

Two Stage Reversible Data Hiding Based On Image Interpolation and Histogram Modification

Nice Mathew¹, Dr. A. Grace Selvarani M. E²

^{1,2}(PG Scholar, Professor & HOD, Department of CSE (PG), Sri Ramakrishna Engineering College, Coimbatore)

Abstract: In this paper a two stage reversible data hiding technique is proposed. At the first stage, an interpolation technique is used to generate a cover image from the input image. The difference values from input image and cover image is used as the carrier to embed data. At the second stage, a histogram modification is applied on a difference image to embed data. The extraction process also works in two stages. The proposed algorithm is expected to increase the embedding capacity as two techniques are combined. The interpolation technique helps to keep the distortion low. Experimental results show that the new method has higher embedding capacity than other existing methods.

Keywords: Reversible data hiding, two stage embedding, image interpolation, histogram modification.

I. Introduction

Digital steganography and watermarking are the two kinds of data hiding technology to provide secure communication and authentication. Steganography means hiding a secret message in such a way that an outsider cannot detect the presence of the hidden message. Today the growth in the IT field, especially in computer networks, has opened new opportunities in scientific and commercial applications. But this progress has also led to many serious problems such as duplications, hacking and malevolent usage of digital information [1]. Steganography is sometimes confused with cryptography, but there are some notable and distinctive differences between the two concepts. In some situations steganography is often preferred to cryptography. This is because in cryptography the cipher text is a scrambled output of the plaintext and the attacker can guess the presence of encryption has been performed and hence can employ decryption techniques to acquire the hidden data. Also, cryptographic techniques often require high computing power to perform encryption which may affect devices that lack enough computing resources to implement encryption. On the other hand, steganography is the process of masking the sensitive data in any cover media like still images, audio or video. This way the attacker does not feel the presence of the data being transmitted since it is hidden to the naked eye and impossible to distinguish from the original media.

Steganography involves 4 steps:

- Selection of the cover media used for data hiding.
- The secret message or information that is needed to be embedded in the cover image.
- A function that will be used to hide the data value in the cover media and its inverse function to retrieve the hidden data.
- An optional key or the password to authenticate or to embed and extract the data.

To avoid the interception of sensitive information by a malicious user during transmission, data hiding has attracted significantly much attention in recent years [2,3,4]. The philosophy of data hiding is to embed the secret data into a cover image and to generate quality stego-image with PSNR higher than 30dB. Since the stego-image with PSNR higher than 30dB is imperceptible to human's recognition, the malicious user will not be able to recognize whether the secret data is embedded in stego-image or not. Only the authorized user can extract the secret data from the stego-image.

Many RDH methods have been proposed in recent years, e.g., the methods based on lossless compression [5,6], difference expansion technique (DE) [7,8,9], histogram shifting (HS) method [14,15,16], and integer transform [18,19], etc. All these methods aim at increasing the embedding capacity (EC) as high as possible while keeping the carrier distortion low. Tian's DE algorithm [7] is an important work in the field of RDH. In DE algorithm, the host image is divided into pixel pairs, and the difference value of two pixels in each pair is expanded to carry one data bit. This method can provide an embedding rate (ER) up to 0.5 bits per pixel (BPP) and outperforms previous compression-based works. In particular, Tian used a location map to record all expandable locations, and afterwards, the technique of location map is widely used by most RDH algorithms. Later on, Tian's method has been improved in many aspects. Alattar [10] generalized DE to a triple or a quad of pixels which can raise the maximum ER from 0.5 BPP to 0.75 BPP. Kamstra and Heijmans [8] utilized low-pass image to predict expandable locations so that the location map can be compressed remarkably. Thodi and

Rodriguez [9] achieved a major improvement by incorporating DE with HS. In addition, instead of the difference value, Thodi and Rodriguez suggested the use of the prediction-error for expansion embedding since this can better exploit local correlations within neighboring pixels. In [11], as an extension of Thodi and Rodriguez's work, Hu et al. proposed a method by constructing a location map dependent on payload. Recently, Hu et al.'s method is further investigated by Li et al. [12] by considering adaptive embedding in which smooth pixels are embedded with more data bits than other pixels. Another improvement of Hu et al.'s method is the work of Zhou and Au [13] by introducing a different capacity parameters determination strategy.

Ni et al.'s HS-based algorithm [14] is another important work of RDH, in which the peak value of image histogram is utilized to embed data. In this method, each pixel value is modified at most by 1, and thus it keeps the visual quality of marked image. In [15], Lee et al.'s proposed a method by utilizing the histogram of difference image. This method performs better than Ni et al.'s by improving both EC and visual quality. The spatial correlation of natural images is exploited in Lee et al.'s method and thus a more correct histogram is obtained. In [16], Hong et al. proposed a new HS-based method by modifying the histogram of prediction-error. This method can well exploit the image redundancy and can thus achieve a much better performance compared with the previously introduced DE-based methods. Recently, Wu and Huang [17] proposed a novel HS-based method where the histogram bins used for expansion embedding are specifically selected to make the embedding distortion minimum as possible. In general, HS-based RDH is implemented by modifying host image's histogram.

Image interpolation [20,21] is usually used to generate a high-resolution image from its low-resolution one, that is, it scales up an image. The original image and interpolated pixels result to some difference values and data hiding can be applied by taking this as an advantage. There are a number of simple interpolation methods, such as nearest neighbor, bilinear, nearest neighbor mean interpolation etc.

II. Related Works

2.1. Neighbor mean interpolation

In 2009, Jung and Yoo [22] proposed a new interpolation technique called "Neighbor mean interpolation (NMI for short)". It is having a low-time complexity and high-calculation speed. The data hiding scheme by Jung and Yoo scales down the input image into 1/4 of its actual size, which then becomes the original image. The scheme then uses NMI to enlarge this original image into a cover image, which has the same size as the original input image. Secret data are embedded into the scaling-up cover image. The authorized receiver can extract embedded secret message from the stego-image and restore the original image from the cover image.

2.2. Ni et al's method

In 2006, Ni [14] proposed a novel reversible data hiding algorithm, which can recover the original image without any distortion from the marked image after the hidden data have been extracted, is presented by Ni et al. This algorithm utilizes the zero or the minimum point of the histogram of an image and slightly modifies the pixel grayscale values to embed data bit into the image. It can embed more data than many of the reversible data hiding algorithms existing at that time.

For a grayscale cover image C with size of $M \times N$, let a grayscale value with the maximum number of pixels in C as the maximum point denoted as P ; let a grayscale value with no pixels or the minimum number of pixels in C as the zero point denoted as Z . Assume $P < Z$, histogram modification method shifts the values within the range $[P + 1, Z - 1]$ of the histogram to the right-hand side by one unit. It leaves the grayscale value $(P + 1)$ empty. A bit b which is extracted from the secret data will be embedded in the selected pixels, and the output stego-image, S is obtained.

2.3. Yongjian Hu's method

In this paper, [23] they investigated the problem of the overflow location map, which is an important issue in DE-based RDH. In order to increase embedding capacity, the compressibility of the overflow location map is improved by designing a new embedding scheme. The overflow location map depends on the payload. Unlike other (overflow) location maps, it contains two types of overflow locations: one from data embedding and the other from pixel shifting. Since the constraint on shifting is looser, it results in fewer overflow locations than that on embedding. Therefore, our overflow location map matrix is often sparser than others.

This feature makes the overflow location map to have higher compressibility. Interleaving histogram shifting scheme also enhances capacity control capability and it benefits image quality. The proposed algorithm often has better resilience to different images and larger embedding capacity.

III. Proposed System

A two-stage reversible data hiding method with high capacity and good visual quality based on image interpolation and histogram modification techniques is proposed. At the first stage, a high-quality cover image using the developed enhanced neighbour mean interpolation is generated. Then take the difference values from input and cover pixels as a carrier to embed secret data. At the second stage, a histogram modification technique is applied on the difference image to further increase the embedding capacity and preserve the image quality without distortion. The extraction is also two staged. After extracting the whole hidden data the cover image is restored. The architecture can be shown as in fig (1).

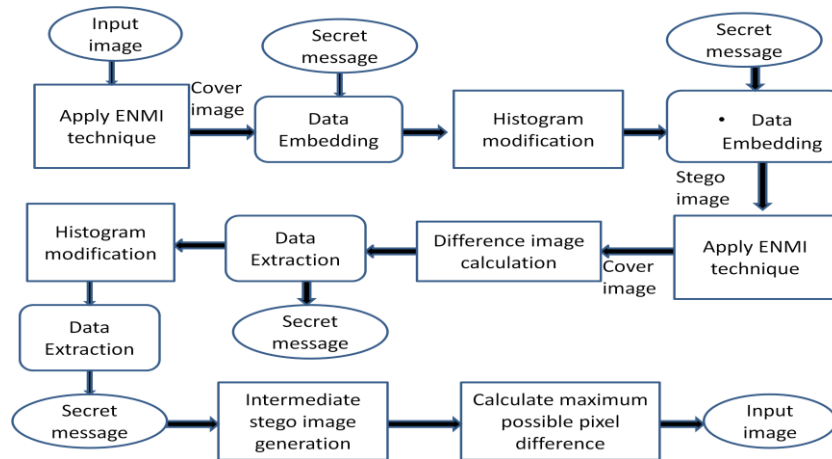


Figure 1: System architecture

3.1. Data embedding

The proposed data hiding algorithm, including two stages with interpolation based data hiding and histogram modification, are shown next in details. The first embedding groups a number of bits and embed its decimal equivalent into pixels. Second embedding is a bit-wise embedding based on histogram modification.

Step 1. For an input image $I(N \times N)$, the ENMI algorithm is applied to generate the corresponding cover image C as given in Eq (1).

for $i = 0$ to $N-1$ do
 for $j = 0$ to $N-1$ do

$$C(i, j) = \begin{cases} I(i, j), & \text{if } i \bmod 2 = 0 \text{ and } j \bmod 2 = 0 \\ I(i, j - 1), & \text{if } j = N - 1 \\ I(i - 1, j), & \text{if } i = N - 1 \\ \frac{I(i, j-1) + I(i, j+1)}{2} & \text{if } i \bmod 2 = 0 \text{ and } j \bmod 2 = 1 \\ \frac{I(i-1, j) + I(i+1, j)}{2} & \text{if } i \bmod 2 = 1 \text{ and } j \bmod 2 = 0 \\ \frac{I(i-1, j-1) + I(i-1, j+1) + I(i+1, j-1) + I(i+1, j+1)}{4}, & \text{otherwise} \end{cases} \quad (1)$$

End for

End for

Step 2. Afterward, a difference image $D0$ can be computed from input image and cover image as shown in Eq (2).

$$D0(i, j) = C(i, j) - I(i, j) \quad (2)$$

Step 3. The number of bits that can be embedded in a pixel, says n is calculated by using Eq (3).

$$n = \lfloor \log_2 |D0(i, j)| \rfloor - 1 \quad (3)$$

Step 4. The first segment of binary secret data $W1$ needs to be partitioned into sub-streams, say w , with n bits before they are hidden. Each sub-stream is extracted out from secret segment $W1$. A single bit "1" called the leading bit is padded to the left-hand side of sub-stream w . An integer value d is calculated from the sub-stream w and its leading bit.

Step 5. Now the integer value d can be embedded into the cover pixel $C(i, j)$. A stego-image $S1$ is generated by adding or subtracting d to or from the cover pixel $C(i, j)$. This helps to make the stego-pixel value closer to the input pixel $I(i, j)$. The embedding rule is shown in Eq (4).

$$S1(i, j) = \begin{cases} C(i, j), & \text{if } i \bmod 2 = 0 \text{ and } j \bmod 2 = 0 \\ C(i, j) + d, & \text{if } I(i, j) \geq C(i, j) \\ C(i, j) - d, & \text{if } I(i, j) < C(i, j) \end{cases} \quad (4)$$

Step 6. The second stage of embedding procedure produces another difference image D1 for increasing the embedding capacity as referred in Eq (5).

$$D1(i, j) = S1(i, j) - C(i, j). \quad (5)$$

Step 7. Then a histogram is constructed based on the difference values D1. Two peak points are chosen where $P2 < 0 < P1$. It is worth mentioning that 0 is not choosing as a peak value as it is necessary to represent the changes of reference pixels when the pixel-shifting in the histogram modification of D1 are carried out. The statute for shifting values in the histogram of D1 is shown in Eq (6).

$$D'1(i, j) = \begin{cases} D1(i, j) + 1, & \text{if } D1(i, j) > P1 \\ D1(i, j) - 1, & \text{if } D1(i, j) < P2 \\ D1(i, j), & \text{otherwise} \end{cases} \quad (6)$$

Step 8. Sequentially scan the difference image D'1 from left-to-right. If the value D'1(i, j) equals to one of the peak values, then one bit w' from the second segment of binary secret data W2 can be embedded in D'1(i, j). The second stage of embedding statute is listed in Eq (7).

$$D''1(i, j) = \begin{cases} D1(i, j) + w', & \text{if } D1(i, j) = P1 \\ D1(i, j) - w', & \text{if } D1(i, j) = P2 \\ D1(i, j), & \text{otherwise} \end{cases} \quad (7)$$

Step 9. The difference value D''1(i, j) can be added to the cover pixel C(i, j) to generate the stego-pixel S(i, j). This is shown in Eq (8).

$$S(i, j) = C(i, j) + D''1(i, j). \quad (8)$$

3.2. Data extraction

Data extraction is the process of separating the secret message from the stego image. There are two stages needed to be performed for extracting the whole hidden data and restoring cover pixels.

Step 1. The procedure first scans the stego-image S and introduces the ENMI algorithm to obtain C by the Eq (9).

$$C(i, j) = \begin{cases} S(i, j) & \text{if } i \bmod 2 = 0 \text{ and } j \bmod 2 = 0 \\ \frac{S(i, j-1) + S(i, j+1)}{2}, & \text{if } i \bmod 2 = 0 \text{ and } j \bmod 2 = 1 \\ \frac{S(i-1, j) + S(i+1, j)}{2}, & \text{if } i \bmod 2 = 1 \text{ and } j \bmod 2 = 0 \\ \frac{S(i-1, j-1) + S(i-1, j+1) + S(i+1, j-1) + S(i+1, j+1)}{4}, & \text{otherwise} \end{cases} \quad (9)$$

Step 2. Now construct a difference image by subtracting C from S by applying Eq (10).

$$D''1(i, j) = S(i, j) - C(i, j). \quad (10)$$

After that, the extraction procedure sequentially scans the difference image D''1 from left-to-right for extracting a segment of secret message by referencing to the values around the peak points.

Step 3. The first stage of extraction initializes W2 as an empty set in advance. A secret bit has been embedded in D''1(i, j) if its value equals to P1, P1 + 1, P2, or P2 - 1. Now the secret data bit w' can be extracted from difference image as in Eq (11).

$$w' = \begin{cases} 0, & \text{if } D''1(i, j) = P1 \text{ or } D''1(i, j) = P2 \\ 1, & \text{if } D''1(i, j) = P1 + 1 \text{ or } D''1(i, j) = P2 - 1 \end{cases} \quad (11)$$

Append the secret bits to W2 in the way that $W2 = W2 || w'$. After that, the histogram modification method is applied on the difference image D''1 as given in Eq (12).

$$D1(i, j) = \begin{cases} D''1(i, j) - 1, & \text{if } D''1(i, j) > P1 \\ D''1(i, j) + 1, & \text{if } D''1(i, j) < P2 \\ D''1(i, j), & \text{otherwise} \end{cases} \quad (12)$$

Step 4. The next is to convert the integer D1(i, j) to its binary format, say b. The first bit which must have been the leading bit "1" of the string b. The secret sub-stream is obtained by discarding the leading bit and takes the rest of the string b. However, when D1(i, j) = 1 is encountered, it means the absence of secret data.

Step 5. Finally, all the secret sub-streams are appended to obtain the secret segment W1. The whole secret message W is obtained by appending W2 to W1. ie, $W = W1 || W2$.

Step 6. The input image can be obtained by applying the ENMI technique on cover image.

IV. Experimental Results

Peak signal-to-noise ratio, often abbreviated PSNR, is an expression for the ratio between original image and noise. It is most commonly used to measure the quality of reconstruction. PSNR measurement is most easily defined via the mean squared error (MSE) value. The mean squared error (MSE) for practical purposes allows comparing the “true” pixel values of our original image to our degraded marked image. The MSE represents the average of the squares of the “errors” between original image and noisy image. The error is the amount by which the values of the original image differ from the noisy image. Given a noise-free $m \times n$ image I and its noisy approximation K , MSE is defined as in Eq. (13) [24].

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \tag{13}$$

The PSNR can be defined as given in Eq. (14):

$$PSNR = 20 \log_{10} \left(\frac{MAX_f}{\sqrt{MSE}} \right) \tag{14}$$

where MAX_f is the maximum possible pixel value of the image. When the pixels are represented using 8 bits per pixel, this is 255. In the absence of noise, the two images I and K are identical, and thus the MSE becomes zero. In this case the PSNR is undefined.

The test images selected for evaluation are shown in Fig:1

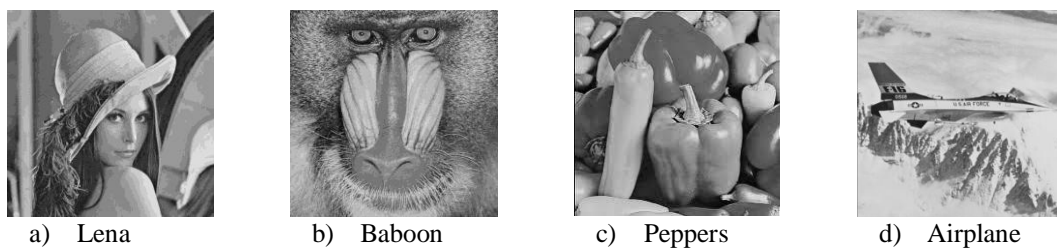


Figure 2: Test images

The algorithm is evaluated by comparing it with some already existing works. Four test images are taken for comparing the algorithms. The PSNR value obtained for different algorithms for each test image can be shown in table I.

Table I. Comparison of PSNR(in dB) for the proposed method with some existing methods for a payload of 20000 bits.

Test image (512x512)	Ni et al.'s method	Jung and Yoo	Proposed system
Lena	43.21	32.94	44.72
Baboon	41.19	35.71	43.86
Air plane	39.93	33.85	42.04
Peppers	43.92	32.77	44.38

V. Conclusion

The proposed method is a two stage reversible data hiding. It combines the image interpolation technique and histogram shifting. The payload embedding capacity can be increased considerably by the proposed work. The secret message can be completely extracted out at the receiver. The scheme ensures that the distortion due to embedding is kept to be low. In future, more advanced histogram shifting techniques can be combined with interpolation of image to achieve high embedding capacity and low distortion. The technique can be also applied to image after partition it into blocks. This can also increase the efficiency of data hiding.

REFERENCES

- [1] C. K. Chan, L. M. Cheng Hiding data in images by simple LSB substitution. Pattern Recognit 37(3), 2004, 469–474
- [2] C. C. Lin, W. L. Tai, and C. C. Chang, Multilevel Reversible Data Hiding based on Histogram Modification of Difference Images, Pattern Recognition, Vol. 41, pp., 2008, 3582-3591.
- [3] H. W. Tseng, Y. M. Tsai, Reversible Data Hiding Scheme Based on Hierarchy Prediction, International Journal of Advanced Information Technologies, Vol.5, 2011, pp.73-90.

- [4] C. C. Chang, W. C. Wu, and Y. C. Hu, Lossless Recovery of a VQ Index Table with Embedded Secret Data, *Journal of Visual Communication and Image Representation*, Vol. 18, 2007, pp 207-216.
- [5] T. Kalker and F. M. J. Willems, Capacity bounds and constructions for reversible data hiding, *Security Watermarking Multimedia Contents V*, vol. 5020, 2003, pp 604-611.
- [6] M. U. Celik, G. Sharma, A. M. Tekalp, and E. Saber, Lossless generalized-LSB data embedding, *IEEE Trans. Image Process.*, vol. 14, no. 2, 2005, pp 253-266.
- [7] J. Tian, Reversible data embedding using a difference expansion, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 8, 2003, pp 890-896.
- [8] L. Kamstra and H. J. A. M. Heijmans, Reversible data embedding into images using wavelet techniques and sorting, *IEEE Trans. Image Process.*, vol. 14, no. 12, 2005, pp 2082-2090
- [9] D. M. Thodi and J. J. Rodriguez, Expansion embedding techniques for reversible watermarking, *IEEE Trans. Image Process.*, vol. 16, no. 3, 2007, pp. 721-730.
- [10] A. M. Alattar, Reversible watermark using the difference expansion of a generalized integer transform, *IEEE Trans. Image Process.*, vol. 13, no. 8, 2004, pp. 1147-1156.
- [11] Y. Hu, H. K. Lee, and J. Li, DE-based reversible data hiding with improved overflow location map, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, 2009, pp. 250-260.
- [12] X. Li, B. Yang, and T. Zeng, Efficient reversible watermarking based on adaptive prediction-error expansion and pixel selection, *IEEE Trans. Image Process.*, vol. 20, no. 12, 2011, pp. 3524-3533.
- [13] J. Zhou and O. C. Au, Determining the capacity parameters in PEE based reversible image watermarking, *IEEE Signal Process. Lett.*, vol. 19, no. 5, 2012, pp. 287-290.
- [14] Z. Ni, Y. Q. Shi, N. Ansari, and W. Su, Reversible data hiding, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 3, 2006, pp. 354-362.
- [15] S. K. Lee, Y. H. Suh, and Y. S. Ho, Reversible image authentication based on watermarking, in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2006, pp. 1321-1324.
- [16] W. Hong, T. S. Chen, and C. W. Shiu, Reversible data hiding for high quality images using modification of prediction errors, *J. Syst. Softw.*, vol. 82, no. 11, 2009, pp. 1833-1842.
- [17] H.-T. Wu and J. Huang, Reversible image watermarking on prediction errors by efficient histogram modification, *Signal Process.*, vol. 92, no. 12, Dec. 2012, pp. 3000-3009,.
- [18] D. Coltuc and J. M. Chassery, Very fast watermarking by reversible contrast mapping, *IEEE Signal Process. Lett.*, vol. 14, no. 4, Apr. 2007, pp. 255-258,.
- [19] X. Wang, X. Li, B. Yang, and Z. Guo, Efficient generalized integer transform for reversible watermarking, *IEEE Signal Process. Lett.*, vol. 17, no. 6, Jun. 2010, pp. 567-570,.
- [20] Jan SR, Hsu SJ, Chiu CF, Chang SL, An improved data hiding method using image interpolation, seventh international conference on intelligent information hiding and multimedia signal processing, 2011, pp 185-188.
- [21] Yalman Y, Akar F, Erturk I, An image interpolation based reversible data hiding method using R-weighted coding, 13th IEEE international conference on computational science and engineering, 2010, pp 346-350
- [22] Jung KH, Yoo KY Data hiding method using image interpolation. *Comput Stand Interfaces* 31(2), 2009, pp 465-470.
- [23] Y. Hu, H. K. Lee, and J. Li, DE-based reversible data hiding with improved overflow location map, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, Feb. 2009, pp. 250-260.
- [24] Salomon, David, *Data Compression: The Complete Reference* (4 ed.). Springer. p. 281. ISBN 978-1846286025, 2007.