IMPLEMENTING DATA HIDING IN STILL IMAGES

Iman Hazwam Bin Abd. Halim and Roshidi Din
Faculty Of Information Technology
Universiti Utara Malaysia 06010 Sintok, Kedah

Abstract: Data hiding refers to the nearly invisible embedding of information within various host data sets and one of them is by using still images. It is a highly multidisciplinary field that combines image and signal processing with cryptography, communication theory, coding theory, signal compression, and the theory of visual perception. There are a number of ways exist to hide information in still images. Common approaches include least significant bit insertion, masking and filtering, algorithm and transformations, and spread spectrum techniques. Each of these techniques can be applied, with varying degrees of success, to different image files formats. The data hiding process of each method will be discussed in this paper and the comparison tables that show the suitability of different image file types with the recent data hiding methods according to some parameters also will be provided. This paper also will discuss about the challenges in implementing data hiding in still images.

Keywords: Data hiding, Digital images, Steganography, Watermarking.

INTRODUCTION

With the development of Internet technologies, multimedia applications have been widely used and can be transmitted conveniently over the networks. Therefore, how to protect secret data during transmission becomes an important issue. One of the possible solutions to cater this problem is by hiding it first before implementing the transmission.

This method is usually known as Data Hiding which refers to the nearly invisible embedding of information within various host data sets and one of them is by using still images. It is a highly multidisciplinary field that combines image and signal processing with cryptography, communication theory, coding theory, signal compression, and the theory of visual perception Data hiding can be classified into a number of application areas but most important sub-disciplines of data hiding are steganography and copyright marking which is closely related with watermarking [4]

To a computer, still image is an array of values that represent light intensities of three colors Red, Green, and Blue (RGB), where a value for each three colours describes a pixel (F.J. Neil, J. Sushil, 1998). There are several still image file types that usually being used for data hiding such as Bitmap, GIF, TIFF, and JPEG.

Besides data hiding in still images, data also can be hidden in other forms of media such as moving images, audio files, text files and file systems. But most of the researches on data hiding are focused on still images. This is due to the facts that still images are one of the most frequently exchanged on the Internet, and the methods for information hiding in them are quite mature. Besides that, an advantage in using still images for data hiding is that they represent a noncausal medium, since it is possible to access any pixel of the image at random [9].

Further on this paper will discuss about image file types that usually used for data hiding, techniques of data hiding which will be focusing on steganography and watermarking, and the discussion and challenges in data hiding.

DISCUSSIONS

Image File Types for Data Hiding

Most data hiding techniques recommends the use of 24-bit images such as BMP files. BMP files use 3 bytes per pixel to represent a color value and it can be represented as hexadecimal, decimal and binary values. (F.J. Neil, J. Sushil, 1998) BMP files can provides the most space for hiding information but because of the large size, it would attract attention of eavesdropper. BMP files also are vulnerable to image compression when it was interfered with hidden information. The next best alternative for data hiding in 24-bit image is 256-color or gray-scale images. The most common found on the Internet is the GIF files. GIF files will decrease the number of bits used to represent each pixel from 24-bits to 8-bits. Many steganography experts recommend using images featuring 256 shades of gray. (F.J. Neil, J. Sushil, 1998) Gray-scale images are preferred because the shades change very gradually from byte to byte, and the less the value changes between palette entries, the better they can hide information. Because GIF format is a lossless compression technique, so the data hidden in the image can be recovered without a problem. Another image file format is TIFF files. The color channel of the TIFF file is organized as Red Green Blue (RGB). The pixels orientation when they are stored in the file is from the upper left corner to the lower right corner of the image. Not all image formats share this property [5]. The fourth image file format is Joint Photographic Experts Group (JPEG). Most images that transferred over the internet are stored in JPEG format because high color quality images can be stored in relatively small files using JPEG compression. That is why most research on data hiding techniques preferred to use JPEG images as their medium. JPEG is a form of lossy compression. It uses the discrete cosine transform to achieve compression. JPEG images offer high compression, but may not maintain the original image's integrity. This can impact negatively on any hidden data in the image. This is due to the lossy compression algorithm, which may ``lose'' unnecessary image data, providing a close approximation to high-quality digital images, but not an exact duplicate. Lossy compression is frequently used on true-colour images, as it offers high compression rates. (F.J. Neil, J. Sushil, 1998)

Data Hiding Techniques for Still Images

A number of ways exist to hide information in digital images. Common approaches include least significant bit insertion, masking and filtering, algorithms and transformations, and spread spectrum techniques. Each of these techniques can be applied, with varying degrees of success, to different image files.

Least significant bit insertion method:

To hide an image in the LSBs of each byte of 24-bit image, 3 bits of hidden messages can be stored in each pixel. A larger amount of information can be stored if the message is being compressed first before embedding it into the image. To the human eye, the resulting stego-image will look identical to the cover image. For example (adapted from the paper by Jajodia and Neil [5, 6]), the letter B can be hidden in three pixels (assuming no compression has been made). The original raster data for 3 pixels (9 bytes) may be

(00100111 11101001 11001000) (00100111 11001000 11101001) (11001000 00100111 11101001)
The binary value for B is 01000010. Inserting the binary value for B in the three pixels would result in
(00100110 11101001 11001000) (00100110 11001000 11101000) (11001001 00100110 11101001)

The underlined bits are the bits that has been changed in the 8 bytes used. On average, the LSB only requires half of the bits in an image to be changed. This example has been adapted from the paper by Johnson and Jajodia.

The 8-bits images are not quite good for LSB manipulation because of color limitations. To implement LSB in 8-bits images, first the cover image must be carefully selected so that the stego-image will not broadcast the existence of an embedded message. When message is inserted into the LSBs of the raster data, the pointers to the color entries in the palette are changed.

The example of LSB insertion for 8-bits images is as follows:

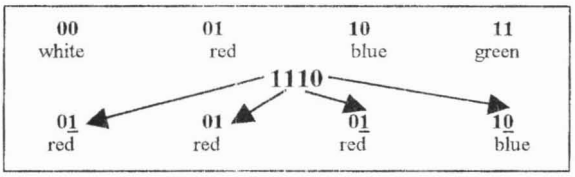

Figure 1: LSB manipulation for 8-bits images

The entries of simple four-color palette of white, red, blue and green has the corresponding palette position entries of 0 (00), 1 (01), 2 (10), and 3 (11), respectively. Hiding number 14 valued 1110 changes the raster data to 01 01 01 10, which is red, red, red, blue. These gross changes in the image are visible and it is clearly shows the weaknesses of the 8-bit images. On the other hand, there is little visible difference noticed between adjacent gray values. (F.J. Neil, J. Sushil, 1998)

The advantage of the LSB insertion method is its simplicity and many techniques use this method. LSB is also allows high perceptual transparency. Modulating the least-significant bit also does not result in a human-perceptible difference because the amplitude of the change is small. However, there are many weaknesses when robustness, tamper resistance, and other security issues are considered. The message can be easily destroyed if scaling, rotation, cropping, addition of noise, or lossy compression are implemented to the stego-image.

Masking and Filtering method:

Masking and filtering techniques are mostly used and restricted on 24 bit and greyscale images. They hide info in a way similar to watermarks on actual paper and are sometimes used as digital watermarks. Masking images entails changing the luminance of the masked area. The smaller the luminance change, the less of a chance that it can be detected. Masking techniques embed hidden information in significant area so that the hidden message is more integral to the cover image than just hiding it in the least significant bit area (See figure 2). Because masking and filtering method are more integrated into the image, they may be applied without fear of image destruction from lossy compression. Masking techniques are more suitable for use in lossy JPEG images than LSB insertion because of their relative immunity to image operations such as image cropping and compression. (F.J. Neil, J. Sushil, 1998)

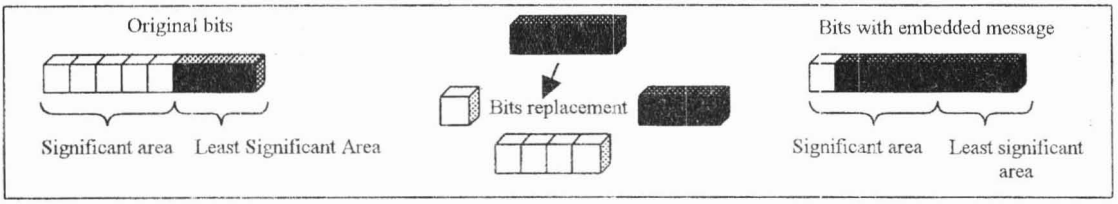

Figure 2: Data masking process

Algorithm and Transformation method:

Algorithm and transformation method embed the message by modulating coefficients in a transform domain, such as Discrete-Cosine Transformation (DCT) that usually used in JPEG compression, and Wavelet Transformation. To explain the encoding and decoding process of algorithm and transformation method, we use Discrete-Cosine Transformation (DCT) as an example. The first step of encoding process is to take the DCT of the image. After that we have to find the coefficients below a certain thresholds. Then these bits will be replaced with the bits that we want to hide. The next step is take the inverse transform and lastly stored it as a regular image (Figure 3). To decode the hidden data first we have to take the transform of the modified image. Then find the coefficients that have been modified and extract the bits data from it. Lastly combine the extracted bits into an actual message (Figure 4). Transform techniques can give robustness against lossy compression because they are

designed to resist or exploit the methods of popular lossy compression algorithm. Transform techniques also can offer increased robustness against image processing such as scaling, rotations or cropping, depending on the invariant properties of the particular transform [2].

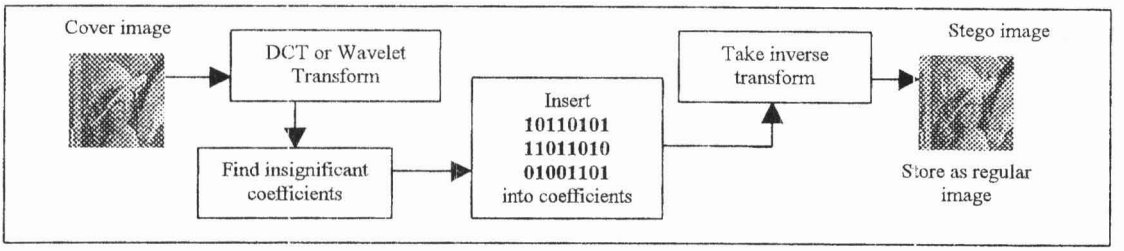

Figure 3: DCT encoding process

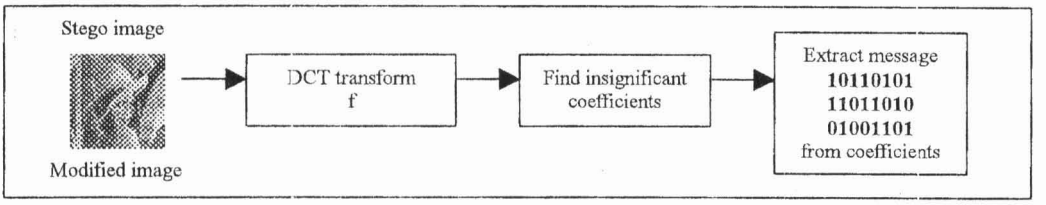

Figure 4: DCT decoding process

Spread Spectrum method:

Spread-spectrum communication describes the process of spreading the bandwidth of a narrowband signal across a wide band of frequencies. This can be done by modulating the narrowband waveform with a wideband waveform, such as white noise. After spreading, the energy of the narrowband signal in any one frequency band is low and therefore difficult to detect. Figure 5 is the example of basic encoding system of data hiding using spread spectrum techniques. First the message to be embedded will be repeated several times and then being modulated with the pseudo-noise signal. After that the spread and modulated message will interleave and added to the image pixels producing the stego image. For decoding process (Figure 6), first the stego image is pre-filtered to remove major components of the image itself. Then the filtered image is demodulated with the pseudo-noise signal that is perfectly synchronized with the one used for embedding. Correlation signal is used to detect any loss of synchronization. After that the process is followed by a summation over a window of length equal to the chip rate, and threshold which yields the message. Data hiding using spread spectrum method is more robust to image processing such as cropping and rotating, but at the cost of message size.

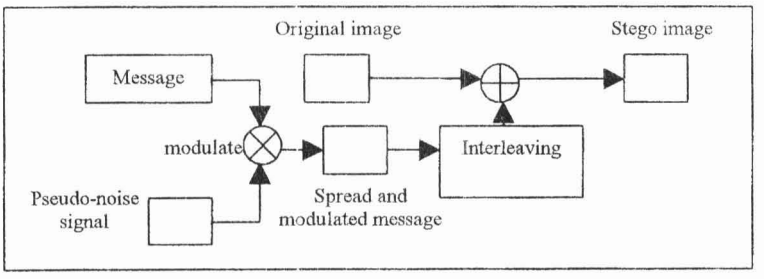

Figure 5: Basic encoding systems of data hiding using spread spectrum techniques

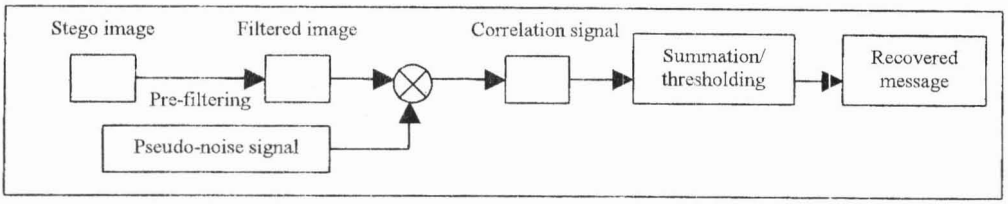

Figure 6: Decoding process of data hiding using spread spectrum techniques

Recent Approach of Data Hiding in Still Images

This section will explain the recent methods used for data hiding in still images.

Data hiding process based on LSB insertion method:

Naval Research Laboratories has proposed NRL Steganographic Method. NRL Steganographic method is based to LSB embedding techniques and was enhanced from the recent Kurak and McHugh method of steganography that can be easily detectable because the entire bit planes have been swapped. (Moskowitz, 2002). The files are saved in uncompressed TIFF format. NRL method concentrate the effort on the red value Rij. It swap the two least significant bits (LSBs) of Rij with a two bit pair from an ASCII character if the image is color, and swap every color's two LSBs if the image is greyscale. The advantage of using the NRL method is its undetectable ability to the human visual system. It is because this method only uses the red value of the cover image to embed the messages and this will reduce the changes of the cover image. The tradeoff off this method is it only use the red value of the cover image, so the capacity of bit planes that can be used to store the hidden message has been reduced.

Data hiding process based on Masking And Filtering Techniques:

Y. K Lee and L.H. Chen have proposed an adaptive image steganographic model based on Minimum-Error LSB Replacement (MELSBR) [7]. This method use 8 bits to store the intensity of each pixel on a grayscale image. By embedding k ($k<8$) bits of message in a pixel, directly replacing the k-LSBs of the pixel will introduce less error than replacing any other k-bits, and the maximum error is 2^k-1.In total of 256 gray levels, there are $2^{(8-k)}$ gray levels with the same value in the k least significant bits as the k message bits. To reduce the embedding error, this method has used an approach of selecting the one that has the minimum error with the original gray level to replace the pixel level. This paper has proposed a simple way to reach the aim by adjusting the $(k+1)^{th}$ LSB, and check its embedding error and after that select the gray-scale with less embedding error to replace the original ones. This model has been tested to different kind of images and the result explains the advantages of this model. The advantage of using this method is it can take off the restriction of fixed embedding size in each pixel. This method also will reduce the embedding error and provide higher embedding capacity.

Data hiding process based on Algorithm and Transformation Techniques:

S. Areepongsa et. al. have proposed a steganography in wavelet-based Homogenous Connected-Region Interested Ordered Transmission (HC-RIOT) for low bit rate transmission over hybrid networks. The goal of this method is to embed the additional information (hidden data) into base layer of HC-RIOT coder providing perceptual invisibility of the message in a lossy transmission environment.This method has many advantages and one of them is the message can be transmitted unique to each transmission event. This allows for authentication of the visual material down to a resolution of each transmitted event of the material. The method is also very robust towards detection or attack by requiring a unique masking value and in addition any tampering of the base layer stream will desynchronize the base layer image which results in an entire loss of image at the decoder.The main disadvantage of the algorithm is the length of the message. Since the base layer is a highly compressed image the number of bits used to transmit the image is small. This also restricts the capacity to hide the information.

Data hiding process based on Spread Spectrum Techniques:

M.M. Lisa and R. T. Charles has proposed a method of embedding information within still images called Spread spectrum image steganography (SSIS). SSIS is a data hiding method that uses digital imagery as the cover signal. The objective of SSIS is to provide the ability to hide a significant amount of information bits within digital images, avoiding detection by the observer and this advocates the maximization of capacity and minimization of perceptibility. A SSIS technique can provide a method of concealing a digital signal within the cover image without increasing the size or dynamic range of the image. Additionally, the original image is not needed to extract the hidden message, and a level of security is provided by the necessity that both the sender and receiver posses the same keys. An eavesdropper will unable to decipher the hidden information without possession of appropriate keys even though the system methodology may be known. Furthermore, the embedded signal power is insignificant compared to that of the cover image, providing low probability of detection and leaving an observer unaware that the hidden data exist.

Selecting Image File Type for Data Hiding

There are some considerations of choosing the image file formats for the use of data hiding. One of them is the dependence on the cover image and the used data hiding technique. Each process has its own method to embed the data in the image. Suitable images file formats for each technique must be distinguished accordingly. (R. Rene, S. Heidrun, 2002) To distinguish the image file formats for each technique, there are some parameters that can be followed. The parameters are the amount of data bits that can be hidden (capacity), the undetectability of the message, and its robustness against removal process. These three parameters are mutually competitive and cannot be clearly optimized at the same time [3].

Image Compatibility with Data Hiding Techniques

From the recent researches and the schematic observations of data hiding parameters that has been reviewed, three tables that show the effectiveness of image files formats comparing to different data hiding techniques used has been produced.

From table 1, we can figured out that for LSB insertion method and bmp type of image will give the highest hidden message capacity because LSB method embeds the hidden message in the least significant area and bmp image type is 24-bit image type and will give the large capacity to hide the message. But LSB method can't give the high capacity if JPEG image is used because JPEG image is the 8-bit image type and the least significant area has been omitted. Masking and filtering method is compatible to all image type because this method embeds the hidden message in the most significant area. And as we can see algorithm and transformation method and spread spectrum method only can embeds a small amounts of message and both methods are usually used for copyright marking.

Table 1: Comparison table of image file formats and data hiding technique based on hidden message capacity

	BMP	GIF	TIFF	JPEG
LSB insertion	high	low	medium	low
Masking and Filtering	high	high	high	high
Algorithm and Transformation	low	low	-	low
Spread spectrum	medium	low	-	low

From table 2, we can assume that most of the methods can give high performance if the undetectability is the main purpose in data hiding. But as we can see in the table, masking and filtering method can't give the high performance for undetectability because the criteria of the method that embeds the hidden message in the most significant bit area could change the colour luminance of the image and this tradeoff would attract the curiosity of the unwanted observer.

Table 2: Comparison table of image file formats and data hiding technique based on undetectability

	BMP	GIF	TIFF	JPEG
LSB insertion	high	high	high	high
Masking and Filtering	medium	medium	-	-
Algorithm and Transformation	low	low	-	low
Spread spectrum	medium	low	-	low

In table 3 it shows that algorithm and transformation method and spread spectrum method are robust against all modifications such as cropping and compression upon all types of images. However LSB methods is vulnerable against all image processing because the hidden message is embedded in the least significant area.

Table 3: Comparison table of image file formats and data hiding technique based on robustness

	BMP	GIF	TIFF	JPEG
LSB insertion	low	low	low	low
Masking and Filtering	medium	medium	-	medium
Algorithm and Transformation	high	high	high	high
Spread spectrum	high	high	high	high

CONCLUSION

We have describes various types of image and possible techniques that can be used for data hiding and from the discussion the comparison tables that shows the performance of each data hiding methods with different types of images has been produced. One of the challenges that has to be cater in data hiding is to choose the most suitable methods and image types that can give the highest performance referring to the parameters that has been discussed and we hope that the comparison tables that has been produced could help in making that decision. One of the possible potential to implement data hiding is by using it as the proof of the copyright. In our nation, we use smart cards as our Identification Card. So the fingerprint templates that have been saved in the smart card should be secured from being forged by the unwanted person. One of the solutions to increase the security is to add some extension signature by using data hiding techniques to keep the originality of the fingerprint templates to the authorized person only.

REFERENCES

1. Bender, W. Gruhl, D. Morimoto, N. and Lu, A. 1996. Techniques for data hiding. *IBM Systems Journal*, Vol. 35, No. 3 and 4. 313-336

2. Eugene, T.L. and Edward, J.D. 1999. A Review of Data Hiding in Digital Images. Video and Image Processing Laboratory (VIPER).

3. Fridrich, J. 1998. Applications of Data Hiding in Digital Images. Tutorial for the ISPACS'98 Conference in Melbourne, Australia.

4. Geruta, K. René , R. 2001. Information Hiding On Wavelet Based Schemes under Consideration of Jpeg2000. University of Rostock, Department of Computer Science, Institute of Computer Graphics.

5. Ira, S.M. Neil, F.J. and Jacobs, M. 2002. A Detection Study of an NRL Steganographic Method. Naval Research Laboratory, Washington.

6. Johnson, N. and Jajodia,S. 1998. Exploring Steganography: Seeing the Unseen. IEEE Computer. 26-34.

7. Lee, Y.K. and Lee, L. H. 2000. An Adaptive Image Steganographic Model Based on Minimum-Error LSB Replacement. Department of Computer and Information Science, National Chiao Tung University, Taiwan.

8. Marvel, L.M. Boncelet, C.G. Jr. and Retter, C.T. 1998. Spread Spectrum Image Steganography. Submitted to the IEEE Transaction on Image Processing.

9. Matteo, F. 2000. *Steganography and Digital Watermarking: a global view.* Laboratory of Advanced Research on Computer Science, University of Bologna. http://lia.deis.unibo.it/Courses/RetiDiCalcolatori/Progetti00/fortini/web/cover.html