Data Hiding Scheme for Digital Images based on Genetic Algorithms with LSBMR

P.M.Siva Raja,
Research Scholar, Sathyabama University,
Chennai, Tamil Nadu, India.

E.Baburaj,
Professor, Department of CSE
Sun College of Engineering and Technology,
Nagercoil,

ABSTRACT

Modern information hiding technology is an important branch of information security. Hiding capacity is very much important for efficient covert communication. The redundancies of digital media as well as the characteristic of human visual system make hiding technology a significant one.Steganography is the Art and Science of writing hidden messages in such a way that no one, apart from the sender and intended recipient suspects the existence of the message. Images are the mostly cover objects used for information hiding schemes. Image steganography is the most popular method for message concealment. Many different carrier file formats can be used, but digital images are the common, because of their frequency in the Internet. In LSBMR, two secret bits can be embedded into each embedding unit and the threshold value for region selection can be determined. The main drawback of this scheme is the absolute difference is taken as the threshold value. In this paper LSB Matching Revisited (LSBMR) image steganography using Genetic Algorithm (GA) is proposed, in which Genetic algorithm is used to select the embedding regions according to the size of the secret message and to optimize the threshold value of the selected image regions. Experimental analysis shows that the proposed algorithm outperforms the existing methods in terms of capacity and security.

Index terms: Steganography, Message concealment, Information hiding, Region selection, Genetic algorithms.

1. Introduction

Information hiding plays a vital role in Multimedia Information Security. In order to protect intellectual properties, The techniques like Image Stegnography is needed. Stegnography is used for information hiding. The aim behind Stegnography is to incorporate secret data into digital cover media (images, Audio, Video) without being suspicious. Exchange of information between two parties over an insecure channel may cause the intruder to intercept, read and compute information [1].Typical Characteristics including invisibility, Storage Capacity and Resilience against attack makes Stegnography an efficient one in hiding technology. Then again the users of digital media are under risk owing to the mounting concern of copyright infringement, illegal distribution, unauthorized tampering and security in Communication [2]. Data hiding that deals with an incorporation of an auxiliary data in the digital media becomes a thinkable solution to the problems over the last era. Quite a lot of data hiding methods have been developed for audio, image, video and graphics etc., and are also versioned in literature issue on copyright and privacy protection 1998;Special issue on enabling security technologies for digital right management 2004(Pan et al. 2004)[3]. Two techniques used under data hiding are stegnography and digital water marking. The anterior setup a covered information channel in point-to-point connection and the subsequent one does not hide the actuality of secret transmission to third party[4]. The degree of requirement may vary for different application but the requirements like visual imperceptibility of the hidden data, security against statistical analysis and robustness to non-malicious operations that a communication channel is to face won't get diverged[5]. The Processing stage includes two steps, compression for efficient storage and transmission and mean/median filtering for noise removal. On the other hand, the depth of the signal processing operation should be restricted to a threshold level, so that the watermarked object can preserve its commercial value[6].

In case of data hiding, Genetic Algorithm (GA) is used for enhancing the fundamentally conflicting requirements of security and robustness. So far , data hiding algorithms focused only on a single requirement or Provides suboptimal solutions to an assembly of requirements based on the applications[7]. The digital information revolution and the thriving progress in network communication provides the advantage of noise free communication, the ease of editing and the internet distribution of digital multimedia data[8]. In this paper LSB Matching Revisited (LSBMR) image stegnography using Genetic Algorithm (GA) is proposed, in which Genetic Algorithm (GA) is used to select the embedding regions according to the size of the secret message and optimize the threshold value of the selected image regions [9]. The rest of the paper is arranged as follows. Section 2 analyzes the limitations of the relevant stegnographic schemes with the other techniques. Section 3 shows the details of data embedding and data extraction in proposed scheme. Section 4 presents Experimental results and discussions. Finally, conclusion and the future enhancement are given in section 5.

2. Related Work

This Section discusses data hiding scheme in order to highlight the effectiveness of this class of tools for performance improvement in data hiding. The objective of this review is to discuss merits and limitations of the relevant work. In the passive warden model, LSB techniques impose a difficulty in differentiating cover images from stego images; given the small changes have been made. In case of active warden, such kind of techniques can be defeated by randomizing the LSB. Embedding process is similar in both

LSB and LSBM approaches [10]. If a secret bit stream is given for incorporation, PRNG generates a traveling order in the cover image and then each pixel is dealt separately. In case of LSB replacement, the LSB of the pixel is overwritten by the secret bit. If the secret bit and the LSB of the given pixel is not equal, then plus or minus one is added randomly [11][22]. At the same time the altered pixel is kept in the range of [0,255].In such a way, the LSB pixels along the traveling order will match the secret bit pattern occurs after the data hiding both the LSB replacement and LSBM. It shows that the extracting process is very similar for both LSB replacement and LSBM approaches [12]. Here, based on the shared key, same traveling order is generated and then by checking the parity bit pixel values, the hidden message can be extracted correctly [13].

In case of LSBM approach, two secret bits can be incorporated into each embedding unit and the threshold value for region selection can be found. After finishing the data hiding process, the resulting image is splitted into non-overlapping blocks[14]. Then the splitted blocks are rotated by a random number of degrees. In data extraction scheme, the side image from the stego image is extracted, then parameter identification process based on the side information and identification of regions for data hiding is performed[15][19]. Finally, the secret message M is obtained. The main disadvantage of this approach is the absolute value is taken as the threshold value T and the threshold for region selection also varies [21]. If Embedding rate increases more regions can be released by decreasing the threshold value.

3. Secret Data Hiding - Issues

In the data embedding stage, the scheme first initializes some parameters, which are used for subsequent data preprocessing and region selection and optimize the threshold value of those selected regions. If the regions are enough for hiding the given secret message M, then data hiding is performed on the selected regions and then repeats the region selection process using Genetic Algorithms and optimize the threshold value of selected regions until the secret message M can be embedded completely[16][18]. In data extraction, the scheme first extracts the side information from the stego image.

Based on the side information, Algorithms does some post processing and identifies the regions that have been used for data hiding [17][20]. Finally, it obtains the secret message M according to the specified extraction algorithm. In this paper, we propose an embedding scheme based on LSBMR with Genetic Algorithms. Genetic Algorithms are used for optimizing the threshold value and the selection of embedding regions.

3.1 Data Embedding

This section describes the process of data hiding method in digital images. The choice of cover images is important and influences the security in a major way. We consider gray scale image as cover and the similar type image like text information as message signal since it can preserve contextual information even after various signal processing operations. Algorithms for embedding process are as follows

Step 1: The cover image of size m x n is first divided into non overlapping blocks of Bz × Bz. For each small block rotate it by a random degree in the range of {0, 90,180,270} as determined by a secret key k1.Two benefits can be obtained by the random rotation, First, it can prevent the detector from getting the correct embedding regions without the secret key k1.and thus the security is increased much more.

Step 2: The resulting image is rearranged as a row vector V by raster scanning and then the vector is divided non overlapping embedding units with every two consecutive pixels (x_i, x_{i+1}).

Step 3: According to the scheme of LSMBR, two secret bits can be embedded into each embedding unit. Therefore, for a given secret message M, the threshold T for region selection can be determined by using genetic algorithm.

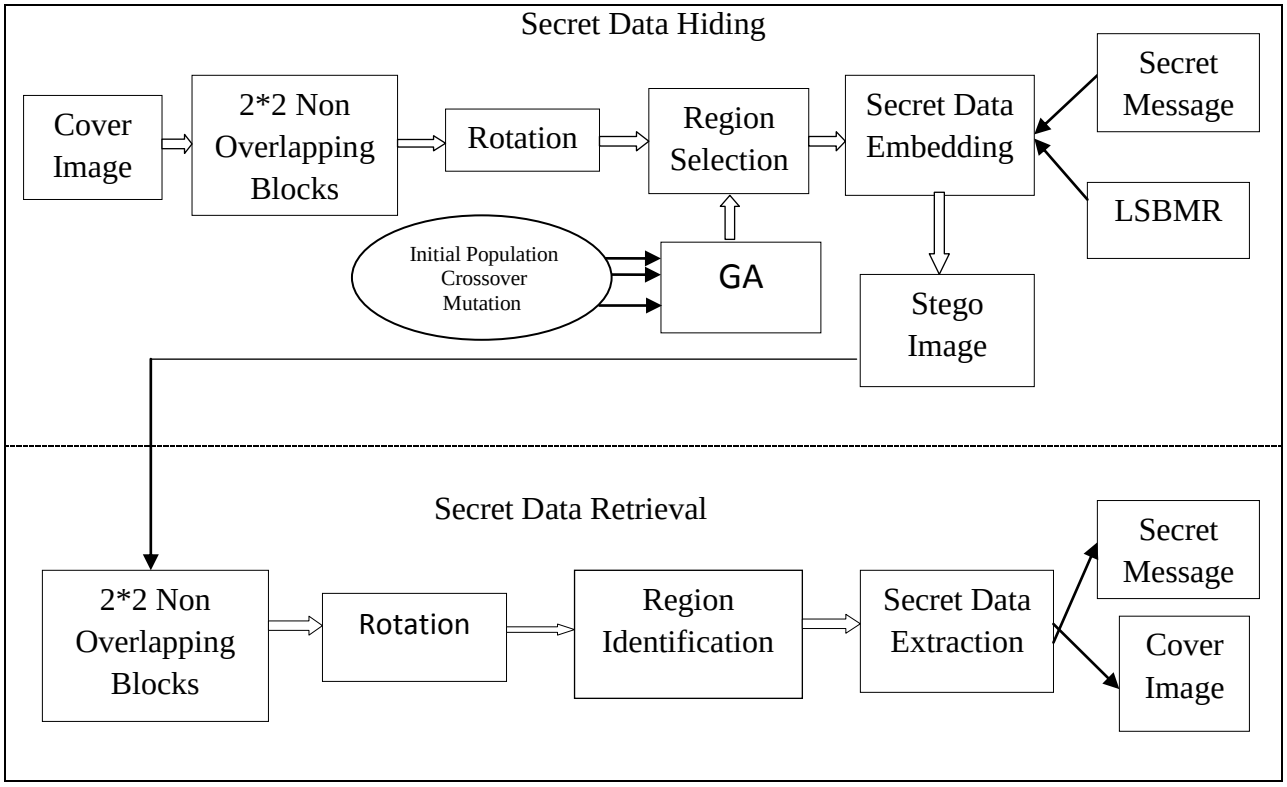

Figure 1. Data Hiding and Data Extraction Scheme using LSBMR-GA

4. GA operations

In our proposed GA, a chromosome is encoded as an array of 64 genes containing quantized DCT coefficients of each 8*8 pixel block of the image. Here we propose a new genetic algorithm approach to find the best position for data embedding and also optimize the quality of the steganographic image. In our optimization we need to handle four conflicting goals-longer hidden message, higher image quality, better robustness and larger data capacity. The first step to model this problem as a GA problem, is determining the chromosome, GA operators and fitness function. The procedure for the selection of embedding regions and optimization of threshold value of selected regions are as follows

4.1 Initialization of Population

Chromosomal representation of the parameter values. The initial population is formed by taking Bz×Bz pixel blocks and Binary Encoding is used for forming the initial population. Binary encoding is the most common and simplest one. In binary encoding every chromosome is a string of bits, 0 or 1.

4.2 Selection

Best fitted pair of individuals is chosen by roulette-wheel selection process by adding up fitness values of the individuals to get fitness. Then randomly select individuals to cross 50% of the fitness value in the cumulative way. The particular individual which crosses 50% criteria in the cumulative process is chosen to be one of the matching pool pair. This process is again carried on to find another individual pool matching pair.

4.3 Crossover

Find the crossover site and to perform crossover between the selected pixel pair to get the new pair of more fitted individuals using arithmetic crossover. The selected matching pool is taken as input and finds the crossover site using arithmetic crossover. Exchange the portions lying on one side of crossover site of those matching pool pair, a new pair of individuals is an outcome.

4.4 Mutation

To mutate or change a particular bit in a pixel block with very small probability. Uniform mutation is used for the mutation process. A very small mutation probability is chosen, depending upon the probability value; change a bit from 1 to 0 or 0 to 1.

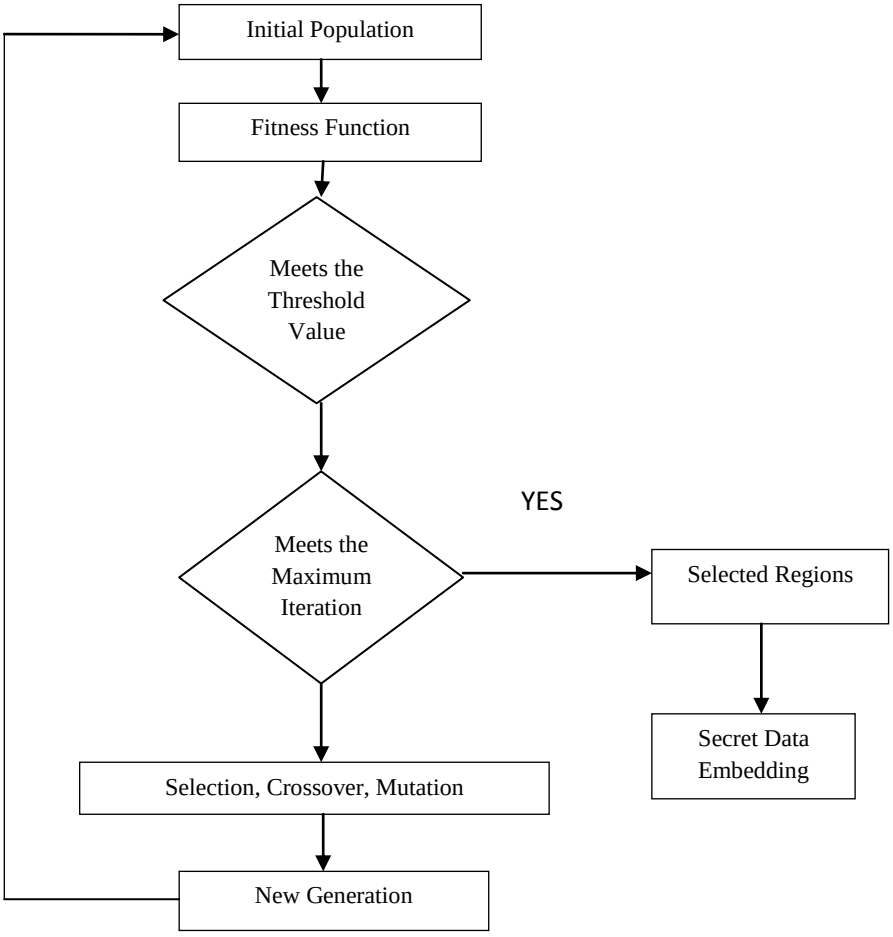

Fig.2 Flowchart for the main module of the problem statement.

4.5 Objective Function

To estimate the fitness value of an individual, the initial population is taken as input, on each population 2D interpolation technique is applied to approximate the original matrix. The mean square error is evaluated by subtracting the interpolated matrix from the original matrix. The square of that MSE is considered to be the fitness value of that particular individual. The fitness function can be evaluated using the equation 1 as

$$f(x)=\left(\frac{1}{m\,x\,n}\right)\sum_{i}^{M}\sum_{j}^{N}(I_{i\,j}\;-\;I'_{i\,j})^{2} \quad (1)$$

Where mxn is the height and width of the cover image, I_{ij} is the pixel value of coordinate (x,y) in cover image, $I^{|}_{ij}$ is the corresponding pixel value in the rotated image.

Step 4: According to LSMBR, two secret bits can be embedded in to each embedding unit. Therefore for a given secret message M, the threshold T for region selection can be determined by using the optimization technique in eqn (1). The embedding regions in a pseudorandom order determined by a secret key k_2. The secret key is the difference between the two pixel values. That value is greater than or equal to the threshold value.

Performing data hiding on the selected embedding regions are as follows. In the LSB matching revisited, the choice of whether to add or subtract one from the cover image pixel is random. This method uses the choice to set a binary function of two cover image pixels to the desired value in eqn (2). The embedding is performed using a pair of pixels as a unit, where the LSB of the first pixel carries one bit of information, and a function of the two pixel values carries another bit of information.

$$f(l,n)=LSB\left(\left[\frac{l}{2}\right]+n\right) \quad (2)$$

Data hiding is performed according to the following two properties:

Property 1: $f(l-1,n)\neq f(l+1,n)$

Property 2: $f(l,n)\neq f(l,n+1)$

Where m_i and m_{i+1} denote two secret bits to be embedded. The function f (l, n) is a random value in {-1, +1}

and (x, xi_{+1}) in the cover image pixel and $(x^{|}, x^{|}_{i+1}) = m_{i+1,}(x^{|}_{I}, x^{|}_{i+1}) = (x, xi_{+1})$

Step 5: After data hiding, the resulting image is divided into non overlapping Bz×Bz blocks. The blocks are then rotated by a random number of degrees based on key $k_{1.}$

5. Data Extraction

The final stage of the algorithm is the retrieval process of the secret message M. To extract the secret message, first extract the side information as the block size Bz and the threshold T from the stego image. For that the stego image is divided into Bz*Bz non overlapping blocks and the blocks are then rotated by random degree based on the secret key k_1.The resulting image is rearranged as a row vector v_i. Finally to get the embedding units by dividing v_i into non overlapping blocks with two consecutive pixels. The travel through the embedding regions whose mean values are greater than or equal to the threshold T according to a pseudorandom order based on the secret key k_2 until all the hidden bits (secret message) are extracted completely based on eqn(3). To extract the two secret bits m_i, m_{i+1} as follows.

$$M_i = LSB(x_i^{||}),\ mi_{+1} = LSB([l^{|}/2] + n_{i+1}) \qquad (3)$$

6. Results and Discussion

This section presents simulation results to demonstrate the effectiveness of the proposed data hiding method compared with existing relevant methods as mentioned in section II.The experiment is carried out using the eight cover image of size (512 x 512), 8 bits/pixel to illustrate the effectiveness of proposed approach and two of them are shown in figure 2.The Peak Signal to Noise Ratio (PSNR) is used to evaluate qualities of the stego images. Experimental results of two stego images are shown in figure. In addition, other noticeable image quality measures, such as Mean Square Error (MSE) and Receiver Optimization Curve (ROC) are also applied to our method indicating the significance to the contribution of this paper. Besides, from MSE values, it shows that the changed value of each pixel is almost same in stego images. To demonstrate the accomplished performance of our proposed approach in capacity and security for hiding secret data in the stegoimage.The analysis of selection technique is based on a comparison as the number of function evaluations. Population size, Crossover, Mutation, criteria for termination, Fitness function and number of generations are the parameters used for analysis.The results of PSNRs and the MSE values obtained from GALSBMR technique is depicted in Table 1.

Lena Baboon Foveon Deer Lichtenstein

Figure 3. (a) Cover Images

Lena Baboon Foveon Deer Lichtenstein

Figure 3. (b) Stego Images

6.1 Histogram Analysis

An image histogram is a type of histogram that acts as a graphical representation of the tonal distribution in a digital image. It plots the number of pixels for each tonal value. By looking at the histogram for a specific image a viewer will be able to judge the entire tonal distribution at a glance. In the field of computer vision, image histograms can be useful tools for thresholding. Because the information contained in the graph is a representation of pixel distribution as a function of tonal variation. Fig.3 shows the histograms of cover and stego images. It is clear that the histograms are almost identical. This is caused by the embedding the message bits in noisy regions of the image.

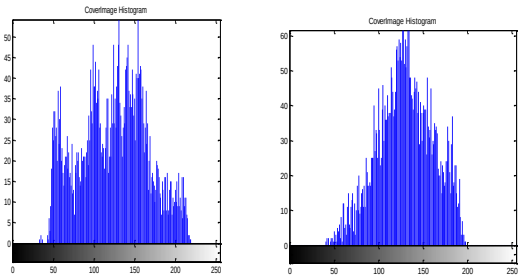

Fig.4 (a) Histogram of Cover Images

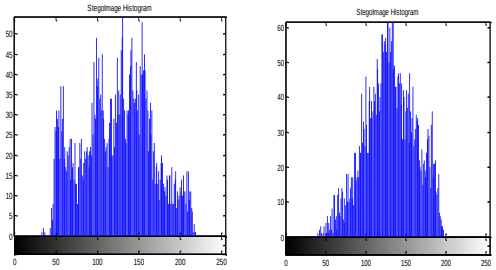

Fig.4 (b) Histogram of Stego Images

6.2 .Optimization of Threshold Value

To evaluate the performance of the proposed selection methods, the experiment analysis of selection techniques based on a comparison of their respective performance estimated as the number of function evaluations. The raw data obtained from the optimization comprised the average, mean and best of 100 generations for each function of the set of functions under consecration. The parameters are used in experiments, The Population size is 100, Selection Methods is an Roulette Wheel selection, Simple Arithmetic Crossover with the probability of 0.08 per generations is used for crossover and the Mutation is the Multi Non Uniform with Probability 5%, criteria for termination is defined as the executions stops on reaching the number of generations, then the Fitness function is an Objective Value. The average, mean, best of the evolutions for each function is evaluated for each number of generations.

6.3 ROC Curve Analysis

For comparing the embedding security of proposed method to that of other methods, Receive Operating Characteristic (ROC) curves have been used, to embed image databases with the minimum value of the maximum capacity of the LSBMR-GA method and the LSBMR method.

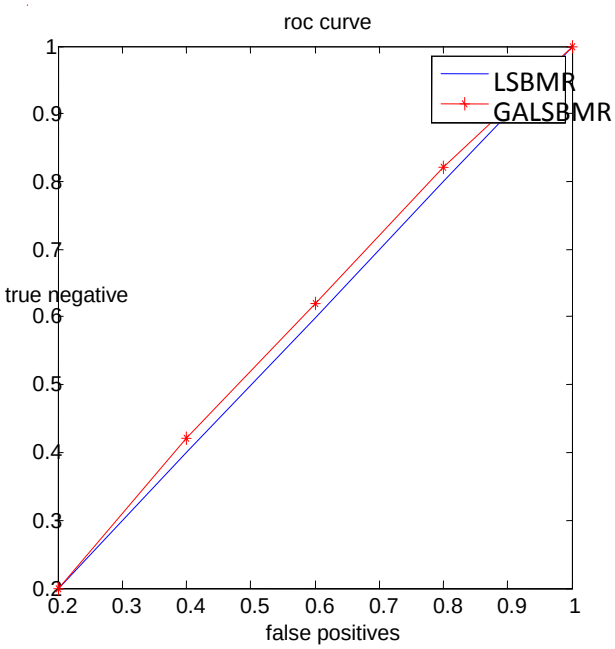

Figure 5.ROC Curve (Embedding)

The Receiver Operating Characteristics (ROC) curves are shown in Fig.5.It can be clearly observed that both specific steganalytic algorithms would fail in detecting the proposed method even when the embedding rate is as high as 75%, while they obtain satisfactory results for detecting stego image using LSBMR and LSBMR-GA methods. As observed from figure the detection accuracy, shown as the area under the ROC curve, is lower for our proposed algorithm as compared to the LSBMR algorithms.

6.4 Comparing Capacity

Figure 6 shows the results of computing capacity for our GA-LSBMR algorithm and the LSBMR method in 800 images. It is clear that the capacity of the proposed method is higher in most images. The mean capacity of the GA-LSBMR method is about 1024 bits higher than the mean capacity given by the LSBMR method.

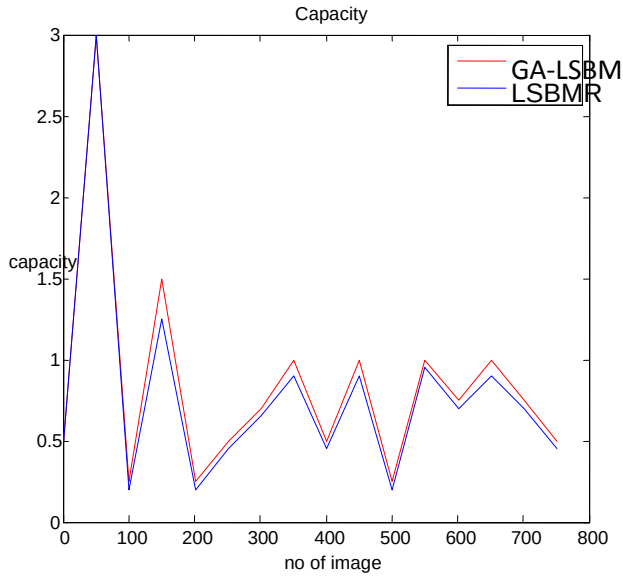

Fig.6 Embedding Capacities of the proposed method (Red) and the LSBMR method (Blue).

The performances of the methods have been evaluated and compared on the basics of two measures which are Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR) .The equations 4 and 5 of these two measures are

$$MSE = \left(\frac{1}{m\,x\,n}\right)\sum_{i=1}^{m-1}\sum_{j=1}^{n-1}(I_{ij} - I'_{ij})^2 \quad (4)$$

$$PSNR = 20\log_{10}\left(\sum_{i=1}^{M}\sum_{j=1}^{N}(S_{ij})^2 \Big/ MSE\right) \quad (5)$$

TABLE I

Results of PSNRS and MSE Values for the proposed algorithm

Image	PSNR	MSE
Lena	48.55	0.00082
Baboon	48.63	0.00079
Foveon	47.89	0.00081
Deer	47.91	0.00089
Lichtenstein	48.92	0.00082

7. Conclusion

Image Steganography systems scale up and become increasingly complex, their optimization techniques face new challenges. Conventional image steganography methods become fragile and show poor performance. Lot of optimization techniques have long been proposed to improve the hiding capacity of the stego images. In this paper, we design on this idea of the optimizing the region using Genetic Algorithm which is a conventional bio-inspired optimization technique used in engineering problems, by which regions in which the message is to be embedded on the cover image. This paper designs GA-LSBMR systems, with detailed design that meets all requirements. The results of our GA-LSBMR evaluation suggest that when the exact regions with correct threshold value are selected, the method offer high embedding capacity. Decoding reliability is improved with the increase in number of iterations, when set of parameter values are fixed. The algorithm is proven to be secured against stego test based on high order statistics. Other steganography methods such as audio/video steganography in the spatial or frequency domains when the embedding rate is less than the maximal amount may be considered as a future work.

8. References

[1] C.H.Yang, C.Y Weng, S.J.Wang et al., ”Adaptive Data Hiding in Edge Areas of Images With Spatial LSB DomainSystems,”IEEE Transactions on Information Forensics and Security, Vol.3, No.3, pp488-497,2008.

[2] Weiqi Luo, Fangjun Huang, Jiwu Huang,”Edge Adaptive Image Steganography Based on LSB Matching Revisited” IEEE Transactions on Information Forensics and Security, Vol.5 No.2, pp 201-214,June 2010.

[3] Mahdi Ramezani, Shahrokh,” Adaptive Image Steganoraphy with Mod-4 Embedding using Image Contrast”, IEEE CCNC 2010 Proceedings.

[4] M.Ramezani and S.Ghaemmaghami, “Towards Genetic Feature Selection in Image Steganalysis,” in 6th IEEE International Workshop on Digital Rights Management, Las Vegas, USA, 2010, PPT.

[5] Mielkainen, ”LSB matching revisited,”IEEE Signal process Lett., vol. 13no. p. 285 – 287, May 2006

[6] S. Dumitrscu, X. Wu . and Z. Wang, Detection of LSB steganography via Sample pair analysis,”IEEE Trans. Signal process., Vol. 51. no. 7. pp.1995 – 2007, Jul. 2003.

[7] Weiqi Luo,Fanggjun Hunang and Jiwu Huang,”Edge Adaptive Image Steganoraphy Based on LSB Matching Revisited”IEEE Transactions on Information Forensics and Security,vol.5, No 2,June 2010

[8] Ran-Zan Wang,Chi-Fang Lin,Ja-Chen Lin “Image Hiding by optimal LSB substitution and genetic algoritm”,Pattern Recognition Society, Published by Elsevier Science Ltd. pp.671-683,2001.

[9] R.Z. Wang and C.F. Lin: ImageHiding by Optimal LSB Substituition and Genetic Algorithm, Pattern Recognition, ELSEVIER, Vol. 34, (2001)671-883.

[10] T. Zhang and X. Ping: A New Approach to Reliable Detection of LSB Steganography in Natural Image,

Signal Processing Journal, ELSEVIER, Vol.83 May(2003) 2085- 2093.

[11] Information Hiding Techniques for steganography and Digital Watermarking. S. Katzenbeizzer and F.A.Petticolas, eds. Artech House, 2000.

[12] R.J. Anderson, "Stretching the Limits of Steganography," proc. First International workshop Information Hiding (IH '96), pp. 39 – 48, 1996.

[13] X. Li. T.Zeng. and B.ang,"Detecting LSB matching by applying calibration technique for difference image", in proc. 10th ACM Workshop on multimedia and security, Oxford, UK., 2008. pp. 133-138.

[14] C.Cachin,"An Information-Theoratic Model for Steganography", proc. Second InInt'l Workshop Information Hiding (IH '98), pp. 306 -318, 1998.

[15] J.Zollner,H.Federrath, H.Klimant, A. Pfitzmann, R. Piotraschke, A. Westfeld, G.Wicke, and G.Wolf, "Modelling the security of steganographic Systems," proc. Second Int'l Workshop Information Hiding (IH '98), pp. 344 – 354, 1998.

[16] C.T.Hsu,J.L.Wu,Hidden digital watermarking in images,IEEE Transactions on Image Processing,vol 8,pp 58-68,1999.

[17] C.K.Chan and L.M. Chen " Hiding data in images by simple LSB substituition,"Pattern Recognit., vol. 37, no. 3 pp 469-474, 2004.

[18] C.C.Chang and H.W.Tseng ,"A steganographic method for digital images using side match,"Pattern Recognit. Lett. vol 25, no, 12 pp. 1431-1437, 2004.

[19] H.C.Wu aN.I. Wu C. S Tsai and M.S. Hwang,"Image Steganographic scheme based on pixel- value differencing and LSB replacement methods", proc. Inst. Elect. Eng. vis Image Digital Process. vol. 152. no. 5, pp. 611-615, 2005.

[20] Y.R.Park , H. H. Kang , S. U. Shin and K. R. Kwon, A steganographic scheme in digital Images Using Informatio of Neighbouring Pixels. Berlin, Germany: Springer-Verlag, 2005, vol. 3612, pp/ 962-968.

[21] K. Hempstalk ,"Hiding Behind corners: Using edges in images for better steganography", in proc. Computing Women's Congress. Hamilton, New Zealand. 2006. .

[22] Y.Q. Shi et al. " Image Steganalysis based on moments of characteristic functions using Wavelet decompostitions prediction-error image. and neural metwork." in proc. IEEE Int. Conf. Multimedia and Expo. Jul. 6-8, 2005. pp. 269-272.