VIDEO STEGANALYSIS OF MULTIPLICATIVE SPREAD SPECTRUM STEGANOGRAPHY

Nematollah Zarmehi and Mohammad Ali Akhaee

School of Electrical and Computer Engineering
College of Engineering, University of Tehran, Tehran, Iran
{zarmehi, akhaee}@ut.ac.ir

ABSTRACT

In this paper we propose a video steganalysis method toward multiplicative spread spectrum embedding. We use the redundancies of the video frames to estimate the cover frame and after extracting some features from the video frames and the estimated ones, the received video is classified as suspicious or not suspicious. In the case that the video declared suspicious, we estimate the hidden message and the gain factor used in the embedder. We also propose a new method for estimating the gain factor in multiplicative spread spectrum embedding. Using the estimated hidden message and gain factor, we are able to reconstruct the original video. Simulation results verify the success of our steganalysis method.

Index Terms— Video steganalysis, spread spectrum steganography, frame estimation

1. INTRODUCTION

Steganalysis is the art and science of detecting concealed information in mediums such as images, videos, and audios, which are imperceptible to human beings. Over the past decade, steganalysis have been studied intensively and many methods are proposed on text, audio, image, and video domains. Most of the work in this area is done on image domain. For instance, in [1–4], some steganalytic methods toward Least Significant Bit (LSB) embedding algorithm are proposed and the steganalysis methods based on OutGuess, F5, YASS, and QIM are studied in [5–8].

There is a need to improve the steganalysis methods in video domain along with image domain. Many of the steganalysis methods for image domain can also be applied to the video domain. However, some researches have also been done only for steganalysis in the video domain. Jainsky *et al.* [9] proposed a video steganalysis method using asymptotic memoryless detection. Their algorithm assumes that the video signal obeys a Gauss-Markov correlation model temporally. In [10] and [11], authors used the temporal redundancy of the video and proposed a method for its steganalysis based on linear and block-based collusion. A video steganalysis method is proposed in [12], which utilizes a three layer feedforward neural network as classifier and extracts features from the discrete cosine transform (DCT) domain in the compressed video frames based on the collusion.

Spread spectrum steganography approach has high robustness against most type of attacks, but this happens at the cost of lower capacity. However, since video signals have higher content and redundancy in comparison to other multimedia, capacity is not an important issue in the video domain. Cox *et al.* [13] proposed using spread spectrum embedding methods, which divide into two major categories: additive and multiplicative. Since the multiplicative method has higher robustness than the additive one, we focus on steganalytic methods toward video multiplicative spread spectrum steganography. We answer the question that whether the received video contains any hidden message or not. First we estimate the cover frames and by comparing them with the received video frames, we compute a residual matrix. We extract some features from the residual matrix, the received video frames, and the estimated ones. Using the extracted features and a support vector machine (SVM) classifier, the video is classified as suspicious or not suspicious. If it is decided that the video contains hidden information, we estimate the hidden message along with the gain factor used for embedding. We also propose a new method for estimating the gain factor in multiplicative embedding. Finally, we reconstruct the original video using the estimated hidden message and the gain factor.

2. MULTIPLICATIVE SPREAD SPECTRUM EMBEDDING METHOD

Cox *et al*. [13] proposed the following multiplicative embedding method,

$$y_i = x_i(1 + \alpha.w_i), \tag{1}$$

where α is the gain factor, w_i, x_i, and y_i are the *i*-th sample of watermark data, cover signal, and watermarked signal, respectively.

Considering the size of each frame $m \times n$, we rewrite the

The work is supported by the Iranian Ministry of Science, Research, and Technology (in Persian ATF) under grant No. SG.110-1-19.

multiplicative embedding method in video domain as follows:

$$\boldsymbol{Y}_l(i,j) = \boldsymbol{X}_l(i,j).\left[1+\alpha_l.\boldsymbol{W}_l(i,j)\right],$$

$$\alpha_l > 0,\ l = 1,\ldots,N,\ i = 1,\ldots,m,\ \text{and}\ j = 1,\ldots,n, \quad (2)$$

where α_l is the gain factor, $\boldsymbol{X}_l(i,j)$ is the pixel of the *i*-th row and the *j*-th column of the *l*-th frame of the cover video, $\boldsymbol{W}_l(i,j)$ is the watermark data, and $Y_l(i,j)$ is the corresponding watermarked pixel. $\boldsymbol{W}_l(i,j)$ takes the values of ± 1 with equal probabilities.

The gain factor controls both robustness and the amount of induced distortion. As an steganalyzer, we are not aware of the watermark data $\boldsymbol{W}_l$ and gain factor. But there is an upper bound for the gain factor based on the acceptable distortion. If we use peak signal-to-noise ratio (PSNR) as a metric for distortion, the maximum value of α_l can be computed as follows:

$$\alpha_{l_{max}} = \frac{255}{10^{PSNR/20}.\sqrt{P_{\boldsymbol{X}_l}}} \quad (3)$$

$$P_{\boldsymbol{X}_l} = \frac{1}{m.n}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|\boldsymbol{X}_l(i,j)\right|^2, \quad (4)$$

where $P_{\boldsymbol{X}_l}$ is the power of the *l*-th frame. For video compression, minimum acceptable PSNR is $30dB$ [14]. Table 1 shows the maximum values of gain factor for eight different video sequences with a PSNR at least $30dB$. The results of Table 1 facilitates the estimating of the gain factor.

Video sequence	$\boldsymbol{\alpha_{max}}$	**Video sequence**	$\boldsymbol{\alpha_{max}}$
Salesman	0.107	*Mobile*	0.053
Akiyo	0.077	*Hall*	0.052
Carphone	0.066	*Foreman*	0.045
Coastguard	0.063	*Bridge-close*	0.043

Table 1: Maximum of gain factor in the multiplicative method for different video sequences.

According to (3), the gain factor is inversely related to the power of the frame. For example the *Salesman* video sequence has a high darkness, which implies a low power and a high gain factor.

3. PROPOSED VIDEO STEGANALYSIS METHOD

Figure 1 illustrates the block diagram of our steganalysis system. The starting point of the system estimates the cover frames. Then a residual matrix is computed and some features are extracted. A classifier uses these extracted features and classifies the video as suspicious or not suspicious. If the video declared as suspicious, we estimate the hidden message along with the gain factor and reconstruct the original video.

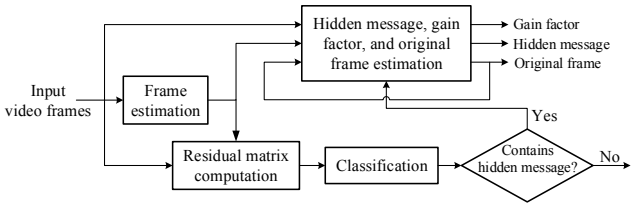


Fig. 1: Block diagram of our proposed steganalysis system.

3.1. Frame Estimation

As shown in the block diagram of our steganalysis system, the starting point is frame estimation. We use the following two methods to serve this purpose.

3.1.1. Denoising

In spread spectrum embedding methods, the stego signal is a noisy version of the host signal. Thus, we can estimate the cover video frames using a method to denoise the received video. Here, we employ the stationary wavelet by soft thresholding [15] for denoising. The estimated frame with this method is denoted as $\hat{\boldsymbol{X}}_l^1$.

3.1.2. Inter-Frame Estimation

Block-matching algorithm is a method to recover a missing block from a video frame, which is widely used in video compression standards [16]. We utilize the same idea as block-matching algorithm to estimate $\boldsymbol{X}_l$. We divide the frame $\boldsymbol{Y}_l$ into 8×8 blocks and for each block, according to the following method, find the best matched block in the previous and next frames. The block-matching criterion is minimum squared error (MSE). Let

$$\boldsymbol{B}_{i,k}^M = \arg\min_{\mathrm{B_{j,k}}\in\boldsymbol{B}_k} MSE\left\{\boldsymbol{B}_i, \boldsymbol{B}_{j,k}\right\},\ k = -1, 1, \quad (5)$$

where $\boldsymbol{B}_i$ is the *i*-th block of frame $\boldsymbol{Y}_l$, $\boldsymbol{B}_{i,-1}^M$ and $\boldsymbol{B}_{i,1}^M$ are the best matched blocks with $\boldsymbol{B}_i$ from the previous and the next frames, respectively, $\boldsymbol{B}_{-1}$ is the set of all 8×8 blocks of frame $\boldsymbol{Y}_{l-1}$, and B_1 is the set of all 8×8 blocks of frame $\boldsymbol{Y}_{l+1}$. The estimated frames from the previous and next frames are denoted as $\hat{\boldsymbol{X}}_{l,-1}^M$ and $\hat{\boldsymbol{X}}_{l,1}^M$, respectively. The estimated frame from this section is formed from the following linear combination of $\hat{\boldsymbol{X}}_{l,-1}^M$ and $\hat{\boldsymbol{X}}_{l,1}^M$:

$$\hat{\boldsymbol{X}}_l^2 = \frac{\rho_{-1}^M\hat{\boldsymbol{X}}_{l,-1}^M + \rho_1^M\hat{\boldsymbol{X}}_{l,1}^M}{\rho_{-1}^M + \rho_1^M} \quad (6)$$

$$\rho_k^M = \frac{cov\left(\boldsymbol{Y}_l, \hat{\boldsymbol{X}}_{l,k}^M\right)}{\sqrt{var\left(\boldsymbol{Y}_l\right).var\left(\hat{\boldsymbol{X}}_{l,k}^M\right)}},\ k = \{-1, 1\}, \quad (7)$$

where ρ_k^M is the correlation coefficient between the current and the estimated frames.

Finally, the estimated frame of $\boldsymbol{X}_l$ is obtained from $\hat{\boldsymbol{X}}_l^1$ and $\hat{\boldsymbol{X}}_l^2$ as follows:

$$\hat{\boldsymbol{X}}_l = \frac{\rho_1 \hat{\boldsymbol{X}}_l^1 + \rho_2 \hat{\boldsymbol{X}}_l^2}{\rho_1 + \rho_2} \tag{8}$$

$$\rho_k = \frac{cov\left(\boldsymbol{Y}_l, \hat{\boldsymbol{X}}_l^k\right)}{\sqrt{var\left(\boldsymbol{Y}_l\right).var\left(\hat{\boldsymbol{X}}_l^k\right)}},\ k = \{1, 2\}. \tag{9}$$

3.2. Residual Matrix Computation

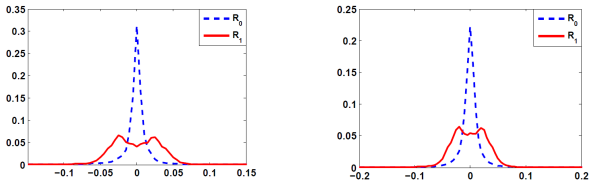
Residual matrix computation is another part of our steganalysis system. To address the classification problem we extract some features from this matrix. Assuming that the frame $\boldsymbol{Y}_l$ is watermarked, the residual matrix can be computed as follows:

$$\boldsymbol{R}_{1,l} = \frac{\boldsymbol{Y}_l}{\hat{\boldsymbol{X}}_l} - \mathbf{1} = \frac{\boldsymbol{X}_l\left(1 + \alpha_l.\boldsymbol{W}_l\right)}{\hat{\boldsymbol{X}}_l} - \mathbf{1} = \hat{\mathbf{0}} + \alpha_l.\boldsymbol{W}_l.\hat{\mathbf{1}}, \tag{10}$$

where $\hat{\mathbf{0}}$ and $\hat{\mathbf{1}}$ are the estimated versions of all-zero and all-one matrices with size $m \times n$, respectively. To avoid division by zero problem, the zeros elements of $\hat{\boldsymbol{X}}_l$ are replaced with a small constant. In the case that $\boldsymbol{Y}_l$ is not watermarked, the residual matrix is

$$\boldsymbol{R}_{0,l} = \frac{\boldsymbol{Y}_l}{\hat{\boldsymbol{X}}_l} - \mathbf{1} = \frac{\boldsymbol{X}_l}{\hat{\boldsymbol{X}}_l} - \mathbf{1} = \hat{\mathbf{0}}. \tag{11}$$

The residual matrix can show the difference between watermarked and non-watermarked video frames. The distribution of residual values for different video sequences are shown in Figure 2.



(a) Akiyo sequence (b) Hall sequence

Fig. 2: Distribution of residual values for two video sequences.

According to Figure 2, the distribution of residual values is different for watermarked and non-watermarked video sequences. This justifies the choice of residual matrix for feature extraction.

3.3. Hidden Message Estimation

The estimation of the hidden message is straightforward. Because the gain factor is non-negative, we can easily estimate the hidden message as follows:

$$\hat{\boldsymbol{W}}_l(i,j) = \begin{cases} 1 & \boldsymbol{R}_{1,l}(i,j) > 0 \\ -1 & \boldsymbol{R}_{1,l}(i,j) < 0 \end{cases}. \tag{12}$$

We assign ± 1 with a probability equal to $\boldsymbol{W}_l(i,j)$, when $\boldsymbol{R}_{1,l}(i,j) = 0$.

3.4. Gain Factor Estimation

The gain factor is estimated based on entropy. From the information theory, we know that conditions reduce the entropy [17]. The following conditional relation between the original and watermarked frames in the multiplicative spread spectrum embedding rule exists:

$$\{\boldsymbol{X}_l(i,j)\} = \{\boldsymbol{Y}_l(i,j) | \alpha_l.\boldsymbol{W}_l(i,j) = 0\}. \tag{13}$$

Therefore,

$$p\left(\boldsymbol{X}_l(i,j)\right) = p(\boldsymbol{Y}_l(i,j) | \alpha_l.\boldsymbol{W}_l(i,j) = 0) \tag{14}$$

and

$$H\left(\boldsymbol{Y}_l(i,j)\right) \geq H\left(\boldsymbol{Y}_l(i,j) | \alpha_l.\boldsymbol{W}_l(i,j) = 0\right) = H\left(\boldsymbol{X}_l(i,j)\right). \tag{15}$$

According to the above facts, we can estimate the gain factor using the following formula:

$$\hat{\alpha}_l = arg \min_{\alpha_i \in A} H\left(\frac{Y_l}{1 + \alpha_i.\hat{W}_l}\right),\ A = (0, \alpha_{l_{max}}], \tag{16}$$

where A is the appropriate interval for gain factor and $\alpha_{l_{max}}$ can be computed from (3), but since $P_{\boldsymbol{X}_l}$ is used in (3) and we not aware of $\boldsymbol{X}_l$, we use the power of its estimation, *i.e.* $P_{\hat{\boldsymbol{X}}_l}$.

Using the estimated hidden message and the gain factor, we can reconstruct the original frame as follows:

$$\hat{X}_{l_R} = \frac{Y_l}{1 + \hat{\alpha}_l \hat{W}_l}. \tag{17}$$

The reconstructed frame can be considered as a new estimation of the received frame. Thus we feed the reconstructed frame back to the system as another estimation of the frame, which can be clearly seen in the block diagram of the steganalysis system shown in Figure 1.

4. CLASSIFICATION

4.1. Feature Extraction

To address the classification problem, we need to extract features which can show the differences between watermarked and non-watermarked frames. In this section, we show how to extract six features for this purpose. Three of the features are

extracted from the video frames and the rest from the residual matrix.

In wavelet transform, the coefficients of different sub-bands are independent, and thus, the extracted features from different sub-bands are independent [18]. Therefore, we can apply one level wavelet transform to each frame and then compute the singular values of each sub-band, except the approximation sub-band. Then, we derive condition number of each transformed sub-band as feature.

According to Figure 2, it is clear that the two depicted distributions have different variances. So we choose the variance of the residual values as feature. Also the peakedness of the distributions of the residual values are different. Therefore, the next two features are the peak value and the kurtosis of the distribution functions.

4.2. Classifier

We extract six features to address the classification problem. Since the two classes with these extracted features are not linearly separable, we use an SVM classifier with (Gaussian) radial basis function (RBF) kernel [19].

5. SIMULATION RESULTS

The simulation results are presented in this section. We watermarked 15000 different video frames based on multiplicative spread spectrum embedding rule. Then, we estimated the hidden message and gain factor using the equations (12) and (16), respectively. The results of estimating hidden message and gain factor are shown in Figures 3 and 4, respectively.

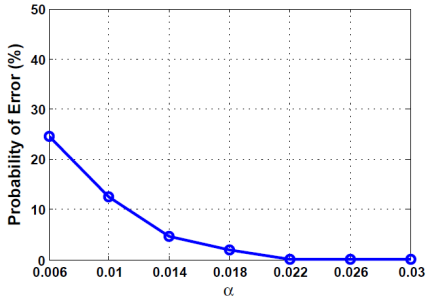


Fig. 3: Error probability of gain factor estimation.

According to Figure 3, the error probability of estimation of α decreases as α increases. By the proposed method for estimating the gain factor, for $\alpha \geq 0.022$, the error probability is less than 0.01% and for $\alpha = 0.006$, which is rarely used in multiplicative steganography, the error probability is 24.51%. Also the probability of correct estimation of hidden message increases as α increases. To achieve a gain factor of 0.025, we should correctly estimate 81.00% of the hidden message.

We also watermarked 40000 different video frames with various gain factors and applied our steganalysis method on them. The probabilities of detection and false alarm are presented here for evaluating the performance of the steganalysis system.

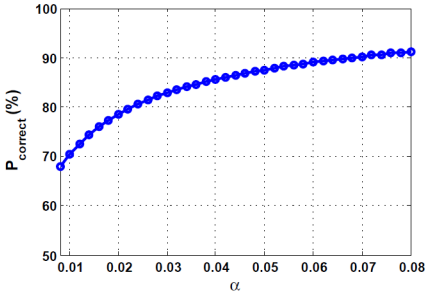


Fig. 4: Probability of correct estimation of the hidden message versus α.

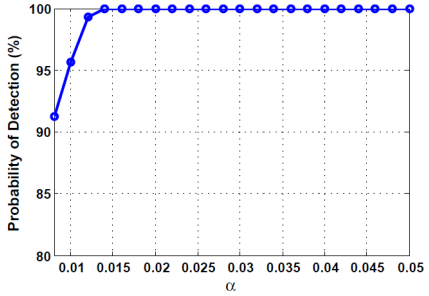


Fig. 5: Probability of detection of stego video versus α.

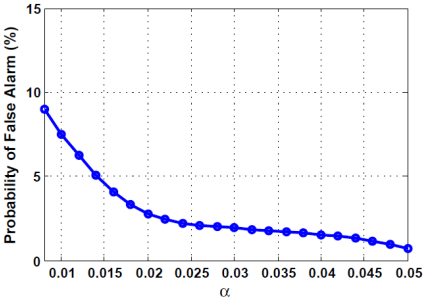


Fig. 6: Probability of false alarm versus α.

Figures 5 and 6 show the probabilities of detection and false alarm, respectively. We can see that for the gain factors higher than 0.012, the probability of detection is greater than 99.33% and the probability of false alarm is less than 6.24%.

In [11], a video steganalysis method toward multiplicative spread spectrum embedding rule is proposed based on linear and block-based collusion. In [11], the results of spatial-based steganalysis method based on Wiener filtering and temporal method based on linear and block-based collusion attacks are compared. Table 2 shows the average of results of [11] over different video sequences for $\alpha = 0.1$.

Method	$P_D(\%)$	$P_F(\%)$
Block-based	100.00	25.00
Averaging	100.00	43.46
Wiener	82.31	84.23

Table 2: Probability of detection and false alarm of proposed method in [11] for $\alpha = 0.1$.

According to the results of Table 2, the proposed method in [11] has a high false alarm probability. In our simulations, we used gain factors less than 0.05 while the results of Table 2 are related to $\alpha = 0.1$. Our steganalysis method has better performance than the method proposed in [11] in terms of probability of detection and probability of false alarm.

6. CONCLUSION

We proposed a video steganalytic method toward multiplicative spread spectrum embedding. We also proposed a new method to estimate the gain factor on the basis of entropy. For the steganalysis problem, we used an SVM classifier that classified the video as suspicious or non-suspicious. In case that the video was suspicious, we estimated the hidden message and the gain factor used in the embedding process and reconstructed the original video. Simulation results verified the success of our proposed steganalysis method. We could correctly estimate 81.00% of the hidden message with a gain factor equal to 0.01. For this value of gain factor, the probabilities of detection and false alarm were 95.64% and 7.50%, respectively.

REFERENCES

[1] A. Westfeld and A. Pfitzmann, "Attacks on steganographic systems," in *Proc. Information Hiding&mdash, 3rd Int',l Workshop*, Springer-Verlag, 1998, pp. 273–289.

[2] S. Dumitrescu, X. Wu, and Z. Wang, "Detection of lsb steganography via sample pair analysis," *IEEE Trans. on Signal Processing*, vol. 51, no. 7, pp. 1995–2007, July 2003.

[3] J. Fridrich, M. Goljan, and R. Du, "Reliable detection of lsb steganography in color and grayscale images," in *Proc. ACM Workshop Multimedia Security*, 2001, pp. 27–30.

[4] J. Fridrich and M. Goljan, "On estimation of secret message length in lsb steganography in spatial domain," in *Proc. IST/SPIE 16th Annu. Symp. Electronic Imaging Science Technology*, 2004.

[5] J. Fridrich, M. Goljan, and D. Hogea, "Attacking the outguess," in *Proc. ACM Workshop Multimedia and Security 2002*, ACM Press, 2002.

[6] J. Fridrich, M. Goljan, and D. Hogea, "Steganalysis of jpeg images: Breaking the f5 algorithm," in *Proc. 5th Int',l Workshop Information Hiding*, Springer-Verlag, 2002.

[7] B. Li, J. Huang, and Y.Q. Shi, "Steganalysis of yass," *IEEE Trans. on Information Forensics and Security*, vol. 4, no. 3, pp. 369–382, Sept. 2009.

[8] K. Sullivan, Z. Bi, U. Madhow, S. Chandrasekaran, and B.S. Manjunath, "Steganalysis of quantization index modulation data hiding," in *International Conference on Image Processing, ICIP '04 2004*, Oct. 2004, vol. 2, pp. 1165–1168.

[9] J.S. Jainsky, D. Kundur, and D.R. Halverson, "Towards digital video steganalysis using asymptotic memoryless detection," in *Proc. MM&Sec07*, 2007, pp. 161–168.

[10] U. Budhia, D. Kundur, and T. Zourntos, "Digital video steganalysis exploiting statistical visibility in the temporal domain," *IEEE Trans. on Information Forensics and Security*, vol. 1, no. 4, pp. 502–516, Dec. 2006.

[11] U. Budhia, "Steganalysis of video sequences using collusion sensitivity," *Ph.D. Dissertation, Texas A&M University, College Station*, 2005.

[12] B. Liu, F. Liu, and P. Wang, "Inter-frame correlation based compressed video steganalysis," in *Congress on Image and Signal Processing, CISP '08*, May 2008, vol. 3, pp. 42–46.

[13] I.J. Cox, J. Kilian, F.T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. on Image Processing*, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.

[14] T.-H. Kim, H. Adeli, C. Ramos, and B.-H. Kang, *Signal processing, image processing and pattern recognition*, Springer, Berlin, 2011.

[15] D.L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. on Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[16] J. Yang, B. Yin, and N. Zhang, "A block-matching based intra frame prediction for h.264/avc," in *IEEE International Conference on Multimedia and Expo*, July 2006, pp. 705–708.

[17] T. Cover and J. Thomas, *Elements of information theory*, Wiley, 1991.

[18] Y.Q. Shi, G. Xuan, D. Zou, J. Gao, C. Yang, Z. Zhang, P. Chai, W. Chen, and C. Chen, "Image steganalysis based on moments of characteristic functions using wavelet decomposition, prediction-error image, and neural network," in *IEEE International Conference on Multimedia and Expo, ICME*, July 2005, pp. 269–272.

[19] S. Theodoridis and K. Koutroumbas, *Pattern recognition*, Academic Press, 2nd. edition, 2003.