Reversible Data Hiding in Encrypted Image

Xinpeng Zhang

***Abstract*—This work proposes a novel reversible data hiding scheme for encrypted image. After encrypting the entire data of an uncompressed image by a stream cipher, the additional data can be embedded into the image by modifying a small proportion of encrypted data. With an encrypted image containing additional data, one may firstly decrypt it using the encryption key, and the decrypted version is similar to the original image. According to the data-hiding key, with the aid of spatial correlation in natural image, the embedded data can be successfully extracted and the original image can be perfectly recovered.**

***Index Terms*—Image encryption, image recovery, reversible data hiding.**

I. Introduction

REVERSIBLE data hiding is a technique to embed additional message into some distortion-unacceptable cover media, such as military or medical images, with a reversible manner so that the original cover content can be perfectly restored after extraction of the hidden message. A number of reversible data hiding methods have been proposed in recent years. In difference expansion method [1], differences between two adjacent pixels are doubled to generate a new least significant bit (LSB) plane for accommodating additional data. A data hider can also perform reversible data hiding using a histogram shift mechanism, which utilizes the zero and peak points of the histogram of an image and slightly modifies the pixel gray values to embed data into the image [2]. Another kind of method makes use of redundancy in a cover by performing lossless compression to create a spare space for data embedding [3]. Furthermore, various skills have been introduced into the typical reversible data hiding approaches to improve the performance [4]–[6].

As is well known, encryption is an effective and popular means of privacy protection. In order to securely share a secret image with other person, a content owner may encrypt the image before transmission. In some application scenarios, an inferior assistant or a channel administrator hopes to append some additional message, such as the origin information, image notation or authentication data, within the encrypted image though he does not know the original image content. For example, when medical images have been encrypted for protecting the patient privacy, a database administrator may aim to embed the personal information into the corresponding encrypted images. It may be also hopeful that the original content can be recovered without any error after decryption and retrieve of additional message at receiver side. That means a reversible data hiding scheme for encrypted image is desirable.

In some existing joint data-hiding and encryption schemes, a part of cover data is used to carry the additional message and the rest data are encrypted. For example [7], the intra-prediction mode, motion vector difference and signs of DCT coefficients are encrypted, while a watermark is embedded into the amplitudes of DCT coefficients. In [8], the cover data in higher and lower bit-planes of transform domain are respectively encrypted and watermarked. In [9], the content owner encrypts the signs of host DCT coefficients and each content-user uses a different key to decrypt only a subset of the coefficients, so that a series of versions containing different fingerprints are generated for the users. In these joint schemes, however, only a partial encryption is involved, leading to a leakage of partial information of the cover. Furthermore, the separation of original cover and embedded data from a watermarked version is not considered. In [10] and [11], each sample of a cover signal is encrypted by a public-key mechanism and a homomorphic property of encryption is exploited to embed some additional data into the encrypted signal. But the data amount of encrypted signal is significantly expanded and the computation complexity is high. Also, the data embedding is not reversible.

This work proposes a novel reversible data hiding scheme for encrypted image, which is made up of image encryption, data embedding and data-extraction/image-recovery phases. The data of original cover are entirely encrypted, and the additional message is embedded by modifying a part of encrypted data. At receiver side, with the aid of spatial correlation in natural image, the embedded data are successfully extracted while the original image is perfectly recovered.

II. Proposed Scheme

A sketch of the proposed scheme is given in Fig. 1. A content owner encrypts the original uncompressed image using an encryption key to produce an encrypted image, and then a data-hider embeds additional data into the encrypted image using a data-hiding key though he does not know the original content. With an encrypted image containing additional data, a receiver may firstly decrypt it using the encryption key, and the decrypted version is similar to the original image. According to the data-hiding key, he can further extract the embedded data and recover the original image from the decrypted version. The detailed procedures are as follows.

Manuscript received December 13, 2010; revised January 26, 2011; accepted February 06, 2011. Date of publication February 14, 2011; date of current version February 28, 2011. This work was supported by the Natural Science Foundation of China (61073190, 60872116, and 60832010), the Shanghai Rising-Star Program (10QH14011), and the Key Scientific Research Project of Shanghai Education Committee (10ZZ59). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. H. Vicky Zhao.

The author is with the School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China (e-mail: xzhang@shu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2011.2114651

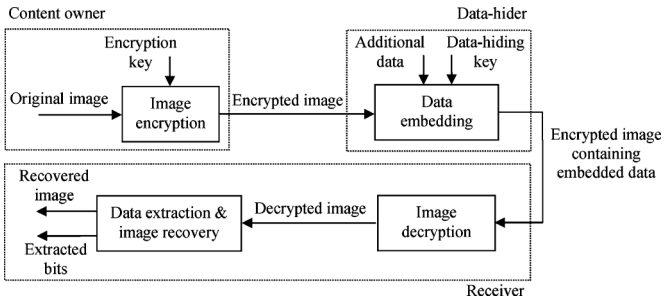

Fig. 1. Sketch of proposed scheme.

A. Image Encryption

Assume the original image is in uncompressed format and each pixel with gray value falling into [0, 255] is represented by 8 bits. Denote the bits of a pixel as $b_{i,j,0}, b_{i,j,1}, \ldots, b_{i,j,7}$ where (i, j) indicates the pixel position, and the gray value as $p_{i,j}$. Thus

$$b_{i,j,k} = \left\lfloor \frac{p_{i,j}}{2^k} \right\rfloor \mod 2, \quad k = 0, 1, \ldots, 7 \tag{1}$$

and

$$p_{i,j} = \sum_{u=0}^{7} b_{i,j,k} \cdot 2^k. \tag{2}$$

In encryption phase, the exclusive-or results of the original bits and pseudo-random bits are calculated

$$B_{i,j,k} = b_{i,j,k} \oplus r_{i,j,k} \tag{3}$$

where $r_{i,j,k}$ are determined by an encryption key using a standard stream cipher. Then, $B_{i,j,k}$ are concatenated orderly as the encrypted data. A number of secure stream cipher methods can be used here to ensure that anyone without the encryption key, such as a potential attacker or the data hider, cannot obtain any information about original content from the encrypted data.

B. Data Embedding

With the encrypted data, although a data-hider does not know the original image content, he can embed additional message into the image by modifying a small proportion of encrypted data. Firstly, the data-hider segments the encrypted image into a number of nonoverlapping blocks sized by $s \times s$. In other words, the encrypted bits $B_{i,j,k}$ satisfying $(m-1) \cdot s + 1 \le i \le m \cdot s$, $(n-1) \cdot s + 1 \le j \le n \cdot s$ and $0 \le k \le 7$ (m and n are positive integers) are within a same block. Then, each block will be used to carry one additional bit.

For each block, pseudo-randomly divide the s^2 pixels into two sets S_0 and S_1 according to a data-hiding key. Here, the probability that a pixel belongs to S_0 or S_1 is 1/2. If the additional bit to be embedded is 0, flip the 3 least significant bits (LSB) of each encrypted pixel in S_0,

$$B'_{i,j,k} = \overline{B_{i,j,k}}, \qquad (i, j) \in S_0 \text{ and } k = 0, 1, 2. \tag{4}$$

If the additional bit is 1, flip the 3 encrypted LSB of pixels in S_1,

$$B'_{i,j,k} = \overline{B_{i,j,k}}, \qquad (i, j) \in S_1 \text{ and } k = 0, 1, 2. \tag{5}$$

The other encrypted data are not changed.

C. Data Extraction and Image Recovery

When having an encrypted image containing embedded data, a receiver firstly generates $r_{i,j,k}$ according to the encryption key, and calculates the exclusive-or of the received data and $r_{i,j,k}$ to decrypt the image. We denote the decrypted bits as $b'_{i,j,k}$. Clearly, the original five most significant bits (MSB) are retrieved correctly. For a certain pixel, if the embedded bit in the block including the pixel is zero and the pixel belongs to S_1, or the embedded bit is 1 and the pixel belongs to S_0, the data-hiding does not affect any encrypted bits of the pixel. So, the three decrypted LSB must be same as the original LSB, implying that the decrypted gray value of the pixel is correct. On the other hand, if the embedded bit in the pixel's block is 0 and the pixel belongs to S_0, or the embedded bit is 1 and the pixel belongs to S_1, the decrypted LSB

$$\begin{aligned} b'_{i,j,k} &= r_{i,j,k} \oplus B'_{i,j,k} \\ &= r_{i,j,k} \oplus \overline{B_{i,j,k}} \\ &= r_{i,j,k} \oplus \overline{b_{i,j,k} \oplus r_{i,j,k}} \\ &= \overline{b_{i,j,k}}, \qquad k = 0, 1, 2. \end{aligned} \tag{6}$$

That means the three decrypted LSB must be different from the original LSB. In this case:

$$b'_{i,j,k} + b_{i,j,k} = 1, \qquad k = 0, 1, 2. \tag{7}$$

So, the sum of decimal values of three decrypted LSB and three original LSB must be seven. The average energy of errors between the decrypted and original gray values is

$$E_A = \frac{1}{8} \cdot \sum_{u=0}^{7} [u - (7-u)]^2 = 21. \tag{8}$$

As the probability of incorrect LSB-decryption is 1/2, when reconstructing an image using the decrypted data, the value of PSNR in the decrypted image is approximately

$$\text{PSNR} = 10 \cdot \log_{10} \frac{255^2}{\frac{E_A}{2}} = 37.9 \text{ dB}. \tag{9}$$

Then, the receiver will extract the embedded bits and recover the original content from the encrypted image. According to the data-hiding key, he may segment the decrypted image into blocks and divide the pixels in each block into two sets in a same way. For each decrypted block, the receiver flips all the three LSB of pixels in S_0 to form a new block, and flips all the three LSB of pixels in S_1 to form another new block. We denote the two new blocks as H_0 and H_1. There must be that either H_0 or H_1 is the original block, and another one is more seriously

Fig. 2. (a) Original Lena, (b) its encrypted version, and (c) a decrypted version containing embedded data.

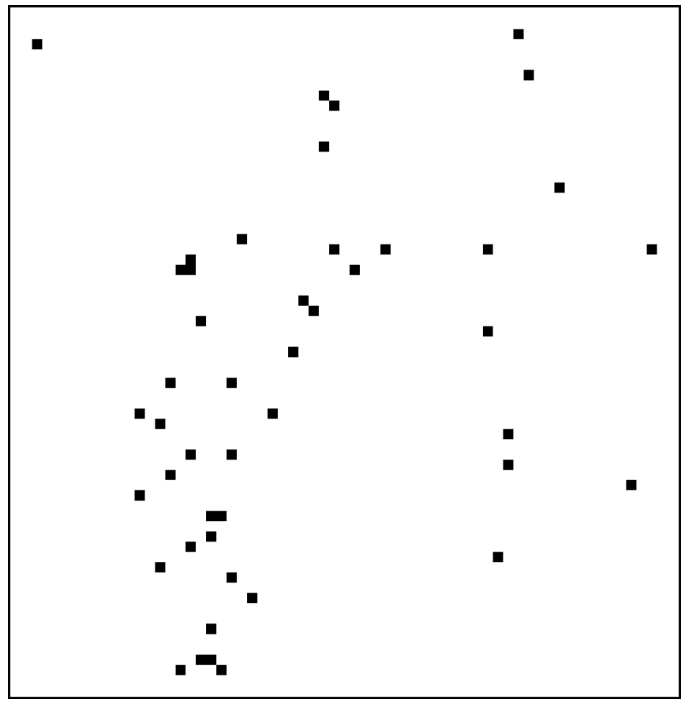

Fig. 3. Blocks of incorrect bit-extraction with the cover Lena and $s = 8$.

interfered due to the LSB flip operation. For the two blocks sized by $s \times s$, define a function to measure the fluctuation in them

$$f = \sum_{u=2}^{s-1} \sum_{v=2}^{s-1} \left| p_{u,v} - \frac{p_{u-1,v} + p_{u,v-1} + p_{u+1,v} + p_{u,v+1}}{4} \right| \tag{10}$$

and denote the values of fluctuation function of H_0 and H_1 as f_0 and f_1, respectively. Because of spatial correlation in natural image, the fluctuation function of original block is generally lower than that of a seriously interfered version. So, the receiver can perform data extraction and image recovery by comparing f_0 and f_1. If $f_0 < f_1$, regard H_0 as the original content of the block and let the extracted bit be 0. Otherwise, regard H_1 as the original content of this block and extract a bit 1. Finally, concatenate the extracted bits to retrieve the additional message and collect the recovered blocks to form the original image.

III. Experimental Results

The test image Lena sized 512×512 shown in Fig. 2(a) was used as the original cover in the experiment. After image encryption, the 8 encrypted bits of each pixel are converted into a gray value to generate an encrypted image shown in Fig. 2(b). Then, we embedded 256 bits into the encrypted image by using the side length of each block $s = 32$. The decrypted image is given as Fig. 2(c), and the values of PSNR caused by data embedding is 37.9 dB, which is imperceptible and verifies the theoretical analysis in (9). At last, the embedded data were successfully extracted and the original image was perfectly recovered from the decrypted image.

In the proposed scheme, the smaller the block size, the more additional data can be embedded. However, the risk of defeat of bit extraction and image recovery rises. Fig. 3 indicates the blocks of incorrect bit-extraction when the original cover Lena and $s = 8$ were used, and most of the blocks are in texture area due to the weak spatial correlation. Fig. 4 shows the extracted-bit error rate with respect to block sizes when four test images Lena, Man, Lake and Baboon sized 512×512 were used as the original covers. These covers are standard test images and freely available in many image databases. Here, the extracted-bit error rate is equivalent to the rate of unsuccessful block recovery. It can be seen that the smoother the cover image, the better is the performance of data extraction and image recovery. When the side length of block s is not less than 32, for most cover images, all the embedded bits can be correctly extracted and the original image can be successfully recovered.

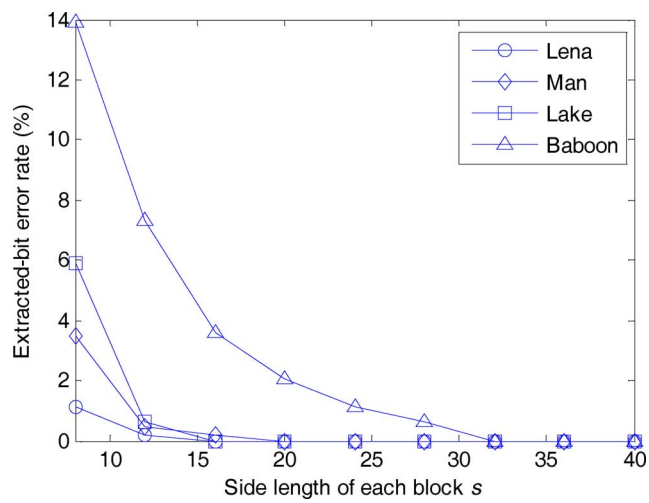

Fig. 4. Extracted-bit error rate with respect to block sizes.

IV. Conclusion and Discussion

In this work, a novel reversible data hiding scheme for encrypted image with a low computation complexity is proposed, which consists of image encryption, data embedding and data-extraction/image-recovery phases. The data of original image are entirely encrypted by a stream cipher. Although a data-hider does not know the original content, he can embed additional data into the encrypted image by modifying a part of encrypted data. With an encrypted image containing embedded data, a receiver may firstly decrypt it using the encryption key, and the decrypted version is similar to the original image. According to the data-hiding key, with the aid of spatial correlation in natural image, the embedded data can be correctly extracted while the original image can be perfectly recovered. Although someone with the knowledge of encryption key can obtain a decrypted image and detect the presence of hidden data using LSB-steganalytic methods, if he does not know the data-hiding key, it is still impossible to extract the additional data and recover the original image. For ensuring the correct data-extraction and the perfect image recovery, we may let the block side length be a big value, such as 32, or introduce error correction mechanism before data hiding to protect the additional data with a cost of payload reduction.

References

[1] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 8, pp. 890–896, Aug. 2003.

[2] Z. Ni, Y.-Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 3, pp. 354–362, 2006.

[3] M. U. Celik, G. Sharma, A. M. Tekalp, and E. Saber, "Lossless generalized-LSB data embedding," *IEEE Trans. Image Process.*, vol. 14, no. 2, pp. 253–266, Feb. 2005.

[4] L. Luo, Z. Chen, M. Chen, X. Zeng, and Z. Xiong, "Reversible image watermarking using interpolation technique," *IEEE Trans. Inf. Forensics Secur.*, vol. 5, no. 1, pp. 187–193, 2010.

[5] W. Hong, T.-S. Chen, Y.-P. Chang, and C.-W. Shiu, "A high capacity reversible data hiding scheme using orthogonal projection and prediction error modification," *Signal Process.*, vol. 90, pp. 2911–2922, 2010.

[6] C.-C. Chang, C.-C. Lin, and Y.-H. Chen, "Reversible data-embedding scheme using differences between original and predicted pixel values," *Inform. Secur.*, vol. 2, no. 2, pp. 35–46, 2008.

[7] S. Lian, Z. Liu, Z. Ren, and H. Wang, "Commutative encryption and watermarking in video compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 6, pp. 774–778, 2007.

[8] M. Cancellaro, F. Battisti, M. Carli, G. Boato, F. G. B. Natale, and A. Neri, "A commutative digital image watermarking and encryption method in the tree structured haar transform domain," *Signal Process.: Image Commun.*, DOI 10.1016/j.image.2010.11.001, to be published.

[9] D. Kundur and K. Karthik, "Video fingerprinting and encryption principles for digital rights management," *Proc. IEEE*, vol. 92, pp. 918–932, 2004.

[10] N. Memon and P. W. Wong, "A buyer-seller watermarking protocol," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 643–649, Apr. 2001.

[11] M. Kuribayashi and H. Tanaka, "Fingerprinting protocol for images based on additive homomorphic property," *IEEE Trans. Image Process.*, vol. 14, pp. 2129–2139, 2005.