D. Zou, J. Gao, C. Yang, Z. Zhang, P. Chai, W. Chen, and C. Chen. Steganalysis based on moments of characteristic functions using wavelet decomposition, prediction-error image, and neural network. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, July 2005.

[6] Y. Wang, J. Osterman, and Y.-Q. Zhang. *Video Processing and Communications*. Prentice-Hall, 2001.

[7] G. Xuan, Y. Q. Shi, J. Gao, D. Zou, C. Yang, Z. Zhang, P. Chai, and C. C. W. Chen. Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions. In *Proceedings of the 7th Information Hiding Workshop*, volume 3727 of *Lecture Notes in Computer Science*, pages 262–277, June 2005.