On Estimation of Secret Message Length in JSteg-like Steganography

Xiaoyi Yu, Yunhong Wang, Tieniu Tan
National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, P.O. Box 2728, Beijing, 100080, P.R. China
E-mails: {xyyu,wangyh,tnt}@nlpr.ia.ac.cn

Abstract

Image steganalysis has attracted more attention recently. In this paper, a new method for detecting secret message and estimating the secret message length of bit-streams embedded using J-Steg like steganography is proposed. First, based on the model of statistical distribution of quantized DCT coefficients, the histogram of cover image is estimated from stego image. Then the secret message is detected and the secret message length is estimated with the estimated cover histogram. Interestingly, our study on steganalysis of J-Steg like embedding offers a proof of Zhang et al.’s method [6]. The methodology described in this paper is a framework which can also be applied to many other stegaographic methods such as JPEG2000 file format embedding.

1. Introduction

Steganography, a science and art of secret communication, aims to hide the very presence of communication. That is to say, the essential goal of steganography is to conceal the facts of a hidden message. The frequently used steganographic method in JPEG format is the J-Steg like algorithm, which proposed by D. Upham [1]. It works by embedding message bits as the LSBs of the quantized DCT (Discrete Cosine Transform) coefficients. The embedding mechanism skips all coefficients with the values of ‘0’ or ‘1’. There are two embedding ways according to the selection of coefficients. One is sequential embedding; the other is random embedding whose coefficients selection is usually determined by a secret stego key shared by the communicating parties.

During the last few years, many powerful steganalytic methods [2,5] capable of detecting J-Steg like embedding were proposed. J-Steg with sequential message embedding is detectable using the chi-square attack [2]. J-Steg with random straddling is detectable using the generalized chi-square attack [5]. Provos point out that the method he proposed can work well [5]. However he hasn’t given any further details for this generalized approach. Both methods can’t estimate the number of changes due to embedding and thus can’t estimate the secret message length. Zhang proposed a method which can reliably detect the existence of hidden message and even could estimate the length of hidden message exactly [6]. Although existing methods can detect hidden messages, these methods have their limitations. They either can’t estimate the length of hidden message or lack theoretical foundation. In this paper, we present a systematic approach for estimating the number of embedding changes due to J-Steg like algorithm steganography in JPEG file formats. Our method can also be applied to many other stegographic methods such as JPEG2000 file format embedding. It is a framework to detect the hidden message and estimate the length of hidden message if one can find a optimal statistical image models for the stegaographic method he want to detect. The detection and estimation algorithm is introduced in Section 2. In Section 3, we apply the method described in Section 2 to steganalysis J-Steg like steganography. The verification of Zhang et al.’s Method is outlined in Section 4. The experimental results and the conclusion are in the last 2 sections of the paper.

2. Method Description

As described in [7], if we have a model which captures some main statistical properties of the carrier media and which will be changed by steganographic process, we can reliably detect the existence of hidden message. We use the terms as [7]. Let \( x \) denote an instance of a class of potential carrier media, such as pixel values or quantized DCT coefficients of an image. If we treat \( x \) as an instance of a random variable \( X \), we can model \( X \) using the probability distribution \( P_X(x) \).

To detect whether there is a hidden message embedded in a suspected media is to perform a hypothesis test to find out whether the instance \( x \) obey the probability distribution \( P_X(x) \). But how to estimate the length of hidden message? In [3], Fridrich summarized the principles to estimate the length of secret message. For most steganographic techniques, it is usually not too difficult to identify a macroscopic quantity \( S(m) \) that predictably changes (for example, monotonically increases) with the length of the embedded secret message \( m \). We can calculate an estimate of the unknown message length \( m \) by solving the equation \( S(m) = S_{\text{stego}} \) for \( m \), where \( S_{\text{stego}} \) is the value of \( S \) for the stego image under investigation. Fridrich call \( S(m) \) the distinguishing
In general, the function \( S \) has several undetermined parameters which can be determined by estimating some extreme values of \( S \), such as \( S(0) \) (i.e. the cover image). To obtain the value \( S(0) \), Fridrich cropping the stego image by four pixels to obtain the macroscopic properties that well approximate the properties of the cover image. In this paper, we apply the same principle to estimate the length of hidden message. Since we can model the carrier media, it is not necessary to crop the stego image to approximate the macroscopic properties of the cover image. If we can model \( x \) precisely, our method can obtain more precise results than Fridrich’s method. We will show how our proposed method works on estimation of the secret message length in JSteg-like steganography.

### 3. Detection of Secret Message & Estimation of Secret Message Length in JSteg-like Steganography

In this section, we will show the process of detection the secret message and estimation the length in JSteg-like steganography. We first model the statistical distribution of quantized DCT coefficients and then use estimated parameters to detect the secret message and estimate the capacity of secret message.

#### 3.1. Model the Statistical Distributions of Quantized DCT Coefficients

In the JPEG compression standard, images are divided into \( 8 \times 8 \) blocks. Each block is passed through a DCT (Discrete Cosine Transform) to produce 64 DCT coefficients and then the coefficients are quantized according to a quantization table and encoded using an entropy encoder. It is generally believed that the distribution of DCT coefficients is Laplacian except for the \([0,0]\) coefficients [3]. These coefficients are called AC coefficients. The Laplacian distribution

\[
p(x) = \frac{\lambda}{2} e^{-\lambda|x|} \tag{1}
\]

If the coefficients are quantized, a generalized Laplacian can still be used to fit the resulting histogram with integer width bins. In [7], Sallee used a specialized form of a generalized Cauchy distribution instead of the generalized Laplacian:

\[
p(x) = \frac{p-1}{2s} \left( \frac{x}{s} \right)^{p} \left( 1 - \left( \frac{x}{s} \right) \right)^{p} \tag{2}
\]

For this distribution, there is a closed form solution for the cumulative density function which makes it easy to integrate the model density for individual histogram bins. As Sallee described, when taking into account a more accurate estimation of the quantization effects, one would find this distribution appears to fit DCT coefficients better than the generalized Laplacian/Gaussian. In this paper, we’ll use this distribution to model quantized DCT coefficient.

Now, we come to the core of the problem. We only know the stego image, how can we model the cover image? As described in Section 1, J-Steg like algorithm works by embedding message bits as the LSBs of the quantized DCT coefficients except ‘0’ and ‘1’. Let \( h(d) \), \( d = \ldots -2, -1, 0, 1, 2, \ldots \), be the histogram of the quantized DCT coefficients from the cover image. Let \( H(d) \) be the histogram of cover image after embedding \( m \) pseudorandom bits in the LSBs, the histograms \( H(d) \) and \( h(d) \) will have relations as follow equations.

\[
H(0) = h(0), \quad H(1) = h(1) \tag{3}
\]

\[
H(2i) = h(2i) - \alpha[h(2i) - h(2i + 1)] \tag{4}
\]

\[
H(2i + 1) = h(2i + 1) + \alpha[h(2i) - h(2i + 1)] \tag{5}
\]

where \( i = \pm 1, \pm 2, \ldots \),

\[
\alpha = \frac{2}{\sum_{i=0}^{m} H(i)}
\]

From (4) and (5), we obtain \( H(2i) + H(2i+1) = h(2i) + h(2i+1) \). The sum of histogram values keeps unchanged after message embedding. This condition must be satisfied for all histogram pairs \( h(2i), h(2i+1) \).

Let \( H_{\alpha}(i) = H(2i) + H(2i + 1) \). We use \( H(0), H(1), H_{\alpha}(i) \) to fit the model (2). In fact, \( H_{\alpha}(i) \) is low precision histogram of AC coefficients from the cover image. The model parameters \( s \) and \( p \) can be determined from \( H_{\alpha}(i) \). The fit process is the same as [7]. Once the model is fit to the histograms for a stego image, it is used to estimate the histogram of the cover image. Let \( \hat{h}(d) \) be the estimated histogram of cover image.

\[
\hat{h}(0) = H(0), \quad \hat{h}(1) = H(1) \tag{6}
\]

\[
\hat{h}(2i) = H_{\alpha}(i) \frac{P(2i)}{P_{\alpha}(i)} \tag{7}
\]

\[
\hat{h}(2i + 1) = H_{\alpha}(i) \frac{P(2i + 1)}{P_{\alpha}(i)} \tag{8}
\]

where \( P_{\alpha}(i) = \int_{2i-0.5}^{2i+1.5} p(x) dx, \)

\[
P(2i) = \int_{2i-0.5}^{2i+0.5} p(x) dx, \quad P(2i + 1) = \int_{2i+0.5}^{2i+1.5} p(x) dx
\]

Figure 1 shows the original coefficient histogram of an image and the estimated histogram after message
embedding. The coefficients are all AC coefficients. It can be seen from the figure that we can almost exactly estimate the histogram of cover image from a stego image. Our estimation method is better than Fridrich’s cropping method, because it isn’t necessary for us to consider double compression effect.

3.2. Detection of secret message and estimation of secret message length

Since we can accurately estimate the histogram of the cover image, the detection of hidden message and estimation of hidden message length becomes easy. The detection is determined using the Chi-square test. The $\chi^2$ statistic is given as $\chi^2 = \sum_{i=1}^{2k}(\hat{h}(i) - H(i))^2$ with 2k-1 degree of freedom. We can perform Chi-square test at significance level $\alpha$ and 2k-1 degree of freedom to decide whether a suspect images contains secret message or not. Furthermore, we can reliably estimate the length of the message bits. For random embedding method, equation (4) and (5) express the distinguishing statistics as a function of secret message $m$ and the original histogram. The original histogram can be estimated using the method described in Section 3.1. We calculate $\alpha$ using the following equation which is derived from (4).

$$\alpha = \arg \min_{\alpha} \sum_{i=1}^{2k} (H(2i) - (1-\alpha)\hat{h}(2i) - \alpha\hat{h}(2i+1))^2$$

The least square approximation in the above equation leads to the following formula for $\alpha$

$$\alpha = \frac{\sum_{i=1}^{2k}(H(2i) - \hat{h}(2i))(\hat{h}(2i) - \hat{h}(2i+1))}{\sum_{i=1}^{2k}(\hat{h}(2i) - \hat{h}(2i+1))^2}$$ (9)

where $k$ is the maximum quantized DCT AC coefficient. Thus the length of unknown message can be calculated as

$$M = 2\alpha \sum_{i=0,j=1} H(i)$$ (10)

Let $C$ denotes the whole coefficients, $S$ the subset of $C$, $\overline{S} = C - S$ the complement of $S$ in $C$, $H_A(d)$ the actual histogram and $\hat{h}_A(d)$ the estimate histogram of set $A$, where $A$ could be $\overline{S}, S$ or $C$. For sequential embedding method, we construct $S$ the same way as sequential embedding method, and ensure $H_{\overline{S}}(d)$ and $\hat{h}_{\overline{S}}(d)$ are approximately equal using Chi-Square test. Then the embedding ratio is

$$\beta = \sum_{i=0,1} H_S(i) / \sum_{i=0,1} H_C(i)$$ (11)

and the length of unknown message is

$$M = \sum_{i=0,1} H_S(i)$$ (12)

3.3. Experimental Results

For testing purpose, we have used the CBIR image database from Washington University (824 JPEG images totally) [8]. Since we can estimate the histogram of the cover image, message embedding in sequential and random ways can be detected and estimated in the way described in Section 3.2. In our experiments, we have generated 8 stego images (4 for sequential embedding and 4 for random) for each image and the length of hidden messages are 20, 50, 80 and 100 percentage of hiding capacity, corresponding to $2\alpha = 0.2, 0.5, 0.8, 1.0$. For the convenience of display, only parts of the results of the experiments are show in Figure 2. ‘△’, ‘*’, ‘o’, ‘+’ and ‘☆’ represent the estimated percentages of message
capacity, corresponding to $2\alpha = 1.0, 0.8, 0.5, 0.2, 0$. Table 1 shows the mean $\mu(2\alpha)/\mu(\beta)$ and the standard deviation $\sigma(2\alpha)/\sigma(2\alpha)$ for each embedding ratio over the whole 824 stego images. From Figure 2 and Table 1, we can see the experimental results are satisfied. And we also can see the standard deviation increases as embedding ratio decreases. This is because less message embedded into the cover image cause little change of DCT coefficients. We should model DCT coefficients more precisely.

4. Verification of Zhang et al.'s Method

As we introduced in Section 1, Zhang et al. proposed a steganalytic technique to estimate the hidden message length using J-Steg like algorithms. Our analysis presented in the preceding section can prove their estimation formula.

From equation (4) and (5), we can derive four equations as follows.

$$\sum_{i>0} H(2i) = (1 - \alpha) \sum_{i>0} h(2i) + \alpha \sum_{i>0} h(2i + 1)$$  \hspace{1cm} (13)

$$\sum_{i>0} H(2i + 1) = (1 - \alpha) \sum_{i>0} h(2i + 1) + \alpha \sum_{i>0} h(2i)$$  \hspace{1cm} (14)

$$\sum_{i>0} H(2i) = (1 - \alpha) \sum_{i<0} h(2i) + \alpha \sum_{i<0} h(2i + 1)$$  \hspace{1cm} (15)

$$\sum_{i<0} H(2i + 1) = (1 - \alpha) \sum_{i<0} h(2i + 1) + \alpha \sum_{i<0} h(2i)$$  \hspace{1cm} (16)

Using the symmetric property of the histogram $h(i)$, we combine (13) and (16) to obtain equation (17), and combine (14) and (15) to obtain equation (18).

$$\sum_{i>0} H(2i) + \sum_{i<0} H(2i + 1) = \sum_{i>0} h(2i) + \sum_{i<0} h(2i + 1) - \alpha h(-1)$$  \hspace{1cm} (17)

$$\sum_{i>0} H(2i + 1) + \sum_{i<0} H(2i) = \sum_{i>0} h(2i + 1) + \sum_{i<0} h(2i) + \alpha h(-1)$$  \hspace{1cm} (18)

Let $f_0 = \sum_{i<0} H(2i)$ and $f_1 = \sum_{i>0} H(2i + 1)$.

We have $f_1 - f_0 = 2\alpha h(-1) = 2\alpha H(1)$, let $\beta = 2\alpha$, we obtain $\beta = (f_1 - f_0)/H(1)$. This is Zhang's estimation formula.

5. Conclusion

We have presented a new approach to steganalysis. The approach, based on statistical model, provides a framework for steganalysis. We have shown that the method can not only detect the existence of secret message, but also estimate the length of secret message. We also offered a proof of Zhang et al.'s method. The methodology described in this paper can also be a framework to be applied to many other steganographic methods such as JPEG2000 file format embedding.

Acknowledgments

The work is sponsored by the Natural Sciences Foundation of China under grants No. 60172054, No. 60121302 and No. 60335010.

6. References


<table>
<thead>
<tr>
<th>$2\alpha / \beta$</th>
<th>For Sequential</th>
<th>For Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu(\beta)$</td>
<td>$\sigma(\beta)$</td>
<td>$\mu(2\alpha)$</td>
</tr>
<tr>
<td>0</td>
<td>0.0748</td>
<td>0.0159</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2562</td>
<td>0.0105</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5365</td>
<td>0.0042</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8160</td>
<td>0.00073</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0009</td>
<td>0.00013</td>
</tr>
</tbody>
</table>