MODEL BASED STEGANALYSIS

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

In this paper, we consider a new method for performing steganalysis using a statistical model of the cover medium. Using model based methodology, examples of detecting secret message and estimating the secret message length of bit-streams embedded using JSteg-like steganography and quantization index modulation are proposed. This steganalysis technique is based on the model of statistical distribution of quantized DCT coefficients. The histogram of cover image and "shrinkage histogram" are estimated from stego image using the statistical model. Then the secret message is detected and the secret message length is estimated. The methodology described in this paper is a framework which can also be applied to virtually any type of media 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. Given the proliferation of steganography tools, there is a growing interest in steganalysis tools, which detect hidden data in multimedia. So far, steganalysis research has lagged behind steganography. Recently, considerable research has been done on steganalysis. Johnson and Jajodia made a careful analysis of signatures introduced by current steganographic softwares [9]. Fridrich developed serial steganalysis methods and a good survey of steganalysis [10]. Universal blind steganalysis include Memon's approach and Farid's approach [6,11]. 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. To solve these problems, we turn to a new methodology based on statistical modeling. We start a systematic study of model-based approach for steganalysis and develop practical schemes based on the statistical modeling. We focus our attention on JSteg-like steganography [1] and quantization index modulation (QIM) [12]. However, in principle our ideas are applicable to steganalysis of any steganography scheme with a statistical description.

The general method description is introduced in Section 2 and JSteg-like steganography and QIM are introduced in Section 3. In Section 4 and 5, we apply the method to steganalysis JSteg-like steganography and QIM. The experimental results and the conclusions are in the last 2 sections of the paper.

2. GENERAL METHOD DESCRIPTION

As described in [7], if we have a model which captures 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 hidden message embedded in a suspected media is to perform a hypothesis test to find out whether the instance x obeys the probability distribution $P_X(x)$. But how to estimate the length of hidden message? Fridrich[4] summarized the principles to estimate the length of secret message. For most steganographic techniques, We identify a macroscopic quantity $S(m)$ that predictably changes (for example, monotonically increases) with the length of the embedded secret message m. Then we can calculate an estimate of the unknown message length m by solving the equation $S(m) = S_{stego}$ for m, where S_{stego} is the value of S for the stego image under investigation. $S(m)$ is called the distinguishing statistics. 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). In this paper, we apply this principle to estimate the length of secret message. If we can model x precisely, our method can obtain more precise results than Fridrich's method, since we need not crop the stego image to approximate the macroscopic properties of the cover image. We will show how our

0-7803-8554-3/04/$20.00 ©2004 IEEE.

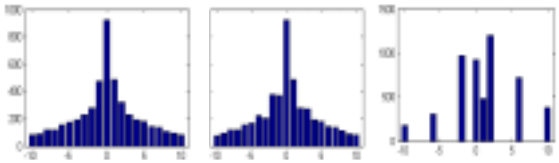
proposed method works on the estimation of the secret message length in JSteg-like steganography and QIM.

3. JSTEG-LIKE STEGANOGRAPHY & QUANTIZATION INDEX MODULATION

The frequently used steganographic method in JPEG format is the JSteg-like algorithm, which is 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. QIM [12] is another useful steganographic method. A simple implementation of QIM using block DCT may be obtained like this: Selected coefficients are taken from JPEG quantized DCT blocks and quantization-index-modulated. Just like JSteg, the embedding process skips all coefficients with the values of '0' or '1'. Although there exist many kinds of QIM method, in this paper, we only use this simple method as an example, because it is simple and representative.

Figure 1 shows the DCT coefficients (3,3) histograms of a test image. Figure 1(a) is calculated from the original image, (b) and (c) from the JSteg and QIM steganography version respectively. From the histograms, we can see data hiding in them will cause histogram changes in image which can be used to steganalysis.

During the last few years, many powerful steganalytic methods [2] capable of detecting JSteg-like embedding were proposed. These methods can't estimate the number of changes due to embedding and thus can't estimate the secret message length. Zhang proposed a method [5] capable of detecting and removing the hidden message from QIM embedded image. In the next 2 sections, we will show the process of model-based steganalysis to detecte the secret message and estimate the length in JSteg-like steganography and QIM.



(a) (b) (c)

Figure 1. (a)Cover, (b)JSteg, (c)QIM histogram

4. STATISTICAL MODEL

It is generally believed that the distribution of 8×8 blocks DCT coefficients is Laplacian except for the [0,0] coefficients [3]. If the coefficients are quantized, a generalized Laplacian can still be used to fit the resulting histogram with integer width bins. Sallee[7] used a specialized form of a generalized Cauchy distribution instead of the generalized Laplacian:

$$p(x)=\frac{p-1}{2s}(|\frac{x}{s}|-1)^{-p} \quad (1)$$

When taking into account a more accurate estimation of the quantization effects, Sallee [7] found 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? Let *h(d)*, *d=...-2,-1,0,1,2,...,* be the histogram of the quantized DCT coefficients from the cover image. Let *H(d)* be the histogram of stego image with *m* pseudo-random bits embedded. We define the embedding ratio

$$2\alpha = m \Big/ \sum_{i\neq 0, i\neq 1} h(i)$$

4.1. Statistical model of JSteg-like steganography

After we embedded *m* pseudo-random bits in the LSBs of quantized DCT coefficients, the histograms *H(d)* and *h(d)* will have relations as the following equations.

$$H(0)=h(0),\quad H(1)=h(1) \quad (2)$$

$$H(2i)=h(2i)-\alpha\left[h(2i)-h(2i+1)\right] \quad (3)$$

$$H(2i+1)=h(2i+1)+\alpha\left[h(2i)-h(2i+1)\right] \quad (4)$$

where $i=\pm 1,\pm 2,...$. From (3) and (4), we obtain *H(2i)+H(2i+1)=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 (1). 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 by maximizing the likelihood $P(h\mid s,p)$ [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.

$$\hat{h}(0)=H(0),\quad \hat{h}(1)=H(1) \quad (5)$$

$$\hat{h}(2i)=H_\alpha(i)(P(2i)/P_\alpha(i)) \quad (6)$$

$$\hat{h}(2i+1)=H_\alpha(i)(P(2i+1)/P_\alpha(i)) \quad (7)$$

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

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

Figure 2 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.

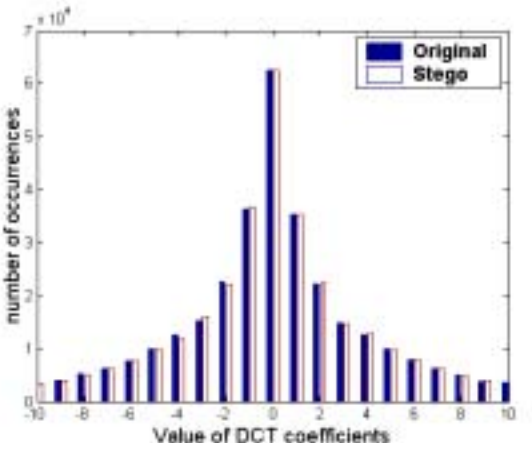


Figure 2. A comparison of coefficient histograms between original histogram of cover image and estimated histogram of stego image

4.2. Statistical model of quantization index modulation

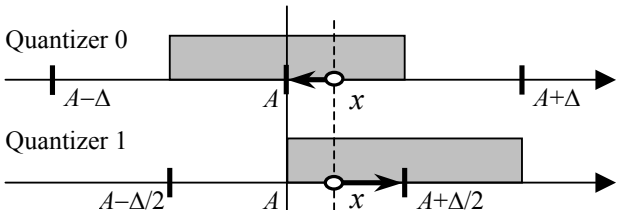


Figure 3. Coefficients transition after QIM

With data embedded using QIM, the coefficients transition can be described in figure 3. In the figure, thick vertical bars indicate quantized values after embedding. With quantizer 0, for example, if a QIM modified coefficient is A, the original coefficient value $x \in [A-\triangle/2, A+\triangle/2]$. With quantizer 1, x is quantized to $A+\triangle/2$. After we embedded m pseudo-random bits using QIM, the histograms $H(d)$ and $h(d)$ will have relations as follows.

$$H(x) = h(x)(1-2\alpha) \tag{8}$$

$$H(A+k\Delta/2) = h(A+k\Delta/2) + \alpha\sum(h(x)+h(y)) + \alpha(h(A+(k-1)\Delta/2) - h(A+(k-1)\Delta/2)) \tag{9}$$

where $A+k\Delta/2 < x < A+(k+1)\Delta/2$

$A+(k-1)\Delta/2 < y < A+k\Delta/2$

The relations between the cover histogram and stego histogram are complex. It is difficult to estimate the cover histogram from stego histogram. But the scaling relation of equation (8) is simple. If we only use $H(x)$ to fit the model (1), we obtain a "shrinkage histogram" which is in proportion to the cover histogram. The scale parameter is $1-2\alpha$. From the estimated "shrinkage histogram" we can estimate the secret message length, which we will show in the next section.

To estimate the "shrinkage histogram", we have to estimate the quantization step. This parameter can be obtained by filtering the stego histogram.

5. MODEL BASED STEGANALYSIS

5.1 Steganalysis of JSteg-like steganography

The detection is determined using the Chi-square test, which is $\chi^2 = \sum_{i=\pm1}^{\pm k} \frac{(\hat{h}(i)-H(i))^2}{\hat{h}(i)}$ with 2k-1 degree of freedom. We can perform Chi-square test at significance level α and 2k-1 degree of freedom to decide whether a suspect images contains secret message or not. Because of the limited space, we only show the estimation of the length of secret message using random embedding method. Equation (4) and (5) express the distinguishing statistics as a function of secret message m and the cover histogram. The cover histogram can be estimated using the method described in Section 4. We calculate α using the following equation which is derived from (4).

$$\alpha = \frac{\sum_{i=\pm1}^{\pm k}(H(2i)-\hat{h}(2i))(\hat{h}(2i)-\hat{h}(2i+1))}{\sum_{i=\pm1}^{\pm k}(\hat{h}(2i)-\hat{h}(2i+1))^2} \tag{10}$$

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

$$M = 2\alpha \sum_{i\neq0, i\neq1} H(i) \tag{11}$$

5.2 Steganalysis of quantization index modulation

From Figure 1(c), we can see the discreteness of histogram. It is easy to detect the artifacts using visual inspection of the histogram. The detection can also be determined by estimating the quantization step, which is half period of FFT of the histogram[5]. For estimation of secret message length, we first estimate the "shrinkage histogram" which is proportional to original histogram. From the "shrinkage histogram", $\tilde{h}(0)$ and $\tilde{h}(1)$ can be obtained. In section 3, we know that the embedding process skips all coefficients with the values of '0' or '1',

which cause *H(0)* or *H(1)* keeps unchanged after message embedding. So Equation (8) expresses the distinguishing statistics as a function of secret message *m* and the "shrinkage histogram". We calculate α using the following equation which is derived from (8).

$$\alpha = (1 - \tilde{h}(m) / H(m)) / 2 \qquad m = 0,1 \qquad (12)$$

The length of secret message can be calculated using (11).

6. 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 JSteg-like steganography and QIM can be detected and estimated in the way described in Section 5. In our experiments, we have generated 8 stego images (4 for JSteg and 4 for QIM) for each image and the length of hidden messages are 20, 50, 80 and 100 percentage of hiding capacity, corresponding to 2α =0.2, 0.5, 0.8, 1.0. For the convenience of displaying, only parts of steganalysis results of JSteg-like steganography are shown in Figure 4. ' △ ', '*','o', '+' and ' ☆ ' represent the estimated percentages of message capacity, corresponding to 2α =1.0, 0.8, 0.5, 0.2, 0. Table 1 shows the mean $\mu(2\alpha)$ and the standard deviation $\sigma(2\alpha)$ for each embedding ratio over the whole 824 stego images. From Figure 4 and Table 1, we can see the experimental results are satisfactory.

Table 1. Statistical Results

$2\alpha / \beta$	For JSteg		For QIM	
	$\mu(2\alpha)$	$\mu(2\alpha)$	$\mu(2\alpha)$	$\sigma(2\alpha)$
0	0.0638	0.0237	0.0038	0.0159
0.2	0.2352	0.0240	0.1680	0.5982
0.5	0.5003	0.0212	0.4919	1.0516
0.8	0.7948	0.0279	0.7995	0.0124
1.0	1.0004	0.0131	1.0000	0.0000

7. CONCLUSIONS

We have presented a new approach to steganalysis. The approach, based on statistical model, provides a framework for steganalysis. It has been demonstrated that the presence of JSteg-like steganography and QIM embedded message can be detected based on analyses of coefficient histograms. We have shown that the method can not only detect the existence of secret message, but also estimate the length of secret message. Our approach is not limited to JSteg-like steganography or QIM, the idea is suitable for any hiding scheme with a good statistical description.

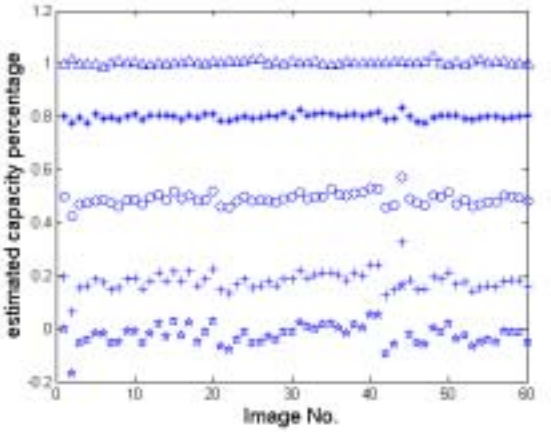


Figure 4. Estimates of embedding ratio on part images from ICBR image database

8. REFERENCES

[1] JPEG-JSteg-V4, http://www.funet.fi/pub/crypt/steganography/jpeg-jsteg-v4.diff.gz.

[2] A.Westfeld and A. Pfitzmann. "Attacks on Steganographic Systems". *IHW 99*, Dresden, Germany, 1999.

[3] S. R. Smoot and L. A. Rowe "Study of DCT coefficient distributions". In *Proceedings of the SPIE Symposium on Electronic Imaging,* volume 2657, San Jose, CA, January 1996.

[4] J. Fridrich, M. Goljan, D. Hogea, and D. Soukal, "Quantitative Steganalysis of Digital Images: Estimating the Secret Message Length", *ACM Multimedia Systems Journal, Special issue on Multimedia Security*, pp. 288-302, 2003.

[5] K. Zhang, S. Wang, X. Zhang, "Detection and Removal of Hidden Data in Images Embedded with Quantization Index Modulation", *MMM-ACNS*,pp.360-370,2003

[6] H. Farid. "Detecting Steganographic Message in Digital Images". *Report TR2001-412*, Dartmouth College, Hanover, NH, 2001

[7] P. Sallee, "Model Based Steganography", In *Proceedings of International Workshop on Digital Watermarking 03*, Seoul, Korea, September, 2003.

[8] CBIR Image Database, University of Washington, http://ww.cs.washington.edu/research/magedatabase/ roundtruth/.

[9] N. F. Johnson, "Sushil Jajodia. Steganalysis of Images Created Using Current Steganography Software". *LNCS Vol.1525*, Springer- Verlag, pp273-289, 1998.

[10] J. Fridrich and M. Goljan. "Practical Steganalysis: State of the Art". In Proceeding of SPIE Vol. 4675, *EI2002*, pp.1-13, January 2002.

[11] N.D.Memon, I.Avcibas, B.Sankur. "Steganalysis Based on Image Quality Metrics". In *Proceeding of SPIE* Vol. 4314, San Jose, California, USA. , 2001

[12] B. Chen, and G. Wornell, "Quantization Index Modulation: a Class of Provably Good Method for Watermarking and Information Embedding", *IEEE Transactions on Information Theory*, 47(4), pp.1423-1443, 2001.