Implementation Of Steganography For Business Documents Security Using Discrete Wavelet Transform Method

Trientje Marlein Tamtelahitu
Department of Information System
Postgraduate Program,
Diponegoro University
Semarang, Indonesia
Email:marlein_juan2@yahoo.com

Eko Sediyono
Department Of Information
Technology
Satya Wacana Christian University
Salatiga, Indonesia
E-mail: ekosed1@yahoo.com

Aris Sugiharto
Department of Information System
Postgraduate Program,
Diponegoro University
Semarang, Indonesia
Email: aris.sugiharto@gmail.com

***Abstract*—Increasingly popular digital media, leading to the emergence of the problem of illegal interception and unauthorized copying of business documents, thus the attention on the security level of business documents becomes very important. A way to protect digital data is steganography. This paper discusses the implementation of DWT-2D steganography. Implementation is done using Matlab R2010a. The experiment performed using the computer with AMD Athlon Neo MV-40 processor, speed 1.6 GHz, and memory (RAM) 2 Gb, resulting stego-image quality/Peak Signal Noise to Ratio (PSNR) between 74.10 dB - 81.79 dB. This experiment uses data correction procedure during extraction process to keep texts returned fully**

Keywords: Steganography, Discrete Wavelet Transform (Dwt), Business Documents, Psnr.

I. INTRODUCTION

Icreasingly popular digital media, resulting in the problem of illegal interception and unauthorized copying of business documents [2], with attention to the level of security will become increasingly important. This led to the need for effective methods of data protection. There are several ways to protect digital data which are transfered to the cryptography [4], steganography [8] or watermarks [6]. This article discusses how to use steganography to protect electronic documents, or PDF format, such as trade transaction documents, correspondence, agreements, exam, financial data and others.

Steganography is a technique of data security by concealing the existence of such information in the media [7]. It is expected not to invite suspicion of the existence of the perception of human observation. It is expected not to invite suspicion of the existence of the perception of human observation. By exploiting the weakness of the human eye, then the digital image is selected as the inserted data carrier to carrier does not damage the quality of the image [1]. Application of digital image steganography techniques have good performance if the operator does not degrade image quality [1][8]. In this paper discussed method Discrete Wavelet Transform steganography as a technique for securing business documents while maintaining the quality of the image carrier/cover image has been interpolated business documents.

II. RELATED RESEARCH

Yue Chen Chi and Tseng (2006), in the publication entitled " Study of steganography database image analysis coefficient of DWT ", proposed that 2-D DWT method using Haar-DWT using information towards neighbouring pixels to assess the capacity of the insertion point. This scheme uses two phases: phase structure prediction and entropy coding stage. secret message is embedded into the image difference values are given after the prediction step done. In the middle of the Edge detector (Mayor) of smart mode, each pixel is processed order scanning previously defined. MED Predictor predictive value for each pixel is processed. Advantages of this method in the frequency domain is to improve the quality of the image by hiding a message in the frequency of the HL, LH and HH. Maximal PSNR value arising from this study is 52.96 DB.

M. Abolfathi and R. Amirfattahi (2010), in the publication entitled "design and implementation of a reliable and authenticated satellite image communication" carry out secure communications from satellite images with a minimum delay that can tolerate. A new method is presented as a design and modern techniques. In this case, using the Discrete Wavelet transform (DWT) as a means of common ground for the compression of satellite imagery and digital watermark. In the process of sowing, the signal from water and mark the DWT coefficients that were the watermark can be selected depending on the value of the statistical function of the image. The proposal for a TPM based on watermark technique is different from the other. Areas of high frequency wavelet coefficients is used as an inconspicuous area to the planting of watermark data. Kim et al., (1999), suggests a number different from the watermark, comparable to the energy contained in each different band. Neural networks used for implementing an automated system that can create watermarks maximum attack power. Maximal PSNR value arising from this study is to 39.332 DB.

Amitava Nag et al (2011), in a publication titled "A Novel Image Steganography Techniques Based on DWT and Huffman Encoding", image-based steganographic techniques proposed DWT, DWT is used to alter the original image (cover image) from the spatial domain into the frequency domain. 2-D DWT first performed at the level grayscale image with size M×N, Huffman encoding is performed on a secret message (picture). Then each bit of the Huffman encoding secret messages that were randomly inserted at a frequency of LH and HL with reference to the frequency of HH as an image mapping. The results showed that the algorithm has a high capacity and secure quality of the image. The maximum PSNR values derived from this study is 54.93 dB.

Steganography is discussed in this article is to hide the business documents doc or pdf to the carrier image/covered two dimensions have to do a better inclusion in such a way that the image quality is maintained through the use of the discrete Wavelet transform.

III. HOW TO RESEARCH AND METHODS

The method proposed in this paper are : the initial stage, the cover image of the original color image converted to grayscale and then converted into DWT using Haar-DWT function which is a simple easy-DWT is applied to get the 4 frequencies[8]. The second stage, determining the embedding assuming a character /byte (= 8 bits) by taking 4 bits LSB position of the character is inserted before one last bit of a byte frequency of HLH, the next 4 bits MSB which is the rest of the characters are placed before one last bit 1 byte at a frequency of HHL. Stage 3, after all the business documents is inserted into the cover image, the DWTcoefficients slightly modified and converted back to spatial domain which produces the stego-image. Stage 4, the stego-image with the color images were merged back and ready for quality evaluation.

Extraction stages of the secret message from the stego-image, done in a way the opposite of embedding. Correction data is done at the time of extraction or store the position of bytes and ASCII values are changed during the embedding process.

Embedding Procedures

Suppose D is the original 8-bit grayscale cover-image with pixel $M_c \times N_c$, denoted as:

$$D=\{x_{ij} \mid 1 \leq i \leq M_c, 1 \leq j \leq N_c, x_{ij} \in \{0,1,\ldots,255\}\} \qquad (1)$$

S is an n-bit secret message represented as:

$$S = \{s_i \; 1 \leq i \leq n, s_i \in \{0,1\}\} \qquad (2)$$

1. Apply D DWT in the frequency domain to obtain H. 4 frequencies obtained matrix is denoted as HLL, HHL, HLH and HHH (All frequencies have the same size $M_c/2 \times N_c/2$).
2. Normalization coefficientat HHL and HLH with equation 3 to achieve the normalization:

   $$N= C_{ij}/255 \qquad (3)$$

   In equation 3, the C_{ij} is the value of the coefficient in the HHL and HLH.
3. Embedding of HLH and HHL based raster scan order can be seen in figure 1 (Yueh Chen and Chi Tseng, 2006), embed the text into bits the original coefficient value, the coefficient of by coefficient.

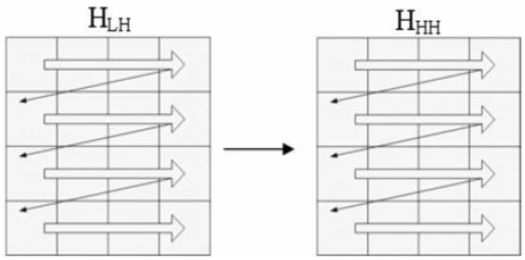


FIGURE 1. EMBEDDING WITH RASTER SCAN ORDER

4. Embedding a character/byte is done by placing 4 bits LSB to MSB HLH and 4 bits to the HHL.

   Assumptions for a character study as follows:

   $$\underbrace{YYYY}_{MSB}\underbrace{YYYY}_{LSB}$$

How to insert a character/byte to the HLH is as follows:

a. Assumption1 byte value frequency HLH before embedded are as follows:

   XXXXXXXX

b. Assumption embedding bit LSB to MSB to HLH and HHL, were as follows:

   XXXYYYYX

   In order for embedding in accordance with the assumption that embedding bits LSB to HLH is 4 bit before one last bit, then the assumption must be zeroed, so that can be inserted YYYY, how during the last 5 bits of HLH slide to the right, take one last bit LSB HLH is stored as a variable P, leaving 3 bits MSB of HLH. The assumptioncan be seen as follows: (1)

   $$XXXXXXXX \longrightarrow P \qquad (2)$$

Scroll to the right 5 times, becomes:

OOOOOXXX

After that restore the position of the initial 3-bit MSB HLH by sliding back to the left 5 times, the results are as R.

$$\text{XXXOOOOO} \longrightarrow R$$

In order for embedding character bit LSB embedding according to the assumption of bit characters in HLH, take bits LSB of secret characters and slide to the left 1 time result as T. Then performed on the HLH bit LSB embedding using the formula:

$$U = (R \quad T) \quad P \tag{4}$$

Assumption embedding character bit HLH applies also to the embedding HHL character is taken from a secret character bits MSB first time sliding to the left, the result as a T. Then do the embedding bit MSB at HHL by using equation 3.

5. After pasting all the bits of text,then modified the coefficient matrix H'. By doing the inverse DWT(IDWT) on H', E is obtained as a stego-image. To normalize there-use the formula:

$$N = 255C_{ij} \tag{5}$$

6. In order to reconstruct the text byte, byte position and the originalvalueis stored as correction data with K symbol required at the time of extraction.
7. Then stored in a special image file format, while K is filled in the tag is not used (e.g., metadata tagged "comments" in the PNG format or metadata tag "correction" tag). Stego-image the entire file is ready for transmission.

Extraction Procedure

$$F = \{y_{ij} | 1 \leq i \leq M_F, 1 \leq j \leq N_F,\ y_{ij} \in \{0,1,\ldots 255\}\} \tag{6}$$

1. Extract the file tag correction data E as:

   $$K = \{K_{ij} | 1 \leq i \leq M_F, 1 \leq j \leq N_F, K_{ij} \in \{\text{position byte}|\text{original value}\}\} \tag{7}$$

   Example K_{ij}= 256|72, meaning that byte position 256 and the original value (ASCII) 72.
2. Get the matrix H' by normalizing DWT in E by using equation 3
3. Specify the bit to be extracted with a raster scan order, extracting bit soft text messages by coefficient value coefficient.
4. Assumed extraction of one frequency H'HL or H'LH:

   XXXYYYYX

   To take 4 bits YYYY H'LH and placed on the LSB way sliding a bit to the right, the result as a W. Decide on H'LH LSB extraction by the formula:

   $$Z = W \wedge 00001111 \tag{8}$$

   To take on H'HL YYYY 4 bits and placed as MSB, how to slide the 3 bits to the left on H'HL, the results as A.Decide on H'HL MSB extraction by the formula:

   $$B = A \wedge 11110000 \qquad (4) \tag{9}$$

   Merger the results of extraction of LSB and MSB to 1 character/byte, is determinedby the formula:

   $$C = Z \vee B \tag{10}$$

Data Correction Procedure

1. Take S, change the corresponding byte position and the original value (ASCII) K.
2. Save the Zip file.

(5)

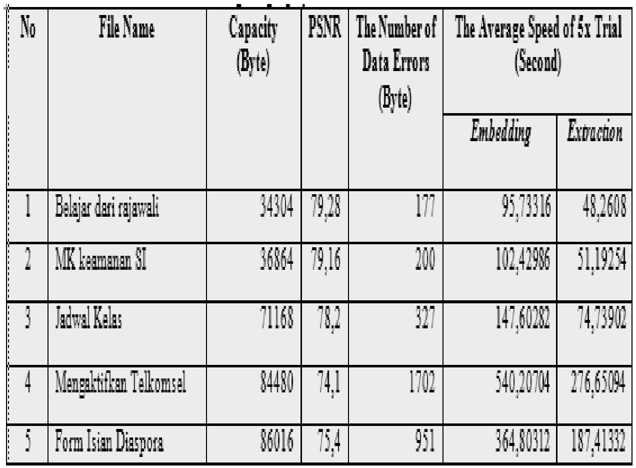
IV. RESEARCH RESULTS

Color images were used as a carrier/cover image jpg format, with the name of the image "Peppers" with size 510x512, 8-bit as shown in Figure 2. This image is the image of testing standard for digital image processing computing [3].

FIGURE 2. COLOR IMAGES “PEPPERS.JPG”

The experiments were performed on a laptop computer with AMDAthlon™ Neo MV-40, 1.6 GHz speed, and memory (RAM) 2Gb. Performance embedding process, PSNR and extraction of digital data (text) can be seen in the capacity of data (bytes), the number of errors of data (bytes), PSNR (dB), embedding time(second) and extraction time(second), each of which is show in table I, table II, table III and table IV.

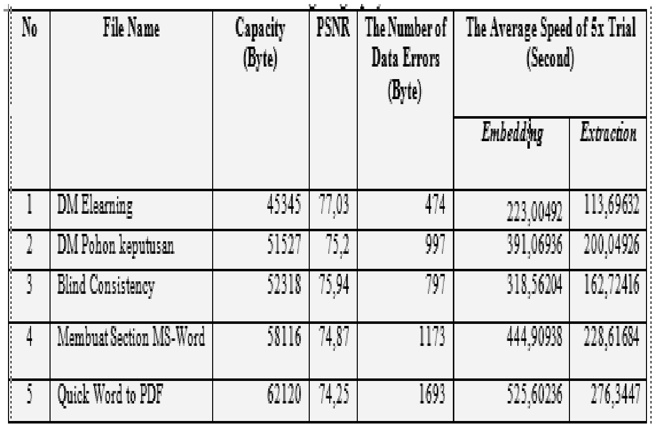
TABLE I. CASE 1



No	File Name	Capacity (Byte)	PSNR	The Number of Data Errors (Byte)	The Average Speed of 5x Trial (Second)	
					Embedding	*Extraction*
1	Belajar dari rajawali	34304	79,28	177	95,73316	48,2608
2	MK keamanan SI	36864	79,16	200	102,42986	51,19254
3	Jadwal Kelas	71168	78,2	327	147,60282	74,73902
4	Mengaktifkan Telkomsel	84480	74,1	1702	540,20704	276,65094
5	Form Isian Diaspora	86016	75,4	951	364,80312	187,41332

FIGURE 3. PERFORMANCE OF STEGANOGRAPHY WITH 2D DWT ON DOC FILE.

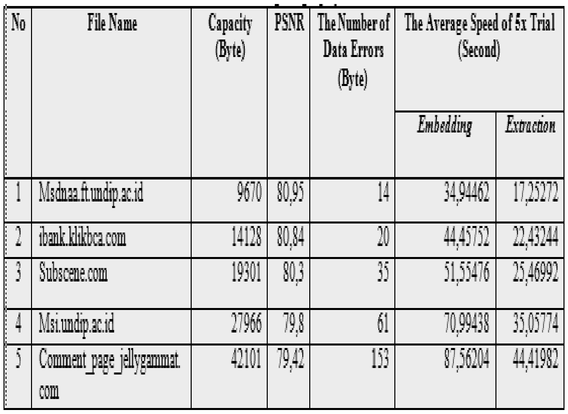
TABLE II. CASE 2



No	File Name	Capacity (Byte)	PSNR	The Number of Data Errors (Byte)	The Average Speed of 5x Trial (Second)	
					Embedding	*Extraction*
1	DM Elearning	45345	77,03	474	223,00492	113,69632
2	DM Pohon keputusan	51527	75,2	997	391,06936	200,04926
3	Blind Consistency	52318	75,94	797	318,56204	162,72416
4	Membuat Section MS-Word	58116	74,87	1173	444,90938	228,61684
5	Quick Word to PDF	62120	74,25	1693	525,60236	276,3447

FIGURE 4. PERFORMANCE OF DWT-2D STEGANOGRAPHY WITH THE PDF FILE.

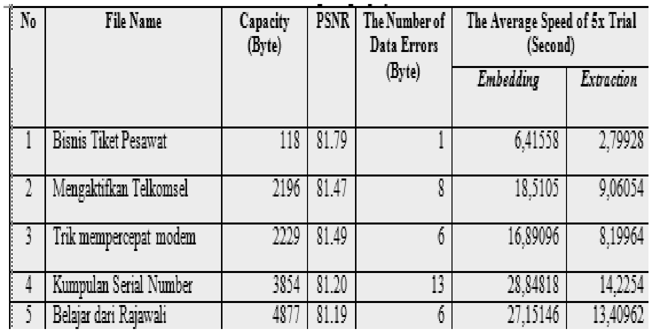
TABLE III. CASE 3



No	File Name	Capacity (Byte)	PSNR	The Number of Data Errors (Byte)	The Average Speed of 5x Trial (Second)	
					Embedding	*Extraction*
1	Msdnaa.ft.undip.ac.id	9670	80,95	14	34,94462	17,25272
2	ibank.klikbca.com	14128	80,84	20	44,45752	22,43244
3	Subscene.com	19301	80,3	35	51,55476	25,46992
4	Msi.undip.ac.id	27966	79,8	61	70,99438	35,05774
5	Comment_page_jellygammat.com	42101	79,42	153	87,56204	44,41982

FIGURE 5. PERFORMANCE OF DWT-2D STEGANOGRAPHY WITH THE HTML FILE.

TABLE IV. CASE 4



No	File Name	Capacity (Byte)	PSNR	The Number of Data Errors (Byte)	The Average Speed of 5x Trial (Second)	
					Embedding	*Extraction*
1	Bisnis Tiket Pesawat	118	81.79	1	6,41558	2,79928
2	Mengaktifkan Telkomsel	2196	81.47	8	18,5105	9,06054
3	Trik mempercepat modem	2229	81.49	6	16,89096	8,19964
4	Kumpulan Serial Number	3854	81.20	13	28,84818	14,2254
5	Belajar dari Rajawali	4877	81.19	6	27,15146	13,40962

FIGURE 6. PERFORMANCE OF DWT-2D STEGANOGRAPHY WITH THE TXT FILE

In the case of 1 on the performance of DWT-2D steganography with the doc files, it seems that if there is an increase in the size of the data capacity, then the stego-image quality or PSNR will decrease because the larger the data inserted will affect the decrease in image quality / PSNR, it is also apply if an increase in the number of data errors. While in certain cases is in table I no.5 which has a size larger data capacity of 86016 bytes, a stego-image quality is better in compare with table I, no.4 which has a data capacity of a smaller size is 84480 bytes, this is because the number of data errors in table I no.4 of 1702 bytes is greater than the number of errors in data embedding tabel1 no.5. namely 951 bytes, in addition to affecting the PSNR this also affects the time of embedding and extraction in table I no.5 is 364,80312 and 187,41332 second faster than their execution of table I no.4 is 540,20704 second and 276,65094 second.

In the case of 2 on the performance of DWT-2D steganography with the pdf files, it seems that if there is an increase in the size of the data capacity, then the stego-image quality or PSNR will decrease because the larger the data inserted will affect the decrease in image quality/PSNR, it is also apply if an increase in the number of data errors. While in certain cases is in table II no. 3 which has a size larger data capacity of 52 318 bytes, a stego-image quality is better in compare with table II No. 2 which has a data capacity of a smaller size is 51,527 bytes, this is because the number of embedding errors of data in table II no .2 that is 997 bytes larger than the number of embedding errors of data in table II no.3. namely 797 bytes, in addition to affecting the PSNR this also affects the time of embedding and extraction in table II no. 3 is 318.56204 and 162.72146 second faster than their execution table II no. 2 is 391.06936 second and 200.04926 second.

In the case of 3 on the performance of DWT-2D steganography with the html files, it appears that if there is an increase in the size of the data capacity, then the stego-image quality or PSNR will decrease because the larger the data inserted will affect the decrease in image quality/PSNR, it is also apply if an increase in the number of data errors. In addition to affecting the PSNR, it also affects the execution time of embedding and extraction due to an increased size of the data capacity and number of errors of data embedding.

In the case of 4 on the performance of DWT-2D steganography with the txt files, it appears that if there is an increase in the size of the data capacity, then the stego-image quality or PSNR will decrease because the larger the data inserted will affect the decrease in image quality / PSNR, it is also apply if an increase in the number of data errors. While in certain cases is in Table IV no. 3 which has a larger data capacity of 2229 bytes, a stego-image quality is better in compare with table IV No. 2 which has a smaller data capacity is 2196 bytes, this is because the number of embedding errors in the data table IV no.2 is 8 bytes larger than the number of embedding errors of data in Table IV no.3. that is 6 bytes, in addition to affecting the PSNR this also affects the time of embedding and extraction in table IV no.3 i.e. 16.89096 8.19964 second and lawyer-second faster than his execution table IV no.2 i.e. 18.5105 second and 9.06054 second. In other cases in table IV no.4 which has a size of 3854 bytes of data capacity and the number of errors affecting the data at 13-execution-time lawyer's embedding and extraction of 28.84818 14.2254 second and second is longer than the table IV no.5 having the time of embedding and extraction 27.15146 13.40962 second and second with the quality of the image/file remains the second PSNR 81.19 dB.

Based on the results of the discussion of cases 1, 2, 3 and 4 show that the stego-image quality PSNR is strongly influenced by the data capacity and the number of data errors at the time of embedding, the resulting PSNR is better if the amount of data capacity and the smaller the amount of data errors. From some test cases demonstrated that the stego-image does not degrade the quality of more than 30dB [8]. The results obtained for PSNR researchers are in the range of 74.10 dB - 81.79 dB, is due prior to the embedding of business documents into a color image, first performed the conversion of color images to grayscale images, the DWT and pasted the text data, further in the results IDWT grayscale image that has been inserted merged back with the color image and its calculated PSNR.

Comparison of embedding and extraction time was 2: 1, can be seen in table IV no.1 95.73316 second time and the embedding of the extraction time is 48.2608 sec. In addition the embedding and extraction time will be faster if the number of data capacity and the smaller the amount of data errors, can be seen in table IV no1 data capacity is 34,304 bytes and the number of errors that is 177 bytes of data produced when embedding is 95.73316 second and extraction time is 48.2608 faster than the data capacity and the amount of other data errors in Table IV.

In terms of security, by exploiting the weaknesses of the human eye, then the digital image is selected as the carrier medium, as long as the inserted data does not damage the quality of the image carrier [1] this can be achieved in this study because the quality of images generated from more than 30dB PSNR [8]. The way has been done in this steganography can store more text by applying the compression of business documents that is zipped before embedding. Data correction procedure is also done at the time of extraction to restore the original value of data (ASCII) which changed during the embedding so that the business documents that is extracted can be returned in their entirety.

V. CONCLUSION

Stego-image quality/Peak Signal Noise to Ratio (PSNR) is influenced by the amount of data capacity and the number of data errors. The smaller the amount of data capacity and the number of data errors then the stego-image quality (PSNR), the better/higher.

PSNR produced by the researcher that is in the range of 74.10 dB - 81.79 dB, because prior to embedding of the text data into a color image, first performed the conversion of color images to grayscale images, grayscale images in-DWT and pasted the business documents, further in-IDWT a grayscale image and the results that have been inserted merged back with the color image and its calculated PSNR.

Time of embedding and extraction time is also influenced by the amount of data capacity and the number of data errors. The smaller the amount of data capacity and the number of data errors then the embedding and extraction times faster.

In terms of security, by exploiting the weaknesses of the human eye, then the digital image is selected as the carrier medium, as long as the inserted data does not damage the quality of the image carrier [1], this can be achieved in this study because the quality of images generated from more than 30dB PSNR [8].

Compressed files do to get the data capacity of a minimum and uniform business documents at the time of embedding and extraction.

Researchers use the data correction procedure at the time of extraction to restore the original value of data (ASCII) which changed during the embedding so that the business documents that is extracted can be returned in their entirety.

REFERENCES

[1] Bender, W; Gruhl, D; Morimoto, N; & Lu, A. 1996. Techniques for data hiding. IBM Sys. J., Vol 35 Nos 3&4, 313–336.

[2] Hui Yu, Yuan; Chen Chang, Chin; & Chen Hu, Yu. 2005. Hiding secret data in images via predictive coding. Science direct. 691-705.

[3] Hui Yu, Yuan; Chen Chang, Chin; & Chang Lin, Iuon. 2007. A NewSteganographic Method For Color And Grayscale Image Hiding. Science direct. 183-194

[4] Schneier, B. 1996. Applied Cryptography. Second ed. Wiley, New York.

[5] Sweene, P. 1991. Error Control Coding (An Introduction). Prentice-Hall, Englewood Cliffs, NJ.

[6] Terzija, Nataša. (2006). Robust Digital Image Watermarking Algorithms for Copyright Protection. Doctoral Disertation. Universität Duisburg-Essen.

[7] Waheed, Qureshi. 2000. Steganography and Steganalysis. As part of GIAC practical repository. SANS Institute.

[8] Yueh Chen Po and chi Tseng yue. 2006. A Study Of Image Steganography Base on DWT Coefficient Analysis. Thesis. Chaoyang University of Technology