UNIVERSAL IMAGE STEGANALYSIS BASED ON GARCH MODEL

Saeed Akhavan, Mohammad Ali Akhaee, Saeed Sarreshtedari

School of Electrical and Computer Eng., College of Eng., University of Tehran, Tehran, Iran
s.akhavan, akhaee, s.sarreshtedari@ut.ac.ir

ABSTRACT

This paper introduces a new universal steganalysis framework. The required image features are extracted based on the generalized autoregressive conditional heteroskedasticity (GARCH) model and higher-order statistics of the images. The GARCH features are extracted from non-approximate wavelet coefficients. Besides, the second and third order statistics are exploited to develop features very sensitive to minor changes in natural images. The experimental results demonstrate that the proposed feature-based steganalysis framework outperforms state of the art methods while running on the same order of features.

Index Terms— GARCH Model; Steganalysis, Higher Order Statistics

1. INTRODUCTION

Steganography techniques are developed to conceal the existence of the secret message during a seemingly normal communication, without raising any doubt for the observers. On the other hand, steganalysis methods are simultaneously evolved to combat this threat and detect the abuse of digital media.

In order to design a general blind steganalyzer, several steganographic methods are applied to a collection of the *training* images. Afterwards, some appropriate features which represent a *digest* of the total information of the images are extracted. At the next step, classifiers are applied to find the optimum classification boundary between the *cover* and *stego* images based on the feature values extracted from the cover and stego training images. New images are classified as cover or stego images, by simply comparing the feature values to the classification boundary. Therefore, there are two important steps for every universal steganalyzer, namely, feature extraction and classification.

The first important stage of every steganalyzer is to extract appropriate features. These features must be designed sophisticatedly to reflect every subtle modification of the original image. The main goal of this paper is to find a set of these appropriate features based on the generalized autoregressive conditional Heteroskedasticity (GARCH) model and the second and third order statistics of the images in both the time and transfer domain.

As mentioned in previous paragraph, the first set of features is based on the GARCH modeling. Autoregressive conditional heteroskedasticity (ARCH) processes were introduced by Engle in 1982 [1]. The samples of the ARCH processes are zero-mean and serially uncorrelated with conditional variances depending on the previous samples. Bollersev in 1986, introduced a generalized ARCH model called as GARCH [2]. In the GARCH model, conditional variances are not only dependent to the previous sample values, but also to their conditional variances. Although this model had been essentially developed to analyze the financial time series [1, 2], it has been shown recently that it is an appropriate tool to model the non-approximate wavelet coefficients with heavy-tailed and non-stationary distributions [3].

In this paper, we apply the GARCH model in the wavelet domain to extract features boosting the steganography footprints. The other set of features are extracted based on the second and third statistics of images in both the time and transfer domain. These types of features have been already exploited in steganalyses literature offering significant results [4, 5]. The next step to design a blind steganalyzer is to select the suitable classifier. Here, we choose the ensemble classifier because of its simplicity and efficiency [6]. This classifier decreases the complexity of the classification algorithm, while keeps the ability of dealing with the large number of features offering almost the same performance.

2. GARCH MODEL

Let ϵ_t represent a 1D stochastic process with zero mean. We define $GARCH(p,q)$ for ϵ_t as below :

$$\begin{aligned} &\epsilon_t = \sqrt{h_t}\eta_t; \quad \eta_t \sim N(0,1); \quad t = 1,2,...,t_{end} \\ &h_t = k + \sum_{i=1}^{p} \alpha_i h_{t-i} + \sum_{j=1}^{q} \beta_j \epsilon_{t-j}^2 \end{aligned} \tag{1}$$

where η_t is a normal process independent of h_t with zero mean and unity variance, h_t denotes the conditional variance of ϵ_t and k, α_i's and β_j's are GARCH parameters that should

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

be estimated. Let ψ_{t-1} denote all information until $t-1$:

$$\begin{aligned}&\psi_{t-1} = \{\epsilon_0, \epsilon_1, ..., \epsilon_{t-1}, h_0, h_1, ..., h_{t-1}\}\\&\Rightarrow \epsilon|\psi_{t-1} \sim N(0, h_t)\end{aligned} \quad (2)$$

In order to estimate the GARCH parameters, we use the maximum likelihood (ML) estimation. Suppose that we want to apply the GARCH model to the process of y_t. For this sake, we define the zero mean process of ϵ_t as below:

$$\begin{aligned}&\epsilon_t = y_t - \mathbf{r}_t.\mathbf{b}; \quad \epsilon_t = \sqrt{h_t}\eta_t\\&\mathbf{r}_t = [y_{t-1}, y_{t-2}, \ldots, y_{t-M}]\\&\psi_{t-1} = \{y_0, y_1, \ldots, y_{t-1}, h_0, h_1, \ldots, h_{t-1}\}\\&\mathbf{b} = [b_1, b_2, \ldots, b_M]\\&t = \{1, 2, \ldots, M, M+1, \ldots, t_{end}\}\end{aligned} \quad (3)$$

where $\mathbf{r}_t$ and $\mathbf{b}$ are the vectors of explanatory variables and unknown parameters respectively. Therefore, we need to estimate the vector $\mathbf{b}$ as well. We can write:

$$f(y_t|\psi_{t-1}) = \frac{1}{\sqrt{2\pi h_t}} \exp(-\frac{(y_t - \mathbf{r}_t.\mathbf{b})^2}{2h_t}) \quad (4)$$

$$h_t = k + \sum_{i=1}^{p} \alpha_i h_{t-i} + \sum_{j=1}^{q} \beta_j (y_{t-j} - \mathbf{r}_{t-j}.\mathbf{b})^2 \quad (5)$$

Then, the Likelihood function is formed as below:

$$LF(\gamma) = \prod_{t=1}^{N} f(y_t|\psi_{t-1}); \gamma = \{k, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q, \mathbf{b}\} \quad (6)$$

We find the parameter set γ that maximizes the likelihood function. Having this optimized parameter set found, the GARCH model can be established completely.

3. FEATURE EXTRACTION BASED ON GARCH MODEL

In this Section, we show that the variances of the GARCH model can be efficiently employed as our steganalysis features. It is noteworthy that nearly all the steganographic algorithms can be modeled as below:

$$\begin{aligned}H_0 &: y_k = S_k \qquad k = 1, 2, ..., n\\H_1 &: y_k = S_k + n_k \quad k = 1, 2, ..., n\end{aligned} \quad (7)$$

where n_k is the additive noise added to the pixel value of S_k due to the embedding and y_k are the pixels of the received image and n is the number of total image pixels. The other assumption here is to model the additive noise as a random variable of zero mean Gaussian distribution. When the embedding is applied to the transform domain such as discrete cosine transform (DCT), the effect of this additive noise back in the time domain is well modeled with a zero mean Gaussian distribution of variance σ_0^2, i.e. $n_k \sim N(0, \sigma_0^2)$. In order to setup a framework in which we are capable to exploit the above-mentioned analyzes, the received image is transformed to the wavelet domain using the Haar or the other filter types. In this domain, we can assume an approximate distribution for the image (S), which enables us to make decision based on detection and estimation techniques. However, since the variance of the additive noise is not large enough, we are interested in the subbands with the least power because in there, the variance difference of the H_0 and H_1 hypotheses are large enough to let us separate them efficiently.

According to the above-mentioned discussion, approximation subband of the wavelet transform is not useful for our application. However, the other subbands are possible candidates since they include highpass filtering and lowering the energy of the image. Applying the wavelet decomposition to each of these non-approximate subbands, results in the signals with lower energy. Working on the horizontal, vertical, and diagonal subbands, we come to a total number of 36 subbands in the second and third level of decomposition. These 36 subbands comprise of 9 and 27 non-approximation subbands from the second and third level of the wavelet decomposition respectively.

Now we rewrite the hypotheses in each new subband:

$$\begin{aligned}&H_0 : y_k = S_k, \qquad \sigma_{y_k}^2 = \sigma_{s_k}^2\\&H_1 : y_k = S_k + n_k, \quad \sigma_{y_k}^2 = \sigma_{s_k}^2 + \sigma_0^2\\&k = 1, 2, ..., n\end{aligned} \quad (8)$$

where each S_k is assumed to be a random variable with Gaussian distribution and the variance of $\sigma_{s_k}^2$. Although we do not need $\sigma_{s_k}^2$ here, it is achievable by having $\sigma_{y_k}^2$ known according to the GARCH model and σ_0^2.

Now we apply the GARCH model to y_k's to find the conditional variances and use the ML estimation for decision making.

$$\begin{aligned}&H_0 : f(\mathbf{y}|H_0)\\&= f(y_1|H_0) f(y_2|y_1, H_0) ... f(y_n|y_1, y_2, ..., H_0)\\&H_1 : f(\mathbf{y}|H_1)\\&= f(y_1|H_1) f(y_2|y_1, H_1) ... f(y_n|y_1, y_2, ..., H_1)\end{aligned} \quad (9)$$

Then the decision is made based on (10):

$$f(\mathbf{y}|H_0) \underset{\text{cover}}{\overset{\text{stego}}{\lessgtr}} f(\mathbf{y}|H_1) \quad (10)$$

According to the GARCH model, we know that the distributions of the above functions are all Gaussian:

$$\begin{aligned}&f(y_1|H_i) \sim N(0, \sigma_{y_1}^2)\\&f(y_2|y_1, H_i) \sim N(0, \sigma_{y_2}^2)\\&\ldots\\&f(y_n|y_1, y_2, \ldots, H_i) \sim N(0, \sigma_{y_n}^2)\end{aligned} \quad (11)$$

We see from (11) that the GARCH variances are very critical to the decision making process. This verifies our previous idea to choose them as an appropriate feature set. By the way, the large number of variances (which is linearly dependent on the size of the subbands) might impose considerable complexity to our steganalysis framework. We know that the GARCH model parameters, i.e. α_i's, β_j's and k, are more limited in number comparing to the variances. They convey also the same amount of information. Therefore, it will be reasonable to replace the variances of each subband with its GARCH parameters as the feature set.

$GARCH(1,1)$ is a prevalent and efficient model to which we stick here. As a result, three features are extracted from each subband giving the total number of $36 \times 3 = 108$ GARCH features.

4. FEATURE EXTRACTION BASED ON THE SECOND AND THIRD ORDER STATISTICS

4.1. Basics

The second or third order statistics reflects the density function regarding to the co-occurrence of two or three different image pixel values. One important challenge of using higher order statistics is their large number. One idea is to consider the differences between consecutive pixels rather than themselves, as mentioned in [4]. In this case, we can restrict the values of statistics to the range of $[-T,T]$ in order to limit the number of second order statistics.

In the next equations, we denote the original matrix with F, which can be the image itself, or its transformed version in the quantized DCT or wavelet subbnads. We also define the difference matrix D which its entries are difference of consecutive entries of the original image. We define and use four types of difference matrices in our work. Suppose that $F^{i,j}$ is the F matrix when shifted by i rows to the right and j columns down. The horizontal, vertical, diagonal and minor diagonal difference matrices are defined based on (12):

$$\begin{aligned} D^h &= F - F^{1,0} \\ D^v &= F - F^{0,1} \\ D^d &= F - F^{1,1} \\ D^h &= F - F^{1,-1} \end{aligned} \tag{12}$$

Now, we have D_h, D_v, D_d and D_{md} with entries limited to the range of $[-T,T]$ from which we want to extract the second order statistics. While the joint second order density function is considered here as the statistics, one can choose the conditional distribution as well. Assuming that M and N represent the number of rows and columns of the original image, the joint distribution function for D_h is calculated as below. Similar processes can be done for the other three matrices.

$$\begin{aligned} NJ^h_{u,v} &= \Pr\left(D^h_{i,j+1} = u, D^h_{i,j} = v\right) \\ &\frac{\sum_i \sum_j \delta\left(D^h_{i,j+1}=u, D^h_{i,j}=v\right)}{M\times(N-1)}; u,v \in [-T,T] \end{aligned} \tag{13}$$

NJ represents the neighboring joint density function and has $(2T+1)^2$ number of entries. In order to further reduce the number of features, NJ^1 and NJ^2 are defined and used in our work:

$$\begin{aligned} NJ^1_{u,v} &= \frac{NJ^h_{u,v} + NJ^v_{u,v}}{2} \\ NJ^2_{u,v} &= \frac{NJ^d_{u,v} + NJ^{md}_{u,v}}{2} \end{aligned} \tag{14}$$

NJ^1 and NJ^2 have $(2T+1)^2$ entries too. The third order horizontal statistics are derived from D_h in a similar way to (13). The same process is applied to the other three matrices.

$$\begin{aligned} NJ^h_{u,v,k} &= \Pr\left(D^h_{i,j+2} = u, D^h_{i,j+1} = v, D^h_{i,j} = k\right) \\ &\frac{\sum_i \sum_j \delta\left(D^h_{i,j+2}=u, D^h_{i,j+1}=v, D^h_{i,j}=k\right)}{M\times(N-2)}; \\ &u,v,k \in [-T,T] \end{aligned} \tag{15}$$

NJ has $(2T+1)^3$ entries here, thus we reduce the number of features in a similar way to the second order statistics:

$$\begin{aligned} NJ^1_{u,v,k} &= \frac{NJ^h_{u,v,k} + NJ^v_{u,v,k}}{2} \\ NJ^2_{u,v,k} &= \frac{NJ^d_{u,v,k} + NJ^{md}_{u,v,k}}{2} \end{aligned} \tag{16}$$

NJ^1 and NJ^2 have also $(2T+1)^3$ entries. Since the process explained above is referred later, we summarize it in the function $NJ(Order,T)$. The input to this function is the original matrix F and outputs are NJ^1 and NJ^2, respectively. The order equals two or three and T is the threshold for clipping. In the next Sections, we discuss about feature extraction based on NJ's in the time and DCT domain.

4.2. Time Domain Feature Extraction

According the results of [5], we use $NJ(3,3)$ in the time domain leading to a number of $2 \times (2 \times 3 + 1)^3 = 686$ features which are called SPAM feature set [5]. One half of these features represent the horizontal and vertical dependencies while the other half describes the correlation in diagonal and minor-diagonal directions.

4.3. DCT Domain Feature Extraction

Considering the works presented in [4] and [7], three types of the second order features are extracted. These three sets are explained in the following subsections separately.

4.3.1. Features Extracted from the Whole DCT Coefficient Matrix

In this case, we regard the matrix of quantized DCT coefficients as F which is defined in the Section 4.1 and extract the D and NJ matrices in four directions similarly. For this sake, we consider $NJ(2,6)$ which leads to a set of $2\times(2\times6+1)^2 = 338$ features.

4.3.2. Features Extracted from Interblock Correlation

These features are extracted through considering the DCT 8×8 blocks as the F matrix. For each block, D and NJ matrices are calculated individually. Apparently, this process results in a very large number of matrices. In order to produce manageable results, we average over all matrices. Thereafter, we have NJ^1 which is the outcome of averaging over all NJ_1^1, NJ_2^1,...,NJ_L^1 matrices, where L stands for the number of 8×8 blocks. Likewise, the NJ^2 matrix is calculated. Since dependencies are considered only inside blocks, this method is called interblock correlation. Here, we choose $NJ(2,4)$ which gives a number of $2 \times (2 \times 4 + 1)^2 = 162$ features.

4.3.3. Features Extracted from Intrablock Correlation

In this case, we put certain entries (frequencies) of all 8×8 blocks excluding the first entry (DC coefficient) together to form a set of 63 new F matrices. The D and NJ matrices are calculated similar to the former sections. The number of matrices is reduced by averaging over all of them similar to the previous section. For example, we have NJ^1 as the average over $NJ_1^1, NJ_2^1, \ldots, NJ_{63}^1$ this time. Since the similar frequencies in different blocks are considered in this method, it is called intrablock correlation. Here, we again choose $NJ(2,4)$ which results in 162 features. All extracted features discussed above result in totally 1456 features.

5. SIMULATION RESULTS

We set up our steganalysis framework based on the features derived in Sections 3 and 4, and the ensemble classifier, to analyze a variety of steganographic methods in the both time and transform domain. 4000 images of the *BOSSbase* database [8] are used for the training and test stages. All images are cropped in case, to the size of 256×256 with the central pixel remaining the same. One half of the images are chosen randomly for the training phase, while the results are derived by testing the classifier over the other half. The comparison criterion is the minimum error probability (P_E) defined in (17):

$$P_E = \min_{P_{FA}} \frac{1}{2}(P_{FA} + P_{MD}(P_{FA})) \qquad (17)$$

P_{FA} and P_{MD} stand for the probabilities of the false alarm and missed detection. The performance of the proposed steganalyzer is examined for several steganography techniques. Embedding rates are presented in bit per pixel (bpp) and bit per AC coefficient (bpac) for the time and frequency domain methods, respectively. In the following, we compare the performance of the proposed steganalysis framework with the state of the art schemes. The first scheme is the steganalyzer proposed by Chen in 2008 which works based on 486 features [7]. CC-PEV steganalyzer introduced in 2007 and improved in 2009 [9], is the second comparing method which exploits a set of 548 features for the sake of classification. The third steganalyzer is the SPAM feature set presented in 2010 with 686 features [5]. Introduced in 2011, is the CCC steganalysis framework which consists of 48600 features based on the *Rich Model* [10]. The CHEN, CC-PEV and CCC steganalyzers are useful just for transform embedding domain but SPAM steganalyzer is suitable for both the time and transform domain embedding methods.

Table 1. Error Probability Performance Comparison

Method	rate	CHEN	CC-PEV	SPAM	CCC	Proposed
LSBM	25 %	-	-	**0.280**	-	0.290
	100 %	-	-	0.113	-	**0.112**
LSBMR	25 %	-	-	**0.285**	-	0.307
	100 %	-	-	0.160	-	**0.150**
HUGO	100 %	-	-	0.300	-	**0.290**
F5	5 %	0.240	**0.100**	0.245	0.115	0.137
	20 %	0.003	0.003	0.004	0.003	**0.002**
nsF5	5 %	**0.157**	0.195	0.365	0.217	0.175
	20 %	0.004	0.004	0.085	0.005	**0.003**
MB1	5 %	0.122	**0.057**	0.285	0.085	0.070
	20 %	0.008	0.008	0.040	0.005	**0.007**
MB2	5 %	0.247	0.230	0.282	0.302	**0.167**
	20 %	0.020	0.009	0.050	0.060	**0.005**
YASS	5 %	0.360	0.362	0.300	0.367	**0.288**
	20 %	0.300	0.245	0.210	0.305	**0.165**

Performance of these steganalyzers are compared in terms of minimum error probability in Table 1 for several time and transform domain embedding techniques. Outcomes show that the proposed framework offers good performance almost for all cases in the transform domain techniques and has better performance for the time domain ones with higher rate of embedding.

6. CONCLUSION

A new steganalysis framework using moderate number of features have been proposed in this work. The feature set is the combination of both GARCH model and higher order statistics features in the time and transform domain. The GARCH model has been used to properly model the heavy-tailed distribution of the non-approximate wavelet or other transform domain. Beside the GARCH features, higher order statistics features help the steganalyzer to detect those embedding methods aiming at keeping the first order statistics (histogram, average, etc.,) as least modified as possible. The experimental results illustrate that the proposed universal steganalysis framework outperforms state of the art schemes while running on the same order of the features.

7. REFERENCES

[1] Robert F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.

[2] Tim Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, no. 3, pp. 307 – 327, 1986.

[3] M. Amirmazlaghani, H. Amindavar, and A. Moghaddamjoo, "Speckle suppression in sar images using the 2-d garch model," *Image Processing, IEEE Transactions on*, vol. 18, no. 2, pp. 250–259, 2009.

[4] Yun Q Shi, Chunhua Chen, and Wen Chen, "A markov process based approach to effective attacking jpeg steganography," in *Information Hiding*. Springer, 2007, pp. 249–264.

[5] T. Pevny, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," *Information Forensics and Security, IEEE Transactions on*, vol. 5, no. 2, pp. 215–224, 2010.

[6] 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, 2012.

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

[8] T. Filler, T. Pevny, and P. Bas, "Boss (break our steganography system)," July 2010, http://boss.gipsa-lab.grenoble-inp.fr.

[9] Jan Kodovský and Jessica Fridrich, "Calibration revisited," in *Proceedings of the 11th ACM workshop on Multimedia and security*, 2009, MM&Sec '09, pp. 63–74.

[10] Jan Kodovskỳ and Jessica Fridrich, "Steganalysis in high dimensions: Fusing classifiers built on random subspaces," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2011, pp. 78800L–78800L.