High Capacity Reversible Data Hiding Based on the Compression of Pixel Differences

Kai-Meng Chen

Computer Engineering College, Jimei University, Xiamen 361021, China; chenkaimeng@jmu.edu.cn

Received: 31 July 2020; Accepted: 24 August 2020; Published: 26 August 2020

Abstract: In this paper, we proposed a novel reversible data hiding method in encrypted image (RDHEI), which is based on the compression of pixel differences. In the proposed method, at the content owner’ side the image is divided into non-overlapping blocks, and a block-level image encryption scheme is used to generate the encrypted image, which partially retains spatial correlation in the blocks. Due to the spatial correlation, in each block the pixels are highly likely to be similar. Therefore, the pixel differences in all blocks are concentrated in a small range and can be compressed. By the compression of pixel differences, the data hider can vacate the room to accommodate secret data in the encrypted image without losing information. At the receiver’s side, the receiver can obtain secret data or retrieve the original image using different keys with no error. The experimental results demonstrate that, compared with existing methods, the proposed method can achieve a higher capacity and visual quality.

Keywords: reversible data hiding; image encryption; compression

1. Introduction

Reversible data hiding (RDH) is a technology which embeds secret data into cover media (such as text, image, and video) imperceptibly and reversibly. The embedded data can be extracted and the cover media can be completely recovered. Because the digital images are used widely in various fields, a series of RDH methods based on images have been proposed [1]. These methods can be classified into two categories: RDH in plaintext images [2–9] and RDH in encrypted images (RDHEI) [10–33].

RDHEI methods embed secret data into an encrypted image in a reversible way without decryption or being aware of the image content. RDHEI technology is useful for applications in which the data managers have no access to the contents of the images for privacy or other reasons. For example, in the application of cloud storage, to protect privacy, the original images are encrypted before storing them in the cloud servers. In RDHEI, there are three parties: the content owner, the data hider, and the receiver. The content owner encrypts the original image and transfers it to the data hider. The data hider embeds secret data into the encrypted image. After data embedding, the receiver can extract the secret data or retrieve the original image from the encrypted image which contains secret data. According to how the data hiding room is vacated, the existing RDHEI methods can be classified into the vacating room after encryption (VRAE) [10–27] and the reserving room before encryption (RRBE) [28–33].
In the VRAE methods, the original image is encrypted with no preprocessing by the content owner. The data hider should vacate room in the encrypted image for data hiding. In Zhang’s method [10], the encrypted image is divided into non-overlapping blocks, and one secret bit is embedded into one block by flipping the three least-significant bits (LSBs) of half the pixels in the block. To extract the secret data and recover the flipped block, a smoothness estimator is used to judge the flipped pixels in each block after image decryption. The methods in [11–13] improve Zhang’s method in terms of the reversibility and visual quality. In Hong et al.’s method [11] and Liao et al.’s method [12], to reduce errors in data extraction and image recovery, different improved smoothness estimators and side matching strategies are used. In Qin et al.’s method [13], an elaborate pixel selection scheme is used to reduce the distortion of the directly decrypted image, and a more precise smoothness estimator is used to reduce errors in data extraction and image recovery. In Wu et al.’s method [14], pixels are pseudo-randomly selected from the encrypted image and divided into same-size groups, and one secret bit is embedded into each group by flipping the \(i\)-th (\(1 \leq i \leq 6\)) LSBs of all pixels in the group. After image encryption, multiple designed content-aware pixel value predictors are used for detecting the bit flipping, so that the embedded bits can be extracted and the image can be recovered. Dragoi et al.’s method [15] extends Wu et al.’s method. In Dragoi et al.’s method, more pixels in the encrypted image can be selected for data hiding so that the embedding rate is improved, and Bose–Chaudhuri–Hocquenghem (BCH) coding is used to reduce errors in data extraction and image recovery. In Zhou et al.’s method [16], the encrypted image is divided into non-overlapping blocks, and a public key modulation scheme is used to embed a certain number of bits into each block. To recover the image and extract the embedded bits, a classifier based on the support vector machine (SVM) is used to judge the embedded bits. In [17], a designed sparse matrix is used to compress the LSB planes of the encrypted image so that the spared bits in the LSB planes can be used to accommodate the secret data. The receiver extracts the embedded data directly from the LSB planes, and uses a specific pixel predictor to recover the compressed LSBs. The methods in [18,19] improve the method in [17]. In Qian et al.’s method [18], the pixels of the encrypted image are divided into three subsets. In the different subsets, the LSBs of the pixels are compressed by different sparse matrices, and are recovered by different pixel predictors. In Qin et al.’s method [19], a specific image encryption scheme is used, so that the encrypted image can be decomposed into smooth regions and complex regions. Only the LSBs of the smooth regions are compressed for data hiding. In [20,21], Low-Density Parity Check (LDPC) code is used to compress the specific bit planes for data hiding. In Zhang et al.’s method [20], the bits of the fourth LSB plane of the encrypted image are compressed. In Qian et al.’s method [21], three quarters of the most significant bits (MSBs) are compressed. To recover the image, the log-likelihood ratio (LLR) and iterative belief propagation algorithm (BPA) decoding algorithm are used to recover the compressed bits.

Some VRAE methods use special encryption schemes to partially retain the spatial correlation in encrypted images, and the data hider uses variants of the different traditional RDH methods (such as histogram shifting, prediction error expansion, and pixel value ordering) to embed secret data. In Huang et al.’s method [22], various histogram shifting-based RDH methods are accomplished in the encrypted domain. In Xiao et al.’s method [23], a variant pixel value ordering scheme is used to embed secret bits into non-overlapping blocks of the encrypted image. In Yi et al.’s method [24], a designed pixel value predictor is used to generate prediction errors, and a variant prediction error expansion scheme is used for data embedding. In Qin et al.’s method [25], the MSBs of each block of the encrypted image are compressed by a sparse matrix compression scheme. In Li et al.’s method [26], based on a different histogram which is generated from the encrypted image, a variant histogram shifting scheme is used to embed secret data. In Ge et al.’s method [27], a multi-level histogram shifting scheme is used to embed secret data into each block of the encrypted image.
In the RRBE methods, before image encryption, the content owner preprocesses the original image to vacate room for accommodating secret data. The vacated room is retained after image encryption, and the data hider can embed secret data into the room directly. In Ma et al.’s method [28], the original image is divided into the smooth regions and the complex regions. To reserve the LSBs of the encrypted image for data hiding, the LSBs of the complex regions are embedded into the smooth regions using a traditional RDH method, such as difference expansion. Image encryption is performed after embedding the LSBs. At the data hider’s side, the secret data can be placed into the original positions of the embedded LSBs directly. In Zhang et al.’s method [29], before image encryption a pixel estimator is used to predict the original values of a portion of pixels, then these pixels are substituted with their prediction errors. After image encryption, based on the histogram of the prediction errors, the data hider uses a variant histogram shifting scheme to embed secret data. In Cao et al.’s method [30], the original image is divided into patches, and each patch is encoded using less bits by a sparse coding technology with an over-complete dictionary. The spared bits in each patch can be used for data embedding after image encryption. In Yi et al.’s method [31], each MSB plane of the original image is divided into non-overlapping blocks and compressed by a designed sparse matrix coding scheme. Then, the LSBs of the image are embedded into the compressed MSB planes to vacate room for data hiding. In Chen et al.’s method [32], a bit plane rearrangement strategy is used to rearrange the MSBs of the original images, then the rearranged MSBs are compressed by an extended run-length coding scheme to vacate room for data hiding. In Qiu et al.’s method [33], the LSBs of the original image are removed for data hiding by using a reversible integer transformation scheme.

Benefiting from the use of the original image’s spatial correlation, the RRBE methods can achieve a much higher capacity than the VRAE methods. However, the RRBE methods require the content owner to handle extra image processing work. In some cases, the content owner may not be able to process images. Therefore, the VRAE methods are more feasible for different applications.

In this paper, we propose a novel VRAE RDHEI method which is based on the compression of pixel differences. In the proposed method, at the content owner’s side a specific image encryption scheme is used to partially retain the spatial correlation in non-overlapping $2 \times 2$ blocks of the encrypted image. At the data hider’s side, for each block the four pixels are divided into two parts—one mark pixel and three replaceable pixels—and three pixel differences between the three replaceable pixels and the mark pixel are collected. Due to the spatial correlation, the pixel differences are highly likely to be concentrated and can be compressed efficiently. By replacing the information of the replaceable pixels with their compressed pixel differences, the data hiding room is vacated in the encrypted image without losing any information. At the receiver’s side, the receiver can extract the secret data with no error and recover the image to its original version.

The rest of the paper is organized as follows. In Section 2, we present the detailed introduction to the proposed method. In Section 3, the experimental results and comparisons are provided. The conclusions are given in Section 4.

2. The Proposed Method

In this section, we introduce the details of the proposed method. Figure 1 illustrates the framework of the proposed method. At the content owner’s side, the content owner encrypts the original image with no preprocessing using two encryption keys, and sends the encrypted image to the data hider. The data hider vacates room in the encrypted image, then uses a data hiding key to embed secret data into the room. The receiver extracts the embedded data from the encrypted image by the data hiding key $K_{en}$, and recovers the original image or generates the marked decrypted image containing the secret data by the encryption keys.
2.1. Image Encryption

For an 8-bit standard gray image sized $H \times W$, we denote the $k$-th bit of the pixel $p_{i,j}$ as $b_{i,j,k}$ ($1 \leq i \leq H, 1 \leq j \leq W$). The $b_{i,j,k}$ is defined as follows:

$$b_{i,j,k} = \left\lfloor \frac{p_{i,j}}{2^{k-1}} \right\rfloor \mod 2, \text{ where } 1 \leq k \leq 8. \quad (1)$$

To partially retain the spatial correlation in the encrypted image for data hiding, a specific image encryption scheme is used to encrypt the original image at the level of the block. The detailed steps of the image encryption scheme are as follows:

**Step 1:** Divide the original image into $N = \lceil H/2 \rceil \times \lceil W/2 \rceil$ non-overlapping $2 \times 2$ blocks $B_r$ ($r = 1, 2, \ldots, N$).

**Step 2:** For each block $B_r$, an 8-bit pseudo-random bit sequence $S_r$ is generated using the block encryption key $K_{en}$. Denoting the four pixels in $B_r$ as $p_r^{(1)}, p_r^{(2)}, p_r^{(3)}$, and $p_r^{(4)}$, each pixel $p_r^{(i)}$ is encrypted into $pe_r^{(i)}$ by $S_r$ as follows:

$$e_r^{(i)} = b_r^{(i)} \oplus bs_{r,k}, \quad i = 1, 2, 3, 4, \quad k = 1, 2, \ldots, 8, \quad (2)$$

$$pe_r^{(i)} = \sum_{k=1}^{8} e_r^{(i)} \times 2^{k-1}, \quad i = 1, 2, \ldots, 8, k = 1, 2, 3, 4, \quad (3)$$

where $b_r^{(i)}$ is the $k$-th bit of $p_r^{(i)}$, and $bs_{r,k}$ is the $k$-th bit of $S_r$.

If $H$ or $W$ is odd, after all the blocks are encrypted, the last row or column should be encrypted at the level of the pixel. For each pixel which does not belong to any block, an 8-bit pseudo-random bit sequence is generated by $K_{en}$ to encrypt the pixel. The encryption process of each pixel is the same as that shown in Equations (2) and (3).

**Step 3:** After all the blocks have been encrypted, to enhance the encryption strength the content owner uses the block permutation key $K_{bp}$ to pseudo-randomly permute all blocks $B_1, B_2, \ldots, B_N$ into $B'_1, B'_2, \ldots, B'_N$ inside the encrypted image.

After image encryption, the encrypted image is sent to the data hider for embedding secret data. Because the four pixels in the block are encrypted by the same bit sequence, the encrypted pixels are still highly likely to be similar. Therefore, the spatial correlation of the original image can be partially retained in the image.
retained in each block. That allows the data hider to vacate room in each block for accommodating secret data.

2.2. Data Embedding

When the data hider receives the encrypted image, first the data hider divides the encrypted image to retrieve the $2 \times 2$ blocks $B'_1, B'_2, \ldots, B'_N$. For each block $B'_r$, the four pixels $p_{r1}, p_{r2}, p_{r3},$ and $p_{r4}$ are divided into one mark pixel and three replaceable pixels. As shown in Figure 2, to simplify the statement $p_{r4}$ is assigned to the mark pixel, and the other three pixels are assigned to the replaceable pixels. Then, three pixel differences $d_r^{(1)}, d_r^{(2)},$ and $d_r^{(3)}$ of block $B'_r$ are calculated as:

$$d_r^{(i)} = p_{r4} - p_{r4}$$

(4)

![Figure 2. The mark pixel and replaceable pixels.](image)

The range of $d_r^{(i)}$ is $[-255, 255]$. However, since the spatial correlation is retained in each block, there is a high probability that the four pixels inside the block are close to each other. Thus, $d_r^{(i)}$ is highly likely to be close to 0.

After all the pixel differences $\{d_r^{(1)}, d_r^{(2)}, d_r^{(3)}, d_r^{(1)}, d_r^{(2)}, d_r^{(3)}, \ldots, d_r^{(1)}, d_r^{(2)}, d_r^{(3)}\}$ are collected from all the blocks, these pixel differences are highly concentrated. Figure 3 shows the histogram of the pixel differences of Lena (in Figure 5). As shown in the figure, most of the pixel differences are concentrated in a small range around 0. Therefore, these pixel differences can be compressed efficiently by Huffman coding. According to the mark pixel and the pixel differences, the three replaceable pixels in each block can be recovered completely. Therefore, the data hider can replace the replaceable pixels with the compressed pixel differences to vacate room in the encrypted image for accommodating secret data.

![Figure 3. The histogram of the pixel differences of Lena.](image)

The procedure of room vacating and data embedding is as follows:

**Step 1:** For each block, calculate the three pixel differences according to Equation (4).

**Step 2:** For all the pixel differences $d_r^{(i)}$ ($r = 1, 2, \ldots, N, i = 1, 2, 3$), calculate the distribution of the difference values and use Huffman coding to encode all $d_r^{(i)}$ into $ed_r^{(i)}$ ($r = 1, 2, \ldots, N, i = 1, 2, 3$).
2.3. Data Extraction and Image Recovery

When the receiver gets the marked encrypted image containing secret data from the data hider, the receiver can obtain different data from the image using different keys.

Data extraction: When the receiver has the data hiding key $K_d$, the receiver can extract the secret data directly from the data hiding room. Beginning from the LSB planes of the replaceable pixels, the receiver uses the data hiding key $K_d$ to retrieve the embedded secret bits from the data hiding room, so that the secret data are extracted.

Image Recovery: When the receiver has the block encryption key $K_{en}$ and the block permutation key $K_{perm}$, the receiver can retrieve the original image from the marked encrypted image or generate the marked decrypted image, which still contains secret data and is highly similar to the original image. The procedure of image recovery is as follows:

Step 1: Extract the length information from the MSBs of the replaceable pixels. Then, extract the Huffman codebook and the bitstream of the compressed pixel differences $BS_d$ according to the length information.

Step 2: According to the Huffman codebook, decompose $BS_d$ into the encoded pixel differences $\{e_d(1), e_d(2), e_d(3), \ldots, e_d(1), e_d(2), e_d(3)\}$, and decode them into the original differences $\{d(1), d(1), d(2), d(1), d(2), d(3), \ldots, d(1), d(2), d(3)\}$.
Step 3: For each block $B_r'$ containing the mark pixel $pe_r^{(4)}$, retrieve the original replaceable pixels $pe_r^{(1)}$, $pe_r^{(2)}$, and $pe_r^{(3)}$ as follows:

$$pe_r^{(i)} = pe_r^{(4)} + d_r^{(i)},$$

where $i = 1, 2, 3$, $r = 1, 2, \ldots, N$. (5)

Step 4: After retrieving the original values of all the replaceable pixels, to recover the original image all the replaceable pixels in the marked encrypted image are substituted with their original values.

To generate the marked decrypted image, for each replaceable pixel the MSBs which are substituted with the auxiliary information are recovered according to the original value, and the LSBs which belong to the data hiding room stay the same.

Step 5: Use the block permutation key $K_{bp}$ and the block encryption key $K_{en}$ to decrypt the processed encrypted image into the original image or the marked decrypted image, which is highly similar to the original image.

3. Experimental Results

In this section, we evaluate the performance of the proposed method in terms of its security, embedding rate, and visual quality. Figure 5 shows the six standard grayscale images which are used to validate the performance of the proposed method [34].

![Six test images](image)

Figure 5. Six test images. (a) Airplane; (b) Baboon; (c) Barbara; (d) Lena; (e) Peppers; (f) Zelda.

3.1. Security Analysis

In the proposed method, the original image is encrypted by two keys: the block encryption key $K_{en}$ which is used to encrypt the pixels at the level of a $2 \times 2$ block, and the block permutation key $K_{bp}$ which is used to pseudo-randomly permute the blocks. For an image sized $H \times W$, $[H/2] \times [W/2] \times 8$ bits are required to encrypt the $[H/2] \times [W/2]$ blocks of the original image. After encrypting the pixels inside blocks, there are $([H/2] \times [W/2])!$ possible situations of block permutation. Therefore, there are $28^{[H/2] \times [W/2]} \times ([H/2] \times [W/2])!$ possible situations of image encryption. For a standard $512 \times 512$ grayscale encrypted image, the number of its possible decryption results is $2^{524288} \times 65536! \approx 2^{1478325}$, which is very large number to ensure computational security. Without the block encryption key and the block permutation key, it is almost impossible to obtain the original image from the encrypted image.
Figures 6 and 7 show the histograms of the six original image and their encrypted versions, respectively. As shown in the figures, the pixel value distribution of each encrypted image is uniform and completely different from that of the original image. Therefore, it is almost impossible to obtain the information of the original image from the encrypted image by statistical attack.

Figure 6. Histograms of the original image. (a) Airplane; (b) Baboon; (c) Barbara; (d) Lena; (e) Peppers; (f) Zelda.

Figure 7. Histograms of the encrypted image. (a) Airplane; (b) Baboon; (c) Barbara; (d) Lena; (e) Peppers; (f) Zelda.
In the proposed method, though the image encryption scheme is based on stream cipher which cannot withstand differential attacks, it is very difficult for an attacker to perform differential attacks in the field of RDHEI. In the application scenarios of RDHEI, different content owners use their own image encryption keys to protect their own privacies, and only share the encryption keys to the receivers who can access the image content. The data hiders or other parties have no access to the image encryption keys and the content owners’ original images, and they cannot provide similar plain images to the content owners or receivers to obtain the corresponding encrypted images for comparisons and differential attacks. Therefore, in general it is not feasible to perform differential attacks in RDHEI scenarios.

3.2. Performance Comparison

In this subsection, we conduct experiments to evaluate the embedding rate and visual quality of the proposed method, and perform a comparison between the proposed method and the related works in [10,12,14,21,23,26].

Table 1 shows the embedding rates of the proposed method and the existing RDHEI methods in [10,12,14,21,23,26]. As shown in the table, in most cases the proposed method can achieve relatively higher embedding rates compared with the other methods. For the smooth images such as Airplane, Lena, Peppers, and Zelda, the embedding rates of the proposed method can exceed 1.3 bpp, which is much higher than that of the other methods. For the complex images such as Baboon and Barbara, the proposed method can still achieve a relatively higher capacity than the other methods (except for Qian et al. [21] on Baboon). According to the experimental results, the data hiding room can be efficiently vacated in each block by Huffman coding.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Airplane</td>
<td>0.0191</td>
<td>0.0709</td>
<td>0.0345</td>
<td>0.2952</td>
<td>0.2799</td>
<td>0.6986</td>
<td>1.6965</td>
</tr>
<tr>
<td>Baboon</td>
<td>0.0047</td>
<td>0.0679</td>
<td>0.0096</td>
<td>0.2952</td>
<td>0.0467</td>
<td>0.2237</td>
<td>0.2902</td>
</tr>
<tr>
<td>Barbara</td>
<td>0.0138</td>
<td>0.0707</td>
<td>0.0239</td>
<td>0.2952</td>
<td>0.1637</td>
<td>0.4049</td>
<td>0.8555</td>
</tr>
<tr>
<td>Lena</td>
<td>0.0243</td>
<td>0.0709</td>
<td>0.0386</td>
<td>0.2952</td>
<td>0.2422</td>
<td>0.7701</td>
<td>1.4122</td>
</tr>
<tr>
<td>Peppers</td>
<td>0.0162</td>
<td>0.0709</td>
<td>0.0320</td>
<td>0.2952</td>
<td>0.2129</td>
<td>0.7418</td>
<td>1.3520</td>
</tr>
<tr>
<td>Zelda</td>
<td>0.0277</td>
<td>0.0709</td>
<td>0.0463</td>
<td>0.2952</td>
<td>0.2654</td>
<td>0.7797</td>
<td>1.6066</td>
</tr>
</tbody>
</table>

In Figure 8, the Peak Signal-to-Noise Ratio (PSNR) values of the marked decrypted images generated by the proposed method are compared with those of the marked decrypted images generated by the other methods. As shown in the figure, the proposed method can achieve higher PSNR values for different images and different embedding rates. That means that the proposed method outperforms the other methods in terms of the visual quality of the marked decrypted image. This is because the proposed method uses the LSB planes to embed secret data, and needs to modify only one bit to embed one secret bit. Compared with the proposed method, the other methods should use higher bit planes or modify more bits to embed one secret bit.
At the receiver's side, the secret data can be extracted and the original image can be retrieved from the encrypted image with no errors. The experimental results:

Table 1. Comparison of the embedding rates (bpp).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Airplane</td>
<td>0.0347</td>
<td>0.0709</td>
<td>0.0345</td>
<td>0.2952</td>
<td>0.2799</td>
<td>0.6986</td>
<td>1.6965</td>
</tr>
<tr>
<td>Baboon</td>
<td>0.0047</td>
<td>0.0679</td>
<td>0.0096</td>
<td>0.2952</td>
<td>0.0467</td>
<td>0.2237</td>
<td>0.2902</td>
</tr>
<tr>
<td>Barbara</td>
<td>0.0138</td>
<td>0.0707</td>
<td>0.0239</td>
<td>0.2952</td>
<td>0.1637</td>
<td>0.4049</td>
<td>0.8555</td>
</tr>
<tr>
<td>Peppers</td>
<td>0.0162</td>
<td>0.0709</td>
<td>0.0320</td>
<td>0.2952</td>
<td>0.2599</td>
<td>0.7418</td>
<td>1.3520</td>
</tr>
</tbody>
</table>

4. Conclusions

In this paper, a novel VRAE RDHEI method is proposed. In the proposed method, the original image is encrypted at the level of a $2 \times 2$ block to partially retain the spatial correlation inside the blocks. At the data hider's side, the pixels in each block are divided into one mark pixel and three replaceable pixels, and the replaceable pixels are replaced with their compressed pixel differences to vacate room in the encrypted image. At the receiver's side, the secret data can be extracted and the original image can be retrieved from the encrypted image with no errors. The experimental results
demonstrate that the proposed method can achieve higher embedding rates and better PSNR values compared to the existing RDHEI methods.

**Funding:** This paper was supported by the National Natural Science Foundation of China (U1936114, 61701191) and the Natural Science Foundation of Fujian Province, China (2019H0021).

**Acknowledgments:** The author would like to thank the anonymous reviewers for their valuable comments and suggestions.

**Conflicts of Interest:** The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

**References**


