2008 International Conference on Advanced Computer Theory and Engineering

JPEG Image Steganalysis Improvement Via Image-to-image Variation Minimization

Chiew Kang Leng
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
94300 Kota Samarahan, Malaysia
kchiew@fit.unimas.my

Josef Pieprzyk
Department of Computer
Macquarie University
NSW 2109, Australia
josef@ics.mq.edu.au

Abstract

In this research, we introduce an approach to enhance the discriminative capability of features by employing image-to-image variation minimization. In order to minimize image-to-image variation, we will estimate the cover image from the stego image by decompressing the stego image, transforming the decompressed image and recompressing back. Since the effect of the embedding operation in an image steganography is actually a noise adding process to the image, applying these three processes will smooth out the noise and hence the estimated cover image can be obtained.

1. Introduction

Image Steganography is a process of embedding a secret message into an image for covert communication to conceal the existence of the communication. The image used as the medium for embedding is called the *cover image* and the generated image from steganography which is embedded with the secret message is called the *stego image*.

On the other hand, steganalysis is a process of detecting a secret message embedded in a multimedia object (such as digital images, video, sound, etc.). The detection of a secret message is equivalent to breaking the steganographic method. Apart from this, steganalysis is also used to measure the security performance of a steganography.

In general, steganalysis can be categorized into two categories, namely, targeted steganalysis and blind steganalysis. Targeted steganalysis normally is designed to detect the existence a of secret message embedded by a specific steganographic technique or its slight variations. While blind steganalysis is a generic technique that can be trained to detect the existence of a secret message embedded by an unknown steganographic technique.

Although the performance of blind steganalysis is often inferior to a targeted one, its flexibility and wide coverage of different steganographic techniques makes blind steganalysis an attractive and practical choice. So in this work we focus on blind steganalysis, specifically we will propose a method to reduce image-to-image variation and to increase feature discriminative ability. In blind steganalysis, feature is a set of descriptors or distinctive characteristic attributes extracted from an image that is discriminative and sensitive to the embedding process. We will illustrate the efficiency of the proposed method by incorporating it in several existing blind steganalysis techniques. The improvements obtained will be presented in this paper as well.

In the next section, we will discuss the related literature. Section 3 will model the steganography artifact as an additive noise. The proposed method will be analyzed in Section 4. Section 5 will elaborate the incorporation of the proposed method in several existing steganalysis techniques and follow by the results of our analysis in Section 6. Finally, the paper is concluded in Section 7. All the notations defined in the equations throughout this paper will be unique unless otherwise stated.

2. Background

The first blind steganalysis is proposed by Avcibas et al. in [1]. In that paper, the authors extracted several features by utilizing image quality metrics. Based on these features, multivariate regression is used to categorize unknown images into either cover images (with no secret message) or stego images (containing a secret message).

Another blind steganalysis technique is proposed by Lyu and Farid in [7]. They use higher-order statistics of a wavelet decomposition as the features. These higher-order statistics are: mean, variance, skewness and kurtosis. They used a support vector machine as the classifier.

Harmsen and Pearlman [5] build a blind steganalysis that

978-0-7695-3489-3/08 $25.00 © 2008 IEEE
DOI 10.1109/ICACTE.2008.94

exploits the first order moment of the characteristic function of an image intensity histogram. The extracted feature is combined with a Bayesian classifier.

Xuan et al. in [12] presented a different approach that uses the characteristic functions as the features and Bayesian classifier. However, instead of using the image intensity histogram, Xuan et al. used wavelet subbands, based upon which the higher order moments are obtained.

Another interesting research direction is an investigation of feature-based steganalysis (see paper [3]). The proposed features are used specifically to detect JPEG image steganography, where Fisher Linear Discriminant is used for classification.

3. Steganography as additive noise

Let X denote an instance of a JPEG cover image and let $P_X(x)$ denote the probability mass function of a cover image. In the JPEG image, the probability mass function can be considered as the frequency count of the quantized discrete cosine transform (DCT) coefficients.

The secret message probability mass function is the distribution of the additive stego noise defined as follows

$$P_N(n) \equiv P(y - x = n), \tag{1}$$

where x and y are quantized DCT coefficients before and after embedding, respectively.

Generally, we can divide a cover image used in a steganography into two parts, x_c and x_s. The part x_c is the unperturbed part and normally consists of a group of the most significant bits. Whereas x_s is the part that will be altered and used to carry the secret message and normally this part contains a group of less significant bits.

Since the additive stego noise is independent of the cover image, perturbing the x_s part by embedding a secret message to it is equivalent to convoluting the additive stego noise probability mass function with the cover image probability mass function. This can be expressed by the following equation

$$P_Y(n) = P_N(n) \quad \Theta \quad P_X(n), \tag{2}$$

where Θ is the convolution and $P_Y(n)$ is the stego image probability mass function.

4 Image-to-image variation minimization

Defining a discriminative feature in image steganalysis is a challenging task because the defined feature should be only sensitive to steganographic alteration and not to image-to-image variation. Image-to-image variation is the difference between one image's underlying statistic to another image's underlying statistic. The underlying statistic can

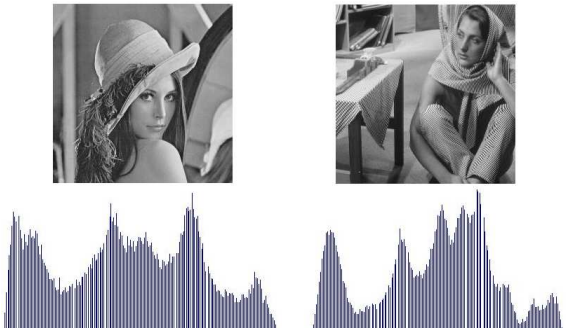

Figure 1. Cover images and histogram distribution

be the histogram distribution of the DCT coefficients or the pixels intensity. For example, the images shown in Figure 1 are obviously different and therefore their underlying statistics (histogram distributions shown below each image) are also different. This difference is the above mentioned image-to-image variation. The question of interest here is how can we categorize these images into either cover images or stego images. It is obvious that there is no consistency to differentiate these two images either is cover image or stego image by just examining at the histogram distribution because the distribution is rather random and different for each of the image. If we were to apply feature extraction directly to the histogram distribution then the extracted feature will have poor discriminative capability because the image-to-image variation is large.

An ideal case will be when the cover image is present together with the stego image during the steganalysis detection process. In this case, the image-to-image variation is at the minimum level because we can subtract the stego image, Y from the cover image, X directly as follows:

$$N = Y - X \tag{3}$$

and the subtraction result obtained will be the stego noise, N. The subtraction can be viewed as pixel-wise subtraction. However, this case is not typical and most of the time we have the stego image only.

Therefore, in order to minimize the image-to-image variation, it is reasonable to estimate the cover image from the stego image. We are going to demonstrate the efficiency of our method by applying it to the existing steganalysis techniques. Thus, we propose two different methods to achieve optimum performance for the respective existing steganalysis techniques. The two proposed methods are defined as follows:

$$\Psi_1 = \Phi(\acute{v}) - \Phi(v) \tag{4}$$

Figure 2. Transformed image by scaling (left) and cropping (right)

$$\begin{aligned} \Psi_2 &= \Phi(\eta) \\ \eta &= \acute{\upsilon} - \upsilon \end{aligned} \tag{5}$$

where υ and $\acute{\upsilon}$ is the stego image and estimated cover image, respectively. The variable η is the additive stego noise generated by the embedding operation and Ψ_i is the features set produced by the steganalysis feature extraction technique, $\Phi(\cdot)$. If υ is a cover image, then it is observed that $\Psi_i \approx 0$ and if υ is a stego image, then it is observed that the absolute value of Ψ_i is always greater than zero and where $i = 1, 2$.

In cover image estimation process, first we will decompress the JPEG images to the spatial domain and apply a transformation to the decompressed images. In our experiments, we employed scaling by bilinear interpolation (shown in the left image of Figure 2) and cropping by 4 pixels in both horizontal and vertical direction (shown in the right image of Figure 2). After that, we recompress the transformed image back to JPEG domain. In the decompress and recompress process, we have used the same JPEG image quality as before the transformation to avoid double compression. Since a steganography can be modeled as additive noise, the effect of the transformation can be attributed as even out the added noise. Hence the cover image estimation is reasonable. We also discuss the improvements obtained and the effectiveness of our method as confirmed by experiments. Similar estimation approaches have proved efficient and can be found in [3, 6].

5 Steganalysis improvement

In this paper, we will select three existing steganalysis techniques to demonstrate the efficiency of the proposed methods. In the following subsection, we will discuss separately the incorporation of the proposed methods with each of the selected steganalysis.

5.1 Moments of wavelet decomposition

Lyu and Farid [7] have proposed to use a higher-order statistic as the features which include mean, variance, skewness and kurtosis. Two sets of these higher-order statistics are obtained and result in 72 features.

The first set is acquired from a wavelet decomposition based on separable quadrature mirror filters. A total of 9 subbands are obtained and the mean, variance, skewness and kurtosis are computed for each of this subband. These 36 higher order statistics Ψ_{w_k} (9 subbands x 4 higher order statistics) and $k = 1, 2, \ldots, 36$ will be used as the features.

The second set of feature is obtained from the log error in the linear predictor (refer to [7] for details) for the same 9 subbands and then the 4 higher order statistics are computed for each of this subband. This will result in another 36 features, Ψ_{e_k} for $k = 1, 2, \ldots, 36$.

Instead of using the 72 features extracted directly from the image in the classification, we use the proposed method from Section 4. Specifically, we improve the features discriminative capability by employing the second proposed method defined in Equation 5. Thus, our improved feature set is defined in the following equation:

$$\begin{aligned} \eta &= \acute{\upsilon} - \upsilon \\ \acute{\Psi}_{w_k} &= \Phi_{w_k}(\eta) \\ \acute{\Psi}_{e_k} &= \Phi_{e_k}(\eta) \\ \acute{\Psi} &= \acute{\Psi}_{w_k} + \acute{\Psi}_{e_k} \end{aligned} \tag{6}$$

5.2 Moment of CF of PDF

The use of characteristic functions (CF) was pioneered by Harmsen and Pearlman in [5]. The CF is obtained by applying a discrete Fourier transform to the probabilistic density function (PDF) of an image. After that, from this characteristic function, the first order absolute moment (called as the *center of mass* in their paper) is computed and used as the feature. Equation (7) shows the calculation of this moment.

$$\Phi(H[k]) = \frac{\sum_{k=0}^{K} k|H[k]|}{\sum_{k=0}^{K} |H[k]|} \tag{7}$$

where $H[\cdot]$ is the characteristic function, $K \in \{0, \ldots, \frac{N}{2} - 1\}$ and N is the width of the domain of the PDF.

The feature proposed in [5] has only the first moment, we further increase to second and third moments according to the following equation for $\alpha \in 1, 2, 3$:

$$\Phi(H[k]) = \frac{\sum_{k=0}^{K} k^{\alpha}|H[k]|}{\sum_{k=0}^{K} |H[k]|} \tag{8}$$

Increasing the moment to a higher order is not always significant and therefore is not necessary [10].

By incorporating the proposed method defined in Equation 4, we obtained a new set of features which is defined as follows:

$$\Psi = \Phi(\acute{H}[k]) - \Phi(H[k]) \tag{9}$$

5.3 Moment of CF of wavelet subbands

The basis of the features proposed in [12] are derived from a Haar wavelet decomposition. The authors have decomposed the wavelet to 12 subbands denoted by LL_i, HL_i, LH_i, HH_i where $i = 1, 2, 3$. The given image histogram denoted as LL_0 is also employed.

Essentially, the wavelet subbands and the image histograms are probability mass functions. Motivated by the characteristic function (CF) from [5], the authors constructed the CF from all the wavelet subbands and the image histograms result in 13 CFs.

After that, the first three moments for each of the 13 CFs are computed as define according to the following relation

$$\Phi_{nm}(H(f_k)) = \frac{\sum_{k=0}^{N/2} f_k^n |H(f_k)|}{\sum_{k=0}^{N/2} |H(f_k)|} \tag{10}$$

where $n = 1, 2, 3$ are the three moments and $m = 1, 2, \ldots, 13$ are the 13 CFs. $H(\cdot)$ is the probability density function of the CF or discrete Fourier transform (DFT) coefficients as discussed in Subsection 5.2.

Next, we will improve this method by incorporating the proposed method as defined in Equation 4. Equation 11 shows the calculation of the new feature set.

$$\Psi_{nm} = \Phi_{nm}(\acute{H}(f_k)) - \Phi_{nm}(H(f_k)) \tag{11}$$

6. Experimental result

In this section, results of experiment will be presented and analyzed. We will explain the experimental setup in Subsection 6.1 and followed by a comparison of the results in Subsection 6.2.

6.1 Experiment setting

Since we are interested in comparing the feature discriminative performance, we will standardize the classification by using support vector machine (SVM [2]) as the classifier in all experiments to guarantee fair comparison.

Three different steganographic methods which are F5 [11], OutGuess [8] and MB1 [9] are employed to create three different types of stego images. In order to have a percentage wise equal number of changes over all images, we define the embedding rate in term of bits per embeddable quantized DCT coefficients of the cover image for each of the steganography. We used four embedding rate which are 5%, 25%, 50% and 100%.

	Steganalysis	**F5**	**OutGuess**	**MB1**
5%	Improved	0.5160	0.5120	0.5165
	Original	0.5077	0.4625	0.5064
25%	Improved	0.5850	0.6460	0.6304
	Original	0.5374	0.5384	0.5379
50%	Improved	0.7236	0.7983	0.7564
	Original	0.5940	0.6606	0.5754
100%	Improved	0.90437	0.8811	0.8815
	Original	0.7218	0.7650	0.6485

Table 1. Performance comparison (AUR) for [7] method

	Steganalysis	**F5**	**OutGuess**	**MB1**
5%	Improved	0.5055	0.4838	0.5040
	Original	0.5022	0.4631	0.5020
25%	Improved	0.5279	0.5261	0.5080
	Original	0.5097	0.4794	0.5029
50%	Improved	0.5654	0.5408	0.5147
	Original	0.5314	0.4799	0.5030
100%	Improved	0.6392	0.5686	0.5228
	Original	0.5971	0.4817	0.5052

Table 2. Performance comparison (AUR) for [5] method

In the dataset construction, 2037 images of four different sizes (512x512, 608x608, 768x768 and 1024x1024) were downloaded from [4]. All images are cropped to get the center portion of the image and are converted to grayscale images. From the constructed dataset, 80% of the dataset is used for the training phase and the remaining 20% is used for the testing phase.

6.2 Result Comparison

We will compare the improved version (IV) by our proposed methods to the original version (OV) for each of the three steganalysis techniques as discussed in Section 5. The detection results were evaluated using the area under the Receiver Operating Characteristic curve (AUR). Higher AUR value indicates better steganalysis performance. The obtained results are tabulated in Table 1, 2 and 3.

To illustrate a clearer picture for the comparison, we select one of the Receiver Operating Characteristic (ROC) curve comparisons obtained from the experiments and display in Figure 3. The selected ROC curves are obtained from the steganalysis technique discussed in Subsection 5.1

	Steganalysis	F5	OutGuess	MB1
5%	Improved	0.5016	0.5087	0.5105
	Original	0.5009	0.4793	0.5042
25%	Improved	0.5514	0.5385	0.5116
	Original	0.5202	0.4927	0.5043
50%	Improved	0.7601	0.5635	0.5648
	Original	0.5625	0.5064	0.5286
100%	Improved	0.8518	0.6213	0.5747
	Original	0.6667	0.5307	0.5520

Table 3. Performance comparison (AUR) for [12] method

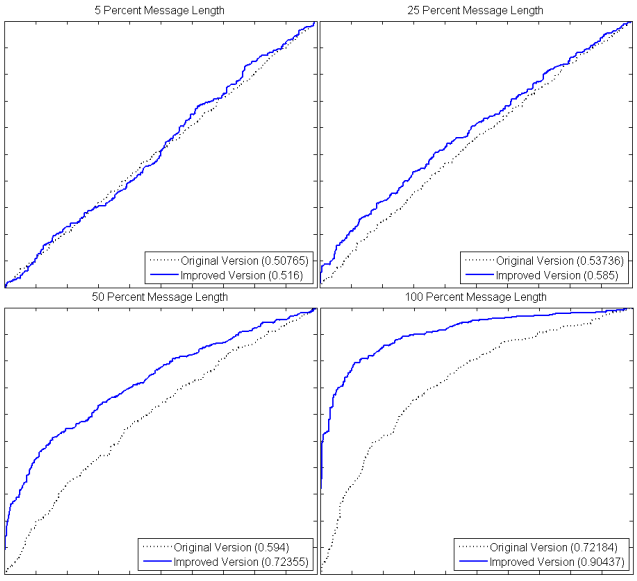

Figure 3. ROC curve and AUR value for the improved version and original version of [7]

in detecting the F5 steganographic model. The Y-axis and X-axis represent the detection rate and false alarm rate, respectively and each ranges from 0 to 1. The value shown inside the bracket is the AUR value indicating the detection accuracy.

From Figure 3 it can be clearly seen that all the IV ROC curves are higher than the OV ROC curves and also each IV AUR value is larger than the corresponding OV AUR value indicating the improved version has outperformed the original version in all the embedded message lengths. As for detecting OutGuess and MB1 steganographic models, one can clearly see from Table 1 that all the improved versions have also outperformed the original versions, verifying the effectiveness of the proposed methods.

Shown in Table 2 is the comparison between the IV and OV of the steganalysis technique in [5]. Although the improvement is not as large as the improvement in [7], overall the performance also has improved.

As for the third improved steganalysis technique [12], Table 3 clearly shows significant improvement in detecting all steganographic models and detecting the F5 steganographic model appeared to be the most improved.

7. Conclusions

In conclusion, our proposed methods have improved the three selected steganalysis techniques by estimating the cover image from stego image to reduce image-to-image variation. In future, we will investigate possible improvements over other steganalysis techniques by using the proposed methods.

References

[1] I. Avcibas, M. Nasir, and B. Sankur. Steganalysis based on image quality metrics. *4th IEEE Workshop on Multimedia Signal Processing*, pages 517–522, 2001.

[2] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm, 2001.

[3] J. Fridrich. Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes. *6th International Workshop on Information Hiding*, 3200:67–81, 2004.

[4] P. Greenspun. Philip greenspun. http://philip.greenspun.com.

[5] J. Harmsen and W. Pearlman. Steganalysis of additive noise modelable information hiding. *Security and Watermarking of Multimedia Contents V*, 2003.

[6] A. Ker. Steganalysis of LSB matching in grayscale images. *IEEE Signal Processing Letters*, 12(6):441–444, 2005.

[7] S. Lyu and H. Farid. Detecting hidden messages using higher-order statistics and support vector machines. *5th International Workshop on Information Hiding*, 2002.

[8] N. Provos. Defending against statistical steganalysis. *Proceedings of the 10th conference on USENIX Security Symposium*, 10:24–34, 2001.

[9] P. Sallee. Model-based steganography. *International Workshop on Digital Watermarking*, 2003.

[10] Y. Wang and P. Moulin. Optimized feature extraction for learning-based image steganalysis. *IEEE Transactions on Information Forensics and Security*, 2(1), 2007.

[11] A. Westfeld. F5 - A Steganographic Algorithm: High Capacity Despite Better Steganalysis. *4th International Workshop on Information Hiding*, 2001.

[12] G. Xuan, Y. Q. Shi, J. Gao, D. Zou, C. Yang, Z. Zhang, P. Chai, C. Chen, and W. Chen. Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions. *7th International Workshop on Information Hiding*, 3727:262–277, 2005.