A new near-lossless image information hiding algorithm with controlled hiding capacity

Yi Zhang*,a, Quan Zhou, Yan-lang Hu, Juan-ni Liu, Rui-hua Liu

National Key Laboratory of Science and Technology on Space Microwave, China Academy of Space Technology, Xi’an, Shaanxi 710100, China
azhangy4@dingtalk.com

Abstract. Image information hiding technology can embed more data without increasing the amount of transmitted data. But in most cases, the information embedding rate is different for different carrier images. To increase the embedding rate, the data will be damaged. In order to improve this problem, a new near-lossless image information hiding algorithm with controlled hiding capacity is proposed, which is referred to as NLH. This method makes the pre-processed data in a specific range to facilitate information hiding by interval transformation and data mapping. According to analysis and calculation, the lossless performance of the proposed algorithm in this paper is better than the LSB algorithm. And the result of simulation shows that this method has a fixed information embedding rate (18.75%/1.5bpp) without more transmitted data, and the original image is restored losslessly or nearly losslessly with the improved transmission efficiency.

1. Introduction

With the development of communication technology, the demand for hidden transmission of different information is increasing, such as, the demand of hiding information in image, voice and text. By hiding information in images, voices and texts, more information can be transmitted without affecting normal communication. At the same time, the original channel data format is not changed, and the amount of transmitted data is not increased.

Image information hiding is one of the current research hot spots in data processing. For the current research on lossless information hiding algorithm, the hiding capacity is not fixed and the embedding rate is low[1][2].Current lossless information hiding algorithms have different embedding rate corresponding to different image, and the information embedding rate is relatively low for lossy algorithms[3][4]. In order to increase the information embedding rate, there are the original image will inevitably have varying degrees of destruction [5].

To solve this problem, a near-lossless image information hiding algorithm with controlled hiding capacities proposed, which is referred to as NLH. By processing image data such as interval conversion and data mapping, the near-lossless information hiding of the original carrier image can be realized, and the information embedding rate is fixed at 18.75%(1.5bpp). This algorithm does not increase the amount of data transmitted, and solves the problem of inconsistent hidden capacity of different carriers in the communication system.

This paper is arranged as follows: the proposed scheme is described in section two, in which the idea of the algorithm is introduced in detail; the performance is analysed in section three. Then the algorithm is simulated in order to show the better performance in cell loss rate in section four. At last, section five concludes the paper [1].
2. Proposed Scheme

2.1. Image Preprocessing

By interval transformation and data mapping, the pre-processed image data is within a certain range, which is convenient for information hiding.

First, divide the image into odd and even bits. Let the maximum and minimum of odd data be Max1 and Min1. Let the maximum and minimum of even data be Max2 and Min2.

1. If the value of ‘max1-min1’ is less than or equal to the threshold T1, then subtract the Min1 from the odd digits of the image data. If the value of ‘max2-min2’ is less than or equal to the threshold T2, then subtract the Min2 from the even digits of the image data. Among them, T1 and T2 are the set thresholds, which both are positive integers, and meet the conditions T1+T2 ≤ 173, (T1+1)(T2+1) ≤ 7614.

2. If the value of ‘max1-min1’ is more than T1 or the value of ‘max2-min2’ is more than T2, divide the value of each byte of the image data by P and get the given interval data value, and then set it as (T1, T2) according to the odd and even bits. Since the value range of the image is between 0 and 255, this algorithm chooses the P to be 3. After calculating each data value, T1 and T2 are both between 0 and 85.

Second, divide T1 and T2 into two parts, those are [0, 80], [81, 85]. According to the different ranges of T1 and T2, transform the given data to get the data (Q1, Q2):

1. If 0<=T1<=80, then Q1=T1, Q2=T2, the range of Q1 and Q2 are [0,80] and [0,85];

2. If 81<=T1<=85 and 0<=T2<=80, then Q2=T1+5, Q1=T2, the range of Q1 and Q2 are [0,80] and [0,85];

3. If 81<=T1<=85 and 81<=T2<=85, then transform T1, T2 into the interval [0, 9] and [91, 93] respectively.

After the above processing, set the data to (Q1, Q2) according to the odd and even bits, where the range of Q1 is 0-80, and the range of Q2 is 0-93.

2.2. Lossless embedded coding method

After pre-processing, the interval of all image data becomes [0, 80] or [0, 93]. Encode the above data to obtain new data, which can be expressed as (H1, H2), the steps are as follows:

1. If the odd byte is between 0 and 80, and the even byte is between 0 and 63, then Q1 and Q2 remain unchanged, that is H1=Q1, H2=Q2. The encoded data is (H1, H2), where H1 is between 0 and 80, and H2 is between 0 and 63;

2. If the odd byte is between 0 and 63, and the even byte is between 64 and 93, then H1=Q2+17, H2=Q1. The encoded data is (H1, H2), where H1 is between 81 and 110, and H2 is between 0 and 63;

3. If the odd byte is between 64 and 80, and the even byte is between 64 and 93, then H1=Q1+47, H2=Q2-64. The encoded data is (H1, H2), where H1 is between 111 and 127, and H2 is between 0 and 29;

Therefore, the range of encoded data H1 is 0-127, and the range of H2 is 0-63.

2.3. Secret information embedding method

After getting the coded data, sort the sensitive information S into groups of 3 bits to get (s1, s2, s3). Embed the sensitive information into the coded data to obtain the encrypted data. The specific embedding method is:

Use s1 to replace the highest bit of the odd byte H1; Use s2 to replace the highest bit of the even byte H2; Use s3 to replace the second highest bit of the even byte H2;

Complete the above operations to get the encrypted data, which is set to (L1, L2) according to the odd and even bits.

2.4. Carrier recovery and secret information extraction method

First, extract secret information from encrypted characters.
Extract the highest bit of the odd byte L1 in the encrypted character as s1, at the same time set the highest bit position of L1 to 0; Extract the highest bit of even byte L2 in the encrypted character as s2, at the same time set the highest bit position of L2 to 0; Extract the second high bit of even byte L2 as S3, at the same time set the second highest bit position of L2 to 0.

After completing the above operations, sensitive information S and encoded characters are obtained, and the encoded characters are expressed as (H1, H2).

Then decode the encoded data (H1, H2) to get the pre-processed data (Q1, Q2), and perform lossless processing on the data (Q1, Q2) to recover the data Ti (i=0~N), which is between 0 and 85.

Finally, let Ci=Ti×3, and Ci is the restored carrier data.

3. Performance analysis

3.1. Analysis of information embedding rate

The proposed algorithm divides the image data into one group according to the parity, and each group of data can embed 3 bits information. In other words, 18.75% of information can be embedded in the entire image data.

If the image size is 256×256, 12288 bytes of data can be embedded;
If the image size is 512×512, 49152 bytes of data can be embedded;
If the image size is 1024×1024, 196608 bytes of data can be embedded;

In summary, the data transmission efficiency of this method can be increased by 18.75%, and sensitive information has certain security and confidentiality.

3.2. Near-lossless performance analysis of carrier data

Peak signal-to-noise ratio (PSNR) is an objective criterion for evaluating images. In order to measure the image quality after processing, we usually refer to the PSNR value to measure whether a certain processing procedure is satisfactory [1]. Peak signal-to-noise ratio (PSNR) is counted and averaged based on the grey value of the image pixels, and is a commonly used indicator to measure signal distortion. Therefore, this paper also uses PSNR as a basis for measuring whether images are similar. The PSNR calculation method is shown in equation (1).

$$PSNR = 10 \log \left( \frac{a_{\text{max}}^2}{MSE} \right)$$  \hspace{1cm} (1)

$$MSE = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} [a(i,j) - \hat{a}(i,j)]^2$$  \hspace{1cm} (2)

In equation (2), MSE is the mean square error. $a(i,j)$ and $\hat{a}(i,j)$ are the corresponding grey value or colour value in image block $B_i$ and image block $\hat{B}_j$ respectively. $mn$ is the total number of pixels of $m \times n$ image block. $a_{\text{max}} = 2^l - 1$, $l$ is the colour depth, which means the number of binary digits occupied by a pixel, where $l = 8$.

In this paper, all data between 0 and 255 are pre-processed to the interval [0, 85]. When restoring the original carrier image, each value is directly multiplied by 3. The restored data value range is [0, 255], and each value is an integer multiple of 3, that is 0, 3...255.

If the original value is not an integer multiple of 3, such as 253,254, it becomes 84 after pre-processing, and the data after restoration is 252, then the difference from the original data is 1,2.

In order to verify the PSNR, we construct some sets of data A of length N, the value of which is between 0 and 255. Pre-process and restore each data in the array according to the method proposed in this paper to obtain array B. According to equation (1), calculate the PSNR between array B and array A. At the same time, this paper also calculated the PSNR of the least significant bit algorithm (LSB) under the same embedding rate(1.5bpp). The specific calculation results of the proposed algorithm NHL and LSB are shown and compared in the following table.
According to analysis and calculation, it can be known that the PSNR between the recovered data and the original data is about 50dB by using the proposed algorithm NLH. At the same embedding rate, the PSNR of the traditional LSB algorithm is 48dB. Therefore, the lossless performance of the NLH algorithm proposed in this paper is better than the LSB algorithm.

Table 1. Theoretical calculation results of PSNR.

| N    | 64  | 256 | 512 | 1024 | 4096 | 16384 | 65536 |
|------|-----|-----|-----|------|------|-------|-------|
| PSNR-NLH(dB) | 49.86 | 49.78 | 49.73 | 49.76 | 49.92 | 49.88 | 49.91 |
| PSNR-LSB(dB)  | 47.87 | 48.19 | 47.99 | 48.06 | 48.04 | 48.10 | 48.12 |

4. Simulation Results and Analysis

For the proposed algorithm in this paper, some different images are simulated and the performances of the proposed algorithm are verified. In this chapter, the information embedding rate is calculated, and the restored carrier image is compared with the original image.

4.1. Simulation model

This section uses a modular approach to construct the simulation model, which mainly includes carrier data generation, secret information generation, data pre-processing, information embedding, secret information extraction, carrier data recovery, information verification and performance calculation module. The specific model diagram is shown in the figure below.

![Simulation model diagram](image)

Figure 1. Simulation model of information hiding transmission method.

The carrier data generation module is responsible for reading various image data and converting the data into a processable format; The secret information generation module is responsible for reading the secret information to be embedded, and processing the secret information according to the result of data pre-processing; The data pre-processing module is responsible for completing the feature analysis, interval transformation and data mapping of image data; The information embedding module completes the lossless encoding of pre-processed data and embedding of secret information; The secret information extraction module is responsible for extracting secret information from the encrypted data after transmission; The carrier data recovery module is responsible for restoring the original carrier image data; The information verification module is responsible for the comparison of the secret information and the comparison between the original carrier data and the restored data; The performance calculation module is responsible for calculating the information embedding rate and the peak signal-to-noise ratio of the encrypted data.

4.2. Simulation results and analysis

According to the above simulation model, multiple sets of image data are simulated in this paper. The original image Lena and its restored image are compared in figure 2. The figure (a) is the original carrier image, and the figure (b) is the restored carrier image.
The following table shows the different image information hiding performance, including PSNR and information embedding rate after restoration of each image, as well as the verification of the secret information.

### Table 2. Embedding rate of different images.

| Image Name   | Length of Carrier data (byte) | Length of Secret information(byte) | Total length of information(byte) | Embedding rate        |
|--------------|-------------------------------|-----------------------------------|----------------------------------|-----------------------|
| Lena.bmp     | 262144                        | 49151                             | 262144                           | 18.75%(1.5bpp)        |
| Airplane.bmp | 262144                        | 49151                             | 262144                           | 18.75%(1.5bpp)        |
| Babara.bmp   | 262144                        | 49151                             | 262144                           | 18.75%(1.5bpp)        |
| Boats.bmp    | 262144                        | 49151                             | 262144                           | 18.75%(1.5bpp)        |
| Girl.bmp     | 262144                        | 49151                             | 262144                           | 18.75%(1.5bpp)        |
| Elaine.bmp   | 262144                        | 49151                             | 262144                           | 18.75%(1.5bpp)        |

### Table 3. Different image information hiding performance.

| Image Name   | PSNR  | Secret information verification |
|--------------|-------|---------------------------------|
| Lena.bmp     | 49.89 | Exactly the same                |
| Airplane.bmp | 49.88 | Exactly the same                |
| Babara.bmp   | 49.87 | Exactly the same                |
| Boats.bmp    | 49.89 | Exactly the same                |
| Girl.bmp     | 49.86 | Exactly the same                |
| Elaine.bmp   | 49.89 | Exactly the same                |

Through simulation, it can be proved that the algorithm proposed in this paper can recover the secret information losslessly. Compared with the original image, the PSNR of the restored carrier image is about 50dB, which can restore the carrier image nearly-losslessly or losslessly.

### 5. Conclusion

This paper proposes a near-lossless image information hiding algorithm NLH with controlled hiding capacity, which can hide a fixed volume of secret information in image data, and has certain versatility. This method does not require any additional information to losslessly hide secret information in an image, and restore the carrier image losslessly or nearly-losslessly. The hidden capacity of NLH can reach 18.75% (1.5bpp), which unconditionally increases the transmission efficiency of image data to 120% of the original. This method not only saves transmission costs, but also has a certain degree of security and confidentiality for sensitive information.
The algorithm NLH increases the capacity of the data transmission system, which can be used in the resource-constrained environments, such as satellite channels, etc. And this algorithm can also be applied to the situations that require convert transmission or require the fix hidden capacity of different carriers. In future, we will study the algorithms that are more suitable for different data types, and balance the relationship between lossless processing and embedding performance to improve the performance of the entire data transmission system.

Acknowledgments
This work is supported by the National Key Laboratory Foundation (No.614241192205) and the National Key Laboratory Foundation (No.6142411204306).

References
[1] Xiang SJ, Luo XR.(2016) Reversible data hiding in encrypted image based on homomorphic public key cryptosystem. Journal of Software, 2016,27(6):1592-1601(in Chinese).
[2] Wang J L, Sun X, Feng X Q.(2018) Adaptive reversible image data hiding using pixel permutation[J]. Journal of Image and Graphics, 2018,23(1): 0001-0008.
[3] Liu Juanni, Zhou Quan.(2016) SMVQ-index class ified coding for loss less data hiding. Journal of Xidian University, 2016, 43(5): 128-132.
[4] Fangjun Huang, Jiayong Li.(2019) JPEG Reversible Data Hiding with Matrix Embedding[C] International Conference on Image and Graphics ICIG 2019:231-243.
[5] Mingming Zhang, Quan Zhou, Yanlang Hu.(2020) A reversible data hiding scheme in JPEG bitstreams using DCT coefficients truncation [J]. KSII Transactions on Internet and Information Systems, 2020, 14(1): 404-421.
[6] Zheng Hui, Quan Zhou.(2020) A Reversible Data Hiding Method for AMBTC Compressed Image without Expansion inside Stego Format [J]. KSII Transactions on Internet and Information Systems (TIIS), 2020, 2020, 14(11): 4443-4462.
[7] Juan-ni Liu, Quan Zhou, Yan-lang Hu, Jia-yuan Wei.(2017) A new SMVQ-based reversible data hiding scheme using state-codebook sorting [J]. Lecture Notes in Computer Science, 2017, Vol.10668: 539-550.
[8] Mingming Zhang, Quan Zhou, Yanlang Hu.(2019) Lossless data hiding in JPEG images with segment coding[J]. Journal of Electronic Imaging, 2019, 28(05):053015(1-14).
[9] Haishan Chen, Jiangqun Ni, Wien Hong, et al.(2016) Reversible data hiding with contrast enhancement using adaptive histogram shifting and pixel value ordering[J]. Signal Process. Image Commun, 2016, 46:1-16.
[10] LI J L, LI J, ZHANG Y, et al.(2021) Load balancing heterogeneous path algorithm for satellite hybrid switching system[J]. Chinese Space Science and Technology, 2021,41(3):105-113(in Chinese).
[11] Muhammad K, Sajjad M, Mehmod I, Rho S, Baik SW.(2016) A novel magic LSB substitution method(M-LSB-SM) using multi-level encryption and achromatic component of an image. Multimedia Tools and Applications. 2016, 75:14867-14893.
[12] Chen H, Ni J, Hong W, Chen T-S.(2016) Reversible data hiding with contrast enhancement using adaptive histogram shifting and pixel value ordering. Signal Processing: Image Communication. 2016, 46:1-16.
[13] Sahu, A.K., Swain, G.(2019) An Optimal Information Hiding Approach Based o Pixel Value Differencing and Modulus Function. Wirel. Pers. Commun. 2019.
[14] Lang Huyan, Yunpeng Bai, Ying Li, etc.(2021) A Lightweight Object Detection Framework for Remote Sensing Images. Remote Sensing, 2021, 13(4):1-25.