Reversible Data Hiding in Encrypted Images

Chaple Gopal¹, G Balram²

¹M.Tech Student, Department of CSE, Anurag Group of Institutions, Hyderabad, India
²Assistant Professor, Department of CSE, Anurag Group of Institutions, Hyderabad, India

Abstract: In recent times, more attention is paid to reversible data hiding (RDH) in encrypted images, since it preserves the excellent property that the original cover can be losslessly improved after embedded data is extracted while protecting the image content’s privacy. All earlier methods embed data by reversibly vacating room from the encrypted images, which may be raise some errors on data extraction and image restoration. In this paper, we proposed a novel method by reserving room before encryption with a traditional RDH algorithm, and thus it is easier for the data hider to reversibly embed data in the encrypted image. Proposed method that can achieve the real reversibility, data extraction and image recovery are free of any error. This Experiments show that this novel method can embed more than 10 times as large payloads for the same image quality as the earlier methods, such as for PSNR dB.

Keywords: Reversible data hiding, image encryption, privacy protection, histogram shift

1. Introduction

Reversible data hiding (RDH) in images is a method, by which the original cover can be losslessly recovered after the embedded message is extracted. This technique is widely used in medical images, military imagery and law forensics, where no less quality of the original cover is allowed. Since the first introduced, RDH method has attracted considerable research interest.

In the theoretical part, Kalker and Willems [1] established a rate-distortion method of RDH, through which they proved the rate-distortion bounds of RDH for memory-less covers and proposed a recursive code construction which, however it does not approach the bound. Zhang et al. [2], [3] was improved the recursive code construction for binary covers and proved that this construction can achieve the rate-distortion bound as long as the compression algorithm reaches the entropy, that establishes the equivalence between the both data compression and RDH for binary covers.

In practical aspect, a lot of RDH techniques have emerged in recently. Fridrichet al. [4] constructed a general framework for RDH. In this first extracting compressible features of original cover and then compressing them lossless and spare space can be saved for embedding auxiliary data. A most popular method is based on difference expansion (DE) [5], in which the difference of each pixel group is expanded, e.g., is multiplied by 2, and thus the least significant bits (LSBs) of the difference are all-zero and can be used for embedding messages. Another hopeful strategy for RDH is histogram shift (HS) [6], in which space is saved for data embedding by shifting the bins of histogram of grey values. The state-of-art methods [7]–[11] usually combined the both DE and HS to residuals of the image, e.g., the expected errors, to achieve better performance.

With regard to providing confidentiality for images, encryption [12] is an effective and popular means as it converts the original and meaningful content to incomprehensible one. Al-though few RDH techniques in encrypted images have been published yet, we have some promising applications if RDH can be applied to encrypted images. In [13], Hwang et al. advocated a reputation-based trust-management scheme enhanced with data coloring (a way of embedding data into covers) and software watermarking, in that data encryption and coloring offer possibilities for upholding the content owner’s privacy and data integrity.

The cloud service provider has no right to introduce permanent distortion during data coloring into encrypted data. Hence, a reversible data coloring technique based on encrypted data is preferred. Suppose a medical image database is stored in a data centre, and a server in the data centre can embed notations into an encrypted version of a medical image through a RDH technique. With this, the server can manage the image or verify its integrity without having the knowledge of the original content, with that he patient’s privacy is protected. On the further hand, a doctor is having the cryptographic key can be decrypt and restore the image in a reversible manner for the purpose of further diagnosing.

Some efforts on RDH in encrypted images have been made. Zhang divided the encrypted image into several blocks by flipping the 3 LSBs of the half of pixels in each block and room can be vacated for the embedded bit. With that the data extraction and image recovery proceed by finding which part has been flipped in one block. This method can be realized with the help of spatial correlation in decrypted image. In Hong et al. [17] ameliorated Zhang’s method at the decoder side by further exploiting the spatial correlation using a different approximation equation and side match technique to achieve much lower error rate. These two methods are mentioned above rely on spatial correlation of original image to extract data. It remains the encrypted image should be decrypted first before data extraction.

To separate the data extraction from image decryption for data embedding following the idea of compressing encrypted images [14], [15]. The compression of encrypted data can be formulated as source coding with side information at the decoder [14], with that the typical method is to generate the compressed data in lossless manner by exploiting the patterns of parity- check matrix of channel codes. In the [18] method compressed the encrypted LSBs to vacate room for

Volume 3 Issue 9, September 2014
www.ijsr.net
Licensed Under Creative Commons Attribution CC BY
additional data by finding syndromes of a parity-check matrix and the side information used at the receiver side is also the spatial correlation of decrypted images.

In all the three methods try to vacate room from the encrypted images directly. However, the entropy of encrypted images has been maximized, with these techniques we can only achieve small payloads [16], [17] or generate marked image with poor quality for large payload [18] and all of them are subject to some error rates on data extraction and image restoration. Even though the methods in [16], [17] can eliminate errors by error correcting codes and also the pure payloads will be further consumed.

In the present paper, we propose a novel method for RDH in encrypted images, for which we do not “vacate room after encryption” as done in [16]–[18], but “reserve room before encryption”. In the recommended method, we first empty out room by embedding LSBs of some pixels into other pixels with a traditional RDH method and then encrypt the image, so that the positions of these LSBs in the encrypted image can be used to embed data. It was not only does the proposed method separate data extraction from image decryption but also achieves excellent performance in two different prospects:

- Real reversibility is realized that the data extraction and image recovery are free of any error.
- For particular embedding rates, the PSNRs of decrypted images are significantly improved; and for the acceptable PSNR, the range of embedding rates is greatly enlarged.

2. Previous Arts

The methods proposed in [16]–[18] can be summarized as the framework, Fig. 1(a) illustrates “vacating room after encryption (VRAE)”.

In this framework, a content owner encrypts the original image with an encryption key using a standard cipher. After producing the encrypted image, owner of the content hands over it to a data hider (e.g., a database manager) and the data hider can embed some auxiliary data into the encrypted image by losslessly vacating some room according to a data hiding key. After that a receiver, maybe the content owner himself or an authorized third party can extract the embedded data with the data hiding key and finally recover the original image from the decrypted version according to the encryption key.

In all methods of [16]–[18], the encrypted 8-bit grayscale images are generated by encrypting every bit-planes with a stream cipher. The method in [16] segments the encrypted image into a number of non-overlapping blocks sized by \( P \times L \) ; each block is used to carry one additional bit. For this, pixels in each block are pseudo-randomly divided into two sets \( S_1 \) and \( S_2 \) according to a data hiding key. If the additional bit 0 is embedded, flip the 3 LSBs of each encrypted pixel in \( S_1 \), otherwise flip the 3 encrypted LSBs of pixels in \( S_2 \).

For data extraction and image recovery the receiver will flips all the three LSBs of pixels in \( S_1 \) to form a new decrypted block. Also the receiver flips all the three LSBs of pixels in \( S_2 \) to form another new block; one of them will be decrypted to the original block. Because of spatial correlation in natural images, the original block is presumed to be much smoother than interfered block and embedded bit can be extracted correspondingly. But, there is a risk of defeat of bit extraction and image recovery when divided block is relatively small (e.g., \( n=8 \)) which has much fine-detailed textures.

Hong et al. [17] reduced the error rate of Zhang’s method [16] by fully exploiting the pixels in calculating the smoothness of each block and using side match. Side match is referred to as the extraction and recovery of blocks are performed according to the descending order of the absolute smoothness difference between two candidate blocks and recovered blocks can further be used to evaluate the smoothness of unrecovered blocks referred as side match.

Zhang’s method in [18] pseudo-randomly permuted and divided encrypted image into a number of groups with size \( L \). The each group of \( L \)-plane images are generated by encrypting every bit-planes with ethash consists of \( P \cdot L \) bits. The data hider generates a parity-check matrix \( G \) sized \(( P \cdot L - S) \times P \cdot L \), and compresses \( C \) as its syndrome \( S \) such that \( S = G \cdot C \). Because the length of it is \( ( P \cdot L - S) \), \( S \) bits are available for data accommodation.

At the receiver side, the \( 8 - P \) most significant bits (MSB) of pixels are obtained by decryption directly. The receiver then estimates \( \chi_i \) by full exploiting the pixels in calculating the cost of each group of \( C \) denoted by \( C^* \). On the other hand, the receiver tests each vector belonging to the cost of syndrome, where \( \Omega(s) = \{ u | G.u= s \} \). From each vector of \( \Omega(s) \), the receiver can get a restored version of \( c' \), and select the one most similar to the estimated version \( c^* \) as the restored LSBs.

3. Proposed Method

Since losslessly vacating room from the encrypted images is relatively difficult and sometimes inefficient, so it is obsessed to find novel RDH techniques working directly for encrypted images. In this RDH technique if we reverse the order of encryption and vacating room, which is reserving room prior to image encryption at content owner side “reserving room before encryption (RRBE)” framework used when the RDH tasks in encrypted images would be more natural and much easier.

As shown in Fig. 1(b), the content owner first reserve enough space on original image and then converts the image into its encrypted version with the encryption key. Then, the data embedding process in encrypted images is inherently reversible for the data hider only needs to accommodate data into the spare space previous emptied out. In which data extraction and image recovery are identical to that of Framework VRAE. Particularly, standard RDH algorithms are the ideal operator for reserving room be-fore encryption and can be easily applied to Framework RRBE to achieve
better performance compared with techniques from Framework VRAE. Because in this new framework, the customary idea that first losslessly compresses the redundant image content (e.g., using excellent RDH techniques) and then encrypts it with respect to protecting privacy.

Next, we elaborate a practical method based on the Framework “RRBE”, which primarily consists of four stages: 1) generation of encrypted image, 2) data hiding in encrypted image, 3) data extraction and 4) image recovery. Here, the reserving operation we adopt in the proposed method is a traditional RDH approach.

3.1 Generation of Encrypted Image

Actually, to construct the encrypted image, first stage can be divided into three steps: image partition and self-reversible embedding followed by image encryption. Beginning, image partition step divides original image into two partsAand B; then, the LSBs of A are reversibly embedded into B with a standard RDH algorithm so that LSBs of A can be used for accommodating messages; latter, encrypt the rearranged image to generate its final version.

1) Image Partition: The operator here for reserving before encryption is a standard RDH technique, so the main goal of image partition is to construct a smoother area B, on which standard RDH algorithms such to achieve better performance. To ensure that, without loss of generality to assume the original image is an 8 bits gray scale image with size MxN and ≤ pixels . Initially, the content owner extracts from the original image, along with the rows, several overlapping blocks whose number is determined by the size of embedded messages, represented by . In detail, each block consists of rows, where , and the number of blocks can be computed through . An important point here is that each block is overlapped by pervious and/or sub sequential blocks along the rows. Each block is define a function to measure its first-order smoothness

Higher relates to blocks which contain relatively more complex textures. The content owner, therefore, selects the particular block with the highest to be A, and puts it to the front of the image concatenated by the rest part B with fewer textured areas, as shown in Fig. 2.

2) Self-Reversible Embedding: The goal of self-reversible embedding is to embed the LSB-planes of A into B by employing traditional RDH algorithms. We simplify the method in [10] to demonstrate the process of self-embedding. This step does not rely on any specific RDH algorithm pixels in the rest of image B are first categorized into two sets: white pixels with its indices i and j satisfying (i + j) mod2 = 0 and black pixels whose indices meet (i + j) mod2 = 1, as shown in Fig. 2. Then, for the each white pixel, i,j which is estimated by the interpolation value obtained with the four black pixels surrounding it as follows

Where the weight is determined by the same method as proposed in [10]. The estimating error is calculated via and then some data can be embedded into the estimating error sequence with histogram shift, which will be defined later. Further we calculate the estimating errors of black pixels with the help of surrounding white pixels that may have been modified. The other estimating error sequence is generated which can accommodate messages as well. Also, we can implement multilayer embedding scheme by considering the modified B as “original” one when needed. In summary, to exploit all pixels of B, two estimating error sequences are constructed for embedding messages in every single-layer embedding process.

Through bidirectional histogram shift, some of the messages can be embedded on each error sequence. That remains, first divide the histogram of estimating errors into two parts left part and the right part and search for the highest point in each part, denoted by and respectively. For typical images, and . Furthermore, search for the zero point in each part, denoted by . To embed messages into positions with an estimating error that is equal to , shift all error values between and with one step toward right and then after we can represent the bit 0 with and the bit 1 with . The embedding process in the left part is similar except that the shifting direction is left and the shift is realized by subtracting 1 from the corresponding pixel values.

The same with other RDH algorithms is overflow or underflow problem occurs when natural boundary pixels change from 255 to 256 or from 0 to -1 . To avoid it, we only embed data into estimating error with its corresponding pixel valued from 1 to 254. Although ambiguities still arise when nonboundary pixels are changed from 1 to 0 or from
254 to 255 during the embedding process. These created boundary pixels in the embedding process are defined as pseudo-boundary pixels. Boundary map is introduced to tell whether boundary pixels in marked image are natural or pseudo in extracting process. It is a binary sequence with bit “0” for natural boundary pixel, bit “1” for pseudo boundary pixel. Later estimating errors of marginal area of $B$ cannot be calculated via (2), to make the best use of $B$ we choose its marginal area shown to place the boundary map we use LSB replacement to embed it and original LSBs of marginal area is assembled with messages, i.e., LSB-planes of $A$, and reversibly embedded into $B$. In most cases, even with a large embedding rate of the length of boundary map is very short the marginal area of $B$ is enough to accommodate it.

Numerous parameters such as $L$, $N$, $M$, $P$, payloads are embedded into the some estimating errors of black pixels $R$, total embedding rounds $S$, start row $S_R$ and end row $E_R$ of $A$ in original image is embedded into marginal area in a similar way these parameters play an important role in data extraction and image recovery process.

3) Image Encryption: After rearranged self-embedded image, denoted by $X$, is generated, we can encrypt $X$ to construct the encrypted image is denoted by $E$. With a stream cipher the encryption version of $X$ is easily obtained. For example, a gray value $X_{i,j}$ ranging from 0 to 255, Those 8 bits denoted by $X_{i,j}(0), X_{i,j}(1),..., X_{i,j}(7)$, such that

$$X_{i,j}(k) = \left[ \frac{X_{i,j}}{2^{k}} \right] \mod 2, \quad k = 0, 1, ..., 7. \quad (3)$$

The encrypted bits $E_{i,j}(k)$ calculated through exclusive-or operation

$$E_{i,j}(k) = X_{i,j}(k) \oplus r_{i,j}(k). \quad (4)$$

Where $r_{i,j}(k)$ is generated via a standard stream cipher determined by the encryption key. At last we embed 10 bits information into LSBs of first 10 pixels in encrypted version of $A$ to tell data hider the number of rows and the number of bit-planes. After image encryption, the data hider or a third party cannot access the content of original image without the encryption key, therefore privacy of the content owner being protected.

B. Data Hiding in Encrypted Image

Once the data hider acquires the encrypted image $E$ embed some data into it, though he does not get access to the original image. The embedding process is starts with locating the encrypted version of $A$, denoted by $A_{E}$. Since $A_{E}$ has been rearranged to the top of $E$, it is effortless for the data hider to read 10 bits information in LSBs of first 10 encrypted pixels. Afterward significantly how many bit-planes and rows of pixels he can modify, simply the data hider adopts LSB replacement to substitute the available bit-planes with additional data $m$. Finally, the data hider sets a label following $m$ to point out the end position of embedding process and further encrypts $m$ according to the data hiding key to formulate marked encrypted image denoted by $E_{F}$. Anyone who does not possess the data hiding key could not extract the additional data.

C. Data Extraction and Image Recovery

The data extraction is completely independent from image decryption, the order of them implies two different practical applications.

Case 1: Extracting Data from Encrypted Images: To manage and update personal information of images which are encrypted for protecting clients’ privacy, a database manager may only get access to the data hiding key and have to manipulate data in encrypted domain. The data extraction before image decryption guarantees the feasibility of our work in this case. When the database manager gets the data hiding key he can decrypt the LSB-planes of $A_{E}$ and extract the additional data $m$ by directly reading the decrypted version. Later requesting for updating information of encrypted images the database manager will updates information through LSB replacement and encrypts updated information according to the data hiding key all over again. The whole process entirely operated on encrypted domain will avoids the leakage of original content.

Case 2: Extracting Data from Decrypted Images: In the first Case, both embedding and extraction of the data are manipulated in encrypted domain. Further, there is a different situation that the user wants to decrypt the image first and extracts the data from the decrypted image when it is needed. We see the following example is an application for such scenario. Assume that Alice was out-sourced her images to a cloud server, and those images are encrypted to protect their contents. That encrypted images the cloud server marks the images by embedding some notation.
includes the identity of the images’ owner the identity of the cloud server and time stamps, to manage the encrypted images. The cloud server doesn’t have any right to do any permanent damage to the images. Therefore authorized user Bob who has been shared the encryption key and the data hiding key, down-loaded and decrypted the images. Bob is hoped to get the marked decrypted images those decrypted images still including the notation, which that it can be used to trace the source and history of the data. The direction of image decryption before/without data extraction is perfectly suitable for this case. Afterward, we describe how to generate a marked decrypted image.

\( a )\) Generate Marked Decrypted Image: To form the marked decrypted image \( X' \) which is made up of \( A' \) and \( B' \) the content owner should do following two steps.

**Step 1.** With the help of encryption key, the content owner will decrypts the image except the LSB-planes of \( AE \). The decrypted version of \( E' \) containing the embedded data can be calculated by

\[
X'_{i,j}(t) = E'_{i,j}(t) \oplus r_{i,j}(k) \tag{5}
\]

and

\[
X'_{i,j} = \sum_{k=i}^{i} X''_{i,j}(t) \times 2^k, \tag{6}
\]

**Step 2.** Extract \( SR \) and \( ER \) in marginal area of \( B' \). By rearranging \( A'' \) and \( B'' \) to its original state that the plain image containing embedded data is obtained.

As can be seen, the marked decrypted image \( X' \) is identical to rearranged \( X \) except LSB-planes of \( A \). At the meantime, it keeps perceptual transparency compared with original image \( C \). More specifically, the distortion is introduced via two separate ways: the embedding process by modifying the LSB-planes of \( A \) and self-reversible embedding process by embedding LSB-planes of \( AE \). The first part distortion is well controlled via exploiting the LSB-planes of \( A \) only and the second part can benefit from excellent performance of current RDH techniques.

\( b) \) Data Extraction and Image Restoration: After generating the marked decrypted image the content owner further extract the data and recover original image. This process is similar to that of traditional RDH methods [10], [11]. The following outlines the specific steps:

**Step 1.** Record and decrypt the LSB-planes of \( AE \) according to the data hiding key; extract the data until the end label is reached.

**Step 2.** Extract \( LRNL+MRML+RP+R \) and boundary map from the LSB of marginal area of \( B' \). Then, scan \( B'' \) to undertake the following steps.

**Step 3.** If \( Rb \) is equal to 0, which means no black pixels participate in embedding process, then go to Step 5.

**Step 4.** calculate the errors \( e_i,j \) and estimates black pixels \( B''_{i,j} \) which belongs to \([1, 254]\) to recover the estimating Error and original pixel value in a reverse order and extract embedded bits when \( e_i,j \) is equal to \( LN, LM \) (or \( LR, RM \)) and \( RN \). Else, if \( B''_{i,j} \in [1, 254] \), refer to the corresponding bit \( h \) in boundary map. If here \( h=0 \), skip this one or else operate like \( B''_{i,j} \in [1, 254] \). Repeat this step until the part of payload \( Rb \) is extracted. If payload \( Rb \) is extracted bits are LSBs of pixels in marginal area restores them immediately.

**Step 5.** Calculate estimating errors \( e_i,j \) of white pixels \( B''_{i,j} \) and extract embedded bits and recover white pixels in the same manner with Step 4. If these extracted bits are LSBs of pixels in marginal area restores them immediately.

**Step 6.** Continue doing Step 2 to Step 5 \( x-1 \) rounds on \( B'' \) and merge all extracted bits to form LSB-planes of \( A \). Until now, we have perfectly recover \( B \).

**Step 7.** Replace marked LSB-planes of \( A'' \) with its original bits extracted from \( B'' \) to get original cover image \( C \). We note that if the content owner wants to retrieve his image in Case 1, will do the procedures are exactly the same in Case 2. So, it is omitted in Case 1 for simplicity.

### 4. Conclusion

Reversible data hiding in encrypted images is a new topic drawing attention because of the privacy-preserving requirements from cloud data management. Earlier methods are implement RDH in encrypted images by vacating room after encryption was opposed to which we proposed by reserving room before encryption. With that the data hider can benefit from the extra space emptied out in previous stage to make data hiding process effort-less. The proposed method can take advantage of all traditional RDH techniques for plain images and achieve excellent performance without loss of perfect secrecy. Moreover, this novel method can achieve real reversibility, data extraction and greatly improvement on the quality of marked decrypted images.

### References


Author Profile

Chaple. Gopal received the B.Tech degree in computer science and Engineering from JNTU Hyderabad in 2011 and pursuing M.Tech Degree in Computer Science and Engineering from CVSR College of Engineering from ANURAG Group of Institutions, JNTU Hyderabad, India

G. Balram working as assistant professor in Computer Science Engineering from CVSR College of Engineering from ANURAG Group of Institutions, Venkatapur(V), Ghatkesar (M), RangaReddy District, Hyderabad-500088, Telangana State.