Reversible Steganography: Data Hiding for Covert Storage

Zhuo Zhang and Weiming Zhang
CAS Key Laboratory of Electro-magnetic Space Information,
University of Science and Technology of China, Hefei , China
E-mail: zhang589@mail.ustc.edu.cn, zhangwm@mail.ustc.edu.cn

***Abstract*—Multimedia based covert storage requires data hiding methods having undetectability and reversibility at the same time, which is not supported by traditional steganography or reversible data hiding methods. To meet the covert storage needs, we present the concept “reversible steganography” in this paper, which has reversibility and moderate undetectability. We proposed a reversible steganographic method based on histogram shifting on prediction errors (PE). This novel method gives priority to PEs with large absolute values for accommodating messages, and thus the modifications will be concentrated in complex areas of the images. To enlarge capacity, a multi-rounds method is used to embed message into some successive bins of PE histogram and a bin selection method is proposed to minimize the embedding distortion. Experimental results show that the proposed method can significantly improve the undetectability of the method in [19].**

I. INTRODUCTION

Nowadays, a large scale of personal data, such as personal images and videos, are stored in the cloud. However the data is under the threaten of some extern invaders such hackers. For example, Google leaked a large number of files in 2009, which brings serious concerns about storage security. Therefore how to protect personal data becomes a serious problem. At present, the most popular technique for protecting privacy is encryption. However, encryption exposes the importance and sensitivity of the data and makes data more likely to attract hackers’ attention. In this context, covert storage, which aims to hide the existence of data, can provide a higher level of protection for data and has been widespread concerned.

There exist two different implementations for covert storage. One is based on the disk on the computer to create hidden space [1], and the other one embeds data into multimedia files (such as images) using specific methods such as data hiding. For covert storage in images, current research is mainly about how to construct the covert file system [3] or how to achieve covert collaboration [4] in the image contents by using steganography.

Usually, steganography is used as a tool for covert communication by embedding the secret data into the cover files such as images, audio and video. In this paper, we only discuss image steganography. For security, steganography should have the ability to resist steganalysis whose purpose is to detect whether a image is modified by steganography. State of the art steganalysis is to extract feature from the image according to the correlation of adjacent pixels and then to train classifier with machine learning [8][9]. To improve security, most recent steganographic methods try to embed data by modifying the complex regions of images that have weak correlation and are difficult to be modeled [7][10].

When using image steganography, such as those in [7][10], for covert storage, we can get high undetectability. However, different from covert communication, the image here is used as a special kind of storage medium that needs to be erasable as traditional storage medium (e.g. disk). In other words, after the stored data being deleted, the storage medium can be restored to its original state. To make the image “erasable”, the data hiding method must be reversible and thus the image can be used repeatedly. From this point, reversible data hiding (RDH), which is another branch of data hiding, is suitable for covert storage. By RDH, the cover image can be losslessly restored after the message being extracted.

So far, many RDH methods on images have been proposed. All these methods are realized through a process of semantic lossless compression [12], [13], in which some space is saved for embedding extra data by losslessly compressing the image. This compressed image should be “close” to the original image, so one can get a marked image with good visual quality. The residual part of images, e.g., the prediction errors (PE), has small entropy and thus can be easily compressed. Therefore, almost all recent RDH methods first generate PEs as the host sequence [14], [15], [16], and then reversibly embed the message into the host sequence by modifying its histogram with methods like histogram shifting (HS) [17] or difference expansion (DE) [18]. Usually, a more accurate prediction technique can generate PEs with a sharper histogram that is more suitable for RDH. Because the pixels in smooth areas can be accurately predicted, so most RDH give priority of modifications to PEs in smooth regions. That’s why traditional RDH cannot resist steganalysis. In fact, traditional RDH mainly used for annotation and authentication in medical images, military images or law forensics.

From above analysis, we can see that traditional steganography try to achieve high undetectability by modifying the complex regions of images but cannot realize reversibility. On the other hand, traditional RDH prefer to modify the smooth regions to obtain larger capacity with reversibility. However, data hiding methods for covert storage require two properties, undetectability and reversibility, at the same time. We call this kind of data hiding method “reversible steganography”, which belongs to adaptive reversible data hiding. Recently, Hong et al. [19] proposed a method of reversible steganogrpahy which can reach higher undetectability than traditional RDH methods. In Hong et al.'s method, the message bits are embedded into PEs with small absolute values, but PEs in complex regions are preferentially modified by a sorting technique.

In the present paper, we improve the method [19] by giving priority to PEs with large absolute values for accommodating messages because PEs with large amplitude usually belong to complex areas. To enlarge capacity, an iterative histogram shifting (HS) method is used to embed message into some successive bins of PE histogram and a bin selection method is proposed to choose bins having smallest distortion. Experimental results show that the proposed method can significantly improve the undetectability of the method in [19].

The rest of this paper is organized as follows. Section 2 describes the proposed method, and experimental results are shown in Section 3, and the paper is concluded in Section 4.

II. Proposed Method

A. Embedding and Extracting

We assume that the cover X is a gray-scale image with pixel $X(i,j) \in [0,255]$, $1 \leq i \leq M, 1 \leq j \leq N$. The proposed method embeds messages by shifting the histogram of PEs. To generate PEs, we randomly select $p\%$, denoted by $\mathbf{I}$, from X according to the encryption key k and calculate their prediction values from the surrounding $1-p\%$ pixels, denoted by $\mathbf{E}$. The estimation of the pixel $X(i,j) \in \mathbf{I}$ is calculated by

$$X'(i,j) = \left\lfloor \frac{\sum\limits_{h=-1}^{1}\sum\limits_{w=-1}^{1} X(i+h,j+w)\mathrm{S}(X(i+h,j+w))}{\sum\limits_{h=-1}^{1}\sum\limits_{w=-1}^{1} \mathrm{S}(X(i+h,j+w))} \right\rfloor \tag{1}$$

where $\mathrm{S}(*)$ represents a sign function defined as

$$\mathrm{S}(X(i,j)) = \begin{cases} 1 & X(i,j) \in \mathbf{E} \\ 0 & X(i,j) \in \mathbf{I} \end{cases} \tag{2}$$

And then, the PE of the pixel $X(i,j) \in \mathbf{I}$ is obtained by

$$e(i,j) = X(i,j) - X'(i,j) \tag{3}$$

Note that, when calculating PEs, the pixel $X(i,j) \in \mathbf{I}$ without surrounding pixels in $\mathbf{E}$ will be skipped. Without loss of generality, we assume all the N pixels in $\mathbf{I}$ can generate PEs, denoted by $\{e_1, \ldots, e_N\}$. An attacker cannot generate these PEs because the pixel set $\mathbf{I}$ is randomly selected by a encryption algorithm according the key k.

As shown in Fig. 1, PEs follow a Laplacian distribution, which is symmetrical. Therefore, without loss of generality, we assume the length of message is $2L$ bits, and embed L bits into the non-negative PEs and the other L bits into the negative PEs in the same manner. Thus, we only describe the embedding process in non-negative PEs. Usually, pixels in smooth regions can be accurately predicted and thus have PEs with smaller magnitude. Therefore modifications on PEs with bigger magnitude are hard to be detected by steganalysis because the corresponding pixels belong to complex areas.

However, the larger PEs are much less than smaller PEs. In order to achieve high payload, we introduce the multiple rounds embedding mechanism. We first select the bin b_{max} as a begin point, and then select several continuous R bins of PEs as carriers. We arrange these bins in descending order such that $\{b_{max}, b_{max}-1, \cdots, b_{max}-R\}$, in which $R+1$ rounds of embedding will be executed. In the ith round, we embed message bits into the bin $b_{max}-i+1$ for $1 \leq i \leq R+1$ Herein, R is determined by the length of message L, such that

$$R = min\{u| \sum_{i=0}^{u} h(b_{max}-i) \geq L\} \tag{4}$$

where $h(b_{max}-i)$ denotes the height of the bin $b_{max}-i$, that is, the number of PEs equal to $b_{max}-i$.

In first round, we scan all the PEs, e_i for $1 \leq i \leq N$, and embed the message bits into the bin b_{max} in following manners:

$$e_i^{(1)} = \begin{cases} e_i + 1 & e_i > b_{max} \\ e_i + m & e_i = b_{max} \\ e_i & e_i < b_{max} \end{cases} \tag{5}$$

where $m \in \{0,1\}$ is the message bit to be embedded. In the rth ($2 \leq r \leq (R+1)$)round, the message bits are embedded into the bin $b_{max}-(r-1)$ as follows:

$$e_i^{(r)} = \begin{cases} e_i^{(r-1)} + 1 & e_i^{(k-1)} > b_{max} - (r-1) \\ e_i^{(r-1)} + m & e_i^{(k-1)} = b_{max} - (r-1) \\ e_i^{(r-1)} & e_i^{(k-1)} < b_{max} - (r-1) \end{cases} \tag{6}$$

After the $R+1$ rounds of embedding, we can generate the stego image by modifying the pixel $X(i) \in \mathbf{I}$ as

$$Y(i) = X'(i) + e_i^{(R+1)}, 1 \leq i \leq N \tag{7}$$

For simplicity, herein, we use 1-dimension indexes to label the pixels, and $X'(i)$ is the predicted value of $X(i)$ obtained with Eq. (1).

As traditional RDH methods, the overflow/underflow may occurs in the embedding process, so a location map is needed to record the positions of overflow/underflow. The location map will be compressed and embedded as a part of the payload.

A pivotal problem is how to choose the beginning bin b_{max}. If b_{max} is too big, the complex regions will carry most of the data, but more PEs will be modified, and some PEs will be modified with a large amplitud. On the other hand, if b_{max} is too small, modifications will be concentrated in smooth areas.

Assume $b_{max} \in [B_{min}, B_{max}]$, and we choose b_{max} with the minimum embedding distortion for some properly defined distortion metric. Since we wish to restrict the embedding changes to those pixels in complex areas, the distortion is required to have the following properties :

1. The bigger the value of the $|PE|$, the smaller the distortion should be.

2. The bigger the changing amplitude of $|PE|$ after R+1 rounds of embedding, the bigger the distortion should be.

To meet both requirements above, when modifying the PE b with a amplitude v after the R +1 rounds of embedding we defined the distortion as

$$d(b, v) = v^2/(|b| + 1) \tag{8}$$

When b_{max} is given, the corresponding R and the target bins $\{b_{max}, b_{max} - 1, \cdots, b_{max} - R\}$ are also determined by Eq. (4). And then we can calculate the total distortion caused by the $R + 1$ rounds of the iterative embedding with the metric (8). We use $D_{b_{max}}$ to denote the total distortion for embedding the L bits of message beginning from the bin b_{max}.

One way to get $D_{b_{max}}$ is to implement the R+1 round embedding above, which will cost much time. And we give a way to estimate the value of $D_{b_{max}}$ quickly. That is, since the message to be embedded usually is encrypted, we assume the number of bit "0" of message embedded into each bin is the same as bit "1", so we can randomly "shift" half PEs of each bin to simulate the embedding procedure, where the operation "shift" means PEs are added by 1. Besides, when R and the target bins are determined, after the R+1 round embedding, half PEs in bin $b_{max} - r, r \in [0, R]$ become $b_{max} + R - 2r + 1$ while others become $b_{max} + R - 2r$. So we can get the total distortion of PEs in bin $b_{max} - r$,denoted by $d_{b_{max}-r}$ with the metric (8):

$$d_{b_{max}-r} = \lfloor h(b_{max} - r)/2 \rfloor \times d(b_{max} - r, R - r) + \\ (h(b_{max} - r) - \lfloor h(b_{max} - r)/2 \rfloor) \times d(b_{max} - r, R - r + 1) \tag{9}$$

and $D_{b_{max}} = \sum_{r=0}^{R} d_{b_{max}-r}$.

For all possible beginning bin $b \in [B_{min}, B_{max}]$, we can get $\{D_{B_{min}}, D_{B_{min}+1}, D_{B_{min}+2}, \cdots, D_{B_{max}}\}$ by any way above, and choose the optimal beginning bin b_{max} such that $D_{b_{max}} = min\{D_{B_{min}}, D_{B_{min}+1}, D_{B_{min}+2}, \cdots, D_{B_{max}}\}$. From the experiment results, using the second way to estimate $D_{b_{max}}$ will save much time and almost get the same optimal b_{max}.

Note that b_{max}, R and the message length L should be embedded into the image as parameters for extraction. For instance, these parameter can be hidden into some pixels by LSB (Least Significant Bit) replacement, and these LSBs can be embedded as one part of the payload.

The image recovery and data extraction procedures are realized in a reverse manner of the embedding procedure. After getting the stego image Y, the receiver first extract the parameters (b_{max}, R, L) and determines the referential set $\mathbf{E}$ with the shared key k. With $\mathbf{E}$, the receiver calculate the prediction value $X'(i)$ of pixels in $\mathbf{I}$ and then generate the PEs such that

$$e_i^{(R+1)} = Y(i) - X'(i), 1 \leq i \leq N \tag{10}$$

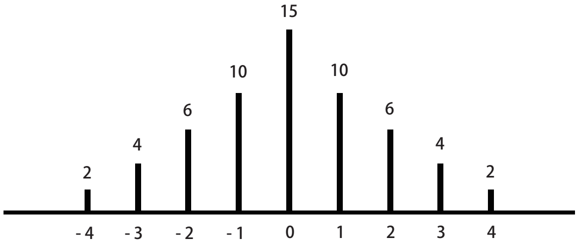

Fig. 1. Original States of the histogram before embedding

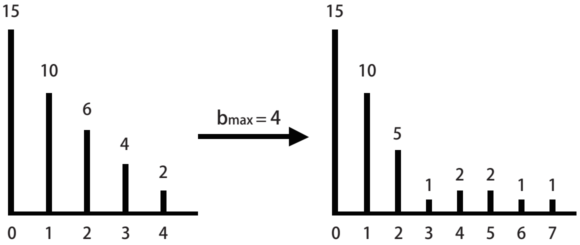

Fig. 2. States of the histogram before and after embedding with $b_{max} = 4$

Then the $R + 1$ rounds extraction is executed. In the rth round for $1 \leq r \leq R+1$, we take the threshold $T_r = b_{max} - R + r - 1$. If $e_i^{(R-r+2)} = T_r$, a bit 0 is extracted; and if $e_i^{(R-r+1)} = T_r$ a bit 1 is extracted. The PE $e_i^{(R-r)}$ is recovered by

$$e_i^{(R-r+1)} = \begin{cases} e_i^{(R-r+2)} - 1 & \texttt{if } e_i^{(R-r+2)} > T_r \\ e_i^{(R-r+2)} & \texttt{if } e_i^{(R-r+2)} \leq T_r \end{cases} \tag{11}$$

After $R+1$ rounds, the receiver can extract all the message bits and get $e_i^{(0)}$ that is just the original PE e_i. With the original values of PEs, the pixels $X(i) \in \mathbf{I}$ can be recovered by

$$X(i) = X'(i) + e_i. \tag{12}$$

B. A Simple Example

To illustrate the embedding procedure, a simple example is given in Fig. 1 and Fig. 2. The histogram of original PEs is shown in Fig. 1. There is 10 bits of message $\{0, 1, 0, 0, 1, 1, 0, 1, 0, 0\}$ to be embedded. The modification result for $b_{max} = 4$ is shown in Fig. 2. Table I records the modification results and distortion, from which we can calculate the total distortion $D_4 = 5.43$. In the same manner, we got distortions for setting $b_{max} = 3, 2, 1$ and 0 respectively by implementing the R+1 rounds embedding, and listed them in Table II, which shows that the minimal distortion is obtained by setting $b_{max} = 3$.

III. EXPERIMENTAL RESULTS

In this section, we compare the proposed method with Hong et al.'s method [19] under different embedding payload rate for resisting steganalysis. Herein, the embedding rate is just relative payload that is measured by bit per pixel (bpp). The performances are evaluated by steganalyzers SPAM[20] with ensemble classifiers[21].

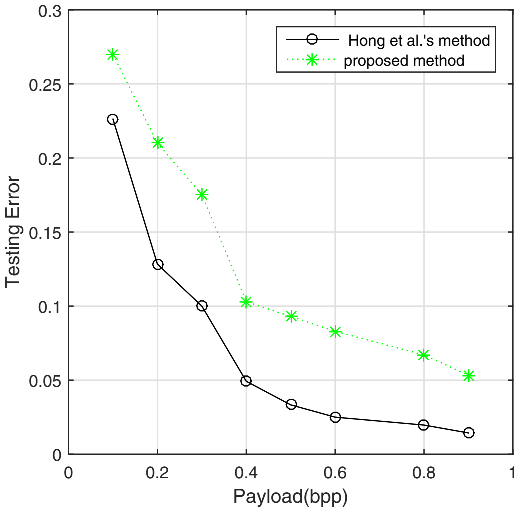

Fig. 3. Comparison between Hong et al.'s method [19] and the proposed method for resisting steganalyzer [20].

All experiments in this section are conducted on BOSSbase ver.1.01 database with an amount of 10,000 gray-scale images with size 512×512, in which 5,000 images are randomly selected for training, and the rest 5,000 images are used for testing. We report the testing error which computes the average of the false positive rate and false negative rate by 10 times of randomly splitting the training and the testing images. Higher detecting error rates means stronger security. It can be observed from Fig. 3, the proposed method significantly outperforms Hong et al.'s method for various kinds of embedding rates.

IV. CONCLUSIONS

In this paper, we present the conception "reversible steganography" for covert storage. Reversible steganography is a special kind of data hiding with reversibility and ability resisting steganalysis, which is necessary for the applications of covert storage.

TABLE I
THE MODIFICATION RESULTS AND DISTORION OF THE EXAMPLE SHOWN IN FIG. 2

original value	final value	count	distortion
4	7	1	$1 \times (4-7)^2 \div (4+1)$
4	6	1	$1 \times (4-6)^2 \div (4+1)$
3	5	2	$1 \times (3-5)^2 \div (3+1)$
3	4	2	$1 \times (3-4)^2 \div (3+1)$
2	3	1	$1 \times (2-3)^2 \div (2+1)$
2	2	3	0

TABLE II
DISTORTION CAUSED BY MULTIPLE ROUNDS EMBEDDING WITH DIFFERENT b_{max}

b_{max}	4	3	2	1	0
distortion	5.43	4.35	11.1	5.4	12.4

We proposed a multi-rounds embedding method for reversible steganography, which gives priority to PEs with larger amplitudes for accommodating data. The main idea behind the proposed method is that bigger PEs come from complex areas of images and modification in complex areas is hard to be detected by steganalysis. Experimental results show that the proposed method can offer higher undetectability and thus is more suitable for the scenario of reversible steganography.

REFERENCES

[1] Anderson R, Needham R, Shamir A. The steganographic file system[C] Information Hiding. Springer Berlin Heidelberg, 1998: 73-82.
[2] Pang H H, Tan K L, Zhou X. StegFS: A steganographic file system[C] Data Engineering, 2003. Proceedings. 19th International Conference on. IEEE, 2003: 657-667.
[3] Lach J. SBFS-steganography based file system[C] Information Technology, 2008. IT 2008. 1st International Conference on. IEEE, 2008: 1-4.
[4] Sosa C, Sutton B C, Huang H H. PicFS: The Privacy-Enhancing Image-Based Collaborative File System[C] Parallel and Distributed Systems (ICPADS), 2010 IEEE 16th International Conference on. IEEE, 2010: 99-106.
[5] Hong W, Chen T S. A novel data embedding method using adaptive pixel pair matching[J]. Information Forensics and Security, IEEE Transactions on, 2012, 7(1): 176-184.
[6] Mielikainen J. LSB matching revisited[J]. Signal Processing Letters, IEEE, 2006, 13(5): 285-287.
[7] Holub V, Fridrich J J. Designing steganographic distortion using directional filters[C]//WIFS. 2012: 234-239.
[8] Goljan M, Fridrich J, Cogranne R. Rich Model for Steganalysis of Color Images[C]//IEEE Workshop on Information Forensic and Security, Atlanta, GA. 2014.
[9] Denemark T, Sedighi V, Holub V, et al. Selection-Channel-Aware Rich Model for Steganalysis of Digital Images[C]//IEEE Workshop on Information Forensic and Security, Atlanta, GA. 2014.
[10] Holub V, Fridrich J, Denemark T. Universal distortion function for steganography in an arbitrary domain[J]. EURASIP Journal on Information Security, 2014, 2014(1): 1-13.
[11] Filler T, Fridrich J. Design of adaptive steganographic schemes for digital images[C]//ISδT/SPIE Electronic Imaging. International Society for Optics and Photonics, 2011: 78800F-78800F-14.
[12] F. Willems, D. Maas, and T. Kalker, "Semantic Lossless Source Coding," 42nd Annual Allerton Conference on Communication, Control and Computing, Monticello, Illinois, USA, pp. 1411-1418, 2004.
[13] W. Zhang, X. Hu, N. Yu, et al. "Recursive Histogram Modification: Establishing Equivalency Between Reversible data Hiding and Lossless Data Compression," IEEE Trans. on Image Processing. vol. 22, no. 7, pp. 2775-2785, Jul. 2013.
[14] D. Thodi and J. Rodriguez, "Expansion Embedding Techniques for Reversible Watermarking," IEEE Trans. Image Processing, vol. 16, no.3, pp. 721-730, Mar. 2007.
[15] B.ou, X. Li, Y. Zhao, R. Ni, Y. Shi, "Pairwise Prediction-Error Expansion for Efficient Reversible Data Hiding," IEEE Trans. on Image Processing, vol. 22, no.12, pp. 5010-5012, Dec. 2013.
[16] Ioan-Catalin Dragoi, Dinu Coltuc, "Local-Prediction-Based Difference Expansion Reversible Watermarking," IEEE Trans. on Image Processing, vol. 23, no. 4, pp. 1779-1790, Apr. 2014.
[17] Z. Ni, Y. Shi, N. Ansari, and S. Wei, "Reversible Data Hiding," IEEE Trans. Circuits Syst. Video Technol., vol. 16, no. 3, pp. 354-362, 2006.
[18] J. Tian, "Reversible Data Embedding Using a Difference Expansion," IEEE Trans. on Circuits System and Video Technology, vol. 13, no.8, pp. 890-896, Aug. 2003.
[19] Hong W, Chen T S, Chen J. Reversible data hiding using Delaunay triangulation and selective embedment[J]. Information Sciences, 2014.
[20] Pevny T, Bas P, Fridrich J. Steganalysis by subtractive pixel adjacency matrix[J]. information Forensics and Security, IEEE Transactions on, 2010, 5(2): 215-224.
[21] Kodovsky J, Fridrich J, Holub V. Ensemble classifiers for steganalysis of digital media[J]. Information Forensics and Security, IEEE Transactions on, 2012, 7(2): 432-444.