Chapter 30
High Capacity Image Steganography Based on Genetic Algorithm and Wavelet Transform

Elham Ghasemi, Jamshid Shanbehzadeh, and Nima Fassihi

Abstract This paper presents the application of wavelet transform and genetic algorithm (GA) in a novel steganography scheme. We employ a GA based mapping function to embed data in discrete wavelet transform coefficients in 4×4 blocks on the cover image. The optimal pixel adjustment process (OPAP) is applied after embedding the message. We utilize the frequency domain to improve the robustness of steganography and, we implement GA and OPAP 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. Our simulation results reveal that the novel scheme outperforms adaptive steganography technique based on wavelet transform in terms of peak signal to noise ratio and capacity, 39.94 dB and 50% respectively.

Keywords Steganography • Discrete wavelet transform • Genetic algorithm • Optimal pixel adjustment process • Image processing

1 Introduction

Steganography is the art and science to hide data in a cover that it can be text, audio, image, video, etc. Data hiding techniques are generally divided in two groups: spatial and frequency domain [1–3]. The first group is based on embedding message in the least significant bit (LSB) of image pixel. The basic LSB method has a

E. Ghasemi (✉) • N. Fassihi
Department of Computer Engineering at Science and Research branch
Islamic Azad University, Tehran, Iran
e-mail: elham_ghasemi@yahoo.com; nimafassihi@gmail.com

J. Shanbehzadeh
Department of Computer Engineering, Tarbiat Moallem University Tehran, Iran
e-mail: jamshid@tmu.ac.ir

S.I. Ao et al. (eds.), *Intelligent Control and Innovative Computing*, Lecture Notes in Electrical Engineering 110, DOI 10.1007/978-1-4614-1695-1_30,

simple implementation and high capacity [4]. However it has low robustness versus some attacks such as low-pass filtering and compression [5]. There are two types of LSB insertion methods, fixed-size and variable-size. In the fixed-size methods are embedded the same number of message bits in each pixel of cover image [6, 7], and the variable-size embeds, the variant number of LSBs in each pixel used for message embedding depends on the image characteristics [8, 9]. A variant of LSB method can be found in [7] that it proposes an optimal pixel adjustment process (OPAP) in which imperceptibility of the stego-image can be improved. Furthermore, this hiding method improved the sensitivity and imperceptibility problem found in the spatial domain.

The second group embeds the messages in the frequency coefficients of images. These hiding methods overcome the problem related to robustness and imperceptibility found in the spatial domain. JPEG, a standard image compression technique, employs discrete cosine transform (DCT). Several steganography techniques for data hiding in JPEG have been proposed; such as JSteg [2], JP Hide&Seek [2] and OutGuess [11]. Most recent researches apply discrete wavelet transform (DWT) due to its wide application in the new image compression standard, JPEG2000. An example is the employment of an adaptive data embedding technique with the use of OPAP to hide data in the integer wavelet coefficients of the cover image [12]. Raja et al. [13] presented a genetic algorithm (GA) based steganography in discrete cosine transforms (GASDCT) domain and, GA based steganography using discrete wavelet transforms (GASDWT). GASDWT has an improvement in bit rate error compared to GASDCT.

The application of GA in steganography can increase the capacity or imperceptibility [14, 15]. Fard, Akbarzadeh and Varasteh [14] proposed a GA evolutionary process to make secure steganography encoding on the JPEG images. Elshafie, Kharma and Ward [15] introduced a parameter optimization using GA that maximizes the quality of the watermarked image. This paper proposes a method to embed data in DWT coefficients using a mapping function based on GA in 4×4 blocks on the cover image and, it applies the OPAP after embedding the message to maximize the Pick signal to noise ratio (PSNR).

This paper is organized as follows: Sect. 2 introduces the proposed algorithm in detail. Section 3 discusses the achieved results and compares the proposed scheme with the state of the art algorithms. Section IV concludes the paper.

2 The Steganography Method

The proposed method embeds the message in DWT coefficients based on GA and OPAP algorithm and then applies on the obtained embedded image. This section describes this method, and embedding and extracting algorithms in detail.

Fig. 30.1 The image Lena after one Haar wavelet transform [17]

2.1 Haar Discrete Wavelet Transform

Wavelet transform has the capability to present data information in time and frequency simultaneously. This transform passes the time domain data through low-pass and high-pass filters to extract low and high frequency information respectively. This process is repeated for several times and each time a section of the signal is drawn out.

DWT analysis divides the discrete signal into two segments (i.e. approximation and detail) by signal decomposition for various frequency bands and scales. DWT utilizes two function sets: scaling and wavelet which associate with low and high pass filters. Such a decomposition manner bisects time separability. In other words, only half of the samples in a signal are sufficient to represent the whole signal, doubling the frequency separability.

Haar wavelet operates on data by calculating the sums and differences of adjacent elements. This wavelet operates first on adjacent horizontal elements and then on adjacent vertical elements. One nice feature of Haar wavelet transform is that the transform is equal to its inverse. Figure 30.1 shows image Lena after one Haar wavelet transform.

After each transformation, the size of the square that contains the most important information is reduced by 4. For detail information on DWT, we can see [16].

2.2 OPAP Algorithm

The main idea of applying OPAP is to minimize the error between the cover and the stego image. For example if the pixel number of the cover is 10,000 (decimal number 16) and the message vector for 4 bits is 1,111, then the pixel number will change to 11,111 (decimal number 31) and the embedding error will be 15, while

Fig. 30.2 A simple chromosome with 16 genes

3	10	4	8	12	7

after applying OPAP algorithm the fifth bit will be changed from 1 to 0, and the embedding error is reduced to 1.

The OPAP algorithm can be described as follows:

Case 1 $(2^{k-1} < \delta_i < 2^k)$: if $p_i' \geq 2^k$, then $p_i'' = p_i' - 2^k$ otherwise $p_i'' = p_i'$;

Case 2 $(-2^{k-1} < \delta_i < 2^{k-1}) : p_i'' = p_i'$;

Case 3 $(-2^k < \delta_i < -2^{k-1})$: if $p_i' < 256 - 2^k$, then $p_i'' = p_i' + 2^k$; otherwise $p_i'' = p_i'$;

P_i, P_i' and P_i'' are the corresponding pixel values of the ith pixel in the three images; cover, stego and the obtained image by the simple LSB method, respectively. $\left(\delta_i = P_i'' - P_i\right)$ is the embedding error between P_i and P_i'' [7]. Therefore after embedding k-LSBs of P_i with k message bits, δ_i will be as follows:

$$-2^{\kappa} < \delta_t < 2^{\kappa} \tag{30.1}$$

2.3 Genetic Algorithm

GA is a technique which mimics the genetic evolution as its model to solve problems. The given problem is considered as input and the solutions are coded according to a pattern. The *fitness* function evaluates every candidate solution most of which are chosen randomly. Evolution begins from a completely random set of entities and is repeated in subsequent generations. The most suitable, and not the bests, are picked out in every generation. Our GA aims to improve the image quality. Pick signal to noise ratio (PSNR) can be an appropriate evaluation test. Thus the definition of fitness function will be:

$$PSNR = 10 \log_{10} \frac{M \times N \times 255^2}{\sum_{\mathrm{ij}} (y_{ij} - x_{ij})^2}. \tag{30.2}$$

Where M and N are the image sizes and, x *and* y are the image intensity values before and after embedding respectively.

A solution to the problem is translated into a list of parameters known as chromosomes. These chromosomes are usually displayed as simple strings of data. At the first step, several characteristics are generated for the pioneer generation randomly and the relevant proportionality value is measured by the fitness function. A chromosome is encoded as an array of 16 genes containing permutations 1 to 16 that point to pixel numbers in each block. Each chromosome produces a mapping function as shown in "Fig. 30.2".

The next step associates with the formation of the second generation of the society which is based on selection processes via genetic operators in accordance with the formerly set characteristics. A pair of parents is selected for every individual. Selections are devised so that to find the most appropriate component. In this way, even the weakest components enjoy their own chance of being selected and local solutions are bypassed. This paper employs Tournament method.

The contents of the two chromosomes which enter the generation process are interacted to produce two newborn chromosomes. In this approach two of the bests are mixed to give a superb one. In addition, during each process, it is likely for a series of chromosomes to undergo mutations and breed a succeeding generation of different characteristics.

2.4 Embedding Algorithm [17]

The following steps explain the embedding process:

Step1. Divide the cover image into 4×4 blocks.

Step2. Find the frequency domain representation of blocks by 2D Haar DWT and get four subbands LL1, HL1, LH1, and HH1.

Step3. Generate 16 genes containing the pixels numbers of each 4×4 blocks as the mapping function.

Step4. Embed the message bits in k-LSBs DWT coefficients each pixel according to mapping function. For selecting value of k, images are evaluated from $k = 3$ to 6. k equal to 1 or 2, provide low hiding capacity with high visual quality of the stego image and k equal to 7 or 8, provide low visual quality versus high hiding capacity.

Step5. Fitness evaluation is performed to select the best mapping function.

Step6. Apply OPAP on the image.

Step7. Calculate inverse 2D-HDWT on each 4×4 block.

2.5 Extraction Algorithm [17]

The extraction algorithm consists of four steps as follows:

Step1. Divide the cover image into 4×4 blocks.

Step2. Extract the transform domain coefficient by 2D HDWT of each 4×4 block.

Step3. Employ the obtained function in the embedding phase and find the pixel sequences for extracting.

Step4. Extract k-LSBs in each pixel.

Table 30.1 Comparison of PSNR of images for variant values of k [17]

Cover image	PSNR $k=3$	$k=4$	$k=5$	$k=6$
Lena	46.83	39.94	32.04	24.69
Jet	51.88	45.20	37.45	29.31
Boat	**48.41**	**40.44**	**31.17**	**23.60**
Baboon	**47.32**	**40.34**	**32.79**	**24.80**

3 Experimental Results [17]

The proposed method is applied on 512×512 8-bit grayscale images "Jet", "Boat", "Baboon" and "Lena". The messages are generated randomly with the same length as the maximum hiding capacity. Table 30.1 shows the stego image quality by PSNR as described in (30.2). Human visual system is unable to distinguish the grayscale images with PSNR more than 36 dB [12]. This paper embedded the messages in the k-LSBs, from $k = 3$ to $k = 6$ and received a reasonable PSNR. Table 30.1 shows PSNR for variant values of k.

This paper embedded the messages in the 4-LSBs and received a reasonable PSNR.

Table 30.1 presents the results and we can see that for k equal to 4 or 5, we obtain the highest hiding capacity and reasonable visual quality. Therefore, we take k equal to 4 as the number of bits per pixel.

Figure 30.3 shows the original cover images along with their histogram analyze which will be used later to compare it with the ones of the resulting stego image to test for imperceptibility.

Figure 30.4 shows images for k equal to 4 that there is no significant change in stego image histogram for 4-LSBs images, thus it is robust against some statistical attacks.

4 Conclusions

This paper introduced a novel steganography technique to increase the capacity and the imperceptibility of the image after embedding. GA employed to obtain an optimal mapping function to lessen the error difference between the cover and the stego image and use the block mapping method to preserve the local image properties. Also we applied the OPAP to increase the hiding capacity of the algorithm in comparison to other systems. However, the computational complexity of the new algorithm is high. The simulation results showed that capacity and imperceptibility of image had increased simultaneity. Also, we can select the best block size to reduce the computation cost and to increase the PSNR using optimization algorithms such as GA (Table 30.2).

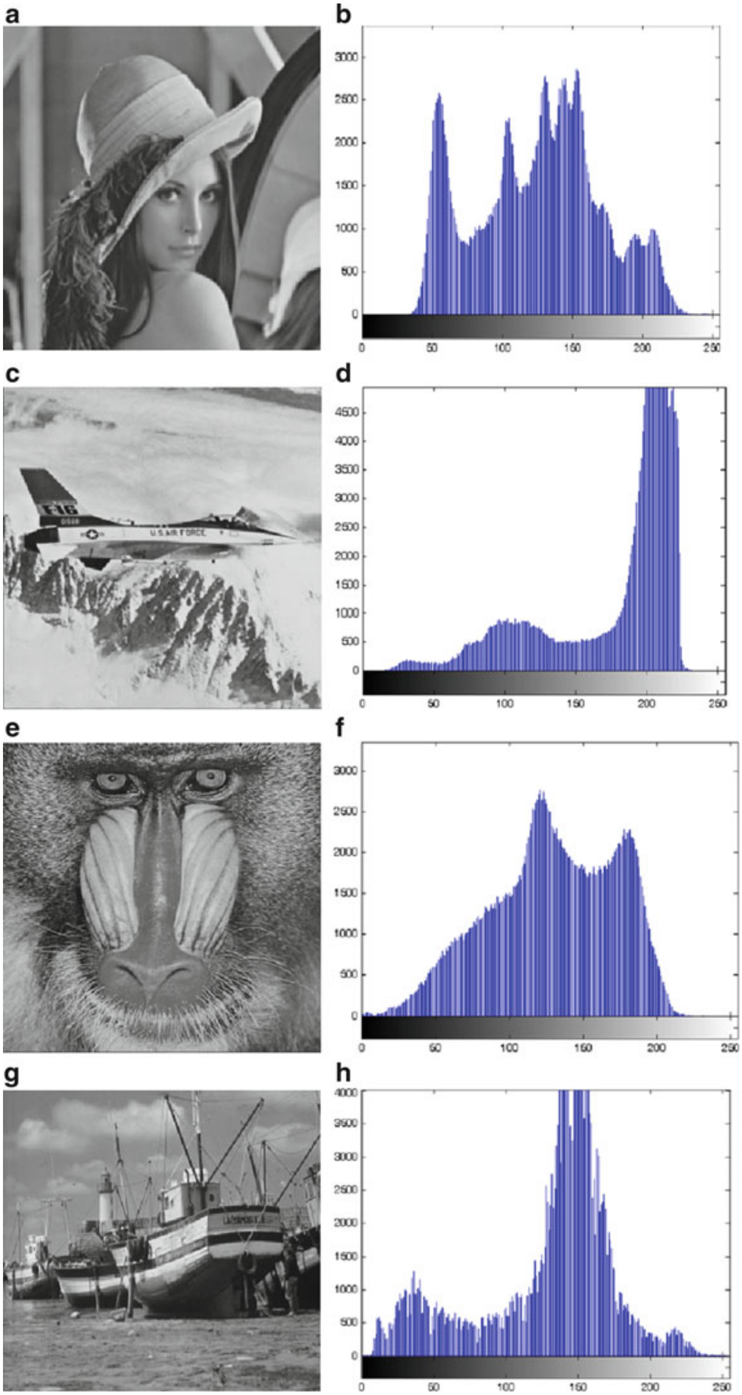

Fig. 30.3 Four cover images used in system simulation and their corresponding histogram: (**a**) cover image Lena; (**b**) Lena histogram; (**c**) cover image Jet; (**d**) Jet histogram; (**e**) cover image Baboon; (**f**) Baboon histogram; (**g**) cover image boat; (**h**) boat histogram [17]

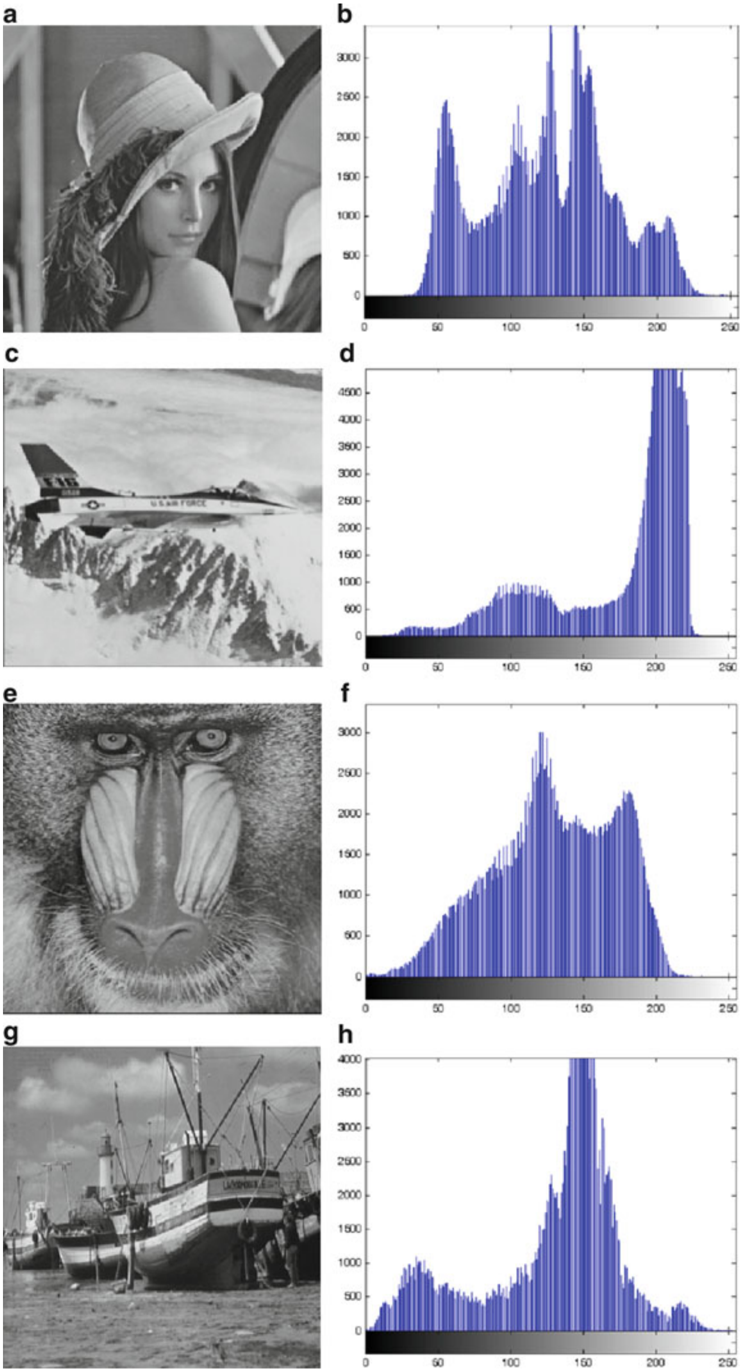

Fig. 30.4 Output stego image of $k = 4$ for embedding data and their corresponding histograms: (**a**) Lena image; (**b**) Lena histogram; (**c**) Jet image; (**d**) Jet histogram; (**e**) Baboon image; (**f**) Baboon histogram; (**g**) Boat image; (**h**) Boat histogram [17]

Table 30.2 Comparison of hiding capacity achieved and the obtained PSNR between our proposed method and methods in [10, 12] and [16, 17]

Cover image	Method	Hiding capacity (bit)	Hiding capacity (%)	PSNR (DB)
Lena	Proposed method	1,048,576	50%	39.94
	adaptive [12]	986,408	47%	31.8
	HDWT [10]	801,842	38%	33.58
	DWT [16]	573,550	27.34%	44.90
Baboon	Proposed method	1,048,576	50%	40.34
	adaptive [12]	1,008,593	48%	30.89
	HDWT [10]	883,220	42%	32.69
	DWT [16]	573,392	27.34%	44.96
Jet	Proposed method	1,048,576	50%	45.20
	DWT [16]	573,206	27.33%	44.76
Boat	Proposed method	1,048,576	50%	40.44
	DWT [16]	573,318	27.33%	44.92

References

1. Inoue H, Miyazaki A, Katsura T (2002) An image watermarking method based on the wavelet transform. In: International Conference on Image Processing, IEEE ICIP, Kobe, vol 1, pp 296–300
2. Provos N, Honeyman P (2003) Hide and seek: an introduction to steganography. In: IEEE Journal Computer Society, ISSN: 1540-7993, 11 June 2003, pp 32–44
3. Provos N (2001) Defending against statistical steganalysis. In: SSYM'01 Proceedings of the 10th conference on USENIX Security Symposium, USENIX Association Berkeley, CA, USA, vol 10, pp 323–335
4. Wu N, Hwang M (2007) Data hiding: current status and key issues. Int J Network Security 4(1):1–9
5. Chen W (2003) A comparative study of information hiding schemes using amplitude, frequency and phase embedding, PhD thesis, National Cheng Kung University, Taiwan
6. Chi-kwong C, Cheng LM (2002) Improved hiding data in images by optimal moderately significant-bit replacement. IEE Electron 37:1017–1018
7. Chan CK, Chang LM (2004) Hiding data in images by simple LSB substitution. Pattern Recogn 37:469–474
8. Chang K-C, Chang C-P, Huang PS, Te-Ming T (2008) A novel image steganographic method using tri-way pixel-value differencing.J Multimedia 3(2):37–44
9. Lee YK, Chen LH (2000) High capacity image steganographic model. IEE Proc Image Signal Process 147:288–294
10. Lai B, Chang L (2006) Adaptive data hiding for images based on haar discrete wavelet transform. In: Lecture Notes in Computer Science, Springer-verlag Berlin Heidelberg, vol 4319, pp 1085–1093
11. Provos N (2001) Defending against statistical steganalysis. In: Proceedings of 10th usenix security symposium, Usenix Assoc, pp 323–335
12. El Safy RO, Zayed HH, El Dessouki A (2009) An adaptive steganography technique based on integer wavelet transform. In: ICNM international conference on networking and media convergence, Cairo 24–25 March, pp 111–117

13. Raja KB, Kiran KK, Satish KN, Lashmi MS, Preeti H, Venugopal KR, Patnaik LM (2007) Genetic algorithm based steganography using wavelets. In: International conference on information system security, ICISS, Springer-Verlag Berlin Heidelberg, vol 4812, pp 51–63
14. Fard AM, Akbarzadeh MR, Varasteh AF (2006) A new genetic algorithm approach for secure JPEG steganography. International conference on engineering of intelligence systems, ICEIS Islamabad, 18 September, pp 1–6
15. ElShafie DR, Kharma N, Ward R (2008) Parameter optimization of an embedded watermark using a genetic algorithm. In: International symposium on communications, control and signal processing, ISCCSP, St Julians12–14 March, pp 1263–1267
16. Chen P, Lin H (2006) A DWT based approach for image steganography. Int J Appl Sci Eng 4(3):275–290
17. Ghasemi E, Shanbehzadeh J, Fassihi N High capacity image steganography using wavelet transform and genetic algorithm. In: Lecture notes in engineering and computer science: proceedings of the international multiconference of engineers and computer scientists 2011, IMECS 2011, Hong Kong, 16–18 March 2011, pp 495–498