An Image High Capacity Steganographic Methods By Modified OPA Algorithm And Haar Wavelet Transform.

1 A.Antony Judice
Assistant Professor,Arunachala College Of Engineering For Women.

UG Students,Arunachala College Of Engineering For Women.

email: pravinhireling@gmail.com

email: dhivyashamimi@gmail.com
3 divyasreedj@gmail.com
4 lekshminairece@gmail.com

Abstract-Steganography is the art and science of hiding communication; a steganographic system thus embeds hidden content in unremarkable cover media so as not to arouse an eavesdropper's suspicion. In the past, people used hidden tattoos or invisible ink to convey steganographic content. Using steganography, information can be hidden in different embedding mediums, known as carriers. These carriers can be images, audio files, video files, and text files. The focus in this paper is on the use of an image file as a carrier, and hence, the taxonomy of current steganographic techniques for image files has been presented.

Today, computer and network technologies provide easy-to-use communication channels for steganography. In most algorithm used to secure information both steganography and cryptography are used together to secure a part of information. Steganography has many technical challenges such as high hiding capacity and imperceptibility. In this paper, we try to optimize these two main requirements by proposing a novel technique for hiding data in digital images by combining the use of adaptive hiding capacity function that hides secret data in the integer wavelet coefficients of the cover image with the optimum pixel adjustment (OPA) algorithm. The proposed system showed high hiding rates with reasonable imperceptibility compared to other steganographic system.

Index Terms- adaptive algorithm, cryptography, discrete wavelet transform, imperceptibility, opa algorithm, steganography, spatial domain.

1.INTRODUCTION

"Steganography" is a Greek origin word which means “hidden writing”. Steganography word is classified into two parts: Steganos which means “secret or covered” (where you want to hide the secret messages) and the graphic which means “writing” (text). Internet users frequently need to store, send, or receive private information. The most common way to do this is to transform the data into a different form. The resulting data can be understood only by those who know how to return it to its original form. This method of protecting information is known as encryption. A major drawback to encryption is that the existence of data is not hidden. A solution to this problem is steganography. The ancient art of hiding messages so that they are not detectable. No substitution or permutation was used. The hidden message is plain, but unsuspecting to the reader. Steganography’s intent is to hide the existence of the message, while cryptography scrambles a message so that it cannot be understood. The use of steganography dates back to ancient times where it was used by romans and ancient Egyptians.

The motivation behind developing image Steganography methods according to its use in various organizations to communicate between its members, as well as, it can be used for
communication between members of the military or intelligence operatives or agents of companies to hide secret messages or in the field of espionage. The main goal of using the Steganography is to avoid drawing attention to the transmission of hidden information. If suspicion is raised, then this goal that has been planned to achieve the security of the secret messages, because if the hackers noted any change in the sent message then this observer will try to know the hidden information inside the message. Each steganographic technique consists of an embedding algorithm and a detector function. The embedding algorithm is used to hide secret messages inside a cover (or carrier) document. The embedding process is usually protected by a keyword so that only those who possess the secret keyword can access the hidden message. The detector function is applied to the carrier and returns the hidden secret message. For secure covert communication, it is important that by injecting a secret message into a carrier document no detectable changes are introduced. The main goal is to not raise suspicion and avoid introducing statistically detectable modifications into the carrier document. The embedded information is undetectable if the image with the embedded message is consistent with the model of the source from which the carrier images are drawn. We point out that the ability to detect the presence does not automatically imply the ability to read the hidden message. We further note that undetectability should not be mistaken for invisibility a concept tied to human perception.

There are a number of steganographic schemes that hide secret message in an image file; these schemes can be classified according to the format of the cover image or the method of hiding. We have two popular types of hiding methods: Spatial domain embedding and transform domain embedding.

The other type of hiding method is the transform domain technique which appeared to overcome the robustness and imperceptibility problems found in the LSB substitution techniques. There are many transform that can be used in data hiding, the most widely used transforms are: the discrete cosine transform (DCT) which is used in the common image compression format JPEG and MPEG, the discrete wavelet transform (DWT) and the discrete Fourier transform (DFT). Most recent researches are directed to the use of DWT since it is used in the new image compression format JPEG2000 and MPEG4. In [2] the secret message is embedded into the high frequency coefficients of the wavelet transform while leaving the low frequency coefficients subband unaltered. While in an adaptive hiding capacity function is employed to determine how many bits of the secret message is to be embedded in each of the wavelet coefficients. The advantages of transform domain techniques over spatial domain techniques are their high ability to tolerate noises and some signal processing operation but on the other hand they are computationally complex and hence slower [2]. In all proposed techniques for steganography whether spatial or transform the key problem is how to increase the size of the secret messages without causing noticeable distortions in the cover object. Some of these techniques try to achieve the high hiding capacity of the cover according to its local characteristics as in [2, 3, 4, 5].

The steganography transform-based techniques have the following disadvantages: low hiding capacity and complex computations [6, 7]. Thus, to get over these disadvantage, the present paper proposes an adaptive data hiding technique joined with the use of optimum pixel adjustment algorithm to hide data into the integer wavelet coefficients of the cover image in order to maximize the hiding capacity as much as possible. We also used a pseudorandom generator function to select the embedding location of the integer wavelet coefficients to increase the system security.

2. THE STEGANOGRAPHY METHOD
The proposed method embed the message in Discrete Wavelet Transform coefficients based on G OPAP algorithm and then applied on the obtained embedded image. This section describes this method, and embedding and extracting algorithms in details.

![Fig.1.New Method in Image Steganography](image)

The main terminologies used in the Steganography systems are: the cover image, secret message, secret key and embedding algorithm. The cover message is the carrier of the message such as image, video, audio, text, or some other digital media. The secret message is the information which is needed to be hidden in the cover image. The secret key is usually used to embed the message depending on the hiding algorithms. The embedding algorithm is the way or the idea that usually use to embed the secret information in the cover message.

### A. Haar Discrete Wavelet Transform.

Wavelet transform has the capability to offer some information on frequency-time domain simultaneously. In this transform, time domain is passed through low and high frequencies respectively. This process is repeated for several times and each time a section of the signal is drawn out.

To understand how wavelets work, let us start with a simple example. Assume we have a 1D image with a resolution of four pixels, having values [9 7 3 5]. Haar wavelet basis can be used to represent this image by computing a wavelet transform. To do this, first the average of the pixels together, pair wise, is calculated to get the new lower resolution image with pixel values [8 4]. Clearly, some information is lost in this averaging process. We need to store some detail coefficients to recover the original four pixel values from the two averaged values. In our example, 1 is chosen for the first detail coefficient, since the average computed is 1 less than 9 and 1 more than 7. This single number is used to recover the first two pixels of our original four-pixel image. Similarly, the Second detail coefficient is -1, since 4 + (-1) = 3 and 4 - (-1) = 5. Thus, the original image is decomposed into a lower resolution (two-pixel) version and a pair of detail coefficients. Repeating this process recursively on the averages gives the full decomposition shown in Table.

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Average</th>
<th>Detail Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>[9 7 3 5]</td>
<td>_</td>
</tr>
<tr>
<td>2</td>
<td>[8 4]</td>
<td>[1 -1]</td>
</tr>
<tr>
<td>1</td>
<td>[6]</td>
<td>[2]</td>
</tr>
</tbody>
</table>

Thus for the one dimensional Haar basis the wavelet Transform of the original four-pixel image is given by [6 2 1 -1]. We call the way used to compute the wavelet transform by recursively averaging and differencing coefficients, filter bank. We can reconstruct the image to any resolution by recursively adding and subtracting the detail coefficients from the lower resolution versions.
Fig. 2. The image Lena after one Haar wavelet Transform

After each transform is performed the size of the square which contain the most important information is reduced by a factor of 4.

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

C. 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.

Peak signal to noise ratio (PSNR) can be an appropriate evaluation test. Thus the definition of fitness function will be:

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

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. 3”.

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.

D. Embedding Algorithm

The details of data hiding steps are described as
follows.
1) Calculate four difference values \( d_i(x, y) \) for four pixel.
2) Using \(|d_i(x, y)| (i = 0, \ldots, 3)\) to locate a suitable \( R_{ki} \) in the designed range table, that is to compute \( J = \min(U_{ki} \cdot |d_{ki}(x, y)|) \) where \( U_{ki} \geq |d_{ki}(x, y)| \) for all \( 1 \leq k \leq n \). Then the located range can be represented by \( R_{ki} \).
3) Compute the amount of secret data bits \( t_i \) can be embedded in each pair by \( R_{ki} \). The value \( t_i \) can be estimated from the width \( W_{ij} \), this can be defined by \( t = \lfloor \log_2 W_{ij} \rfloor \).
4) If \( t_i \) of \( P_i \) \( (i = 0,1,2) \) satisfies branch conditions, two pixel pairs of \( P0 \) and \( P3 \) are processed by the original PVD method. Otherwise, the proposed triway scheme is used to process \( P_i \) \( (i = 0,1,2) \).
5) Read \( t_i \) bits from the binary secret data and transform the bit sequence into a decimal value \( b_i \).
6) Calculate the new difference value \( d'_i(x, y) \).
7) Modify the values of \( P_n \) and \( P_{n+1} \) by the following formula:
\[
(P'_n, P'_{n+1}) = (P_n - \lfloor m/2 \rfloor, P_{n+1} + \lfloor m/2 \rfloor)
\]
where \( P_n \) and \( P_{n+1} \) represent two pixels in \( P_i \) and \( m = d' - d \).
8) Using the selection rules to choose the optimal reference point \( P'_i(x, y) \) with minimum MSE, then this selected point is used to offset the other two pixel pairs.
9) Now, the new block constructed from all pixel pairs and embedded with secret data is generated.

An illustration of the data embedding process is shown in Fig. 2. In Fig.3, suppose that the sample block is comprised by \( (P_{k1}(x, y), P_{k1+1}(x, y), P_{k1+1}(x, y), P_{k1+1+1}(x, y)) \) and the gray values are \((100,126,115,107)\). At first, we set the pixel values of four pairs to be \( P0 = (100,126), P1 = (100,115), P2 = (100,107), \) and \( P3 = (115,107) \), respectively. Then, difference values calculated from four pairs are 26, 15, 7. The amount of bits can be embedded into each of four pairs are \( t0 = 4, t1 = 3, t2 = 3, \) and \( t4 = 3 \), respectively. By considering the branch conditions in this block, pixel pair \( P3 \) is discarded. Suppose that the binary secret data to be embedded is \( 11010101001101 \). Therefore, the individual binary bit streams of \( P_i \) \( i \in \{0,1,2\} \) are 1101, 010, and 100. The corresponding decimal values of those bit streams are 13, 2, and 4, respectively. By following the definition of \( d'_i = l_i + b_i \), the new difference values of those three pairs are computed as \( d'_0 = 16 + 13 = 29, d'_1 = 8 + 2 = 10, \) and \(-8\), respectively. Based on Step 7, we can obtain the new gray values of three pixel pairs, where \( P'0 = (98,127), P'1 = (103,113), \) and \( P'2 = (102,106) \). Then \( m0 = 3, m1 = -5, \) and \( m2 = -3 \) can be computed. Now, the embedded block is obtained and given by \((102,131,112,106)\).

**E. The Extraction algorithm**

To retrieve the embedded secret data from the stegoimage, the extraction algorithm is described in the following steps.

1) Partition the stego-image into \( 2 \times 2 \) pixel blocks, and the partition order is the same as that in the embedding stage.
2) Calculate the difference values separately for each block in the stego-image.
3) If \( t_i \) satisfies the branch conditions, two independent pixel pairs are selected; otherwise, three pixel pairs are used for further processing.
4) After \( R_{ki} \) is located, \( l_{ki} \) is subtracted from the selected \( |d^\wedge_{ki}(x, y)| \) and \( b^\wedge_i \) is obtained. If the stego image is not altered, \( b^\wedge_i \) is equal to \( b \). Finally, \( b^\wedge_i \) is
converted from a decimal value into a binary sequence with the bit stream is only one part of the secret data before embedding.

3. STEGANOGRAPHY PROCEDURE

![Steganography Procedure](image)

All of the approaches to steganography have one thing in common that they hide the secret message in the physical object which is sent. The following figure shows the steganography process of the cover image being passed into the embedding function with the message to encode resulting in a steganographic image containing the hidden message. A key is often used to protect the hidden message. This key is usually a password, so this key is also used to encrypt and decrypt the message before and after the embedding. Secrets can be hidden inside all sorts of cover information: text, images, audio, video and more. However, there are tools available to store secrets inside almost any type of cover source. The most important property of a cover source is the amount of data that can be stored inside it, without changing the noticeable properties of the cover.

4. PROPOSED SYSTEM

The proposed system is an adaptive data hiding scheme, in which randomly selected integer wavelet coefficients of the cover image are modified with secret message bits. Each of these selected coefficients hide different number of message bits according to the hiding capacity function, the capacity function used is a modified version of the one in . After data insertion we apply optimum pixel adjustment algorithm in to reduce the error induced due to data insertion. The block diagram is shown in "Fig. 3". We can say that the proposed system is classified into three cases of operation according to different applications; Low hiding capacity with good visual quality (high value of peak signal to noise ratio "PSNR"), average hiding capacity with reasonable visual quality and high hiding capacity with low visual quality.

F. Embedding algorithm

The blocks of the embedding algorithm is explained in the following steps:

**Step 1:** Read the cover image file into a two dimensional decimal array to handle the file data more easily.

**Step 2:** histogram modification it is used to prevent overflow/underflow that occurs when the changed values in integer wavelet coefficients produce stego-image pixel values to exceed 255 or to be smaller than O[5,9]. This problem was found to be caused by the values near 255 or 0. The problem can be solved by mapping the lowest 15 grayscale levels to the value of 15 and the highest 15 grayscale levels to the value 240.

**Step 3:** divide the cover image into 8x8 non overlapping blocks. By this division each 8x8 block can be categorized as a smooth or complex block.

**Step 4:** (Integer wavelet Transform): transform each block to the transform domain using 2D Haar integer wavelet transform resulting LLI, LHI, HLI and HHI.

**Step 5:** Calculate hiding capacity (number of bits to be used in hiding message bits) of each coefficient, we used a
modified version of the hiding capacity function in [5]. The length of LSBs of wavelet coefficients (L) is determined according to [5]:

From experiments we found that as we lower the bits used to hide the secret message in the LL subband the resulted distortion in the stego-image becomes lower; so that we modified this hiding capacity function by using different ranges for k for the LH, HL and HH subbands where its values are form 1 to 4. For the LL subband the value of k is equal to 0 and in some cases the bits used is fixed to only bits to enhance the stego-image quality.

Form experiments of different values of k we divided the system into 3 cases of operation depending on the requirements of the user; these cases are:

**Case 1**: k 1 for LH, HL and HH subbands, while using 2 bits for embedding data in LL subband.
This case provides low hiding capacity with high visual quality of the stego-image.

**Case 2**: k 3 for LH, HL and HH subbands, while using 2 bits for embedding data in LL subband.
This case is for applications requiring average hiding capacity with reasonable visual quality.

**Case 3**: k 4 for LH, HL and HH subbands, while k = 0 for LL subband.
Case 3 is considered as the worst case of data embedding where it is used when the high visual quality of the stego-image is not important and the user requires only high hiding capacity.

Note that we dropped the case of k=2 because it provided no significant improvement to the result obtained by k=1 or k=3.

**Step 6**: Embed L bits of message into the corresponding randomly chosen coefficients. Random selection of coefficients provides more security where the sequence of the message is only known to both sender and receiver by using a previously agreed upon secret key.

**Step 7**: Apply optimal pixel adjustment algorithm, while taking into consideration that each modified coefficient stays in its hiding capacity range where each value of L is calculated according to the absolute value of the wavelet coefficients any significant change in this value will produce different value of L to be calculated at the receiver.

The main idea of using the optimum pixel adjustment (OPA) algorithm is to minimize the error difference between the original coefficient value and the altered value. For example, if a binary number 1000 (decimal number 8) is changed to 1111 (decimal number 15) because its three LSB's were replaced with embedded data; the difference from the original number is 7.

The algorithm we used in Step 8 is the final step in the proposed scheme, where it can minimize the error by half. The main idea of OPA is to check the bit right next to the last changed LSBs is used to decrease the error.

**Step 8**: Finally, calculate the inverse integer wavelet transform on each 8x8 block to restore the image to spatial domain.
G. The Extraction Algorithm

At the receiver uses the extraction algorithm to obtain the secret message. The block diagram of the extraction algorithm is shown in "Fig. 7".

5. EXPERIMENTAL RESULTS

The proposed system was applied to two typical 512x512 8-bit grayscale images shown in figure "baboon" and "barb"; it achieved satisfactory
results against other systems using wavelet transform.

Fig. 8. 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.

Human visual system is unable to distinguish the grayscale images with PSNR more than 36 dB. This paper embedded the messages in the 4-LSBs and received a reasonable PSNR.

Fig. 9. 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.

1. Imperceptibility/Steego-image quality

This aspect measures how much difference (distortion) was caused by data hiding in the original cover, where the higher the stego-image quality, the more invisible the hidden message. We can judge the stego-image quality by using Peak Signal to Noise Ratio (PSNR). The PSNR for an image of size MxN is calculated by

$$PSNR = 10 \log_{10} \left( \frac{255^2}{MSE} \right)$$

The $MSE$ is the Mean Square Error, $P(x, y)$ stands for the image pixel value in the cover image and $P'(x, y)$ is for the pixel value at position $x$, in the image after inserting secret message. A high value of PSNR means better image quality (less distortion), it is recorded that in grayscale images that the human visual system (HVS) can not detect any distortions in stego-images having PSNR that goes beyond 36 dB.
2. Payload /Hiding Capacity

The hiding capacity indicates of how much data can be hidden within a cover image without making obvious degradation in the cover image quality. Due to the importance where it has no meaning that an algorithm hides large amount of data and produce large distortion in image quality. So we can say that a steganographic technique is an addition if it proves increase in payload while maintaining an acceptable visual quality of stego-image or improve the stego-image quality at the same hiding capacity level or if it can improve both[11].

The values of H.C. ranges from (22% to 30%). Also the PSNR is calculated for each stego-image and it ranges form (37 dB to 40 dB); which are far above the threshold for the HVS of 36dB.

Comparing the resulting stego images and their histograms with the ones in Fig we can see that there is no significant change in Barb histogram. Unlike barb image we can see that the Baboon histogram is changed significantly due to the large number of edges in the original image although it does not affect the visual quality of the resulting stego-image.

Figure shows that the proposed system can give a high hiding capacity of 48% of the cover image size with a PSNR of about 31 dB which gives a reasonable visual quality of the stego-image.

The histogram analysis for both stego-images shows that when the size of secret data increases, the histogram tends to be smoother. This is clear when comparing the histograms in "Fig. 7", "Fig. 8" and "Fig. 9" with the corresponding ones of the original images in "Fig. 6".

To further investigate the imperceptibility of the proposed system we compared the hiding capacity of our system with other systems at the same PSNR value and it showed better results. For example, the system in [12] showed a maximum hiding capacity of 36% of the cover image at a PSNR value of 34.63 dB while our system showed a hiding capacity of 38% of the cover image at the same PSNR.
The proposed method is applied on 512x512 8-bit grayscale images “Jet”, “Boat”, “Baboon” and “Lena”. The messages are generated randomly with the same length as the maximum hiding capacity. Table I shows the stego image quality by PSNR by the formula. Human visual system is unable to distinguish the grayscale images with PSNR more than 36 dB [5].

6. CONCLUSION

In this paper we proposed a novel data hiding scheme that hides data into the integer wavelet coefficients of an image. The system combines an adaptive data hiding technique and the optimum pixel adjustment algorithm to increase the hiding capacity of the system compared to other systems. The proposed system embeds secret data in a random order using a secret key only known to both sender and receiver. It is an adaptive system which embeds different number of bits in each wavelet coefficient according to a hiding capacity function in order to maximize the hiding capacity without sacrificing the visual quality of resulting stego image. The proposed system also minimizes the difference between original coefficients values and modified values by using the optimum pixel adjustment algorithm. There was no error in the recovered message (perfect recovery) at any hiding rate. From the experiments and the obtained results the proposed system proved to achieve high hiding capacity up to 48% of the cover image size with reasonable image quality and high security.

REFERENCES