Journal of Advanced Computer Science & Technology, 3 (2) (2014) 169-178
©Science Publishing Corporation
www.sciencepubco.com/index.php/JACST
doi: 10.14419/jacst.v3i2.3404
Research Paper

A new high capacity steganography based on bit-inverting method in DWT domain

Mahnoosh Sadat Azaminezhad *, Amir Massoud Bidgoli

Department of Computer and IT Engineering, Engineering Faculty, Tehran North Branch, Islamic Azad University, Tehran, Iran
**Corresponding author E-mail: M_azaminezhad@yahoo.com*

Copyright © 2014 Mahnoosh Sadat Azaminezhad, Amir Massoud Bidgoli. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Steganography is a technique which embeds the secret messages in the cover image and transmitted in such a way that the existence of information is undetectable. In this paper, we present a new technique that consists of two encoding algorithm such as Huffman encoding and bit inverting algorithm. In this technique, we proposed a modified secure and high capacity based steganography scheme of hiding a secret message such as text or image into a cover image. Our simulation results show that the algorithm has a good perceptual invisibility and high capacity in addition to be secure.

Keywords: *Bus Inverting, Discrete Wavelet Transform (DWT), Huffman Coding, Steganography, Wavelet.*

1. Introduction

There have been considerable progresses in all domains of communications and technology nowadays. Thanking to these improvements, we'd better to find a secure and safe method for data communication. Accordingly, the information security has transformed to one of the most important requirements in digital communication. Information hiding is one of the major solutions for confidential communication. Some of the methods have prepared security and safety for information transformation and reception. Cryptography, watermarking, and steganography are some the useful methods of this domain [1].
Cryptography refers to the information protection through transforming it into a new format that is called “cipher text”. The purpose is to preserve the massages contents from unauthorized accesses. Cryptography methods use some keys for transferring data. Some of algorithms use a key, which is called “public key”, and some others use two keys, which are called “public key and private key”. The second algorithm is persistent enough for all types of attacks [2].
Watermarking is another method similar to cryptography, which is used, for copyright preservation. This is a protecting technique, which embeds a watermark about copy right in digital media. Watermarking has major features such as imperceptibility, security, robustness, and blind detection. For instance, imperceptibility indicates this subject that the human eye cannot distinguish the difference between the watermarked type picture and the original version. The most important purpose is the recognition of copy right through watermarking [1], [3].
Steganography refers to those algorithms, which hide the message inside a carrier media [2]. It is a technique for hiding a message, image, or a file inside a message, image or other file. The data steganography system is specified by three different parameters, which are related to each other. Capacity, security, and persistence are the aforesaid parameters that are important in steganography method. Capacity refers to the amount of data that can be assuredly hidden inside the media. Security mentions the eavesdropper inability for recognizing the hidden information. In fact it is impossible for the attackers to recognize and extract the confidential information. Robustness is the amount of changes that can be done in stego-media without eradication of the confidential information. Practically the competence of the embedded data to be intact provided that the stego image undergoes transformation and evolution caused by the intelligent stego attacks [1].
The information is obscurant in cryptography while the steganography attempts to hide the message existence that means the human eye can't distinguish the difference between the main media and stegoed one.

The primary goal of steganography is to avoid the detection or even raising the suspicion that a secret message is being passed on. Steganography is applicable to (i) Confidential communication and secret data storing, (ii) Protection of data alteration, (iii) Access control system for digital content distribution, (iv) Media Database systems and etc. In image steganography, the information hides exclusively in an image which is called cover image. After embedding the secret message, the cover image is called the stego-image. The steganography system can be useful provided that had prepared a method for invisible information embedding and the hidden messages must be meaningful after extraction. The hidden messages might be a text or images [4].

The image steganography schemes can be divided into two categories: spatial-domain based [5] and transform-domain based [1], [5].

In spatial domain the message embedded directly without any changes or transformation in the hidden and carried data. One of the simplest and most prevalent methods in this group is the Least Significant Bit (LSB) that has been used as a pioneer of the advanced methods [5], [6].

The main privilege of this group is gaining better quality and more capacity but they are not strong in attack confrontation. The embedding process has been accomplished directly in this method. Therefore the stego image had been sensitive and is not resistant against operations such as lossy compression, cropping, blurring and etc [7].

The first stage in transformation domain technique is the conversion of the carrying picture and the hidden messages into a set of frequency domain coefficients. After conversion to frequency coefficients, embedding in new domain, there is a need for reverse reformed coefficients transformation to make image stego.

There are several techniques in transformation domain such as Discrete Cosine Transform (DCT), Fourier Transform (FT) and Discrete Wavelet Transform (DWT) that are the most important and applicable algorithms in this domain.

The transformation domain techniques have overcome the weakness of robustness and imperceptibility contrary to the spatial domain.

The DWT is also the best method in transformation domain for the integration and a combination of time and frequency. We have proposed a DWT steganography technique which achieves to considerable capacity and PSNR in this algorithm.

The rest of the essay is as the following:

Section 2 describes the existing works through evaluating several algorithms that are in spatial domain and especially in transformation domain. In section 3 the steganography method has been proposed with more details and the experimental results and analysis have been shown in section 4. Finally the work summary and ideas for future researches have been represented in section 5.

2. Related work

In this section we will display a background of steganography algorithms especially the transformation domain. As it mentioned before, there are two major types of steganography methods: spatial and transformation. In this essay we have represented a method based on transformation and ignore the previous work evaluation based on the local domain.

Most of the researchers have used spatial domain techniques for their algorithm in the last investigations, but the privileges of the transformation domain techniques convinced them to use the transformation domain techniques specially Discrete Wavelet Transform (DWT) instead of spatial domain. We mentioned in the previous section that transformation domain techniques have considerable features versus the spatial domain techniques such as improvement in robustness and imperceptibility.

There are several methods which use DWT for implementation of steganography but some disadvantages such as low robustness against attackers and low security forces the researchers to use other methods of information hiding techniques like encoding, cryptography, and compression. We can achieve better robustness, quality improvement, and higher capacity using these methods.

Wavelet is a combination of sine and cosine functions which contains the basic signal information in both time and frequency domain. This competence has most important data versus other transformation domain techniques such as Fourier (DFT) and Cosine (DCT). Therefore in the latest researches the DWT is the main choice in transformation domain techniques [3], [5].

According to the data format either the cover objects or the hidden messages that are used in steganography method; we can use 1-Dimensional or 2- Dimensional DWT method. The main part in DWT computations relates to the wavelet filters. In fact they show whatever computations need for transformation. One of the most important, simplest, and common methods in this domain is Haar DWT. If the data be the vector (1 D array), 1-D DWT will be used. This is one step in 1-D DWT method. We have computed the addition of all data pair by pair in the first half of the data and will compute the subtraction of all data pair by pair in the next half. Fig. 1 shows a diagram for 1-D DWT [8].

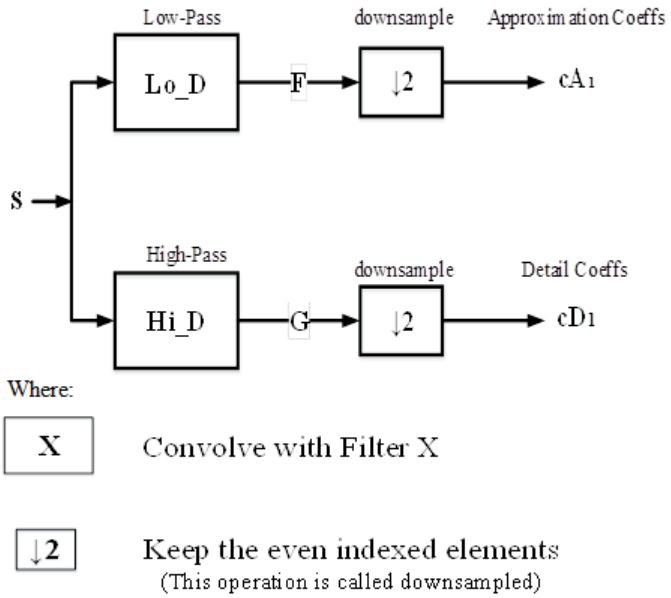


Fig. 1: 1D-DWT Algorithm Diagram

This method is used for both vertical and horizontal dimension in 2-D DWT. Fig. 2 shows the 2-D DWT using transformed image. There are four sub-bands in DWT; LL, LH, HL, and HH. The LL sub-band is the main features of data and will be ignored from any changes in embedding method, but other sub-bands include high frequency coefficients which are the purpose of embedding method [9], [10].

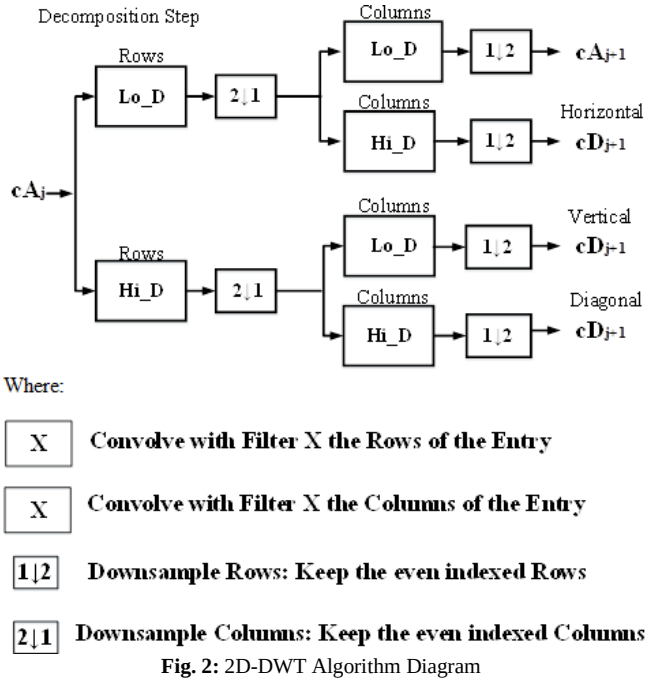
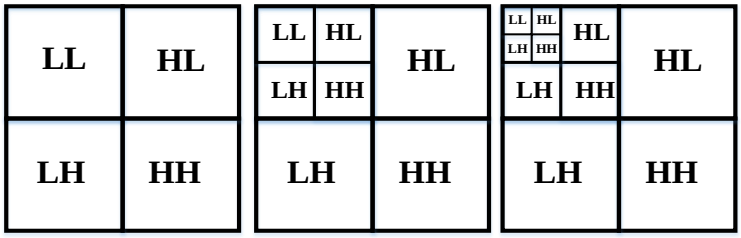


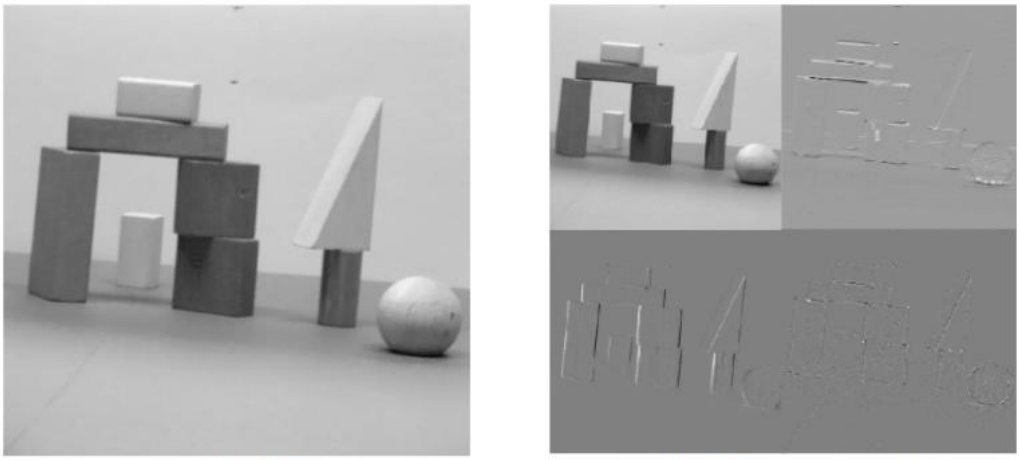
Fig. 2: 2D-DWT Algorithm Diagram

We can use a single-level and multi-level DWT in 2-D DWT. Each transformation is being used as LL in the main sub-band. In single-level DWT it will be utilized in the whole image, but in multi-level DWT it will be applied on the LL sub-band. Fig. 3 demonstrates 1, 2, and 3-levels of 2-D DWT. It improves the robustness and capacity transformation through enhancing more depth and layer but it reduces the algorithm's quality. Generally, it is being used almost in single-level, however in some algorithms a 2-level has been used too. Also Fig. 4 shows a standard picture which has been analyzed by a single-level of 2-D DWT [11].



(a) Single Level (b) Two Level (c) Three Level

Fig. 3: Decomposition of Image



(a) Original Image (b) After the one-level

Fig. 4: 2D-DWT of an Image

3. Proposed method

In this part we introduce our proposed steganography method for hiding a large amount of high security, good vision data without losing hidden information. The represented method includes two techniques:

1) Huffman algorithm
2) Bit-inverting method

In the next sections we will explain the principles, process and diagram of these methods.

3.1. Huffman encoding

Huffman encoding is one of the popular encoding algorithms for data compression. The main idea of Huffman encoding is to find an optimum code word for compressing a set of data. This expression indicates the use of variable length instead of a fixed length for encoding the main data and the criterion for the length of each data is the frequency of each character. As an instance, in an encoding algorithm that uses a permanent length 8 bits is used for each symbols (like a pixel in a picture) but in variable form, shorter length codes are dedicated to the more frequent characters and longer length codes belong to the less frequent characters in the string. Therefore we can receive desirable result and total size of the data with variable length will decrease in encoding.

The Huffman encoding consists of two major stages. The first step is the calculation of Huffman table and Huffman tree. The Huffman encoding algorithm initiates with the Huffman table construction. In this stage the main data character with their frequency are ordered and stored in a list accompanying by their ascending arrangement. This algorithm is being done by performing similar stages till it finishes. In any condition, two symbols of lowest probability/frequency are selected and then being replaced by an auxiliary node which is the sum of two selected probability nodes. The list will end while it reaches just an auxiliary symbol. The probability of symbol is 1.0 in this situation. In the next stage the tree will be constructed in bottom-up manner. In fact the two smallest symbols are the leaves of the tree and the auxiliary node must be add upon them as a parent up to the evacuation of the entering list when the binary tree is been constructed and shows all the characters frequency numbers in a file.

Afterward, we need to calculate the corresponding code word of each symbol by traversing the tree from the root to the leaf assigning 0 for left analysis and 1 for right. Suppose that we have five characters with their probabilities represented in the table 1.

Table 1: symbols and probabilities

Symbols	a_1	a_2	a_3	a_4	a_5
Probabilities	0.4	0.2	0.2	0.1	0.1

They are shown as the following and are combined in the represented result in Fig.5:

1) A4 and a_5 are combined with each other and have created a_{45} axillary name with 0.2 repetitions. The symbols a_4 and a_5 have been dropped in this step and the a_{45} has been added instead.
2) There are four sign in the table including a_1 with 0.4 probability, a_2, a_3, and a_{45} with 0.2. Two signs of a_3 and a_{45} are arbitrarily selected with the least probability, combined with each other and an auxiliary name as a_{345} with probability of 0.4 will be replaced in the table instead.
3) Three signs of a_1, a_2, a_{345} with probabilities of 0.4, 0.2, and 0.4 respectively have been omitted till now. We arbitrarily select a_2 and a_{345} and put it after combination in the table with an auxiliary name as a_{2345} with probability of 0.6.
4) Finally two a_1 and a_{2345} remained symbols are combined together and a_{12345} auxiliary name will be replaced by a probability of 1. Whenever just a symbol by 1 probability remains the tree will be complete. We arbitrarily assign 0 to the right edge and 1 for the left edge to identify the code word of each symbol finally. So the results of code words are 0, 10, 11, 1101, and 1100 respectively.

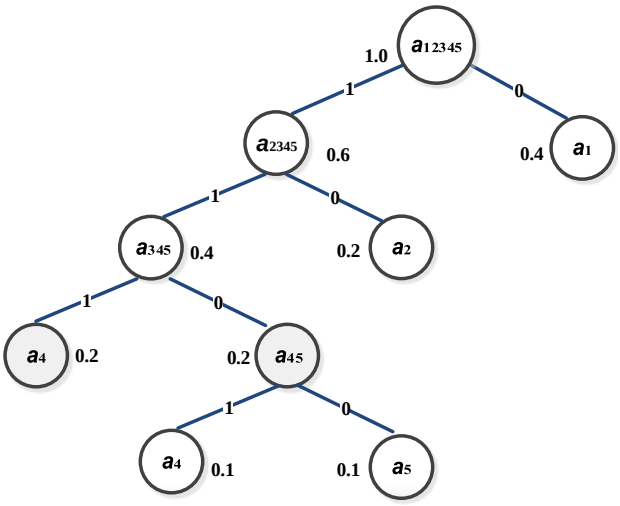


Fig. 5: Huffman Tree

It is necessary to mention that the assignment of the codes to the edges is arbitrary. So the mean of the code word's length size equals to:

$0.4\times1+ 0.2\times2 + 0.2\times3 + 0.1\times4 + 0.1\times4 = 2.2$ bits/symbol (a)

However what is more important is that the Huffman code is not unique for the symbols' selection. The trend being arbitrary and the reason in this event derive from the existence of two symbols with the same minimal frequency values (probability). For better conception of the previous example, five symbols can be combined differently with each other and various Huffman codes (e.g. 01, 11, 00, 101, and 100) can be obtained. The mean of the new code word's length size has not changed and will remains as the previous code:

$0.4\times2 + 0.2\times2 + 0.2\times2 + 0.1\times3 + 0.1\times3=2.2$ bits/symbol (b)

Suppose that we must have at least 3 bits in each symbol but we achieved just 2.2 bits. The accumulation ratio over the main scheme is (2.2/3)*100 = 73%. It can be shown that Huffman provides better compression for any communication issue in comparison to the common state [12], [13].

3.2. Bit-inverting method

The more important method in our mind is bit-inverting usage in the proposed method. This is a simple and effective method for reducing the power consumption in digital systems.

The most important feature achieved from this algorithm is writing a data with lowest value obtained from the difference of carrier picture pixel's value and the message pixel's value or difference of carrier picture pixel's value and the message complemented pixel's value [14].

Let's take a parameter 'D' into consideration as the distance between the current message and the corresponding pixel amount for embedding and a parameter '$\overline{D}$' as the distance between the message pixel value complement and the corresponding pixel amount. In the second stage we can choose between D and $\overline{D}$. The one with lower amounts must be

selected and the amount of corresponding pixel has to be embedded in the carrying pixel. Take the following example into consideration for better conception of the algorithm.
Suppose as an instance that the purpose is embedding data in an eight bit pixel image frame. Also, it is intended to place steganography data inside each pixel in the image with 4 bit width. According to primitive LSB law, these 4 bits must be placed in the least value 4 bits. So the 4 least value bits of each pixel are compared with the amount of considered 4 bits due to submergence. Now if we suppose that the amount of least significant 4 bits in a pixel be equal to 1001 (9) and the amount of 4 bits for embedding equal to1010 (10), we have to perform the two following steps:

1) D computation.

The first step is to write without change that in this situation the message pixel value (9) and the cover pixel value (10) amount of difference (deviation) equal 1.
D=10-9=1

2) Complement amount computation.

In the next step the amount of message data complement must be computed. Therefore, since the message is 4 bit, the maximum amount is$(2^4) - 1 = 15$.
Then if we subtract the message amount from the maximum amount the complement amount will be achieved. The complement amount becomes (6).
15-9=6

3) $\overline{D}$ Computation.

In this state the amount of difference (deviation) equals 4.
$\overline{D}$=10-6=4

4) D and $\overline{D}$ comparison.

D=1
$\overline{D}$=4
As it can be seen, the first state of D has less difference and writing it will produced a better results.
Now suppose that the amount of 4 bits for embedding equals 0110(6). In this situation the difference in real amount equals 3 and the difference in complement amount equals 1. So writing the complement in this state is better.

3.3. Encoding algorithm

The block diagram of the proposed steganography system is depicted in Fig.6. According to this Fig., the process of the embedding secret data can be described as follow:

1) Read the secret-object as a text or image.
2) Prepare the secret object as bit stream.
3) Apply Huffman Encoding method on bit stream input.
4) Read the cover image.
5) Decompose the cover image by using Haar wavelet transform.
6) Apply bit inverting method on the output of the Huffman encoding and embed the message bits in 4 least value bits of each pixel from the one of the approximate sub bands as HH.
7) Apply inverse DWT.
8) Prepare stego image to display.

3.4. Decoding algorithm

The following steps explain the decoding process:

1) Read the stego-image
2) Decompose the stego-image by using Haar wavelet transform
3) Extract the decoded message from HH sub band
4) Apply inverse Bit-inverting method on encoded message.
5) Apply inverse Huffman Encoding method on output of inverse bit inverting algorithm.
6) Apply inverse Bitwise-input algorithm
7) Prepare secret -object to display.

The schematic representation of decoding process was given in the Fig .7.

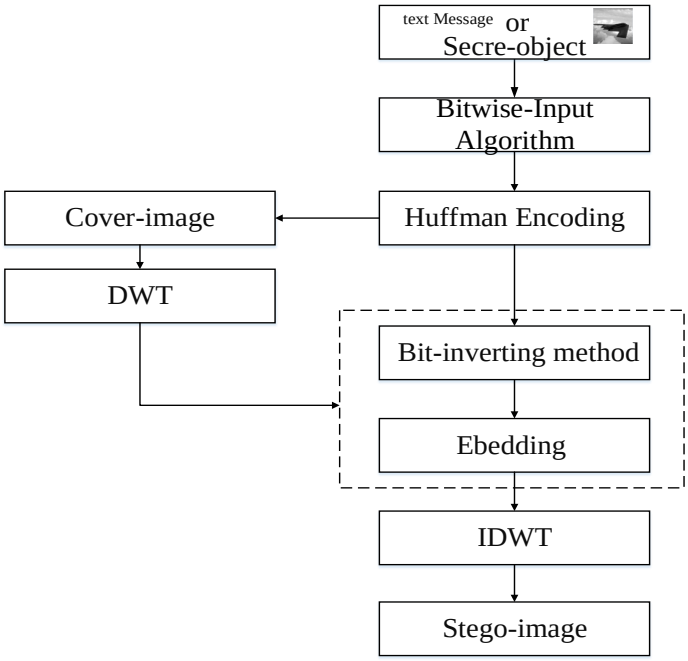


Fig .6: Encoding Process of Proposed Algorithm

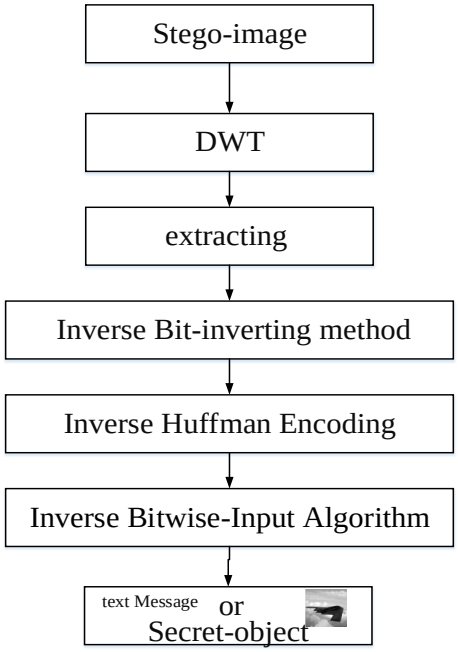


Fig.7: Decoding Process of Proposed Algorithm

4. Experimental result

The parameters of steganographic system, such as the number of data bits that can be hidden, the invisibility of the message, and its resistance to removal, can be related to the characteristics of communication system such as capacity and peak signal-to-noise ratio (PSNR) [15].

In our study, we use peak signal to noise ratio (PSNR) to measure the distortion between an original cover image and stego image. The PSNR and MSE of cover image verses stego image are defined as follows:

$$PSNR = 10log_{10}\frac{255^2}{MSE}$$

Where the Mean Square Error (MSE) defined as:

$$MSE = \frac{1}{M * N}\sum_{i=1}^{M}\sum_{j=1}^{N}(a_{ij} - b_{ij})^2$$

MSE is the mean square error representing the difference between the original cover image, a, sized M*N and the stego image, b, sized M*N, and the a_{ij} and b_{ij} are pixels located at the ith row the jth column of image a and respectively. A large PSNR value means that the stego image is most similar to original image and vice versa. It is hard for human eyes to distinguish between original cover image and stego image when the PSNR is larger than 30dB [16].
To evaluate the performance of the proposed method, we implemented the proposed method by using MATLAB.
The Fig. 8. (a) and 8. (b) Show the secret image as text or image and Fig. 9. (a) Shows the original cover image and Fig. 9. (b) Shows the stego image.

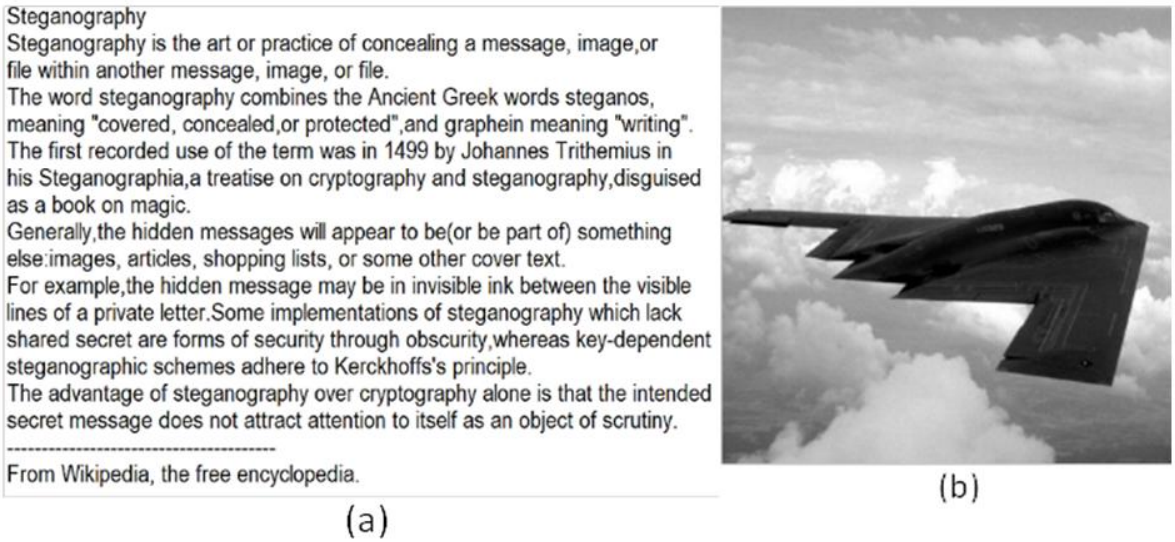


Fig .8: Secret Object to Be Embedded. (A) Secret Text, (B) Secret Image

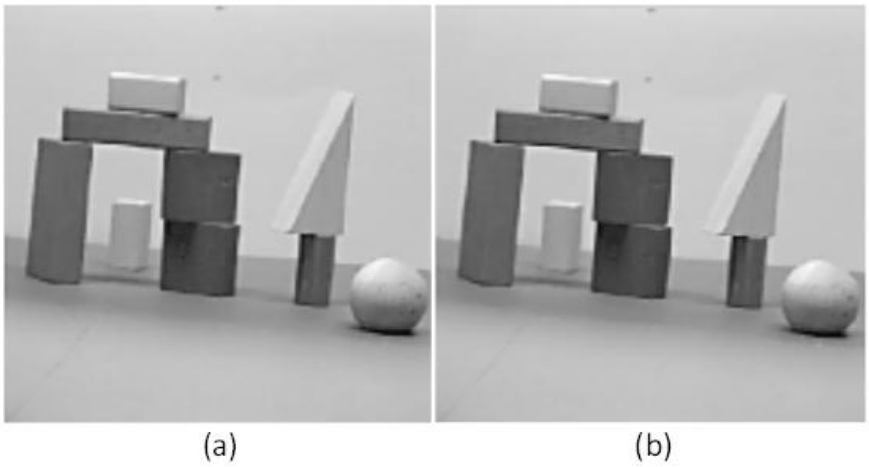


Fig .9: (A) Cover Image, (B) Stego Image

For some simulation results, 14 images with different characteristics are used to examine and compare the performance of the proposed algorithm .The message was hidden within each of cover images to study the influence of cover images nature.
Table2 shows the experimental results of the PSNR between the cover image and stego image computed for both of secret text and secret image.
From table.2 we observed that the PSNR values of proposed method are within the acceptable range (higher than 30 (dB)) which indicates that the quality is good.
The other parameter of steganographic system, such as the number of data bits that can be hidden (capacity) is taken for our experiment.
As mentioned in the section of Huffman encoding, this algorithm is an effective and popular method for variable length data compression. Our method produces a set of codewords with variable length that has the lowest average length and they are used instead of fixed length symbols.
In table 3, steganographed data size value for a message or an image after the applied data conversion (primary string bit) and steganographed data size after the applied Huffman algorithm with percentage of improvement is shown.
These results demonstrate that the proposed method improves the capacity more than 35%.

Table 2: PSNR of Proposed Method for Different Cover Image

Cover Image	Secret Text	Secret Image
Airplane	43.288	43.101
Arch	58.51	57.649
Baboon	38.292	38.187
Barbara	32.089	32.059
Boat	42.036	41.891
Cameraman	44.837	44.608
Fingerprint	37.046	36.959
Jet	48.89	48.217
Lena	40.095	39.968
Lighthouse	37.393	37.31
Man	36.901	26.660
Peppers	42.481	42.275
Rino	44.358	44.037
Woman	33.558	33.511

Table 3: Percentage Improvement in Steganography Data Size in Our Proposed Algorithm

percentage Improvement	Huffman	Bit Stream	Algorithm
56.71%	5050	8904	Secret Text Size
35.86%	12458	34740	Secret Image Size

Usually, the high capacity requirement will conflict with the high PSNR requirement. Generally speaking, when the size of the message will decrease so the capacity is increase, and this will affect the PSNR inversely. So, a trade-off should be made between capacity and PSNR requirement.
From the results shown in table 2 and table 3, it can be concluded that to progress in decrease the size of secret message and so achieve the higher capacity for embedding is better to satisfied to get this PSNR.

5. Conclusion

In this paper a secure image steganography technique is proposed to hide secret object as a text or image, which also tells how to hide data bits. The experimental results show that the technique produced good quality stego images with good PSNR values with reduced size of the secret object.

References

[1] K. Lee and H. Chen, A High Capacity Image Steganographic Model, In IEEE Proceedings on Vision Image and Signal Processing, China, 147:3, (2000), 288-294.

[2] P. V. Nadiya, B. M. Imran, Image steganography in DWT domain using double-stegging with RSA encryption, Signal Processing Image Processing & Pattern Recognition (ICSIPR), International Conference on, DOI: 10.1109/ICSIPR.2013.6497941, (7-8 Feb. 2013), 283-287. http://dx.doi.org/10.1109/ICSIPR.2013.6497941.

[3] G. Prabakaran and R. Bhavani, A Modified Secure Digital Image Steganography Based on Discrete Wavelet Transform, International Conference on Computing Electronics and Electrical Technologies (ICEET), DOI:10.1109/ICCEET.2012.6203811, (2012), 1096 - 1100. http://dx.doi.org/10.1109/ICCEET.2012.6203811.

[4] H. Reddy and K. Raja, High Capacity and Security Steganography Using Discrete Wavelet transform, International Journal of Computer Science and Security (IJCSS), 3: 6, (2008), 462-472.

[5] A. Nag, S. Biswas, D. Sarkar and P. Sarkar, Novel Technique for Image Steganography Based on DWT and Huffman Encoding, International Journal of Computer Science and Security (IJCSS), 4:6, (2008).

[6] P. C. Mandal, Modern Steganographic technique: A Survey, International Journal of Computer Science & Engineering Technology (IJCSET), 3: 9, (Sep 2012), 444-448.

[7] A. Abdelwahab and L. A. Hassaan, A Discrete Wavelet Transform Based Technique for Image Data Hiding, 25th National Radio Science Conference, DOI: 10.1109/NRSC.2008.4542319, (March 18-20, (2008), 1–9. http://dx.doi.org/10.1109/NRSC.2008.4542319 .

[8] B. Banik and Prof. Samir K. Bandyopadhyay, A DWT Method for Image Steganography, International Journal of Advanced Research in Computer Science and Software Engineering, 3:6, (June 2013), 983-989.

[9] S. Liu, H. Yao, W. Gao, Steganalysis of Data hiding Techniques in Wavelet Domain, Proc. of Int, Conf. on Information Technology: coding and computing, (2004), 751-754.

[10] M. Misiti, Y. Misiti, G. Oppenheim and J. Poggi, Wavelet Toolbox for Use with MATLAB, User Guide Math Works Inc., (2000).

[11] A. Al-Ataby and F. AL-Naima, A Modified High Capacity Image Steganography Technique Based on Wavelet Transform, The Arab journal of Information Technology, 7: 4, (2010), 358-364.

[12] G. Satyavathy and M. Punithavalli, LSB, 3D-DCT and Huffman Encoding based Steganography in Safe Message Routing and Delivery for Structured Peer-to-Peer Systems, IJCA Special Issue on Artificial Intelligence Techniques, (2011), 1-5.

[13] M. Sharma, Compression Using Huffman Coding, IJCSNS International Journal of Computer Science and Network Security, 10:5, (May 2010), 133-141.

[14] R. B. Lin and C. M. Tsai, Weight-based Bus-Invert Coding for Low-Power Applications, Processing of the 7th Asia and South Pacific Design Automation Conference and 15th International Conference on VLSI Design, DOI:10.1109/ASPDAC.2002.994897, (Jan. 2002), pp. 121-125. http://dx.doi.org/10.1109/ASPDAC.2002.994897.

[15] A. Chawla and P. Shukla, A Modified Secure Digital Image Steganography Based on DWT Using Matrix Rotation Method, International Journal of Computer Science and Communication Engineering, 2:2, (2013), 20-25.

[16] H. Motamedi, A. Jafari, A New Steganography Based on Denoising Methods in Wavelet Domain, International ISC Conference on Information Security and Cryptology, 9th, DOI: 10.1109/ISCISC.2012.6408185, (2012), 18-25. http://dx.doi.org/10.1109/ISCISC.2012.6408185.