Improved Image Steganography Technique for Colored Images using Huffman Encoding with Symlet Wavelets

Saddaf Rubab[1], Dr. M. Younus[2]

[1&2]Department of Computer Engineering, College of Electrical & Mechanical Engineering, National University of Sciences & Technology (NUST), Islamabad, Pakistan

Abstract

Steganography is data hiding technique that is mostly used. It uses cover object of type text, images, and videos. The paper presents a new devised algorithm to hide text in any colored image of any size using Huffman encryption and 2D Wavelet Transform. We presented simulated results which prove that there is very negligible image quality degradation. We used PSNR metric for this purpose. The subject algorithm also proved secure as Huffman table is required to decode the information.

Key Words: *Image steganography, Huffman, Wavelet transform*

1. Introduction

We have many techniques to imply data hiding including cryptography, steganography and watermarking. In this paper steganography is discussed only. Steganography is evolved from two Greek words that mean "covered writing". It is used to hide data within anther one either of same type or of different. It maintains the beauty of data hiding by providing better security and imperceptibility of hidden message in cover object. The main process of steganography is done by hiding the secret content in cover media by using any steganography method, so that any other party or enemy is not able to even guess that any message is embedded in [1]. To get back the secret message the reverse of steganography, called steganalysis is used. Figure 1 explains the above in pictorial form.

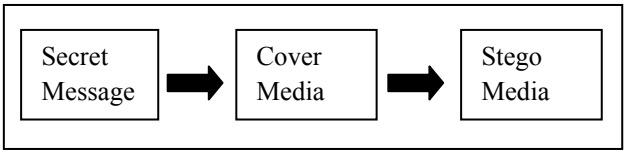

Figure 1: Image Steganography Block Diagram

Encryption is different from steganography. Encryption involves plain text and convert it to encrypted or cipher text using unique encryption key [2]. Encrypted text can only be converted back to plain text if key is known to the recipient. In both encryption and steganography we hide our original message, but output of encryption is detectable by reader and output of steganography is not detectable by reader by just seeing the text. It is clear from the difference that encrypted text is more vulnerable to security attacks by opponents than steganography.

Steganography use different cover media like, text, image, video etc. Image is most common cover media for it and we call it Image Steganography. We divide it in two domains:

a. Spatial Domain
b. Transform Domain

In spatial domain secret messages are inserted in the Least Significant Bit (LSB) of image or by bit shifting. In transform domain, secret message is embedded in the coefficients of cover media in frequency domain. In transform domain cosine, wavelet etc are used as transforms. Section 2 explains wavelet transform in more detail and Huffman algorithm in Section 3. Devised method is explained in Section 4 and simulated results are shown in Section 5. Conclusions and future work are illustrated in Section 6.

2. Wavelet Transform

In comparison to other transforms, wavelet transforms proved to be the best for image transformation [3]. In its basic operations, it decomposes the input signal into set of functions which are called wavelets. For image applications in transform domain, wavelet transform of image is computed, then modifications are made and at final step, inverse of wavelet is taken to get resulted image. We can select any family of wavelet from discrete or continuous wavelets like Haar, Coiflet,Symlet, Daubechies [4]. In discrete wavelets, we have different levels like 1-D, 2-D … n-D. This work presents 2D wavelet transform and symlet wavelets. Original signal is decomposed twice in 2-DWT in a way that makes use of scaling and wavelet functions of level 1 or 1-DWT. Figure 2 explains it in detail.

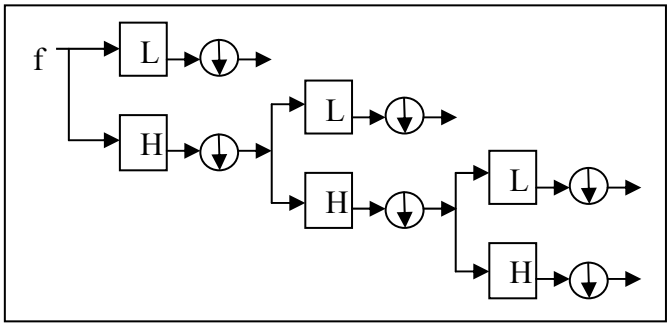

Figure 2: 2DWT

3. Huffman (Encryption Algorithm)

Several encryption algorithms including DES [5], AES [6], Triple DES [5], were used in past few years. But variable length codes get more attention from researchers in little time. It is a code that maps source symbols to variable length of bits like Lempel-Ziv coding, Huffman codes and arithmetic coding.

Huffman was introduced by David A. Huffman in 1952 [7] when he was a PhD student at MIT. It is used for lossless data compression. It represents the data in few bits and in few memory locations. It creates binary tree. The process starts by set of symbols/letters and their respective frequencies in ascending or descending order. Each symbol with its frequency is a leaf node at start. Next step is to select two symbols with smallest frequencies, add their frequencies and assign it to parent node, until only one node remains which is called the root node. Assign 0's and 1's to all the nodes and translate the codes by reading from leaf to root node. This way we construct a Huffman Table.

3.1 Huffman Table

The second part of Huffman is to decode the tree generated in compression of data. In this, we create Huffman table, which is used to decode symbols/letters in original data using the generated codes. That's the reason it implements more security because Huffman table is required at the recipient side. It prevents enemy attacks on secret data.

4. Proposed Steganography Method

The proposed algorithm was implemented in MATLAB 7.9. It is divided into two main phases.

Following are the steps of devised method:

START

Phase I: Huffman Encoding

Step I: Obtain secret Message

Step II: Perform Huffman encoding on the secret message.

Step III: Create Huffman Table.

Step IV: Perform the binary conversion.

Step V: Make groups of 3bits each

Check last group (if number of bits! = 3)

Add zero at end of bits (0)

End

End

Phase II:

Step I: Select colored Cover Image.

Step II: Separate Red, Green and Blue Components (RGB).

Step III: Apply 2DWT on each component separately.

Step IV: Embed Huffman codes (calculated in Phase I)

Step V: Add RGB Components

Step VI: Apply Inverse 2DWT

End

END

To embed Huffman codes, firstly divide Huffman groups by 3. For example we have 12 groups formed by Huffman encryption phase. By dividing it by three (03), we get 3 sets of 4 groups each. Now select 3 consecutive bits of first set and insert into the Least Significant Bit (LSB) of 2DWT coefficient of red component. Do this for the rest of three groups in set 1. Repeat this with set 2 for 2DWT coefficient of green component of and set 3 for 2DWT coefficients of blue component.

5. Experimental Results

5.1 Peak Signal To Noise Ratio (PSNR)

In this work PSNR values are used to show the image imperceptibility [8].

$$MSE = \frac{1}{MN} \sum_{i=0}^{M} \sum_{j=0}^{N} (X_{ij} - \overline{X_{ij}}) \quad (1)$$

$$PSNR\ (dB) = 10\log_{10} \left(\frac{I^2}{MSE}\right) \quad (2)$$

X_{ij} represents the i^{th} rows and j^{th} columns of original image and $\overline{X_{ij}}$ represents the i^{th} rows and j^{th} columns of transformed image. Higher the PSNR value means more difficult to perceive that any hidden message is hidden.

5.2 Results

Devised method is implemented in MATLAB 7.9 using Windows 7 platform. Results are carried out with a Microsoft Word file of 1 page varying number of letters. This work is simulated using jpeg format colored cover

images of size 512x512. Following are the three images of size 512x512 to discuss the results.

Figure 3a: Lena

Figure 3b: Boats

Figure 3c: Mask

Table 1 shows the PSNR values against the numbers of characters embedded. It is clear from the results that by increasing the number of characters the PSNR value decays. All the images are of same size, if we increase image size and number of characters remains constant, then the PSNR value will improve.

Table 1: PSNR values for 512x512 images with respect to number of characters

Cover Image (.jpeg)	Number of characters	PSNR
Lena	3584	64.79
	4915	63.46
	6809	60.53
Boats	3584	64.27
	4915	63.07
	6809	60.36
Mask	3584	63.84
	4915	62.96
	6809	60.10

6. Conclusions

The presented algorithm is for any size of colored image. It gives more capacity for larger image sizes. It enhances security and also preserves the image quality. By inserting Huffman codes into the three components of colored image it becomes complicated but it provides us with better security measures, we can say it provides triple security. In terms that someone if find Huffman table it is not possible even then to decode the full message. This new method also improves the PSNR values. In future work may be carried out to increase the capacity and encode the secret message with more number of words.

7. References

[1] N. Provos, P. Honeyman, "Hide and seek: An introduction to steganography", IEEE Security Privacy Magazine (2003), Volume: 1, Issue: 3, Publisher: IEEE Security & Privacy, Pages: 32-44
[2] Postnote October 2006 Number 270 Data encryption
[3]http://scholar.lib.vt.edu/theses/available/etd-12062002-152858/unrestricted/Chapter4.pdf
[4] C. E. Heily and D. F. Walnut, "Continuous and discrete wavelet transforms", SIAM Review 1989 Society for Industrial and Applied Mathematics, Vol. 31, No. 4, pp. 628-666, December 1989 007
[5] National Institute of Standards and Technology, "Data Encryption Standard (DES)", FIPS PUB 46-3, 1999 October 25
[6] NIST, "Report on the Development of the Advanced Encryption Standard (AES)", October 2, 2000
[7] David A. Huffman, "A Method for the Construction of Minimum - Redundancy Codes", Proceedings of the I.R.E., September 1952, pp 1098–1102
[8] Almohammad, A.; Ghinea, G.; "Stego image quality and the reliability of PSNR", 2nd International Conference on Image Processing Theory Tools and Applications (IPTA), 2010, Pages: 215 – 220