Appl. Sci. **2015**, *5*, 1033-1049; doi:10.3390/app5041033

OPEN ACCESS

applied sciences

ISSN 2076-3417

www.mdpi.com/journal/applsci

Article

Multi-Bit Data Hiding Scheme for Compressing Secret Messages †

Wen-Chung Kuo [1,†,*], Shao-Hung Kuo [2] and Lih-Chyau Wuu [1]

[1] Department of Computer Science and Information Engineering, National Yunlin University of Science & Technology, No.123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan; E-Mail: wuulc@yuntech.edu.tw

[2] Graduate School of Engineering Science and Technology Doctoral Program, National Yunlin University of Science & Technology, No.123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan; E-Mail: g9810814@yuntech.edu.tw

† This paper is an extended version of our paper published in The International Multi-Conference on Engineering and Technology Innovation 2015, Kaohsiung, Taiwan, 30 October–3 November 2015.

* Author to whom correspondence should be addressed; E-Mail: simonkuo@yuntech.edu.tw; Tel.: +886-5-534-2601 (ext. 4515); Fax: +886-5-531-2170.

Academic Editors: Wen-Hsiang Hsieh and Takayoshi Kobayashi

Received: 12 August 2015 / Accepted: 23 October 2015 / Published: 4 November 2015

Abstract: The goal of data hiding techniques usually considers two issues, embedding capacity and image quality. Consequently, in order to achieve high embedding capacity and good image quality, a data hiding scheme combining run-length encoding (RLE) with multi-bit embedding is proposed in this paper. This work has three major contributions. First, the embedding capacity is increased 62% because the secret message is compressed before embedding into the cover image. Secondly, the proposed scheme keeps the multi-bit generalized exploiting modification direction (MGEMD) characteristics, which are effective to reduce modified pixels in the cover image and to maintain good stego image quality. Finally, the proposed scheme can prevent modern steganalysis methods, such as RS steganalysis and SPAM (subtractive pixel adjacency matrix), and is compared to MiPOD (minimizing the power of the optimal detector) scheme. From our simulation results and security discussions, we have the following results: First, there are no perceivable differences between the cover images and stego images from human inspection. For example, the average PSNR of stego images is about 44.61 dB when the secret message (80,000 bits) is embedded for test cover images (such as airplane, baboon, Lena) of size 512×512. Secondly,

on average, 222,087 pixels were not modified after embedding for the cover image. That is to say, 12% less pixels are modified as compared to the MGEMD method. From the performance discussions, the proposed scheme achieves high embedding capacity and good image quality, but also maintains stego image security.

Keywords: RLE; MGEMD; RS steganalysis; embedding capacity

1. Introduction

Networks are ubiquitous in modern life. More and more things are increasingly digital, such as photos, videos, music, documents, personal information, and so on. Therefore, how to protect digital information is a hot issue. Cryptography and steganography are two popular technologies used to protect digital products. For cryptography, a key is used to encrypt data into meaningless numbers, then we can use the same key or another to decrypt. Common encryption methods are advanced encryption standard (AES), data encryption standard (DES), RSA and MD5. In general, cryptographic technologies provide a certain level of security, but cannot maintain security when the ciphertext is decrypted. Therefore, steganography technologies have been developed. Steganography technologies can be classified into watermarking and data hiding [1]. Digital watermarking technology, in general, can be divided into two categories [2], visible and invisible watermarking. A visible watermark's advantage is the human eye can discern it. No algorithm is needed to view the information that represents data sources or the owner. The disadvantage of a visible watermark is that the image is changed by the watermark. It is easily overwritten or removed by signal processing technology. Watermarking techniques can be divided into two types: fragile and robust watermarks. A fragile watermark is primarily used to protect the integrity of the image. The slightest modification to the media with a fragile watermark results in the destruction of the watermark. A missing watermark denotes tampering. Robust watermarking can survive a designated class of transformations. An example of a robust watermark application is a watermark to carry copy and access control information. The media may be compressed, cropped or otherwise transformed, but the watermarked information survives.

Utilizing digital signal processing and digital imaging technologies to hide secret data without reducing the quality of the cover image is called data hiding. This technique is not readily apparent and hides information in any form (text, images, video). Data hiding has two techniques: spatial domain and transform domain. This kind of data hiding technology has very high image quality and is undetectable by the human eye. The technology of data hiding can be classified into two types: one is irreversible data hiding, and the other is reversible. The difference is reversible data hiding is lossless and can reconstruct the original cover image from the stego image after the secret message is extracted.

The watermarking technology always modifies pixels in the frequency domain [3], but pixels modified in the frequency domain contribute to more distortion. Therefore, data hiding technology usually occurs in the spatial domain for less distortion than watermarking technology and quick processing. For data hiding in the spatial domain, least significant bit (LSB) replacement is a classic scheme [4], but the

capacity of embedded data is one bit per pixel (bpp), and it has no security. The exploiting modification direction (EMD) [5] scheme can embed 1.16 bpp for two pixels in each group.

Data compression [6] can save storage space and speed up network transmission. In other words, the raw data are processed through various mathematical algorithms to reduce data storage space. This reduced amount of data is transmitted. Then, the decompression operation can recreate the original data at the receiver. There are two types of data compression. One is lossless compression, such as PCX, GIF, TIFF, TGA and PNG image formats, ZIP, RAR data compression technology, run-length encoding, Huffman coding and Lempel–Ziv–Welch (LZW). The other is lossy compression, such as JPEG (Joint Photographic Coding Expert Group), VQ (Vector Quantization) [7–9] and SMVQ [10,11] (Side Match Vector Quantization). Therefore, to achieve the smaller secret message size, the data compression technology is a good solution.

To leverage the advantages of compression, we will propose a data hiding scheme that can embed more secret data, *i.e.*, secrets are pre-compression, and then uses multi-bit data hiding. The proposed scheme effectively reduces secret messages size to improve embedding capacity and also combines multi-bit generalized exploiting modification direction (MGEMD) [12] to increase the embedding capacity.

In Section 2, we review previous work of RLE (run-length encoding) technology and the multi-bit embedding scheme. Section 3 gives a detailed introduction of the proposed method and then proposes a modified speed up method for MGEMD. In addition, there is also a discussion on the overflow/underflow problems and solutions. Experiments are given in Section 4. Finally, some conclusions are given in Section 5.

2. Related Work

In this section, there are two main related works. In the data compression part, we will review the RLE method. Then, three data hiding methods [5,12,13] based on the extraction function are introduced.

2.1. Run-Length Encoding

Run-length encoding (RLE) [14] is a well-known, simple and quick form of data compression in which sequences of the same data value are found in many consecutive data elements. The RLE applications of this encoding are when the source information comprises long substrings of the same character or binary digit. For this reason, using RLE to compress the binary secret message is very applicable. For example, secret message 00001011101 will encoded into $0(4)1(1)0(1)1(3)0(1)1(1)$.

In 2006, Chang *et al.* [15] proposed two new image steganographic methods using the run-length approach. There are two methods, one is BRL (hiding bitmap files by run-length), which focuses on binary images, and the other is GRL (hiding general data files by run-length). The major idea of these methods is to use RLE to increase the SMVQ method embedding capacity [16]. For binary images, Agaian and Cherukuri [17] also proposed run-length based steganography. Their proposed algorithm is dependent on their run length characteristics and characteristics values of the block and alters pixels of the cover image's embeddable blocks. Simultaneously, this scheme also enhances the security of the embedded data and the capacity of the embedding method. In addition, steganographic access control

in data hiding using run-length encoding and modulo operations was proposed by Lee *et al.* [18] in 2011. In their scheme, a high capacity steganographic with access control modifies sharp bitstreams into smooth bit streams and embedded into the cover image. The modulo value is fixed in this scheme, meaning the embedding capacity is limited.

Accordingly, RLE to increase embedding capacity for data hiding is important. In particular, it will increase the compression ratio when there are many continuities of ones and zeros in these binary images (black and white picture). We use binary images and also gray-scale images for the experiments in this paper. The results reveal a good compression ratio and improved embedding capacity.

2.2. Exploiting Modification Direction Method

The exploiting modification direction (EMD) [5] method was proposed by Zhang and Wang in 2006. This method can embed more secret message capacity than the 1-LSB replacement data hiding method. In EMD, two pixels in each group and each pixel value in the image only change once (-1, 0 or $+1$). Therefore, to achieve this condition, the following extraction function as Equation (1) is given in the Zhang and Wang scheme.

$$f(g_1, g_2, \ldots, g_n) = [\sum_{i=1}^{n}(g_i \times i)] \bmod (2n+1) \quad (1)$$

where g_i is the value of the pixel i and n is the number of pixels. For example, when $n = 2$, two pixels, g_1 and g_2, are considered. Therefore, the extract function is $f(g_1, g_2) = (1 \times g_1 + 2 \times g_2) \bmod 5$. According to their analysis, the best hiding bit rate is in five-ary. However, the secret embedding capacity decreases when the pixel number increases for each group. Specifically, the embedding capacity is less than 1 bpp (bits per pixel) when the pixel numbers are more than three for each group.

2.3. Generalized Exploiting Modification Direction

In order to improve the secret embedding capacity and to embed the binary secret data directly, Kuo and Wang proposed the data hiding method based on generalized exploiting modification direction (GEMD scheme) [13]. The main idea of the GEMD scheme is that each $(n + 1)$-bit binary secret message can be hidden into n adjacent pixels in the cover image. The new extraction function $f_b(g_1, g_2, \ldots, g_n)$ is defined as Equation (2):

$$f_b(g_1, g_2, \ldots, g_n) = [\sum_{i=1}^{n}(g_i \times (2^i - 1))] \bmod 2^{n+1} \quad (2)$$

2.4. Multi-Bit Generalized Exploiting Modification Direction

In 2012, Kuo *et al.* also proposed the multi-bit GEMD (MGEMD) [12] method to increase embedding capacity by using adaptive k. MGEMD can also choose different values of n to determine how many pixels in a group are used to hide secrets in k bits of each pixel and the ability to hide an extra pixel group

into one-bit information, *i.e.*, it can embed the secret messages' $(nk+1)$ bits. MGEMD's extraction function is shown as Equation (3):

$$f_c(g_1, g_2, \ldots, g_n) = [\sum_{i=1}^{n}(g_i \times c_i)] \bmod 2^{nk+1} \tag{3}$$

where the weight value of c_i is:

$$c_i = \begin{cases} 1, & i = 1. \\ 2^k c_{i-1} + 1, & i \neq 1 \quad and \quad i > 0 \end{cases} \tag{4}$$

For example, $c_1 = 1$, $c_2 = 9$, $c_3 = 73$, $c_4 = 585$ when $k = 3$, $n = 4$ from Equation (4). Obviously, the difference between the MGEMD scheme and GEMD scheme is that the modulus is changed from 2^{n+1} to 2^{nk+1} in order to increase embedding capacity for the MGEMD scheme.

3. The Proposed Scheme

As a rule, the goals of data hiding techniques are security, capacity, robustness, imperceptibility, unambiguousness and non-removability, respectively. Data hiding techniques are focused on increasing embedding capacity and high stego image quality. Obviously, significant differences between the original cover image and stego image will be generated when the capacity of embedded secrets increases. Thus, how to enhance the embedding capacity while still maintaining the original stego image quality is a very important issue. In order to give a solution to this issue, a high embedding capacity and good image quality scheme is proposed in this section.

3.1. Multi-Bit Data Hiding Scheme for Compressing Secret Messages

In data hiding schemes, unambiguousness means that the stego image was securely transmitted to the receiver and extracts the same secret message as embedded by the sender. To support this attribute, we need to employ a lossless compression method to allow us access to the original data. Fortunately, RLE is suitable, since it is simple, quick and lossless. There are three phases included in the proposed scheme. The flowchart of three phases (secret image compression phase, MGEMD phase and embedding phase) are shown in Figure 1.

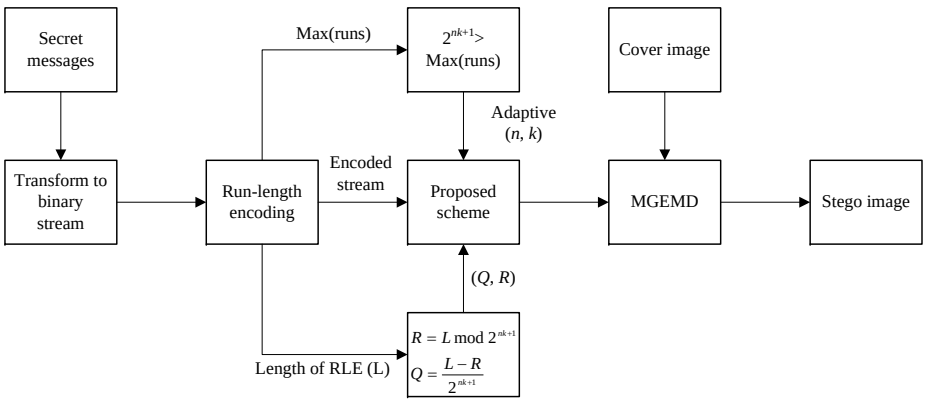


Figure 1. Flowchart of the proposed scheme.

3.2. Secret Image Compression Phase

In the proposed scheme, data compression is used to decrease the secret message size, which effectively increasing the embedding capacity. In order to minimize the cost of transmission in a limited bandwidth network, information needs to be compressed before delivery to improve transport efficiency. The proposed scheme is combined with MGEMD to increase embedding capacity. As a result, the multi-bit data hiding scheme for compressing secret messages has twice the embedding capacity. The maximum runs and total length information of RLE will be regarded as secret messages to hide in the cover image.

Algorithm 1 . The multi-bit data hiding scheme for compressing secret messages.

Input: A cover image and secret image (S) with gray level image
Output: A stego image ($I^{'}$)

Step 1. The gray level secret image is transformed into a binary stream.

Step 2. Compressing the binary stream by RLE for the new secret message ($S^{'}$), which includes maximum runs, total length information and the embedding secret messages (s).

Step 3. Check if the new secret is zero or one in the high-order bit. This information tells us the beginning bit is zero or one by using RLE. Then, the first pixel's LSB of the cover image will be changed. Simultaneously, we also count the zeros and ones to record the total length information.

Step 4. Find parameters (n and k), such as 2^{nk+1} > Max(runs) and $n \geq k$. Quotient (Q) and remainder (R) are calculated from total length (L) information using Equation (5).

$$\begin{aligned} R &= L \bmod 2^{nk+1} \\ Q &= \frac{R-L}{2^{nk+1}} \end{aligned} \tag{5}$$

Step 5. For the second pixel to the last, decision variables n and k divide the pixels into n adjacent pixels $(x_1, x_2, \ldots, x_n)$ as a non-overlapping group.

Step 6. Compute the value $t = f_c(x_1, x_2, \ldots, x_n)$ and the difference D, *i.e.*, $D = (s - t) \bmod 2^{nk+1}$.

Step 7. If $D = 0$, then cover pixels do not change;
else if $D = 2^{nk}$, then $x_n^{'} = x_n + 2^{nk}$ and $x_1^{'} = x_1 + 1$.
else if $D < 2^{nk}$, then $D = (d_{n-1}, d_{n-2}, \ldots, d_0)_{2^k}$, $x_{i+1}^{'} = x_{i+1} + d_i$
and $x_i^{'} = x_i - d_i$ for $i = n-1, n-2, \ldots, 0$.
else if $D > 2^{nk}$, then $D = 2^{nk+1} - D = (d_{n-1}, d_{n-2}, \ldots, d_0)_{2^k}$, $x_{i+1}^{'} = x_{i+1} - d_i$
and $x_i^{'} = x_i + d_i$ for $i = n-1, n-2, \ldots, 0$.

Step 8. Repeat Step 6 to Step 7 until all secret messages are hidden.

Example 1: Let $n = 3$ and $k = 3$. Given the cover image's pixels $(155, 155, 155, 158, \ldots)$, secret messages $s : (164, 91, 155, 247, \ldots)_{10}$. The secret messages are compressed with RLE and hidden in the cover pixels. Finally, we get the stego pixels = $(156, 156, 161)$ from Algorithm 1 by the following steps.

Step 1. Convert the secret message $s = (164, 91, 155, 247, \ldots)_{10}$ into the binary stream $(1010010001011011 \ldots)_2$.

Step 2. Use RLE to compress the secret stream, $s = [1(1)0(1)1(1)0(2)1(1)0(3)1(1)0(1)$
$1(2)0(1)1(2)0(1)1(1)0(1)1(2)0(1)1(3)0(2)1(2) \ldots]$.
The new secret s' = $(1112131121211121322 \ldots)$ from the s value sequence and the RLE begins at one.

Step 3. The least significant bit of the first pixel is equal to one, meaning it is not modified, *i.e.*, the first pixel is still 155.

Step 4. Compute the value $t = f_c(155, 155, 158) = 796$ and $D = (1 - 796) \bmod 2^{10} = 229 = (345)_8$.

Step 5. Since $229 < 2^9$ and $(d_2, d_1, d_0)(3, 4, 5)$, we can compute the stego pixels by using the following equation.
For $d_2 = 3$, compute $x_3^{'} = 158 + 3 = 161$ and $x_2^{'} = 155 - 3 = 152$;
For $d_1 = 4$, compute $x_2^{'} = 152 + 4 = 156$ and $x_1^{'} = 155 - 4 = 151$;
For $d_0 = 5$, compute $x_1^{'} = 151 + 5 = 156$.

The stego pixels are $(156, 156, 161)$.

3.3. Data Embedding

Before embedding, secret messages must be transformed to a binary stream. Then, the binary stream uses RLE lossless compression to reduce the data size. Finally, the compressed binary stream is hidden by the MGEMD scheme. The algorithm of the proposed scheme is shown in Algorithm 1.

3.3.1. Speeding up the Modified Method

In this subsection, we describe MGEMD features and then use these characteristics to speed up the embedding process. The MGEMD scheme groups the cover pixels into three categories for computation, *i.e.*, $D < 2^{nk}$, $D = 2^{nk}$ and $D > 2^{nk}$. In order to speed up the embedding speed, we propose the embedding formulas shown as Tables 1 and 2 for $D < 2^{nk}$ and $D > 2^{nk}$, respectively. Now, we assume that $s_1 = 5429$ and $s_2 = 6643$ when $k = 3$, $n = 4$ and have both cover pixels of $(10, 19, 5, 9)$ in Tables 1 and 2, respectively.

Table 1. Speeding up the embedding method when $D > 2^{nk}$.

Item	Formula	Example
$s_1 = 5429$	$D = (s - f_c) \bmod 2^{nk+1}$	$D = 7810$
$D > 2^{nk}$	$D = 2^{nk+1} - D$	$382 = 8192 - 7810$
$D_{10} \to d_2$	$d = (d_3 d_2 d_1 d_0)_8$	$(0576)_8$
(x_1, x_2, x_3, x_4)	(x_1, x_2, x_3, x_4)	$(10, 19, 5, 9)$
$d_3 d_2 d_1 d_0$	$(0 - d_3)(d_3 - d_2)$ $(d_2 - d_1)(d_1 - d_0)$	$(0-0)(0-5)(5-7)(7-6)$ $0, -5, -2, 1$
$(x_1^{'}, x_2^{'}, x_3^{'}, x_4^{'})$	$(x_1 + d_0, x_2 + d_1, x_3 + d_2, x_4 + d_3)$ $(10 + 1, 19 - 2, 5 - 5, 9 + 0)$	$(11, 17, 0, 9)$

Table 2. Speeding up the embedding method when $D < 2^{nk}$.

Item	Formula	Example
$s_2 = 6643$	$D = (s - f_c) \bmod 2^{nk+1}$	$D = 832$
$D < 2^{nk}$	$D = D$	832
$D_{10} \to d_2$	$d = (d_3 d_2 d_1 d_0)_8$	$(1500)_8$
(x_1, x_2, x_3, x_4)	(x_1, x_2, x_3, x_4)	$(10, 19, 5, 9)$
$d_3 d_2 d_1 d_0$	$(d_3)(d_2 - d_3)$ $(d_1 - d_2)(d_0 - d_1)$	$(1)(5-1)(0-5)(0-0)$ $1, 4, -5, 0$
$(x_1^{'}, x_2^{'}, x_3^{'}, x_4^{'})$	$(x_1 + d_0, x_2 + d_1, x_3 + d_2, x_4 + d_3)$ $(10 + 0, 19 - 5, 5 + 4, 9 + 1)$	$(10, 14, 9, 10)$

3.3.2. The Solution for the Overflow/Underflow Problems

Unfortunately, in the embedding process, overflow/underflow problems may occur in the stego pixel values after applying MGEMD. That is to say, the stego pixel values may exceed the maximal value 255

or may be smaller than the minimal value zero for the gray-level image. Therefore, the proposed scheme provides a scheme to address this problem. In our scheme, 2^k bits are embedded in each cover pixel, where the pixel value of the cover image can fall between zero and $2^k - 1$. In order to avoid problems, we can modify the pixel value to $[0, 1, \ldots, 2^{k-2}]$ and $[255 - (2^k - 2), \ldots, 255]$, respectively. Therefore, the value of the cover pixel will be modified to $2^k - 1$ when its value is between zero and $2^k - 2$. Similarly, the value of the cover pixel will be modified to $255 - (2^k - 1)$ when its value is between 255 and $255 - (2^k - 2)$. Hence, this solves the overflow/underflow problems during the embedding phase.

3.4. Data Extracting Phase

In the data extraction process, some information needs to be coordinated with the sender, such as the value of n and k. Next, the stego image can be transmitted from the sender to the receiver. The receiver is able to recover the secret message using the following steps:

Algorithm 2 : Data extracting.

Input: A stego image ($I^{'}$)
Output: The secret image (S)

Step 1. Check the first pixel's LSB for zero or one.

Step 2. The second pixel to the last pixel of the stego image are divided into non-overlapping groups of n adjacent pixels $(x_1, x_2, \ldots, x_n)$; then, we compute secret value t by using Equation (3).

Step 3. The decoded binary stream (RLE) will be transformed into an eight-bit gray-level value to reconstruct the secret image.

Example 2: From the results of Example 1, the receiver receives the stego image $(155, 156, 156, 161, \ldots)$ and $(n, k) = (3, 3)$ from the sender. After the data extraction process, the secret message can be recovered, and the secret image can be reconstructed. After following these steps, we can extract the hidden secret images of the stego image.

Step 1. Let the pixels of the stego image be $(155, 156, 156, 161, \ldots)$. The first pixel 155 is odd, meaning the LSB of 155 is one.

Step 2. From Step 1, we know the sequence of RLE begins at one.

Step 3. All stego pixels are divided into three adjacent pixels as a non-overlapping group; then, each group will use Equation (3) to compute the value t. The extracted value is $(1112131121211121322\ldots)$. Therefore, the RLE sequence is $[1(1)0(1)1(1)0(2)1(1)\ 0(3)1(1)0(1)\ldots]$.

Step 4. Decoding RLE sequence $[1(1)0(1)1(1)0(2)1(1)0(3)1(1)0(1)\ldots]$ into a binary stream $(1010010001011011\ldots)$. Transform the binary stream into an eight-bit gray-level image to reconstruct the secret image.

4. Experimental Results

In this section, experimental results are given to verify the proposed scheme's embedding capacity in terms of the increased rate, the image quality of the stego images, the decrease of pixels modified rate and image steganalysis. From these experiments, eight common eight-bit gray-level images and binary images of size 512×512 are tested. These eight cover images are airplane, baboon, boat, Elaine, Gold Hill, Lena, peppers and Tiffany, as shown in Figure 2. Two gray-level secret messages (bridge and pentagon) of size 100×100 are shown in Figure 3. Similarly, the eight binary images are shown in Figure 4. In the maximum embedding capacity test, we also use the random number generator (PRNG) to generate random bits as the secret message to be hidden.

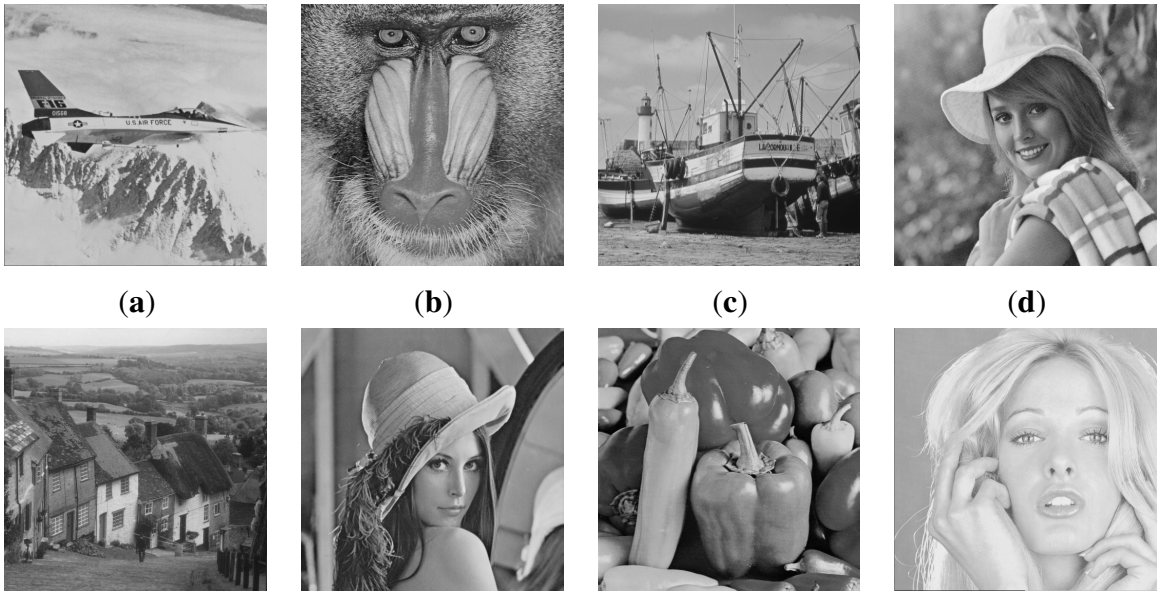


Figure 2. Eight 512×512 gray test images. (**a**) F16; (**b**) baboon; (**c**) boat; (**d**) Elaine; (**e**) Gold Hill; (**f**) Lena; (**g**) pepper; (**h**) Tiffany.

Figure 3. Secret images.

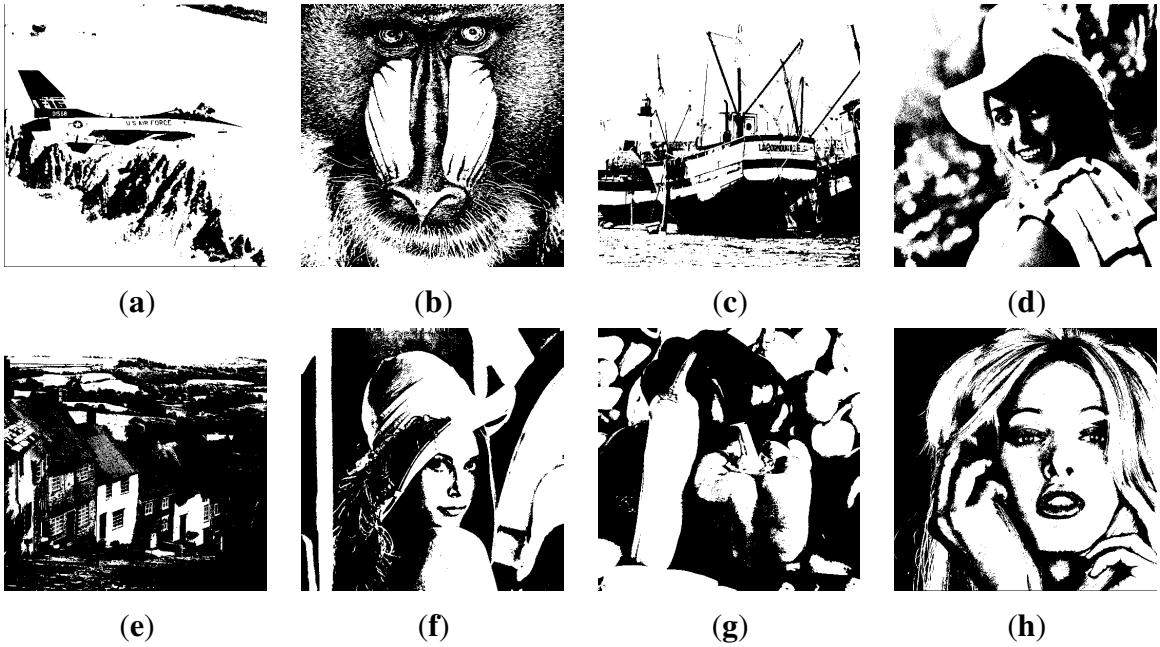
The embedding capacity (bpp) and image quality (PSNR: peak signal-to-noise ratio) are two important criteria to evaluate the stego image in the data hiding system. The PSNR is defined as Equation (6):

$$PSNR = 10 \times \log_{10}(\frac{255^2}{MSE}) \quad (6)$$

where the MSE (mean square error) is defined as Equation (7).

$$MSE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (C(i,j) - S(i,j))^2 \tag{7}$$

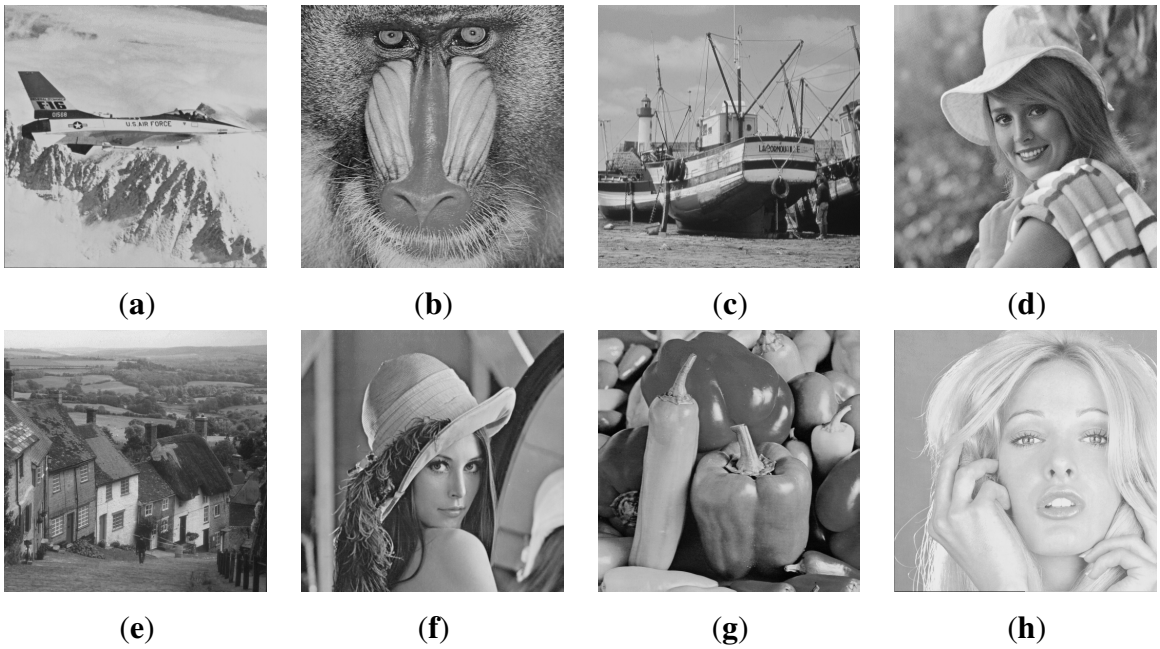
where M, N is the image size and $C(i,j)$ and $S(i,j)$ are the pixel values of the stego image and the original image (cover image), respectively.



(**a**) (**b**) (**c**) (**d**)

(**e**) (**f**) (**g**) (**h**)

Figure 4. Eight binary images. (**a**) F16; (**b**) baboon; (**c**) boat; (**d**) Elaine; (**e**) Gold Hill; (**f**) Lena; (**g**) pepper; (**h**) Tiffany.



(**a**) (**b**) (**c**) (**d**)

(**e**) (**f**) (**g**) (**h**)

Figure 5. Stego images (proposed scheme $n = 3, k = 3$). (**a**) F16; (**b**) baboon; (**c**) boat; (**d**) Elaine; (**e**) Gold Hill; (**f**) Lena; (**g**) pepper; (**h**) Tiffany.

The eight stego images are processed by the proposed method and MGEMD as shown in Figures 5 and 6, respectively. Simultaneously, we compare the PSNR and non-modified pixels between our proposed scheme and MGEMD. The results are shown in Tables 3 and 4 when the embedding capacity is $80,000$ bits and $262,144$ bits, respectively. Furthermore, the proposed scheme also compares to [15,18] for the payload of about one hundred thousand, two hundred thousand and four hundred thousand binary images, respectively, shown in Table 5. From the comparison tables, we can find that our proposed scheme has high capacity and good image quality compared to the other schemes. According to Figure 5, there are no human perceivable differences between the cover images and stego images using our proposed scheme. From Table 3, the proposed scheme has better image quality than the MGEMD scheme. Conversely, the modified pixels of the stego image used by the proposed scheme are less than the MGEMD scheme.

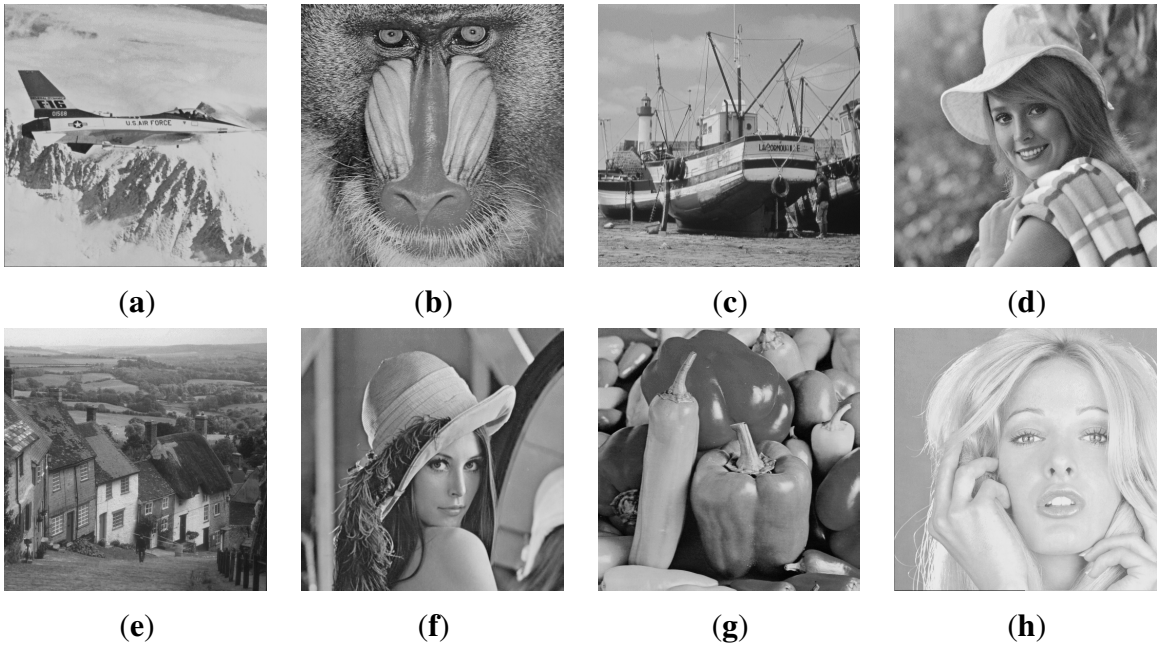


Figure 6. Stego images (MGEMD $n = 3, k = 3$). (**a**) F16; (**b**) baboon; (**c**) boat; (**d**) Elaine; (**e**) Gold Hill; (**f**) Lena; (**g**) pepper; (**h**) Tiffany.

Table 3. Comparison table for the PSNR and non-modified pixels.

Method	Item	F-16	Baboon	Boat	Elaine	Gold Hill	Lena	Pepper	Tiffany
Proposed	PSNR (dB)	46.93	44.62	44.84	44.68	44.56	44.53	44.57	44.52
Scheme	non-modified (pixel)	238,872	221,987	222,598	222,046	221,895	222,121	222,054	221,880
MGEMD	PSNR (dB)	42.21	42.20	42.30	42.22	42.20	42.15	42.14	42.08
Scheme	non-modified (pixel)	192,284	191,912	192,564	191,993	192,151	192,253	191,920	191,885

Table 4. PSNR and non-modified pixels of binary image hiding.

Binary Image	Item	Method	F-16	Baboon	Boat	Elaine	Gold Hill	Lena	Pepper	Tiffany
F-16	PSNR (dB)	Proposed	51.55	51.59	51.59	51.61	51.42	51.55	51.66	51.29
		MGEMD	41.79	41.80	41.96	41.83	41.84	41.76	41.77	41.69
	non-modified pixel	Proposed	254,072	254,067	254,093	254,079	254,023	254,082	254,093	253,976
		MGEMD	185,830	185,397	186,226	185,592	185,790	185,795	185,414	185,328
Baboon	PSNR (dB)	Proposed	44.59	44.58	44.79	44.64	44.47	44.53	44.53	44.53
		MGEMD	41.79	41.80	41.95	41.82	41.83	41.77	41.76	41.76
	non-modified pixel	Proposed	221,746	221,671	221,948	221,743	221,569	221,799	221,548	221,548
		MGEMD	185,796	185,485	186,242	185,542	185,774	185,735	185,456	185,456
Boat	PSNR (dB)	Proposed	50.65	50.56	50.58	50.65	50.14	50.48	50.59	50.33
		MGEMD	41.79	41.81	41.97	41.82	41.82	41.76	41.77	41.69
	non-modified pixel	Proposed	251,898	251,957	251,933	251,995	251,808	251,920	251,888	251,747
		MGEMD	185,811	185,418	186,225	185,512	185,650	185,834	185,503	185,366
Elaine	PSNR (dB)	Proposed	48.13	48.20	48.23	48.25	48.02	48.08	48.23	47.97
		MGEMD	41.80	41.80	41.96	41.83	41.82	41.76	41.76	41.70
	non-modified pixel	Proposed	244,532	244,614	244,661	244,653	244,657	244,657	244,548	244,340
		MGEMD	185,791	185,309	186,177	185,497	185,667	185,831	185,396	185,367
Gold Hill	PSNR (dB)	Proposed	48.51	48.59	48.60	48.59	48.37	48.51	48.61	48.36
		MGEMD	41.79	41.80	41.93	41.83	41.83	41.79	41.76	41.70
	non-modified pixel	Proposed	246,027	246,120	246,198	246,057	246,062	246,159	246,088	245,940
		MGEMD	185,802	185,420	186,190	185,555	185,742	185,829	185,333	185,409
Lena	PSNR (dB)	Proposed	51.64	51.65	51.65	51.66	51.34	51.51	51.61	51.50
		MGEMD	41.79	41.80	41.93	41.83	41.83	41.78	41.76	41.70
	non-modified pixel	Proposed	254,132	254,123	254,126	254,127	254,072	254,110	254,107	254,041
		MGEMD	185,775	185,401	186,198	185,520	185,773	185,795	185,311	185,388
Pepper	PSNR (dB)	Proposed	52.66	52.76	52.60	52.69	52.42	52.65	52.64	52.47
		MGEMD	41.80	41.81	41.95	41.82	41.82	41.76	41.76	41.70
	non-modified pixel	Proposed	255,739	255,816	255,810	255,764	255,790	255,796	255,801	255,748
		MGEMD	185,743	185,409	186,155	185,516	185,764	185,756	185,400	185,367
Tiffany	PSNR (dB)	Proposed	49.11	49.13	49.14	49.16	48.91	49.04	49.11	48.90
		MGEMD	41.80	41.80	41.93	41.82	41.84	41.76	41.75	41.71
	non-modified pixel	Proposed	247,911	247,934	247,975	247,967	247,840	247,955	247,911	247,732
		MGEMD	185,824	185,433	186,138	185,567	185,819	185,760	185,264	185,386

Table 5. Comparison of the PSNR values using binary image to hiding with different payloads.

Payload	Proposed Scheme	BRL-Scheme [15]			LSW [18] Scheme
		$^{\star}k = 4$	$k = 6$	$k = 8$	
100,000	60.12	51.13	48.46	46.32	49.85
200,000	56.85	48.12	45.46	43.32	48.85
400,000	53.69	45.11	42.45	40.31	43.85

$^{\star}k$ is the runs length of the bitmap files by run-length (BRL) scheme.

4.1. Maximum Embedding Capacity

In this subsection, we allow all pixels of the cover image to be embedded. From the simulation results, the RLE compression ratio of the binary image is better than the gray-scale image, because the binary image has many continuous zeros or ones to increase the compression ratio. The test cover

images are shown as Figure 2, and the secret messages were generated from PRNG. For example, if we use the MGEMD scheme (when $n = 3$ and $k = 3$) with non-compressed secret messages, then the group numbers are 87,381 ($512 \times 512 \div 3$) with the image size of 512×512. The embedding fixed secret capacity is 878,310 (87,381 $\times(nk+1) = 87{,}831 \times 10$) bits. However, according to our simulation results, the secret embedding capacity is 1,757,500 bits using our proposed scheme. That is to say, the embedding performance of the proposed scheme is 200% better than without using RLE. Furthermore, using the GMEMD method, we compare the embedding capacity with the compressed secret data or those that are not. The result is shown in Figure 7.

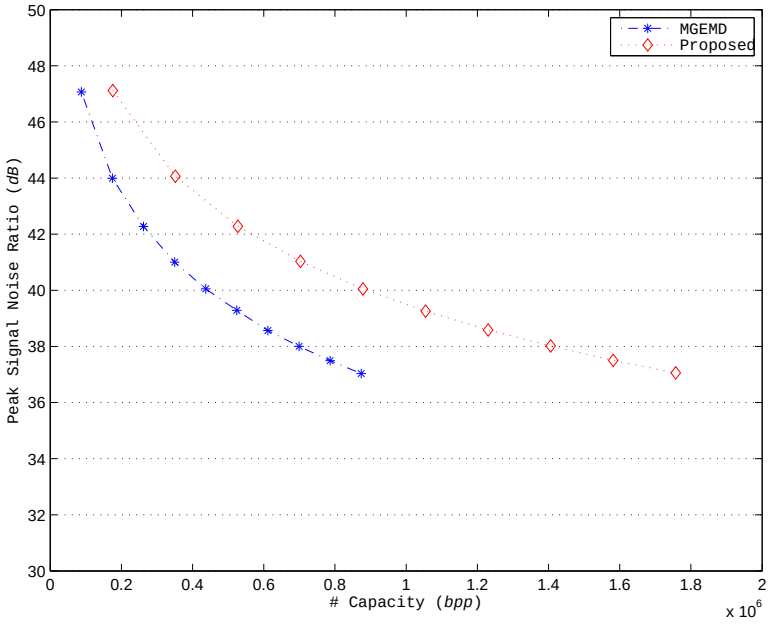


Figure 7. PSNR comparison result.

4.2. Image Steganalysis

In general, steganalysis for the stego image cannot be judged by the human eye. Because the stego image's quality is more than 33 dB, the human eye is unable to detect if the stego image has secrets or not. Therefore, it is necessary to have an effective way to detect whether a secret message is embedded. For image steganalysis, the regular singular steganalysis (RS steganalysis) is a common method to detect the stego image. In this paper, in order to prove that our proposed scheme has good security, we propose this detection method to test the stego image generated by our scheme. In RS steganalysis, n adjacent pixels $(x_1, x_2, \ldots, x_n)$ are selected as a pixel group. Then, the discrimination function DF, defined as $DF(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} |x_{i+1} - x_i|$, is applied to quantify the smoothness or regularity of each pixel group. The flipping function of RS steganalysis is used to define three types of pixel groups: regular (R), singular (S) and unusable (U). The percentages of all groups of regular and singular with masks $M = [1001]$ and $-M = [-100-1]$ are represented as R_m, R_{-m}, S_m and S_{-m}. The statistical hypotheses of RS steganalysis are $R_m \cong R_{-m}$ and $S_m \cong S_{-m}$, meaning the test image can pass steganalysis. In other words, R_m overlaps with R_{-m}, and S_m overlaps S_{-m}. Using the RS steganalysis method, we test a method based on LSB and our proposed method. The simulation results are shown in

Figure 8. According to Figure 8a to 8d, the distribution of $R_m \cong R_{-m}$ and $S_m \cong S_{-m}$ for our proposed scheme is normal, but 1-LSB, 2-LSB and 3-LSB denote hidden information; because the 1-LSB, 2-LSB and 3-LSB methods mean that the secret message was hidden in the position of 1-LSB, 2-LSB and 3-LSB for each cover image's pixel, respectively. In addition, we evaluate our proposed method with the modern steganalysis tool SPAM (subtractive pixel adjacency matrix) [19]. For the SPAM test, we use 500 image data from [20]. There are 250 cover images and 250 stego images used for training. The test results between minimizing the power of the optimal detector (MiPOD) [21] and our proposed scheme are shown in Figure 9. The MiPOD scheme was proposed by Sedighi *et al.* in 2015. The major contribution of the MiPOD scheme is that the pixel values are changed by at most ± 1 when a secure message is embedded into the cover image with the Gaussian cover model [22,23], and Sedighi *et al.* also considered a novel detectability-limited sender and estimated the secure payload of each cover image. In Figure 9, the vertical axis represents the error rate obtained by the SPAM method, and the horizontal axis represents the embedding rate (bpp), which ranges from 0.1 bpp to 6.7 bpp. From Figure 9, we can find that the error rate of the proposed method is better than the MiPOD scheme when the embedding capacity is over 0.6 bpp. In comparison, the proposed method has a higher embedding rate and a lower error rate. Therefore, if we can adapt the secret data embedding rate (such as MiPOD scheme); this can provide a certain probability to avoid detection attack. Furthermore, to understand steganalysis and stego image security, the spatial rich model (SRM) [24] or maxSRM [25] method can be used to analyze the proposed method in the future.

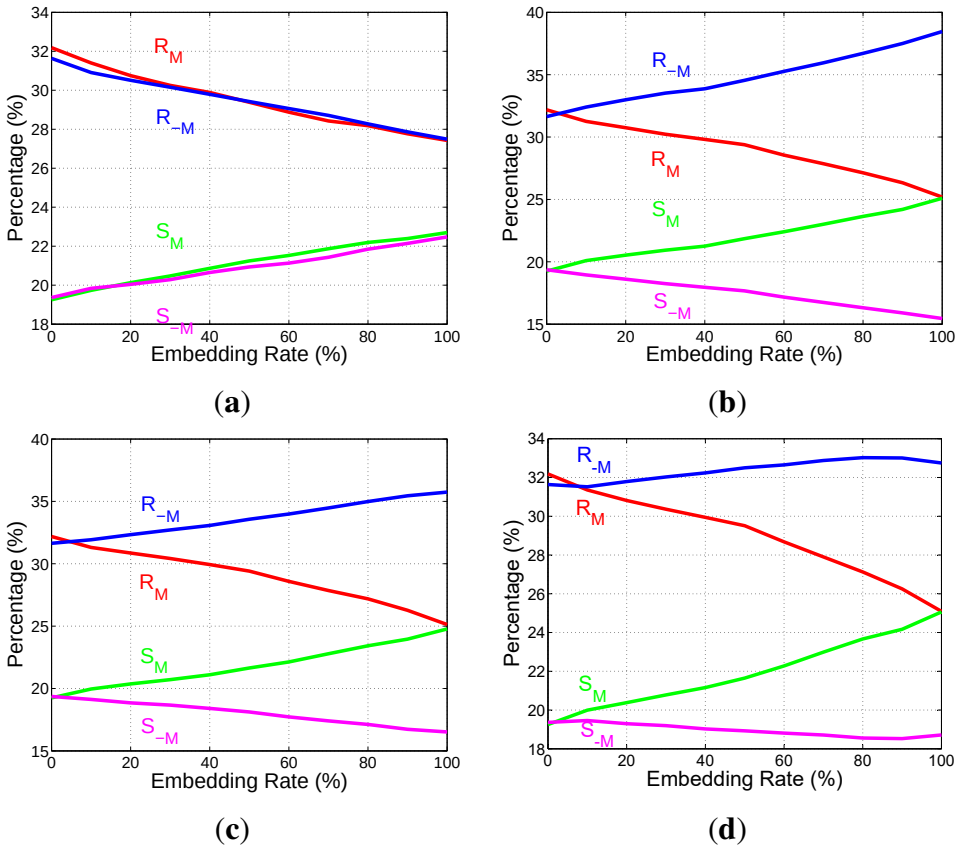


Figure 8. RS steganalysis. (**a**) Proposed scheme; (**b**) 1-least significant bit (LSB); (**c**) 2-LSB; (**d**) 3-LSB.

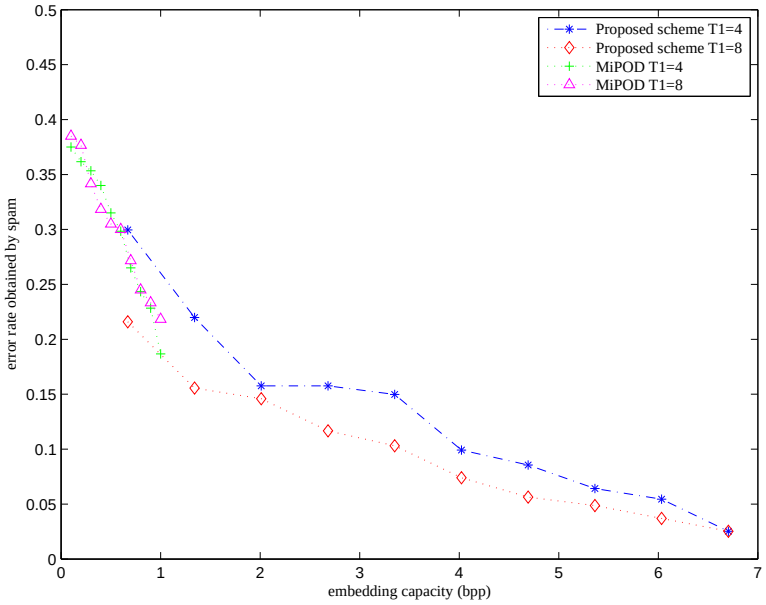


Figure 9. The simulation error rate between minimizing the power of the optimal detector (MiPOD) and the proposed scheme for subtractive pixel adjacency matrix (SPAM).

5. Conclusions

A new data hiding technology is proposed in this paper, which combines the multi-bit data hiding scheme for compressing secret messages and a quicker operation for MGEMD. From the simulation results, we show that the proposed scheme can increase embedding capacity because the RLE compression ratio is high, though dependent on the source data, *i.e.*, the binary images had a better compression ratio than the gray-scale images. Consequently, the quality of the stego image generated by our proposed scheme is not only better than the MGEMD method, but also modifies fewer pixels when the length of the secret message is the same. Additionally, according to our steganalysis and performance discussion, our scheme provides a higher embedding capacity than previous approaches, and it also can resist visual attack and RS steganalysis and SPAM.

Acknowledgments

This work was supported in part by the Ministry of Science and Technology of the Republic of China under Contract No. MOST 104-2221-E-224-023 and MOST 104-2221-E-492 -014 -MY2.

Author Contributions

All authors discussed the contents of the manuscript and contributed to its preparation. Wen-Chung Kuo and Lih-ChyauWuu conceived and designed the proposed scheme; Shao-Hung Kuo performed the experiments and wrote the paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Petitcolas, F.A.P.; Anderson, R.J.; Kuhn, M.G. Information Hiding – A Survey. *Proc. IEEE* **1999**, *87*, 1062–1078.
2. Cox, I.J.; Miller, M.L.; Bloom, J.A.; Fridrich, J.; Kalker, T. *Digital Watermarking and Steganography*, 2nd ed.; Morgan Kaufmann: San Francisco, CA, USA, 2008.
3. Swanson, M.D.; Kobayashi, M.; Tewfik, A.H. Multimedia Data-Embedding and Watermarking Technologies. *Proc. IEEE* **1998**, *86*, 1064–1087.
4. Ker, A. Improved Detection of LSB Steganography in Grayscale Images. *Inf. Hiding Lect. Notes Comput. Sci.* **2004**, *3200*, 97–115.
5. Zhang, X.; Wang, S. Efficient Stegano graphic Embedding by Exploiting Modification Direction. *IEEE Comm. Lett.* **2006**, *10*, 1–3.
6. Sayood, K. *Introduction to Data Compression*, 4th ed.; Morgan Kaufmann: San Francisco, CA, USA, 2012.
7. Ma, X.; Pan, Z.; Hu, S.; Wang, L. Reversible Data Hiding Scheme for VQ Indices based on Modified Locally Adaptive Coding and Double-Layer Embedding Strategy. *J. Vis. Commun. Image Represent.* **2015**, *28*, 60–70.
8. Qin, C.; Chang, C.C.; Chen, Y.C. A Novel Reversible Data Hiding Scheme for VQ-Comp ressed Images Using Index Set Construction Strategy. *KSII Trans. Internet Inf. Syst.* **2013**, *7*, 2027–2041.
9. Wang, W.J.; Huang, C.T.; Yang, C.H.; Wang, S.J. VQ-based Algorithms Extended to Non-Embedded Watermarking for Multimedia Ownership Prevention Systems. *Peer-to-Peer Netw. Appl.* **2014**, *7*, 676–686.
10. Chen, L.S.T.; Lin, J.C. Steganography Scheme Based on Side Match Vector Quantization. *Opt. Eng.* **2010**, *49*, 037008.
11. Shie, S.C.; Jiang, J.H. Reversible and High-Payload Image Steganographic Scheme based on Side-Match Vector Quantization. *Signal Process.* **2012**, *92*, 2332–2338.
12. Kuo, W.C.; Wuu, L.C.; Kuo, S.H. The High Embedding Steganographic Method based on General Multi-EMD. In Proceedings of the 2012 International Conference on Information Security and Intelligent Control (ISIC'12), Yunlin, Taiwan, 14–16 August 2012; pp. 286–289.
13. Kuo, W.C.; Wang, C.C. Data Hiding based on Generalised Exploiting Modification Direction Method. *Imaging Sci. J.* **2013**, *61*, 484–490.
14. Run-length encoding. Wikipedia Archive. Available online: http://en.wikipedia.org/wiki/Run-length_encoding (accessed on 1 July 2015).
15. Chang, C.C.; Lin, C.Y.; Wang, Y.Z. New Image Steganographic Methods Using Run-Length Approach. *Inf. Sci.* **2006**, *176*, 3393–3408.
16. Chang, C.C.; Tseng, H.W. A Stegano graphic Method for Digital Images Using Side Match. *Pattern Recognit. Lett.* **2004**, *25*, 1431–1437.
17. Againan, S.S.; Cherukuri, R.C. Run Length Based Steganography for Binary Images. *Pattern Recognit. Mach. Intell. Lect. Notes Comput. Sci.* **2005**, *3776*, 481–484.
18. Lee, C.F.; Weng, C.Y.; Sharma, A. Stega nographic Access Control in Data Hiding Using Run Length Encoding and Modulo Operations. *Secur. Commun. Netw.* **2011**, doi:10.1002/sec.333.

19. Pevn, T.; Bas, P.; Fridrich, J. Steganalysis by subtractive pixel adjacency matrix. *IEEE Trans. Inf. Forensics Secur.* **2010**, *5*, 215–224.
20. The USC-SIPI Image Database. Available online: http://sipi.usc.edu/database/ (accessed on 1 July 2015).
21. Sedighi, V.; Cogranne, R.; Fridrich, J. Content-Adaptive Steganography by Minimizing Statistical Detectability. *IEEE Trans. Inf. Forensics Secur.* **2015**, doi:10.1109/TIFS.2015.2486744.
22. Fridrich, J.; Kodovsky, J. Multivariate Gaussian model for designing additive distortion for steganography. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, BC, Canada, 26–31 May 2013; pp. 2949–2953.
23. Sedighi, V.; Fridrich, J.; Cogranne, R. Content-adaptive pentary steganography using the multivariate generalized Gaussian cover model. In *SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics*, San Francisco, CA, USA, 4 March 2015; pp. 94090H.
24. Fridrich, J.; Kodovsk, J. Rich models for steganalysis of digital images. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 868–882.
25. Denemark, T.; Sedighi, V.; Holub, V.; Cogranne, R.; Fridrich, J. Selection-channel-aware rich model for steganalysis of digital images. In Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS), Atlanta, GA, USA, 3–5 December 2014; pp. 48–53.