HIDING INFORMATION IN IMAGES

[1]Pınar Çivicioğlu [2]Mustafa Alçı
e-mail: civici@erciyes.edu.tr *e-mail: malci@erciyes.edu.tr*
[1]*Erciyes University, Civil Aviation School, Avionics Department, 38039, Kayseri, Turkey*
[2]*Erciyes University, Faculty of Engineering, Department of Electronics Engineering, 38039, Kayseri, Turkey*

Key words: Steganography, encryption, information hiding

ABSTRACT

In this study, a new algorithm is proposed for the cryptography and steganography of an image or text. There is a variety of steganography and cryptography tools which are commonly used for these purposes but the proposed algorithm, which can be coded by any appropriate programming language, can be used instead of these expensive tools. The decription algorithm of the encrypted steganography algorithm is also given.

I. INTRODUCTION

Steganography is the art of hiding information so that its presence cannot be detected [1]. Classical steganography concerns the ways of embedding a secret message which can be a copyright mark or a serial number in a cover message such as a computer code, video film or an audio recording. The embedding is typically characterized by a key and it is difficult for a third one to detect or remove the embedded information without this key. When the information is embedded in the cover object then it is called a stego-object. Thus for example a mark can be embedded in a covertext giving a stego-text, a text can be embedded in a cover image giving a stego-image [2]. The purpose of the steganography in general is to hide data well enough that unintended recipients do not suspect about the steganographic medium of containing hidden data. The target of steganography is not to keep the others from learning that there is an information hidden. If a steganography method causes someone to suspect about the carrier medium, then it means that the method has failed [3].

Steganography literally means, “covered writing” and includes the methods of transmitting secret messages through cover carriers in such a manner that the existence of the embedded messages is undetectable. Carriers of such messages may be any digitally represented code of transmission such as innocent images, video, audio or text. The hidden message may be a plaintext, a ciphertext, an image or anything that can be represented with a bit stream [4].

The concepts of steganography and data hiding are not new. It is believed that steganography was first used during the Golden Age in Greece. Early in the 2nd World War, steganographic technology consisted almost only of the invisible inks with which a common form of invisible writing was possible. The sources for invisible inks were vinegar, fruit juices, and milk, all of which darken when heated. An apparently innocent letter could contain a very different message written between the lines with invisible inks. Chemicals were also used for preparing these inks.

Steganography and cryptography are relatives in the technology hiding family. Cryptography encrypts a message so that it cannot be understood while steganography hides the message in order to make it not seen. Cryptography encodes data so that an unintended recipient cannot understand its intended meaning but steganography does not alter data to make it unusable to an unintended recipient. Most steganographic methods also encrypt the message so even if the presence of the message is detected, deciphering the message will still be required. Steganography is complementary to cryptography because it adds an extra layer of security.

Another form of steganography, called watermarking, is being used increasingly by the bussinesses. Watermarking is used for identification and entails embedding a unique piece of information within a medium without altering the medium noticeably. Watermarking is commonly used for protecting copyrighted digital media, such as web pages and audio files [3]. The only difference between watermarking and steganography is that in watermarking the cover is the object of communication while in steganography the hidden message is the object of communication. Typically, steganography comprises the conversion of point-to-point communication between two parties and therefore does not necessarily need the robustness which is required in watermarking. Watermarking however needs an extra robustness against manipulations that may attempt to remove it.

Digital watermarks can be visible (perceptible) or invisible (imperceptible) to human vision. Visible watermarks are typically confined to an area of the image and they can be compared to traditional paper watermarks and logos seen on TV broadcast stations. Attackers to the watermarks can remove these watermarks by cropping the image. Visible watermarks are technically not approved as steganography. Invisible watermarks have an advantage over visible watermarks because their location may be unknown. A common application is to distribute the watermark across the whole image. This can supply some protection against cropping attacks [5].

There has been a rapid growth of interest in the subjects of hiding information over the last years [1-12] and for two main reasons: Firstly the broadcasting and publishing industries have become interested in techniques for hiding encrypted serial numbers in digital films, copyright marks, books and audio recordings [2]. Once a cover medium is selected, a technique for embedding the message must be decided on. There exist many different methods for hiding information in images. Some of the common approaches of hiding information in digital media involve [6]:

- Least significant bit (LSB) insertion
- Masking and filtering
- Algorithms and transformations

In this study, two different types of information have been used as the message information that will be hidden. In the first simulations, the Mandrill image and in the second simulations a text has been encrypted and then hidden in the Lena image, therefore steganography and cryptography were both used. There is a variety of steganography and cryptography tools which are commonly used for these purposes but in this study, a new method is proposed which can be coded by any appropriate programming language.

II. PROPOSED ALGORITHM FOR ENCRYPTION

Most steganographic methods also encrypt the message so even if the presence of the message is detected, deciphering the message will still be reqired. Steganography is complementary to cryptography because it adds an extra layer of security. Therefore both steganography and cryptography were used in this study. The proposed algorithm is explained step by step below:

1. Select a *numeric* key (This key may be equal to *M*x*N*, where *M* and *N* denote the total number of the pixels in the horizontal and the vertical dimensions of the image).
2. Initialize the *uniformly distributed random permutation of the integers producer algorithm* (RIP) by using the selected *numeric* key. The RIP(*m;n*) corresponds to random permutations of integers between *m* and *n* (*m* and *n* are both integers).
3. Obtain a *message scattering matrix* by using the RIP(1;*M*x*N*).
4. Replace the positions of the pixels of the message by using the elements of RIP(1;*M*x*N*) in order to obtain the *scattered message matrix* (SM).
5. Run Step-2 and Step-3 in order to obtain *pixel data scattering in binary bits matrix* (PC), PC=RIP(k;8). ($6 \le k \le 8$ is recommended)
6. Take the i^{th} pixel from the *Cover Image* and convert its gray value to 8-bit binary (P). Take the i^{th} element of PC corresponding to the position of the bit which will be replaced in P with the binary value of the i^{th} pixel of the SM.
7. Repeat Step-6 for each pixel of the scattered message.

III. PROPOSED ALGORITHM FOR DECRYPTION

The proposed algorithm for the decription of the encrypted steganography algorithm is given below step by step:

1. Select the *numeric* key used at encryption phase.
2. Initialize the *uniformly distributed random permutation of the integers producer algorithm* (RIP) by using the selected *numeric* key.
3. Obtain the *message scattering matrix* by using the RIP(1;*M*x*N*).
4. Take the *pixel data scattering in binary bits matrix* (PC), PC= RIP(k;8).
5. Take the i^{th} pixel from the *Encrypted Image* and convert its gray value to 8-bit binary (P). Take the i^{th} element of PC corresponding to the position of the bit which will be taken as the i^{th} pixel of the original binary converted message (T) in P.
6. Replace the positions of the i^{th} pixel in T with the i^{th} element of PC.

IV. IMAGE QUALITY MEASURES (IQMS)

In order to quantitatively evaluate the success of the proposed method, some of the well-known IQMs are employed: Mean Squared Error (MSE) [13], Correlation Coefficient (Corr) [14] and Peak Signal to Noise Ratio (PSNR). The quality measure of PSNR is defined with,

$$\text{PSNR}=10\ \log_{10}\left(\frac{I_{\max}{}^{2}}{MSE}\right)\text{dB} \tag{1}$$

where $I_{\max}$ is equal to 255 for 8 bit gray scale images. The MSE is calculated by using the Eq. (2) given below:

$$\text{MSE}=\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(Y_{i,j}-S_{i,j})^{2} \tag{2}$$

M and *N* denote the total number of the pixels in the horizontal and the vertical dimensions of the image. $S_{i,j}$ represent the pixels in the original image and $Y_{i,j}$ represent the pixels of the stego-image [15]. Pearson Correlation Coefficient (Corr) is given by,

$$Corr = \frac{\sum\sum(S-\overline{S})\ \ (Y-\overline{Y})}{\sqrt{\sum\sum(S-\overline{S})^2\sum\sum(Y-\overline{Y})^2}} \qquad (3)$$

where

$$\overline{S} = \frac{\sum\sum S}{M\ N} \quad \text{and} \quad \overline{Y} = \frac{\sum\sum Y}{M\ N} \qquad (4)$$

S represents the pixels of the original image and Y represents the pixels of the stego-image.

V. SIMULATIONS

Simulations are conducted on the images shown in Figures 1 and 3. The images shown in Figure 1-(a) and Figure 3-(a) have been used as *Cover Images* and the images shown in Figures 1-(b) and 3-(b) have been used as *Message Images.* The simulations were realized for 6 different situations of embedding. In the first realizations, the message information is embedded only in the last bit (LSB) of the cover image. In the other realizations, the message information is embedded in one of the last 2, 3, 4, 5, 6 bits of the cover image randomly. As can be seen from Figure 2 and Figure 4, embedding information into cover image causes some distortions. The effect of the distortions are evaluated with IQMs and tabulated in Tables 1-2.

Table 1. IQMs of the stego-image in which the Mandrill is hidden.

Randomly selected bit from the last *n* bits	**MSE**	**Corr**	**PSNR**
1	0.499	1.000	51.146
2	1.241	1.000	47.192
3	3.102	0.999	43.214
4	8.801	0.998	38.686
5	26.919	0.994	33.830
6	85.509	0.982	28.811

Table 2. IQMs of the stego-image in which a text is hidden

Randomly selected bit from the last *n* bits	**MSE**	**Corr**	**PSNR**
1	0.504	1.000	51.106
2	1.231	1.000	47.227
3	3.079	0.999	43.247
4	8.767	0.999	38.702
5	26.836	0.995	33.844
6	85.487	0.985	28.812

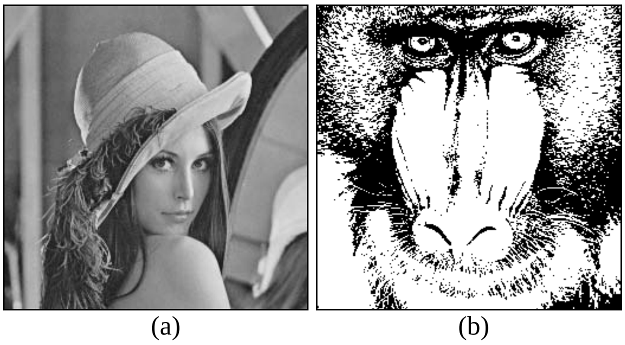


Figure 1. (a) The well-known Lena image which is used as the cover image
(b) The well-known Mandrill image which is used as the first message information

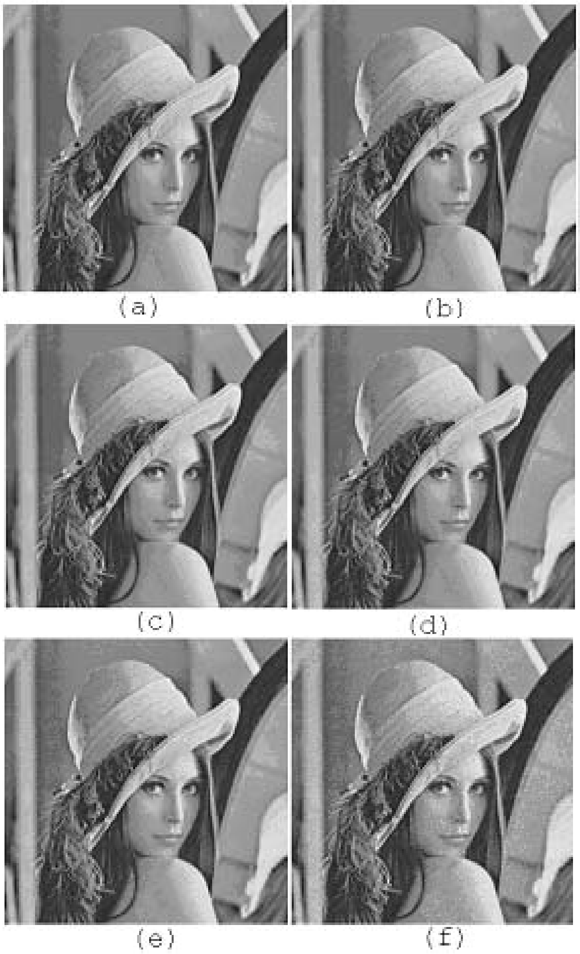
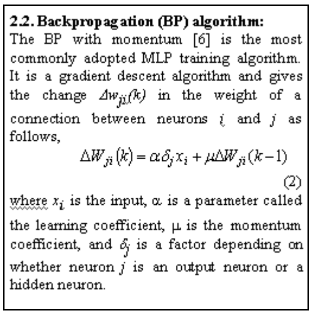


Figure 2. The stego-images: Message information is hidden randomly in one of the last (a) 1 bit, (b) 2 bits, (c) 3 bits, (d) 4 bits, (e) 5 bits, and (f) 6 bits of the cover image pixels.

(a)



2.2. Backpropagation (BP) algorithm:

The BP with momentum [6] is the most commonly adopted MLP training algorithm. It is a gradient descent algorithm and gives the change $\Delta w_{ji}(k)$ in the weight of a connection between neurons i and j as follows,

$$\Delta W_{ji}(k) = \alpha\delta_j x_i + \mu\Delta W_{ji}(k-1) \qquad (2)$$

where x_i is the input, α is a parameter called the learning coefficient, μ is the momentum coefficient, and δ_j is a factor depending on whether neuron j is an output neuron or a hidden neuron.

(b)

Figure 3. (a) The Lena image as the cover image
(b) A text as the second message information

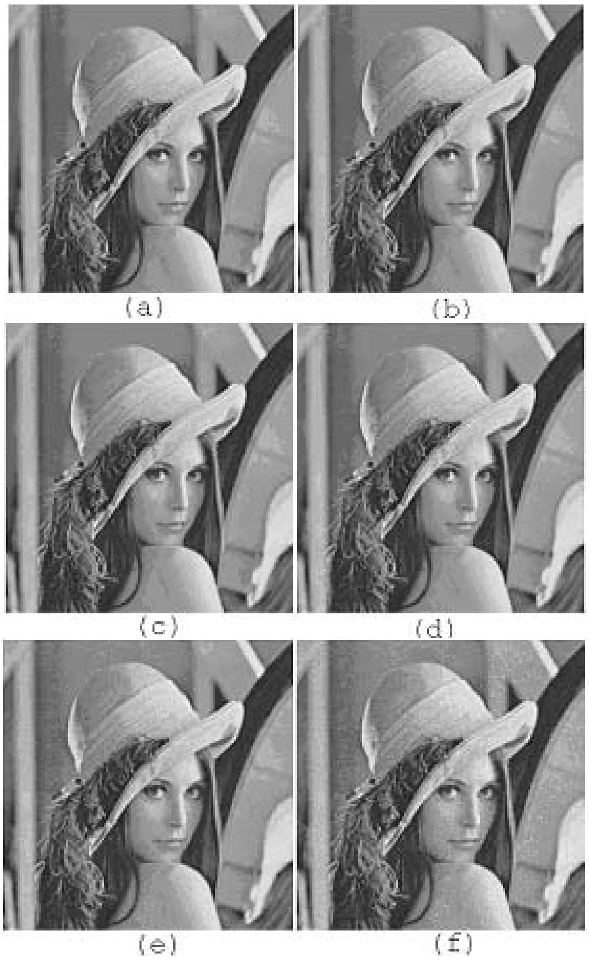


Figure 4. The stego-images: Message information is hidden randomly in one of the last (a) 1 bit, (b) 2 bits, (c) 3 bits, (d) 4 bits, (e) 5 bits, and (f) 6 bits of the cover image pixels.

VI. CONCLUSION

In this paper, a new method is introduced for the steganography of an image or text. The proposed method can be coded by any appropriate programming language. The most important property of the proposed method is that the message information is scattered randomly over the last *n* bits of the cover image pixels. By evaluating the Figures 2, 4 and Tables 1, 2, it can be seen that the proposed method is very successful at hiding information into a cover image. The MSE, Corr, and PSNR IQMs show that the best results are achieved when the message information is hidden in the LSB and worst results are obtained when the message information is hidden in one of the last 6 bits of the cover image. As the number of the pixels into which information is hidden increases, it gets more difficult for the unintended ones to decrypt the meaning of the hidden message.

REFERENCES

1. C. Cachin, An Information-Theoretic Model for Steganography, Proc. of 2nd Workshop on Information Hiding, pp. 1-12, 1998.
2. R. J. Anderson, F.A.P. Petitcolas, On the Limits of Steganography IEEE Journal of Selected Areas in Communications, Vol. 16, No. 4, pp. 474-481, 1998.
3. D. Artz, Digital Steganography: Hiding Data within Data, IEEE Internet Computing, Vol. 5, No.3, pp. 75-80, 2001.
4. N. F. Johnson, S. Jajodia, Steganalysis: The Investigation of Hidden Information, Proc. IEEE Information Technology Conf., pp. 113-116, 1998.
5. N. F. Johnson, An Introduction to Watermark Recovery from Images, Proc. SANS Conf. and Workshop on Intrusion Detection and Response, pp. 10A1-10A6, 1999.
6. N. F. Johnson, S. Jajodia, Exploring Steganography: Seeing the Unseen, Vol. 31, No. 2, IEEE Computer, pp. 26-34, 1998.
7. X. G. Xia, C.G. Boncelet, G.R. Arce, A Multiresolution Watermark for Digital Images, IEEE Int. Conf. on Image Proc., Vol. 3, pp. 548-551, 1997.
8. R. J. Anderson, Stretching the Limits of Steganography, Information Hiding, Springer Lecture Notes in Comp. Science, Vol. 1174, pp. 39-48, 1996.
9. W. Bender, D. Gruhl, N. Morimoto, A. Lu, Techniques for Data Hiding, IBM Systems Journal, Vol. 35, No. 3-4, pp. 313-336, 1996.
10. F. M. Boland, O. Ruanaidh, C. Dautzenberg, Watermarking Digital Images for Copyright Protection, Proc. of IEE Int. Conf. on Image Processing and its Applications, pp. 326-331, 1995.
11. L. M. Marvel, C. G. Boncelet, C. T. Retter, Reliable Blind Information Hiding for Images, Proc. of Information Hiding Workshop, pp.48-62, 1998.
12. M. D. Swanson, M. Kobayashi, A. H. Tewfik, Multimedia Data-Embedding and Watermarking Technologies, Proc. of the IEEE, Vol. 86, No. 6, pp. 1064-1087, 1998.
13. Matlab 6.2 Software, Mathworks Inc, 2000.
14. Microsoft Office, Excel software, 2000.
15. H. L. Eng, K. K. Ma, Noise Adaptive Soft-Switching Median Filter, IEEE Transactions on Image Processing, Vol.10, pp.242-251, 2001.