A Modification-Free Steganography Method Based on Image Information Entropy

Xia ShuangKui\textsuperscript{1} and Jianbin Wu\textsuperscript{2}

\textsuperscript{1}School of Electronic Information, Wuhan University, Wuhan, China
\textsuperscript{2}College of Physical Science and Technology, Huazhong Normal University, Wuhan, China

Correspondence should be addressed to Jianbin Wu; wujianbin@mail.ccnu.edu.cn

Received 29 September 2017; Revised 4 December 2017; Accepted 6 February 2018; Published 26 March 2018

1. Introduction

At present, the transmission of important data on the secure communication mainly relies on the cryptography. Cryptography techniques are aimed at encrypting the data in order to ensure its security, thus making it incomprehensible for an adversary. However, the encryption technology has an unavoidable shortage that it clearly indicates the existence of the important data and then easily attracts the attacker’s attention.

On the other hand, steganography seeks to provide a covert communication channel between two parties. A common class of steganographic algorithms embeds the secret message in cover works such as images, video, audio, or text. The combination of cover work and secret message is referred to as the stego work and a goal of all steganographic algorithms is to ensure imperceptibility, robustness, capacity, and security. Digital images as important carrier information are widely used in steganography. Presently steganography has achieved lots of research results, mainly in spatial and frequency domain of the image [1–7]. LSB (Least Significant Bit) replacement algorithm is proposed in [1]. The author of [2] proposes an algorithm based on gray level modification and multilevel encryption, which improves the image quality of the carrier and presents a greater challenge to the steganalysis. In literature [3], a steganography algorithm based on run length is proposed. The algorithm has good properties to avoid detection of LSB. Some literature [4–10] discusses the steganography of the frequency domain. In literature [5], the author combines the chaotic sequences and information hiding to propose a DCT (Discrete Cosine Transform) domain method based on Logistic map and achieves good robustness. The algorithm of information hiding based on the discrete wavelet transform is studied in [6]. Although the above steganography research has achieved some good properties, there is a fatal defect that if the adversaries obtain the original image, they can judge whether the carrier image has been modified and then detect the secret information because all algorithms discussed in references must modify the original image according to their principles.
In this paper, the modification-free steganography algorithm based on image information entropy (MFSA, modification-free steganography algorithm) is proposed, which integrates big data and turbo coding technology. The MFSA establishes a relationship between the image features and the payload information. Through collecting, filtering, and cleaning of big data of network image, these images were divided into different classes according to the relevance of the content of the image, such as landscapes, cars, and small animals. The purpose of classification is to avoid noncooperative concerns. Then, we extract the entropy of these selected images and establish entropy matrix. Finally, the random occult information is mapped onto images according to certain algorithms, so as to construct a complete feature library; thus, we can conduct secret communication without modifying the original images. Because the selected images are hot, highly relative, they will not draw the attention of noncooperative side, and a truly safe and secret communication can be achieved. At the same time, the MFSA also uses error correction coding to further improve the robustness of the secret communication system.

The rest of the paper is organized as follows. Section 2 proposes the MFSA algorithm and elaborates the principle and process of the algorithm from the feature libraries establishment and information extracting, respectively. Section 3 includes the experiments and discusses the performance of the MFSA algorithm, in which the imperceptibility, robustness, and safety are discussed, respectively. Finally, conclusions are drawn in Section 5.

2. The Proposed Algorithm

Given the lack of imperceptible yet robust steganography algorithms, the proposed algorithm adopts a different approach to achieve steganography. The overall block diagram of random number visualization representation system is shown in Figure 1, which includes a complete sender processing flow, a communication channel, and a receiver processing flow.

The random number at the sending end is encrypted to the corresponding ciphertext information. Because the information may be affected by the noise in the process of transmitting, which will result in the error of the binary information, the turbo error correction coding technique is used for correction in the system. Then the turbo encoded information sequence will be the business information of the sending side. Big data acquisition system based on the Java environment is a system that acquires the hot image from network, using network crawler [11] to search and filter the big data. The proposed algorithm is based on a hybrid feature extraction mechanism of the hot image. A complete image feature library is constructed; the follow-up process only needs to update the feature library from time to time. The selected image set is turned into a visual animation in accordance with the control information; then the animation is sent to the receiver through the public communication, and the receiver gets the complete secret information to be delivered by using image resolution, control information extraction, algorithm analysis, turbo decoding, decryption, and other operations. The mapping algorithm module is one of the core parts of the whole system. The modification-free steganography algorithm based on the information entropy of the image not only has the characteristics of high security and robustness, but also establish a mapping relationship between the feature of image and the secure data. Based on the idea of zero-steganography secret communication, the algorithm considers the image information entropy as the starting point and implements the grid description of the image, the extraction of the entropy, the reduction of the entropy matrix, quantization, and a series of mathematical
operations in order to establish a mapping relationship between images and the secure data for secret communication. The use of this kind of hybrid approach for characteristic matrix generation provides a decent robustness against common attacks as well as sufficient capacity. The following section will focus on the principles and implementation of MFSA algorithm.

Suppose that a single image has a capacity of \( C \) bits; the number of images required to build a complete image library is \( 2^C \). The length of the payload information is \( L \) bits, the number of images sent each time is \( N = L/C \) (\( N \) upload rounding), the size of each image is \( p \) (MB), and the capacity of complete image library is \( S = p \cdot 2^C \) (MB). In this paper experiment, the value of parameter \( c \) is 32; with consideration of computing power and practical application, the value of parameter \( p \) is 0.1.

2.1. The Principle of MFSA Algorithm. There are some features of the image such as color, brightness, histogram, and entropy; by extracting and quantifying, the certain image can express some binary bit sequences. However, some image feature space such as histogram has high dimensionality and poor antinoise ability. The image information entropy is a quantitative description of the image characteristics. It can be seen from the information theory that the information entropy can represent the amount of information contained in the image. From the perspective of image information entropy, the mapping relation between the entropy and the payload information is constructed in this paper, the information entropy of the image is used to represent the payload information, and the zero-steganography covert communication is achieved. Later, the construction method of image information entropy matrix is introduced in detail and the basic principles and implementation of the algorithm will be discussed.

The information entropy is defined as the mathematical expectation of random variables \( I(x) \) in the set \( \{X, q(x)\} \).

\[
H(x) = -\sum_{x \in X} q(x) \log q(x) .
\]

(1)

\( H(x) \) is called the information entropy; \( q(x) \) is the probability of occurrence of \( X \). In the grayscale image, each pixel can be treated as an argument \( m \) (0 to 255). The pixels of the entire image can be viewed as a collection \( \{m, p(m)\} \), \( p(m) \) is the probability density of the point where the gray value is \( m \), and the expression of the information image entropy \( F \) can be obtained, as shown in

\[
F = \sum_{m=0}^{k} p(m) \log (p(m)) .
\]

(2)

\( k \) represents the gray value of the pixel, \( k = 255 \), \( p(m) \) represents the probability density of the pixel values appearing in the entire image.

In (2), the information entropy of the image discussed refers to the global information entropy, which characterizes the statistical distribution of all the pixels of the entire image. The spatial distribution of the image pixels is not taken into account; this will result in different images with the same probability distribution having the same information entropy. In order to solve the problem and use the spatial image information, this paper presents the concept of unit entropy. The grid descriptor is introduced, and Figure 2 shows a grid description of size 16, which maps the original image to the grid, so that any image with different size becomes a \( n \times n \) (\( n = 16 \) phalanx. The information entropy of each grid unit in Figure 2 is obtained by using formula (2); finally we get a \( 16 \times 16 \) entropy matrix \( E \), which is shown in formula (3).

The \( n \times n \) entropy matrix needs dimensionality reduction in order to reduce the redundancy, and the reduced eigenvalue vector \( (\alpha_1, \alpha_2, \ldots, \alpha_n) \) is obtained. Since the resulting eigenvalue vector is a floating-point number, we can get the binary bit sequence \( \beta \) after quantization. The index is the binary vector which the carrier image can express actually.

2.2. The Implementation of MFSA Algorithm. Section 2.1 describes the basic principles of the algorithm, which plays an important role in the construction of the complete feature library shown in Figure 1. The implementation flowchart of the algorithm is shown in Figure 3. The size of the selected single image will not exceed 100 kb and the format of the selected images adopts the most popular JPEG in the network. And the measurement of the relevance of images according to the fractal feature of images classification algorithm proposed in [12] is used to select image needed. The algorithm can classify landscape images, artificial drawing images, and computer-generated images, so that the required images achieve a high degree of relevance, giving a very
natural feeling. Referring to the flowchart in Figure 3, the following steps are operated to the filtered picture to obtain the random number that needs to be expressed.

1. The image is mapped onto a grid with a resolution of $16 \times 16 (n = 16)$, and the entropy matrix is obtained by using formula (2).

2. Derive the eigenvalue of the entropy matrix, and take the eight largest values to get the eigenvalue vector $(\alpha_1, \alpha_2, \ldots, \alpha_8)$.

3. The eigenvalue vector is quantized to obtain random number $\beta$. The quantitative formula is shown in

$$\beta_i = \text{dec2bin} \left( \text{mod} \left( \text{round} (\alpha_i), 16 \right) \right), \quad (i = 1, 2, \ldots, 8).$$

In the equation, mod() represents the congruent operation, round() represents rounding operation, dec2bin() represents integer conversion to vector operation, $\alpha_i$ represents eigenvalue vector, $\beta_i$ represents a 4-bit row vector, and finally the line vector $\beta = (\beta_1, \beta_2, \ldots, \beta_8)$.

4. The extracted random number $\beta$ is compared with the random number $\delta$ to be expressed. If they are matched and there is no image that is able to express the random number in the complete library, the image is added to the complete feature library. If they are not matched, the image is discarded and then the algorithm reselects an image from the image buffer library to repeat the above operations until the completion of the establishment of a complete feature library. Once the complete feature library is established, all we have to do in the calling process is to update it from time to time instead of creating one.

3. Turbo Error Correction Coding

Generally, covert communication implements transmission with the help of the public network link. Although the MFSA algorithm itself has a high fraudulence and concealment because of its characteristics of modification-free steganography, this algorithm incorporates turbo error correction coding technology to further improve the robustness of the system considering the security and complexity of public links.

The turbo code is used to improve the robustness performance of the system and reduce the bit error rate of transmission due to its excellent error correction performance. Turbo code is a high-performance error correction. The principle and performance of turbo code coding are described in detail in [13, 14]. In order to make the turbo code suitable for the system better, this paper improves the QPP (Quadratic Polynomial Permutation) interweaver used in turbo and gives some parameters of turbo coding.

The equation $F(x) = f_0 + f_1x + \cdots + f_mx^m$ is the permutation polynomial for permuting $\{0, 1, 2, \ldots, N - 1\}$. The derivative of $F(x)$ is as $F'(x) = f_1 + 2f_2x + \cdots + mf_mx^{m-1}$. The selected $N$'S permutation polynomial structure can be applied to the turbo code interweaver by selecting the appropriate $N$ and polynomial coefficients $f$. In this system, we select $m = 2$, and the resulting polynomial is $F(x) = f_0 + f_1x + f_2x^2$. For the constant term $f_0$ in the quadratic polynomial only affects the shift in the interweaver and does not work on decoding performance, $F(x) = f_0 + f_1x + f_2x^2$ can be further reduced as $F(x) = f_1x + f_2x^2$ to simplify the calculation.
The QPP interweaver uses some particular quadratic polynomial to make them satisfy certain conditions and become QPP structure. So the most critical problem is to solve the polynomial coefficients $f_1$ and $f_2$. The polynomial coefficients $f_1$ and $f_2$ of the system are satisfied and the length $k$ of the interweaver can be obtained using computer technology, as shown in Table 1. Table 2 shows the turbo encoder and the operating parameters of others used in the paper.

### Table 1: The coefficient of DPP interweaver permutation multinomial.

<table>
<thead>
<tr>
<th>$k$</th>
<th>32</th>
<th>64</th>
<th>128</th>
<th>256</th>
<th>512</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_1$</td>
<td>3</td>
<td>7</td>
<td>19</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>$f_2$</td>
<td>10</td>
<td>12</td>
<td>42</td>
<td>16</td>
<td>18</td>
</tr>
</tbody>
</table>

4. **Experiment and Analysis**

It is known that the statistical analysis of the original carrier is a security risk of the covert communication, and the robustness of the carrier is also an important factor in the secure communication. So this paper designed three experiments and an analysis; the antistatistical analysis ability, the antiscale interference ability, and the antiorotation attack capability of the MFSA algorithm are tested, respectively. The safety performance of this MFSA algorithm is analyzed. In the experiment, the MFSA algorithm uses a size of $16 \times 16$ grid descriptor. The basic characteristics of the image are mainly concentrated in the larger 5 to 8 eigenvalues, which contains information accounted for more than 90% of the total eigenvalue vector. So the experiment takes the larger 8 eigenvalues; after quantization, the length of the binary vector that can be mapped to a single image is 32 bits.

#### 4.1. Experiment 1

In experiment 1, the immune to statistical analysis was tested between the CSFA algorithm and the proposed algorithm in the literature [15], under the premise of known original “carrier.” The result is shown in Figure 4. Figures 4(a) and 4(b) are the original carrier images and histograms, respectively. Since the MFSA algorithm proposed in this paper does not embed the data in the original image, Figures 4(a) and 4(b) are the “carrier” images and histograms of the MFSA algorithm; Figures 4(c) and 4(d) are the secret carrier and its histogram; the embedding algorithm is proposed by the literature [15].

Due to the little payload, from the visual point of view, no difference was found between Figures 4(a) and 4(c). But there is a slight change in Figures 4(b) and 4(d). As indicated by the red arrow in the figure. The arrows on the left of Figure 4(d) indicate that the statistical value is smaller than the corresponding position in Figure 4(b), and the arrow on the right indicates that the statistical value is larger than the corresponding position in Figure 4(b). Other subtle differences with the naked eye cannot be directly found.

The matrix of the original image and confidential image in literature [15] can be expressed as $E_{ij}$ and $F_{ij}$ ($i, j \in (1, 512)$), as shown in
Table 2: The parameters of turbo coding.

<table>
<thead>
<tr>
<th>Code length (interleaving length)</th>
<th>Bit rate</th>
<th>Interweaving method</th>
<th>Decoding method</th>
<th>Iterations</th>
<th>Production matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>32 64 128</td>
<td>0.8</td>
<td>Pseudo-random intertwined</td>
<td>Log-MAP</td>
<td>12</td>
<td>(7, 5), (15, 17)</td>
</tr>
<tr>
<td>256 512 variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5: Gray scale image of different scale. (a) Image resolution, 1024 × 1024. (b) Image resolution, 512 × 512. (c) Image resolution, 256 × 256. (d) Image resolution, 128 × 128.

\[
E_{ij} = \begin{pmatrix}
E_{11} & \cdots & E_{ij} \\
\vdots & \ddots & \vdots \\
F_{11} & \cdots & F_{ij}
\end{pmatrix}
\]

\[
F_{ij} = \begin{pmatrix}
F_{11} & \cdots & F_{ij} \\
\vdots & \ddots & \vdots \\
F_{1i} & \cdots & F_{ij}
\end{pmatrix}
\]

(5)

Since the essence of the literature [15] is still the alternative method, it is possible to obtain the embedded secret information by performing mathematical operations on \(E_{ij}\) and \(F_{ij}\). According to formula (6), the element “1” position and number can be obtained; we can get the length and location of the secret information embedding.

\[
Q_{ij} = E_{ij} \odot F_{ij}, \quad (i, j \in (1, 512)).
\]

(6)

However, the MFSA algorithm is based on the idea of no zero modification through the mapping method to express the secret information. Because it does not make any changes, it is very effective in antistatistical analysis, especially the statistical analysis of known vectors. In addition, because there is no embedded information, the original picture will not show any change, so it will not be due to visual anomalies caused by noncooperation side of the vigilance.

4.2. Experiment 2. In experiment 2, the MFSA algorithm immune to scale attack was tested. As shown in Figure 5, Figure 5(b) is the original “carrier” image, resolution size is 512 × 512, and some different scale images are given as follows: Figure 5(a) is the original image amplification twice the image, and the resolution becomes 1024 × 1024; Figure 5(c) is the image where the original image was shrunk twice and so on. The attack comes from altering the image scale.

The MFSA algorithm is used to demapping the original image and images with varying degrees of scaling attack, as much as possible to restore the random number it represents.

Figure 5: Gray scale image of different scale. (a) Image resolution, 1024 × 1024. (b) Image resolution, 512 × 512. (c) Image resolution, 256 × 256. (d) Image resolution, 128 × 128.

\[
\begin{align*}
E_a &= \frac{p}{n}, \\
E_a' &= \frac{p}{n} = \sum_{i=1}^{n} a_i \odot b_i.
\end{align*}
\]

(7)

The 32-bit random number vector was recovered from graph (b) \((b_1, b_2, \ldots, b_{32})\); Figures 5(a), 5(c), and 5(d) are similar. The bit error rate is used as its evaluation. Equation (7) gives the solution of the bit error rate \(E_a\) (\(E_a\) is the bit error rate that is not added to turbo error correction coding; \(E_a'\) is the error rate of adding turbo error correction code) for the random number shown in Figure 5(a). Replace \(a_i\) with \(c_i\) and \(d_i\), respectively; we can find the bit error rates \(E_c\) and \(E_d\) of graphs (c) and (d).

Table 3: Statistical analysis of attack at different scale.

<table>
<thead>
<tr>
<th></th>
<th>Figure 5(a)</th>
<th>Figure 5(b)</th>
<th>Figure 5(c)</th>
<th>Figure 5(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E) (%)</td>
<td>0</td>
<td>0</td>
<td>3.125</td>
<td>6.25</td>
</tr>
<tr>
<td>(E') (%)</td>
<td>0</td>
<td>0</td>
<td>1.25</td>
<td>3.125</td>
</tr>
</tbody>
</table>

As can be seen from Figure 5, although the scale of the original “carrier” image has been attacked, the main content of the image is not changed. The spatial position information of the image and the statistical distribution information of the pixel values are not changed. From the statistical results of Table 3, although the original “carrier” image size has undergone great changes, the correct rate of demapping recovery is also about 95%, and error correction coding can
Figure 6: Different angle rotation of the original secret image. (a) Original secret image. (b) Rotation of 90 degrees anti-clockwise of the original secret image. (c) The original secret image horizontally mirrored and vertically mirrored. (d) Rotation of 270 degrees anti-clockwise of the original secret image. (e) The original secret image horizontally mirrored. (f) Rotation of 90 degrees anti-clockwise and the original secret image vertically mirrored. (g) The original secret image vertically mirrored. (h) Rotation of 270 degrees anti-clockwise and the original secret image vertically mirrored.

Table 4: Results of statistical analysis of attack at different rotation angles.

<table>
<thead>
<tr>
<th></th>
<th>Figure 6(a)</th>
<th>Figure 6(b)</th>
<th>Figure 6(c)</th>
<th>Figure 6(d)</th>
<th>Figure 6(e)</th>
<th>Figure 6(f)</th>
<th>Figure 6(g)</th>
<th>Figure 6(h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E$ (%)</td>
<td>0</td>
<td>7.375</td>
<td>0</td>
<td>5.375</td>
<td>3.125</td>
<td>7.375</td>
<td>3.125</td>
<td>8.375</td>
</tr>
<tr>
<td>$E'$ (%)</td>
<td>0</td>
<td>4.125</td>
<td>0</td>
<td>3.125</td>
<td>2.125</td>
<td>5.225</td>
<td>1.125</td>
<td>6.125</td>
</tr>
</tbody>
</table>

further improve the correct rate of demapping. So it can be concluded that the MFSA algorithm has a good effect on the attack defense.

4.3. Experiment 3. In experiment 3, the MFSA algorithm immune to the rotation attack was tested. As shown in Figure 6, the attack comes from image rotation, horizontal flip, horizontal mirror, and vertical mirror of the original image.

Perform 100 independent tests on different "carrier" images; according to the method in experiment 2, the inverse mapping of the original "carrier" image is subjected to different angles of rotation after the attack image is obtained. The result of statistical analysis is shown in Table 4.

From the table, the original "carrier" images subjected to varying degrees of rotation attacks can be seen, the demapping data was recovered, and the data bit error rate remained at about 95%. In particular, as shown in Figure 6(c), the image is horizontally mirrored and vertically mirrored and the random number can be 100% recovered by demapping. Because the "carrier" image suffered a different degree of rotation attack after the basic characteristics did not change, the statistics of the image are not changed. So the feature matrix and its eigenvalue vector remain essentially unchanged. This shows that the MFSA algorithm has a good effect on the anti-rotation attack.

4.4. Analysis of Safety Performance of MFSA Algorithm. The proposed MFSA algorithm is based on zero modification of the carrier image and it expresses the secure information by using the mapping relationship rather than embedding method, which conceals the existence of the covert communication and has a decisive effect on avoiding the nonpartner's doubt and monitoring. Besides, the MFSA algorithm itself has a certain level of safety.

The safety performance of the MFSA algorithm is dominated by two keys: $K_1$ is a 6-bit expression algorithm control variable, which controls when to choose what kind of random number expression algorithm; $K_2$ is the 20-bit quantization coefficient which determines the relative parameters in the quantization process. When extracting the feature from the picture, it is necessary to add the number of the floating-point type to the quantization coefficient and then transform it into the integer number. On the one hand, if the nonpartner obtains the secret key $K_1$ only, they would not break the MFSA algorithm. The nonpartner cannot obtain the quantified binary vector through iterative computation. On the other hand, even if the nonpartner gets the secret keys $K_1$ and $K_2$, they still cannot get the secure information, because the parameters and formulas of the secret key are shared only between the sending and receiving end.

5. Conclusion

Based on the idea of zero modification of the carrier image, the MFSA algorithm is proposed in this paper; the mapping relationship between the image information entropy and information is used to express the secure information. And the network image search engine is used to obtain the "appropriate" image from the big data of image; the big data parallel processing method as the technical support. The mathematical reasoning and the result of experiment demonstrate the proposed MFSA algorithm which has a good performance on the immune statistical analysis and some attacks. Turbo coding technology improves the system robustness and security; the simulation results showed the
noise immunity of our method and this method can resist the scale and rotation attacks. Even if the nonpartner cracks the secret communication channel, and the communication “carrier” is intercepted, the nonpartner cannot judge whether the “carrier” is secret image and cannot get the content of communication. It is very suitable for the covert communication with smaller capacity and high security level, such as the transmission of security system keys, key figures, time, location, and other information transmissions.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Natural Science Foundation of China and China General Technology Research Institute (U1736121).

References
