Data Hiding in Still Images Based on Blind Algorithm of Steganography

Gabriel BUGÁR, Vladimír BÁNOCI, Martin BRODA, Dušan LEVICKÝ, Denis DUPÁK
Dept. of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic
gabriel.bugar@tuke.sk, vladimir.banoci@tuke.sk, martin.broda@tuke.sk, dusan.levicky@tuke.sk, denis.dupak@tuke.sk

***Abstract*— Steganography is the science of hiding secret information in another unsuspicious data. Generally, a steganographic secret message could be a widely useful multimedia: as a picture, an audio file, a video file or a message in clear text - the covertext. The most recent steganography techniques tend to hide a secret message in digital images. We propose and analyze experimentally a blind steganography method based on specific attributes of two dimensional discrete wavelet transform set by Haar mother wavelet. The blind steganography methods do not require an original image in the process of extraction what helps to keep a secret communication undetected to third party user or steganalysis tools. The secret message is encoded by Huffman code in order to achieve a better imperceptibility result. Moreover, this modification also increases the security of the hidden communication.**

***Keywords*—steganography, DWT, message hiding.**

I. INTRODUCTION

Steganography is the art of concealing the existence of information within seemingly innocuous carriers. Images are the most popular cover media for steganography. In the last decade, several steganography schemes have been developed to solve the privacy problem. Among these schemes is an approach that hides a secret message in the spatial domain of the cover image.

In Lee and Chen's method [2], the least significant bit (LSB) of each pixel in the cover image is modified to embed a secret message. In Wang et al.'s method [3], the optimal substitution of LSB is exploited. These schemes have a higher quality of stego image but they are very sensitive to modification. In Chung et al.'s method [4], singular value decomposition (SVD) based hiding scheme is proposed. In Tsai et al.'s scheme [5], the bit plane of each block truncation coding (BTC) block is exploited to embed a secret message. For such reasons, many imagery steganography methods have been invented. Here we briefly review research carried out particularly in the Discrete Cosine Transform (DCT) domain as JSteg method that hides information sequentially in LSBs of the quantized DCT coefficients (qDCTCs) while skipping 0's and 1's; OutGuess method scatters information into the LSB of qDCTCs [6]. Another method employs the technique of matrix encoding to hold secret information using LSB of qDCTCs in F5. Others researches of using DCT transform are mentioned in [7], [8].

In this paper we propose a novel method, which allows hiding secret information in still gray scale image while maintaining good visual quality and high capacity of secret information. Advantage of this method is full reconstruction of secret data without having original image present on the recipient side. The organization of the paper is as follows. In the next section, we describe the fundamentals of discrete wavelet transformation in images, the Huffman encoding and objective visual quality measurements to simulate human perception which are PSNR [dB] and SSIM. In Section III. we give overview of proposed steganography system design. In Section IV. are given results of proposed method in manner of PSNR and SSIM values using with different detail levels cover images. Finally, we conclude our paper in Section V. with discussion and contribution of our proposed method to the image steganography.

II. DISCRETE WAVELET TRANSFORM IN IMAGES

A Haar wavelet represents the elementary type of wavelet. In discrete form, Haar wavelets are related to a mathematical operation called the Haar transform. The Haar transform serves as a prototype for all other wavelet transforms. The Haar transform, like all wavelet transforms, decomposes a discrete signal $f = (f_1, f_2,\ldots,f_N)$ into two subsignals $(a_1|d_1)$ of half its length, where $a^1 = (a_1, a_2,\ldots, a_{N/2})$ represents the approximation (average) coefficients and $d_1 = (d_1, d_2\ldots d_{N/2})$ the detailed (difference) coefficients. The first value of approximation coefficient a1 is computed by taking the average of the first pair of values of $f : (f_1 + f_2) / 2$; and then multiplying it by $\sqrt{2}$. where formula for a^1 values is given as follows [10]:

$$a_n = \frac{f_{2n-1} + f_{2n}}{\sqrt{2}} \tag{1}$$

The other sub-signals are referenced as detailed (difference) coefficients. The first fluctuation of the signal f, which is denoted by $d^1 = (d_1, d_2\ldots d_{N/2})$ is calculated according to formula

$$d_n = \frac{f_{2n-1} - f_{2n}}{\sqrt{2}}, \tag{2}$$

where $n=1, 2, \ldots, N/2$.

When DWT is applied to an image it is decomposed into 4 sub bands: LL_{ij}, HL_{ij}, LH_{ij} and HH_{ij}. LL_{ij} part contains the most significant features. So if the information is hidden in LL_{ij} part the stego image can withstand compression or other

978-1-4799-3715-8/14/$31.00 ©2014 IEEE

manipulations. But sometimes distortion may be produced in the stego image and then other sub bands can be used.

Application of 2D-HT to a grayscale static image will retrieve transformation coefficients that can be defined by the next formulas:

$$LL_{i,j} = \left(I_{2i-1,j} + I_{2i,j}\right)\frac{1}{\sqrt{2}} = = \left[f(2i-1,2j-1) + f(2i-1,2j) + f(2i,2j-1) + f(2i,2j)\right]\frac{1}{2} \quad (3)$$

$$LH_{i,j} = \left(I_{2i-1,j} + I_{2i,j}\right)\frac{1}{\sqrt{2}} = = \left[f(2i-1,2j-1) + f(2i-1,2j) - f(2i,2j-1) + f(2i,2j)\right]\frac{1}{2} \quad (4)$$

$$HL_{i,j} = \left(I_{2i-1,j} - I_{2i,j}\right)\frac{1}{\sqrt{2}} = = \left[f(2i-1,2j-1) - f(2i-1,2j) + f(2i,2j-1) - f(2i,2j)\right]\frac{1}{2} \quad (5)$$

$$HH_{i,j} = \left(I_{2i-1,j} - I_{2i,j}\right)\frac{1}{\sqrt{2}} = = \left[f(2i-1,2j-1) - f(2i-1,2j) - f(2i,2j-1) - f(2i,2j)\right]\frac{1}{2}, \quad (6)$$

where i = 1, 2, ... M/2 and j = 1, 2, ..., N/2.

A. Huffman encoding

Huffman coding is a popular method for compressing data with variable-length codes. Given a set of data symbols (an alphabet) and their frequencies of occurrence (or, equivalently, their probabilities), the method constructs a set of variable-length codewords with the shortest average length and assigns them to the symbols. Huffman coding serves as the basis for several applications implemented on popular platforms. Making use of these properties, there is the possibility of using Huffman coding to encode the secret message before it is being embedded to the cover image. This modification reduces the message size and enhances its security as well [5].

B. Visual quality of stego-image

The objective measures, which stem from statistical approach, are used in general. Their application is based on measuring of the distortion between the cover data $f(i, j)$ with resolving capacity $M \times N$ and modified data $f(i, j)$ with same resolving capacity. To the most frequent used criterions for evaluating the quality of reconstructed information are Peak Signal Noise Ratio (PSNR) and Structural SIMilarity Index (SSIM), which has better reflection of Human Visual System (HVS) model.

Parameter	
Peak signal to noise ratio	$PSNR = 10\log_{10}\frac{\left(2^n-1\right)^2}{MSE} \quad [dB]$
SSIM	$SSIM(A,B) = \frac{\left(2\mu_A\mu_B + C_1\right)\cdot\left(2\operatorname{cov}_{AB} + C_2\right)}{\left(\mu_A^2 + \mu_B^2 + C_1\right)\cdot\left(\sigma_A^2 + \sigma_B^2 + C_2\right)}$

where MSE is the Mean squared error, SSIM (A, B) represents the measure between two windows A and B of common size.

III. Steganography System Design

The proposed steganography method is based on minimal changes of 2D-HT transformation coefficient's values of the cover image. High correlation between the transformation coefficients on same position in different decomposition areas is a very important feature. If coefficient $LL_{i,j}$ is even integer number then also coefficients $LH_{i,j}$, $HL_{i,j}$ and $HH_{i,j}$ are even integer numbers. Hence, if a coefficient value is modified in a decomposition area than there is also needed the change of the coefficient value at the same position accordingly in other three areas. Otherwise, reconstructed image pixel values may not be integer numbers. It can cause; stego image would not be reconstructed. So there are only four suitable modifications of transformation coefficients for embedding. Each modification affects a different set of image pixels' values after applying 2D-IHT. Tab. I. presents all type of possible coefficient modifications.

TABLE I. Modification of Transformation Coefficient

	$I_{2i-1,\,2j-1}$	$I_{2i-1,\,2j}$	$I_{2i,\,2j-1}$	$I_{2i,\,2j}$
$LL_{i,j}$	*1*	*1*	*1*	*1*
$LH_{i,j}$	*0*	*1*	*0*	*1*
$HL_{i,j}$	*1*	*0*	*0*	*1*
$HH_{i,j}$	*0*	*0*	*1*	*1*

where "*1*" and "*0*" represent the transformation coefficients modifications on the same position in areas according to the secret message value.

A. Embedding process

The process of embedding begins with a choice of compliant cover and secret image. Cover image is an ordinary grayscale images given by size of M×N pixels. According some steps we create auxiliary picture matrices from the selected cover image with same dimension. Picture elements of $I^1_{i,j}$ matrix are equal to the picture elements of every odd column and odd lines of cover image on same position. Picture elements of $I^2_{i,j}$ matrix are equal to the picture elements of every even column and odd lines of the cover image on same position. Picture elements of $I^3_{i,j}$ matrix picture elements are equal to the picture elements of every odd column and even lines of the cover image on same position. Picture elements of $I^4_{i,j}$ picture elements are equal to the picture elements of every even column and even lines of the cover image on same position. Other picture elements of $I^1_{i,j}$, $I^2_{i,j}$, $I^3_{i,j}$ and $I^4_{i,j}$ matrixes are zeros.

$$I^1{}_{2i-1,2j-1} = I_{2i-1,2j-1} \quad (7)$$

$$I^2{}_{2i-1,2j} = I_{2i-1,2j} \quad (8)$$

$$I^3{}_{2i,2j-1} = I_{2i,2j-1} \quad (9)$$

$$I^4{}_{2i,2j} = I_{2i,2j} \quad (10)$$

where i=1, 2, 3…M and j=1, 2, 3…N.

Each matrix was transformed by third level DWT, where we obtain the third decomposition transformation coefficients. From equations (3) – (6), the values of the components $LL_{i,j}$, $LH_{i,j}$, $HL_{i,j}$ and $HH_{i,j}$ are different only in sign:

$$I_{ij}^1 : LL_{ij} = LH_{ij} = HL_{ij} = HH_{ij} \quad (11)$$

$$I_{ij}^2 : LL_{ij} = LH_{ij} = -HL_{ij} = -HH_{ij} \quad (12)$$

$$I_{ij}^3 : LL_{ij} = -LH_{ij} = HL_{ij} = -HH_{ij} \quad (13)$$

$$I_{ij}^4 : LL_{ij} = -LH_{ij} = -HL_{ij} = HH_{ij} \quad (14)$$

In the next steps the secret image picture elements are adjusting to a column vector and converting to binary values (*0*, *1*). Subsequently, these values are rearranged to a row vectors so the first bit shall be denoted as MSB bit and likewise, end bit as LSB bit of picture elements. Through these changes we obtain a vectors, which contain significant groups of zero and unit bits. These modifications actually represent the strong correlation between adjacent pixel elements. In the next step, the secret message vector is encoded by Huffman coding. Taking advantage of modifications we made, the efficiency of the resulting code is increased, i.e. we achieved a greater data compression.

In the next step, the secret message matrix is divided to 64 matrices. This is followed by the process of secret message embedding. The transformation coefficients of the first matrix area I_{ij}^1 will be change according to the bits of the adjusted secret message. Modified coefficients are obtained as follows:

$$I^1(i,j) = floor(I^1(i,j)) \quad (15)$$

$$\begin{aligned} &if \quad m(2i-1,2j) == 1 \\ &\quad then\ \hat{I}^1(i,j) = I^1(i,j) + 0.5 \end{aligned} \quad (16)$$

$$\begin{aligned} &if \quad m(2i,2j-1) == 1 \\ &\quad then\ \hat{I}^1(i,j) = I^1(i,j) + 0.25 \end{aligned} \quad (17)$$

$$\begin{aligned} &if \quad m(2i,2j) == 1 \\ &\quad then\ \hat{I}^1(i,j) = I^1(i,j) + 0.125 \end{aligned} \quad (18)$$

$$\begin{aligned} &if \quad m(2i-1,2j-1) == 0 \wedge I^1(i,j) == 2k+1 \\ &\quad then\ \hat{I}^1(i,j) = I^1(i,j) + 1 \end{aligned} \quad (19)$$

$$\begin{aligned} &if \quad m(2i-1,2j-1) == 1 \wedge I^1(i,j) == 2k \\ &\quad then\ \hat{I}^1(i,j) = I^1(i,j) + 1 \end{aligned} \quad (20)$$

$$\begin{aligned} &if \quad abs(I^1(i,j) - \hat{I}^1(i,j)) > 1 \\ &\quad then\ \hat{I}^1(i,j) = I^1(i,j) - 2\,, \end{aligned} \quad (21)$$

where $m(i,j)$ represents the modified secret message, $I^1(i,j)$ represents the adapted cover image DWT transformation coefficients and $\hat{I}^1(i,j)$ represents the DWT transformation coefficients of stego image.

The TABLE II. illustrates the all possible outcomes of these operations, where for the labeling of even and odd numbers are used symbols ***e*** and ***o***. We apply the component modification of $LH_{i,j}$, $HL_{i,j}$ and $HH_{i,j}$ areas according to the rules described above.

TABLE II. TRANSFORMATION COEFFICIENTS MODIFICATION

***m*(*x*, *y*)**	***LL'*(*x*, *y*)**	***m*(*x*, *y*)**	***LL'*(*x*, *y*)**
0000	± ***e***.000	**1**000	± ***o***.000
0001	***e***.125 or –***e***.875	**1**001	***o***.125 or –***o***.875
0010	***e***.250 or –***e***.750	**1**010	***o***.250 or –***o***.750
0011	***e***.375 or –***e***.625	**1**011	***o***.375 or –***o***.625
0100	± ***e***.500	**1**100	± ***o***.500
0101	***e***.625 or –***e***.375	**1**101	***o***.625 or –***o***.375
0110	***e***.750 or –***e***.250	**1**110	***o***.750 or –***o***.250
0111	***e***.875 or –***e***.125	**1**111	***o***.875 or –***o***.125

Every transformation coefficient of cover image can hide four bits of secret message in third decomposition level, what is obvious from Tab. II. The first bit (marked as bold in Tab. II) of secret message is related to coefficient modification according to change to even or odd number. The remaining three secret message bits are hidden into the decimal part of the transformation coefficient value. In third decomposition level the transformation coefficient decimal part values can have eight possible outcomes ($8 = 2^3$ » 3 bits).

B. Extraction process

On the receiver side, it is necessary to have the stego image and know the imbedding algorithm's modifications of cover image in order to successful reconstruction of the secret message. Firstly, we create four matrices based on stego image in the same way as they were created during the imbedding process from cover image. Then we apply third level 2D-HT to each matrix. From each transformation coefficient matrix is extracted a part of secret message. The proposed method retrieves the first part of information from the LL_{ij} transformation coefficients according to Tab. II.

IV. EXPERIMENTAL RESULTS

In this section we present performance of proposed method based on different features' length of code word of secret message. For testing purposes the cover and secret images were selected with different detail levels and image resolution. All experiments were performed in MATLAB R2010a on a PC with hardware configuration: Intel(R) Core(TM)2Quad CPU Q8400 @ 2.66GHz, 4.00GB RAM. Besides a proper selection of cover image, there is also an important the selection of the Huffman coding alphabets size. The Tab. III. contains a Huffman coding efficiency, secrete message coding duration and duration of the image decoding in case of different alphabets size. Testing of Huffman coding efficiency was performed with the same size of images however with different image's detail levels. The coding efficiency is defined as the ratio of the symbol size and the average length of code word.

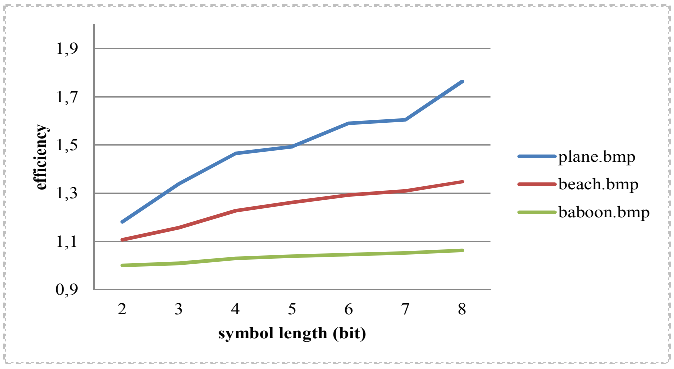

Fig. 1. Symbol length Huffman encoding efficiency

The Fig. 1. shows the character of code efficiency for different images. It can be observed that with increasing character's size the coding efficiency is also increasing. Before encoding the message, the secret message bits are rearranged. This step will cause that Huffman coding would be more effective in selected images that usually posses with significant amount of monotone surfaces.

TABLE III. SECRET MESSAGE HUFFMAN ENCODING EFFICIENCY

	baboon.bmp			*beach.bmp*			*plane.bmp*		
symbol *(bit)*	***efficiency***	***coding duration*** *(sec)*	***decoding duration*** *(sec)*	***efficiency***	***coding duration*** *(sec)*	***decoding duration*** *(sec)*	***efficiency***	***coding duration*** *(sec)*	***decoding duration*** *(sec)*
2	1.00	3.17	35.78	1.11	3.31	32.59	1.18	3.02	32.41
3	1.01	2.41	14.40	1.16	2.37	13.63	1.34	2.33	12.68
4	1.03	2.22	13.03	1.23	2.27	11.49	1.47	2.21	10.31
5	1.04	2.54	15.56	1.26	2.55	13.17	1.49	2.51	11.41
6	1.05	3.35	22.80	1.29	3.46	18.77	1.59	3.28	15.24
7	1.05	4.96	39.19	1.31	5.17	32.30	1.61	4.91	25.10
8	1.06	8.27	71.84	1.35	8.70	56.63	1.76	8.28	41.99

The Fig. 2 shows duration of the embedding and extraction process. Experiments have shown that secret message encoding and the decoding execution is reaching its maximum for character length equals to 4 bit/symbol. For this reason further experiments are executed with characters 4 bit/symbol length.

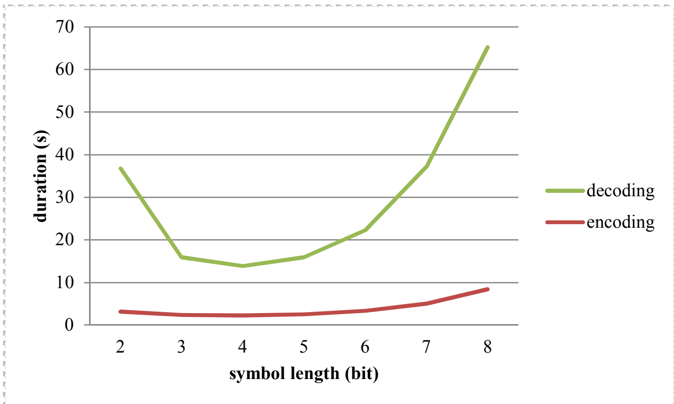

Fig. 2. Encoding and decoding execution time

The advantage of this method is a large capacity. In order to hide one bit of secret messages, only 2 bits of cover image's picture element are needed. In special cases we are able to hide one bit of secret message to one bit of cover image's picture element. This can be achieved if the size of the encoded message (including the header) is smaller than the original size of the secret message.

TABLE IV. EMBEDDING AND EXTRACTION RESULTS

Cover Images *(384×384)*	***PSNR (dB)***	***SSIM***	***Capacity utilization (%)***	***Embedding duration (s)***	***Extraction duration (s)***
Forest	34.993	0.7917	95.2099	38.9712	181.1879
Lena	35.064	0.8723	95.2099	37.4675	178.3249
Desert	34.982	0.9465	95.2099	38.8029	177.2396

To ensure this conditions; it is necessary that the secret image must contains sufficiently amount of monotone areas. The Tab. IV. contains results of insertion and extraction of a secrete image with the same dimension as the cover image. Alphabet character's length of Huffman code is 8 bit / symbol.

V. CONCLUSION

In this paper we design steganography method that uses the properties of Haar transform coefficients. The secret message is compressed before insertion in order to enlarge the capacity of the proposed system. The great advantage of this method is that the extraction does not require the presence of the original - cover image what increases the safety of steganography system. However, increasing transmission capacity has a negative impact on the secret messages perceptibility in stego image. The selection of a suitable cover image is highly important in method's capacity as well as imperceptibility of stego image due to cover images degradation on large monotonous areas that are affected by embedding.

ACKNOWLEDGMENT

The paper was supported by Ministry of Education of Slovak Republic VEGA Grant No. 1/0386/12.

REFERENCES

[1] D., Levický and S. Šurin: Codebook Code Division Multiple Access Image Steganography. In I. S. Radioengeneering, Vol. 14, No. 1, APRIL 2005.
[2] Lee, Y. K., Chen, L. H., High capacity image model. *IEEE Proc. Vision, Image Signal process. 147(3) (2000) 288-294.*
[3] Wang, R. Z., Lin C. F., Lin, J. C. Image hiding by optimal LSB substitution and genetic algorithm. Pattern Recogn. 34 (3) (2001) 671-683.
[4] Chung, K. L., Shen, C. H., Cang, L.C. *A novel SVD- and VQ-based image hiding scheme.* Pattern Recogn. Lett. 22 (9) (2001) 1051-1058.
[5] Tsai, P., Hu, Y. C., Cang, C. C. An image hiding technique using block truncation coding. *Proceedings of Pacific Rim Workshop on Digital Steganogarphy.* Kitakyaushu, Japan, July 2002, pp. 54-64.
[6] D. Solomon, A concise Intorduction to Data Compression, Springer, Sofcover 2008, ISBN: 978-1-84800-071-1, pp. 62–91.
[7] Sslle, P. Model based steganography, in: *International Workshop on Digital Watermarking*, 2003, pp. 154–167.
[8] Miyake, K., Iwata, M., Shiozaki, A. Digital steganography utilizing features of JPEG images. *IEICE Trans. Fundam.* E87-A, April 2004, pp. 929–936.