Digital Multimedia Archiving based on Optimization Steganography System

Raniyah Abdullah Wazirali
University of Technology, Sydney
Raniyah.A.Wazirali@student.uts.edu.au

Zenon Chaczko
University of Technology, Sydney
Zenon.chazco@uts.edu.au

Anup Kale
University of Technology, Sydney
Anup@uts.edu.au

***Abstract*— As soon as digital artifacts have become a part and parcel of everyday life, the need for digital media archives with the capacity of preserving the given metadata has risen impressively. The process of converting the digital metadata to archives, however, is fraught with a number of difficulties, the key one concerning the methodology for embedding high payload capacity information into the digital multimedia and at the same time retains high quality of the image. The given paper will consider steganography as a possible solution to the aforementioned issue. Allowing for detecting the genetic algorithm for boosting the PSNR value with the information of high capacity will help solve the issue regarding the digital multimedia archiving. Many sizes of data are embeded inside the images and the PSNR (Peak signal-to-noise ratio) is also taken for each of the images verified.**

Keywords—Steganography; LSB; Genetic Algorithm; Block mapping, message segmentation

I. INTRODUCTION

The present-day world of media and communication is dominated by digital multimedia contents. More to the point, the use of digital content has finally become an available option for most people. However, apart from the everyday use of digital media, its significance in research and academic life of a number of students must be mentioned. Media allows for accessing important resources, training essential skills and managing time more efficiently. The information processing that such media performs,, however, still needs major improvement.

Using an embedding process in order to connect two types of data and enhancing the capacity of metadata, steganography can be used to handle the aforementioned issue. By encompassing the spatial domain of steganography in the given research, the author of the given research is going to conduct an analysis of the use of digital multimedia and its effect on various spheres of people's life. For the given purpose, the technique of the least significant bit subscription scheme (LSB) is going to be utilized. The choice of the research technique was predetermined by its high rates of uncomplicatedness [1] and impressively big hiding capacity, as well as the fact that it hinges on such specifics of a human eye as the insensibility of low bit images [2]. For the stereo-image not to be blocked by the SDDS system, it is required that messages should be embedded in a particular spot, therefore, providing the maintenance of the cover image statistical features. Seeing how the spatial-domain embedding approach, which does not allow for choosing the given position due to the regularity of message distribution, embedding the message in a certain location will be fraught with a number of difficulties, such as the definition of changes in the spatial project. In the project in question, the mapping technique is going to be used in order to split the cover image under consideration into several major blocks and, thus, perform an analysis, since the aforementioned approach will help place every single element of the message in its place. The abovementioned operation will contribute to defining the most reasonable frequency domain position and, thus, drive the number of static feature disturbances on the chosen spatial domain slot to minimum, which can be attained by using the Genetic Algorithm (GA). Therefore, the goal of the given paper is to research the GA usage for concealing metadata. Section Two reviews the stenography methods. Section Three shows how to utilize GA for data hiding. Section Four evaluates the matrix method, while Section Five offers an analysis of the proposed algorithm. Section Six provides a list of results and concluding remarks. Block substitution is going to be used in the given paper as opposed to the bit based one because of the inferior properties of the latter [3, 4].

II. REVIEW OF STEGANOGRAPHY METHODS

Steganography is traditionally defined as the skill of concealing specific data in traditional messages for users to make it harder to decipher these implications. Some of the steganography techniques have been used so widely that they have turned into widely accepted and, therefore, frequently used formats [5].

As a rule, steganography methods are split into two major categories, i.e., Transform Domain and Spatial Design approaches [6]. The former approach presupposes that images should be converted before the message is implanted into it [7]. Speaking of the former method, one must mention that it allows for hiding the key message of the image in its essential parts, which is why the Transform Domain requires differentiating between high, middle and low frequency elements of the image. The fact that the signal energy rates are especially high in the lower frequencies, explains why the visibility of the image is of an especially high quality. Judging by the above-mentioned specifics of the Transform Domain, it is reasonable to assume that planting secret data in higher frequencies will suffice to prevent the image distortion. In addition, Transformation Domain methods win by comparison with the rest of the methods due to the fact that the TD does not depend on the format of the image; quite on the contrary, because of the message embedded into the image, TD can

withstand both a lossy and a lossless compression [8].

Contrasted to the TD, the Spatial Domain, these are the pixels that the key messages are embedded into. It offers such methods of planting messages into the image as the LSB plane direct manipulation and the replacement of the cover image with the bits to be concealed. [9, 10]. The given method, however, is rather easy to spot [11]. Instead of it, LSB matching based on planting the data into cover image has been suggested recently. Due to the matching process that it involves, high similarity rates and, therefore, low detectability is exercised [12].

The given study is not the first instance of using GA as the mean to address the LSB issues. GA has also been used in a number of studies in order to conceal the data in the rightmost k LSBs of the host image [13], Mielikainen's method [14] being one of the most recent updates of the technique in question.

III. GA Approach

The genetic algorithm (GA) approach is an exploration and optimization technique built on the knowledge of genetics and natural collection. A GA permits a population collected of many individuals to change under listed variety rules to a state that exploits the "fitness" (i.e., reduces the cost function). The approach was established by John Holland (1975) [15]. The genetic algorithm begins with no information of the exact solution and based totally on replies from its progress operators such as reproduction, crossover and mutation to get the most suitable solution. By beginning at some independent ideas and examining in parallel, the GA approach prevents local minima and meets to achieve optimal solutions. Therefore, GAs have been used to be accomplished of finding high performance ranges in multifarious domains without suffering the challenges related with high dimensionality, as may happen with rise decent techniques or approaches that trust on imitative information [15,16]. Figure 1 shows the basic process of the Genetic Algorithm.

Process of Concealing the Metadata: GA Stages

1. Creating a random population of chromosomes;
2. Assessing the objective (fitness) function;

The evaluation of the given function hinges on the PSRN criterion, which is supposed to reach its minimal value for the process to have any meaningful results.

PSRN, or Peak Signal to Nose Ratio, being the criterion as the foundation for the fitness function, it is traditionally defined with the help of the following functions:

$$PSNR = 10 * log_{10} \frac{Max^2}{MSE} \quad [1]$$

$$MSE = \frac{1}{m * n} \sum_{i-1}^{m} \sum_{j-1}^{n} (A_{ij} - B_{ij})^2 \quad [2]$$

The Peak Signal to Noise Ratio PSNR value is considered to be the fitness function. Any value under 30 dB of PSNR values indicate low quality (i.e., distortion caused by embedding is high). A high and acceptable quality stego image should strive PSNR value of 40 dB, or greater [17]. The higher PSNR value means the minimum changes and the higher quality. The score matrix will evaluate the changes of the cover image for each block.

3. Repeating the steps a–c until the new population is made:
   a. Choosing a pair of chromosomes (probability increasing together with the function of fitness);
   b. Forming two new strings with a crossover of chromosomes;
   c. Mutating the newly obtained chromosomes and plant the new strings into the population.
4. Swapping the new and the previous population;
5. As long as the optimum solution can be provided by correcting the value of error with the amount of generations or the maximum amount of generations is attained before it ceased to grow at the point where it serves as the location of the best chromosome, the experiment can be considered successful.

IV. Hiding Process Using Score Matrix

The combination of the secret image and cover image is assessed with the help of the so-called matrix M [18]. Ti, also known as a double state scoring, works in the following way:

1. T_1 is scored when the cover and stego image pixels are similar or identical;
2. T_2 is scored when the stego image data is dependent on the cover image.

To demonstrate the significance of maintaining the cover pixel values at the same level, fewer alterations must be made to the cover image. Supposing, T_1 isgreater than T_2 ($T_1 > T_2$). Thus, it becomes obvious that the pixels, which do not demand the LSB alterations area better option than the ones that do.

Supposing that S and H represent the matrix of the hidden data and the cover image correspondingly, the stream value of T_1 , which allows for providing the formula for teh score matrix M:

$$M_{L_s * L_h} = \begin{cases} m(i,j) | m(i,j) \in \{T_1, T_2\} \\ 1 \le i \le L_s \quad , 1 \le j \le L_h \end{cases} \quad [3]$$

(m (i, j)) equals i -th row (M, L_s) ;

-th column (M, L_s) equals the length of the hidden stream groups;

L_2 equals the length of the cover stream groups).

By choosing the M and L element from each row and column, one will be able to define the order that ha the best matching potential [8]. The given formula is used for adjustment list calculation:

$J=[j_1, j_2, \ldots, j_k, \ldots, j_{L_h}]$ [4]

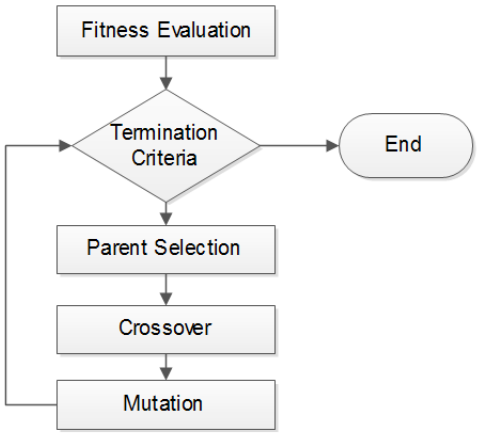

Fig. 1. General Flowchart for Genetic Algorithm

V. PROPOSED METHOD

Supposing, the hidden message in the image is split into N similar parts, the mapping function connecting the hidden and cover image blocks can be technically defined as the N permutation. The best mapping function selection being $P_N^N * T$ in the given case, N increases exponentially; for instance, a 512*512 image will consist of 4096 elements, computational complexity being equal to O(4096!)*T. Therefore, it will take too much time to calculate the N – in fact, the number that will be obtained in the course of the multiplication will verge towards infinity.

The given issue, however, can be resolved efficiently, as the research results show; by using a score matrix as the method for carrying out the mapping function procedure, one is capable of assessing the performance of any matching order of both the secret image and the cover image. In its turn, the normalized total score of J is calculated with the help of the following formulas [5,6]:

$$f(J) = \frac{1}{f^m} \sum m(i,j) \quad [5]$$

$$f^m = L_s . T_1 \quad [6]$$

f^m being the maximum possible value used in the adjustment list.

A. Embedding Procedure

The cover image in the proposed method is the multiple images. The system divides the cover images to equal blocks and based on the size of the blocks, the secret message will divide too. The fitness function in this case is the best value of PSNR. It is done in order to obtain an optimal mapping function to reduce the difference error between the cover and the stego-image, therefore improving the hiding capacity with low distortions.

According to the system, in accordance with which the secret image, or message, is split into N blocks, the following step must be taken in order to create the pseudo code for a stego image:

[1] Splitting the cover image into several blocks (e.g., 8*8 blocks) and give each block label;
[2] Assessing the score matrix represented in each block;
[3] Calculating the adjustment list value by locating and stating the fitness function;
[4] Locating the most efficient and reasonable adjustment list with the help of the GA strategy;
[5] Implanting the cover data into the image in question by following the adjustment list provided.

B. Extracting Procedure

The inverse steps will be taken to extract the secret message by the receiver. The receiver will receive the stego-images and the key. The key contains the positions in order of the secret message which indicate the image name and the block number. The key can be sent in an encrypted manner to increase the security.

The pseudo code for constructing a stego image is explained in the following steps:

[1] divide the cover image into blocks. For example, 8*8 blocks.
[2] Evaluate the score matrix in each block.
[3] Define the fitness function in order to calculate the adjustment list value.
[4] Find the best adjustment list by using the genetic algorithm strategy.
[5] Extract the secret data into the cover image in accordance with the adjustment list.

VI. EXPERIMENT RESULTS

According to the results acquired in the course of the research, the introduction of the Graphical User and Matlab R2013a improves the efficacy of the code a few notches. As the table provided below shows, the average value of the PSNR reaches 72.15533 dp and the related capacity of the 152*512 Lenna.jpg and papper.jpg images increases, as Figure 2 shows. Thus, the proposed method turns out to be more efficient than the traditional LSB method for increasing the capacity of stego images.

Table 1 below shows the PSNR value of the proposed system with its related capacity (text file size in kilo byte) in 512*512 images for peper.jpg. The improvement of the proposed method is noticeable. It is obvious from table 1 that the system can embed 54KB which is almost 16 pages in Microsoft Word in one picture with a very high quality. Therefore, Figure 2 shows comparison of proposed method with normal LSB method in term of the embedded capacity in in Kilo Byte (KB) and PSNR Value.

Bitmap (BMP) format is the format of the stego image file utlized in proposed algorithm. The BMP file format delivers visuals files inside the Microsoft Windows OS. Characteristically, the BMP files are large because they are uncompressed. The benefit of utilising BMP files is the effortlessness and extensive approval of BMP files in

Windows programs. Consequently, this kind of image is selected to be used in our proposed method.

The system is verified using the different images as showed in Fig. 3 and 4. Fig. 3-4 (A) shows the original image before the message is stored inside the image and Fig. 3-4 (B-F) shows the stego images after the message is stored inside the image with various sizes of secret data. We found that the stego image does not have a noticeable distortion on it (as seen by the naked eyes).

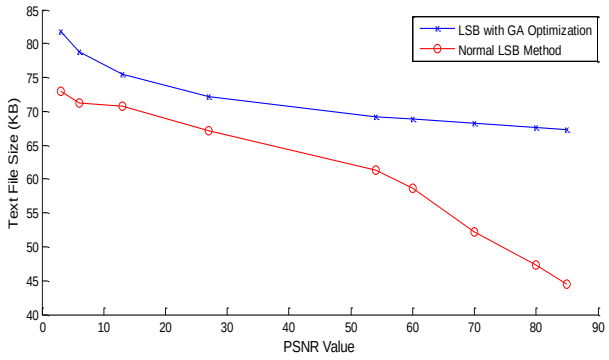

Fig. 2. Comparison of proposed method with normal LSB method

Fig 3. The obtained result form (A) the original image, (B) the stego image with hiding 3KB, (C) the stego image with hiding 6KB, (D) the stego image with hiding 13KB, (E) the stego image with hiding 27KB, (F) the stego image with hiding 54KB,

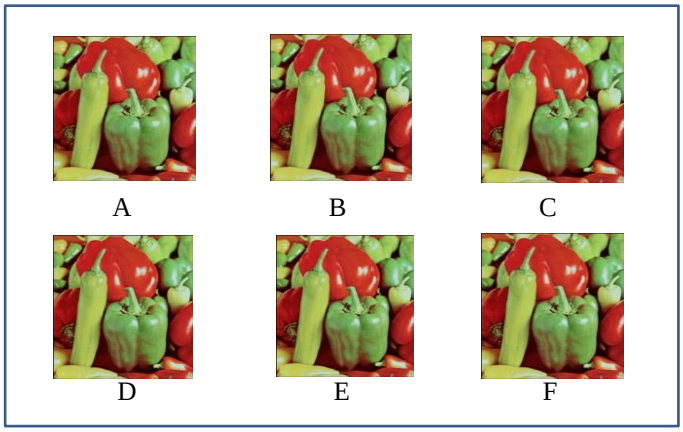

Fig 4. The obtained result form (A) the original image, (B) the stego image with hiding 3KB, (C) the stego image with hiding 6KB, (D) the stego image with hiding 13KB, (E) the stego image with hiding 27KB, (F) the stego image with hiding 54KB,

TABLE 1. THE EMBEDDED CAPACITY IN IN KILO BYTE (KB) FOR THE PROPOSED METHOD AND ITS RELATED PSNR IN PEPPER.JPG

Text File Size	PSNR Value For the proposed	PSNR for Normal LSB
3 KB	81.8165	72.9199
6 KB	78.8355	71.1563
13 KB	75.4804	70.6875
27 KB	72.1530	67.1538
54 KB	69.1431	61.2597
60 KB	68.8443	58.5787
70 KB	68.1788	52.1589
80 KB	67.5968	47.2575
85 KB	67.3496	44.4554

CONCLUSION

The task of preserving digital media is extremely important, since it allows for its more efficient use. However, retaining the capacity of images, as well as their imperceptibility, is rather difficult. With the help of steganography, however, the process of conserving digital media can be upgraded a few notches and made times easier. Moreover, the quality of the images remains just as high.

Experiments have been done on the idea of choosing the optimal locations of the metadata in order to obtain rightmost position. A new system based on the genetic algorithm has been developed. A few images have been tested using various size of text to be concealed. The stego images do not have any noticeable distortion on it that can be realized by the naked eyes. The PSNR value is consider as a fitness function for the proposed method. The proposed method has a high value of the PSNR value which indicates high quality of the stego images. Hence, the proposed method is very efficient to hide even large size of metadata and therefore, provide an excellent environment for creating digital media archiving to merge the digital media with the related description

REFERENCES

[1] A. D. Ker, "Steganalysis of LSB matching in grayscale images," *Signal Processing Letters, IEEE,* vol. 12, pp. 441-444, 2005

[2] L. Shao-Hui, C. Tian-Hang, Y. Hong-Xun, and G. Wen, "A variable depth LSB data hiding technique in images," in *Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on*, 2004, pp. 3990-3994 vol.7

[3] W. Yi-Ta and F. Y. Shih, "Genetic algorithm based methodology for breaking the steganalytic systems," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 36, pp. 24-31, 2006

[4] E. Kawaguchi and R. O. Eason, "Principles and applications of BPCS steganography," 1999, pp. 464-473

[5] Fabien A. P., Ross J. Anderson and Markus G., "Information Hiding - A Survey", *Proceedings of the*

IEEE, special issue on Protection of Multimedia Content, pp. 1062-1078, 1999.

[6] Silman J., "Steganography and Steganalysis: An Overview", *SANS Institute*, 2001.

[7] Lee Y. K. and Chen L. H., "High capacity image steganographic model", *IEEE Proceedings of Visual Image Signal Processing*, Vol. 147, No. 3, pp. 288-294, 2000

[8] Shi Y. Q. and Sun H., *Image and Video Compression for Multimedia Engineering*, CRC Press, Boca Raton London New York Washing, D.C., 2001.

[9] Ker A., "Improved detection of LSB steganography in grayscale image", *Lecture Notes in Computer Science*, pp. 97-115, 2005.

[10] Mahdavi M., Samavi Sh., Zaker N. and Modarres-Hashemi M., "Steganalysis Method for LSB Replacement Based on Local Gradient of Image Histogram", *Iranian Journal of Electrical & Electronic Engineering*, Vol. 4, No. 3, pp. 59- 70, 2008.

[11] Chan C. K. and Chan L. M., "Hiding data in image by simple LSB substitution", *Pattern Recognition*, Vol. 37, pp. 469-467, 2004.

[12] Chang C. C., Hsiao J. Y. and Chan C. S., "Finding optimal least-signification-bit substitution in image hiding by dynamic programming strategy", *Pattern Recognition*, Vol. 36, pp. 1583-1595, 2003.

[13] Wang, R.Z., Lin, C.F., Lin, J.C.: `Image hiding by optimal LSB substitution and genetic algorithm', Pattern Recognit., 2001, 34, (3), p. 671-683

[14] Mielikainen J., "LSB matching revisited", *IEEE Sigal Processing Letters*, Vol. 13, No. 5, pp. 285- 287, 2006.

[15] H Holland. "Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology". *Control and Artificial Intelligence. second edition. Cambridge*, MA: MIT Press, 1992.

[16] Yi-Ta Wu, Frank Y Shih. Genetic Algorithm Based Methodology for Breaking the Steganalytic Systems. IEEE Transactions on systems, man, and cybernetics Part B: cybernetics, February 2006, vol 36, no 1.

[17] J. Fridrich, M. Goljan, and D. Rui, "Detecting LSB steganography in color, and gray-scale images," *MultiMedia, IEEE,* vol. 8, pp. 22-28, 2001

[18] Soleimanpour M., Talebi S. and Azadi-Motlagh H., "A Novel Technique for Steganography Method Based on Improved Genetic Algorithm Optimization in Spatial Domain", *Iranian Journal of Electrical & Electronic Engineering*, Vol. 9, No. 2, pp. 67-75, June 2013.