ABSTRACT

Owing to the popular usage of JPEG images, the steganographic tools for JPEG images emerge increasingly nowadays, among which OutGuess, F5, and the model based steganography are the most advanced. Advancing the previous work, we present in this paper a new universal steganalysis method based on statistical moments derived from both image 2-D array and JPEG 2-D array. In addition to the first order histogram, the second order histogram is considered. Consequently, the moments of 2-D characteristic functions are also used for steganalysis. The extensive experimental works have shown that the proposed method outperforms in general the prior-arts of steganalysis methods in attacking the three aforesaid steganographic schemes.

1. INTRODUCTION

The popularity of computer utilization accelerates the wide spread of the Internet. As a result, millions of pictures flow on the Internet everyday. Nowadays, the interchange of JPEG (Joint Photographic Experts Group) images becomes more and more frequent. Many steganographic techniques operating on JPEG images have been published and become publicly available. Most of the techniques in this category modify the $8 \times 8$ block discrete cosine transform (BDCT) coefficients in the JPEG domain to embed data.

Among all of the steganographic techniques, the recent published schemes, OutGuess [1], F5 [2], and the model-based steganography (MB) [3] are the most advanced. OutGuess embeds the to-be-hidden data using the redundancy of the cover image. For JPEG images, OutGuess preserves statistics based on the BDCT histogram. It does two things to reduce the possible changes on the cover image caused by data embedding. That is, OutGuess identifies the redundant BDCT coefficients and embeds data into these coefficients causing the least effect during the data embedding. Furthermore, it adjusts the untouched coefficients during the embedding procedure to preserve the original BDCT histogram as much as possible.

F5, developed from Jsteg, F3, and F4, takes two measures to resist steganalysis: straddling and matrix coding. Straddling scatters the message as uniformly distributed as possible over a cover image. With matrix embedding, F5 improves the embedding efficiency (defined as the number of embedded bits per change of the BDCT coefficient).

MB embedding tries to make the embedded data correlated to the cover medium. This is realized by splitting the cover medium into two parts, modeling the parameter of the distribution of the second part given the first part, encoding the second part by using the model and to-be-embedded message, and then combining the two parts to form the stego medium. Specifically, the Cauchy distribution is used to model the JPEG BDCT mode histogram and the embedding keeps the lower precision histogram of the BDCT modes unchanged.

To detect the very existence of hidden information in a stego image, many steganalysis methods have been proposed. A universal steganalysis method using higher order statistics has been proposed by Farid [4]. Quadrature mirror filters are used to decompose a test image into wavelet subbands. The higher order statistics are calculated from wavelet coefficients of each high-frequency subband to form a group of features. Another group of features is similarly formulated from the prediction errors of wavelet coefficients of each high-frequency subband.

In [5], a universal steganalysis system has been proposed by Shi et al. The statistical moments of characteristic functions of a test image, its prediction-error image, and their discrete wavelet transform (DWT) subbands are selected as features. It has been shown that the usage of the moments of characteristic functions, the moments from all of wavelet subbands including the low-low (LL) subbands, and the absolute moments have made the steganalysis effective.

Different from the above two universal steganalysis methods, a specific steganalysis method specifically designed for attacking the JPEG steganographic schemes has been proposed by Fridrich in [6]. With a relatively small-size set of well-selected features, this method outperforms both universal steganalysis methods [4, 5] in attacking OutGuess, F5 and MB according to [7] and our experimental work.

In this paper, from [5], we develop a new universal
steganalysis method based on statistical moments derived from both image 2-D array and JPEG 2-D array. In addition to the first order histograms, the second order histograms are considered. Consequently, the moments of 2-D characteristic functions are also utilized for steganalysis. The extensive experimental results have shown that this steganalysis method outperforms in general the prior-arts [4, 5, 6] in attacking OutGuess, F5 and MB.

The rest of this paper is organized as follows. In Section 2, the feature generation procedure is addressed. The classifier used in our work is briefly described in Section 3. Experimental works are presented in Section 4. Discussion is made and conclusion is drawn in Section 5.

2. FEATURE GENERATION

In this work, steganalysis is considered as a task of two-class pattern recognition, i.e., a test image needs to be classified as either a stego image (with hidden data) or a non-stego image (without hidden data). Therefore, feature generation is a key step in our steganalysis.

The major characteristics of the previous work [5] have been described in Section 1. Specifically, 78-D feature vectors used in steganalysis are constructed as follows. The first half of features are generated from the given test image and its 3-level Haar wavelet decomposition. The second half of features are from the prediction-error image and its 3-level Haar wavelet decomposition. Denoting the test image and the prediction-error image as the LL subbands, we have in total 26 subbands. The characteristic function (CF) (i.e., the discrete Fourier transform of the histogram) [8] of each of these subbands is calculated. The first three absolute moments of these CF’s are used to form the 78-D feature vectors. The absolute moments are defined as follows:

$$M_n = \frac{\sum_{i=1}^{N/2} x_i^n |H(x_i)|}{\sum_{i=1}^{N/2} |H(x_i)|},$$

(1)

where $H(x_i)$ is the CF component at frequency $x_i$ and $N$ is the total number of different value level of coefficients in a subband under consideration. The prediction-error image is the difference between the original image and the prediction image. The prediction is depicted in Equation (2) and Figure 1. For more details, readers may refer to [5].

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

(2)

Fig. 1. The context of prediction

The method [5] performs well in steganalyzing several major data embedding schemes. However, our experimental results have indicated that its performance on attacking OutGuess, F5 and MB is not as good as that achieved by [6]. In particular, it does not perform well in detecting MB. This has motivated our work reported in this paper. Towards this direction, two measures have been taken in this new steganalysis method as shown below.

2.1. JPEG 2-D array

If we consider features used in [5] as statistical moments derived from the given test image 2-D array, we now include the statistical moments derived from the JPEG 2-D array associated with the test image. Namely, we first consider a 2-D array generated by applying 8x8 block DCT to the test image followed by quantization using JPEG quantization table. Note that these quantized JPEG BDCT coefficients can be either positive, or negative, or zero. In addition, the steganographic schemes operating in the JPEG images do not touch the DC coefficients nor change the sign of the AC coefficients [2, 3] (note that a DCT coefficient changing to zero is not sign change). We then take the absolute values of all these coefficients. This resultant 2-D array is referred to as JPEG 2-D array in this paper.

Applying the procedure described in [5] to this JPEG 2-D array results in another set of 78 features. This has been shown in the block diagram of our proposed method, Figure 2. The moments of the CF’s are defined as in Equation (1), where $N$ is the total number of different absolute values of JPEG quantized BDCT coefficients in a subband under consideration.

In computing the prediction-error 2-D array from the JPEG 2-D array, however, we do it in a slightly different way. Firstly, for those zero elements in the JPEG 2-D array, we simply set the prediction values as zero as shown in Equation (3), implying that the zero DCT coefficients remain zero in the prediction-error 2-D array. Secondly, taking absolute-value operation is applied to the prediction error 2-D array as shown in Equation (3).

$$\hat{x} = \begin{cases} 0, & \text{if } x = 0 \\ \max(a, b), & \text{if } c \leq \min(a, b) \\ \min(a, b), & \text{if } c \geq \max(a, b) \\ a + b - c, & \text{otherwise} \end{cases}$$

(3)

2.2. Second-order histograms

Modern steganographic schemes, such as OutGuess and MB, try to keep the histogram change as less as possible in order to resist the steganalysis as much as possible. For instance, MB embedding keeps the lower precision histogram of the BDCT modes unchanged. In order to break this strategy, we propose to use higher order histograms in the steganalysis. It is known that the computational complexity will increase dramatically as the order of histogram increases. In the
specific steganalysis algorithm reported in this paper, we only use the second-order histogram.

The second-order histogram [9] is a measure of the joint occurrence of pairs of pixels separated by a specified distance and orientation. Denote the distance by \( \rho \), and the angle with respect to the horizontal axis by \( \theta \). The second-order histogram is defined as

\[
h_{2}(j_{1}, j_{2}; \rho, \theta) = \frac{N(j_{1}, j_{2}; \rho, \theta)}{N_{r}(\rho, \theta)},
\]

(4)

where \( N(j_{1}, j_{2}; \rho, \theta) \) is the number of pixel pairs for which the first pixel value is \( j_{1} \), while the second pixel value is \( j_{2} \), and \( N_{r}(\rho, \theta) \) is the total number of pixel pairs in the image with separation \((\rho, \theta)\). The second-order histogram can also be used in JPEG 2-D array. The second-order histogram corresponds to a 2-D array, often called dependency matrix or co-occurrence matrix.

For each wavelet subband derived from JPEG 2-D array, we generate three second-order histograms with the following three separations:

\[
(\rho, \theta) = \{(1,0), (1,\frac{\pi}{2}), (1,\frac{-\pi}{2})\},
\]

(5)

which are called horizontal 2-D histogram, vertical 2-D histogram, and diagonal 2-D histogram, respectively. For example, looking at Figure 1, the pair \((x, a), (x, b), (x, c)\) are separated by \((1, 0), (1,\frac{-\pi}{2}), (1,\frac{-\pi}{4})\), respectively. After applying 2-D DFT to the second-order histograms to obtain the 2-D CF’s, the two marginal moments of 2-D CF’s are calculated by

\[
M_{\rho,a} = \frac{\sum_{j_{1}}^{\frac{N}{2}} \sum_{j_{2}}^{\frac{N}{2}} u_{j_{1}} v_{j_{2}} H(u_{j_{1}}, v_{j_{2}})}{\sum_{j_{1}}^{\frac{N}{2}} \sum_{j_{2}}^{\frac{N}{2}} H(u_{j_{1}}, v_{j_{2}})},
\]

(6)

\[
M_{\phi,a} = \frac{\sum_{j_{1}}^{\frac{N}{2}} \sum_{j_{2}}^{\frac{N}{2}} u_{j_{1}} v_{j_{2}} H(u_{j_{1}}, v_{j_{2}})}{\sum_{j_{1}}^{\frac{N}{2}} \sum_{j_{2}}^{\frac{N}{2}} H(u_{j_{1}}, v_{j_{2}})},
\]

(7)

where \( H(u_{j_{1}}, v_{j_{2}}) \) is the 2-D CF component at frequency \((u_{j_{1}}, v_{j_{2}})\) and \( N \) is the total number of different absolute values of coefficients in a subband under consideration. Since for each of these three directions, we can generate two marginal moments according to Equations (6) and (7), and thus have additional \(78 \times 3 = 234\) feature components. Hence, we have in total \(390-D\) feature vectors for steganalysis.

The block diagram of feature generation procedure is shown in Figure 2.

3. CLASSIFICATION

Instead of neural network (NN) used in [5], the support vector machine (SVM) is used in this work as classifier considering its comparable performance to and more efficient calculation than that of the NN. Due to the constraint of paper length, the more detailed description about the SVM is omitted here. Readers please refer to [10] for more details. Note that the polynomial kernel [11] is used in our investigation.

4. EXPERIMENTAL WORK

4.1. Image database

An image database consisting of 7,560 JPEG images with quality factors ranging from 70 to 90 is used. One third of these images were taken by members of our research group in different places, at different times, and with different digital cameras. The other two thirds were downloaded from the Internet. Each image was cropped (central portion) to the size of either \(768 \times 512\) or \(512 \times 768\). Some sample images are shown in Figure 3.

4.2. Stego image generation

The experiments reported here focus on detecting the Outguess, F5, and MB1. The codes for these algorithms are publicly available [12, 13, 14]. Since there are quite a few zero BDCT coefficients in the JPEG images and the quantity of zero coefficient varies, the data embedding capacity differs from image to image. The common practice is to use the ratio between the length of hidden data and the number of non-zero BDCT AC coefficients as the measure of data embedding capacity for JPEG images. For OutGuess, we embed 0.05, 0.1, and 0.2 bpc (bits per non-zero BDCT AC coefficients).
In [5], absolute moments of characteristic functions of a test image, its prediction-error image, and their wavelet subbands are used for universal steganalysis. In this new steganalyzer, we resort to the JPEG 2-D array and second-order histograms. Consequently, moments generated from the JPEG 2-D array and from 2-D characteristic functions are included for steganalysis, resulting in superior performance to that achieved by the prior-arts in general.

(2) In particular, the detection rates for MB1 have been raised significantly. They are not only much higher than that achieved by [4, 5] but also rather higher than that by [6].

(3) The detection rates for OutGuess are much higher than [4, 5] and slightly higher than [6].

(4) The detection rates for F5 are much higher than [4, 5] and comparable with [6].

(5) It is noted that the detection rates for F5 are higher than for MB1 if we do not take absolute values in forming the JPEG 2-D array. But, at this circumstance, the detection rates for both F5 and MB1 will decrease. Further investigation in this regard is under way.

ACKNOWLEDGEMENT

Appreciation goes to Mr. Jianjiong Gao for some of his programming work in related experimental investigation.

REFERENCES