Integrity Verificationn of Secret Information in Image Steganography

[1]Amit Khare
Transmission Engineer, WTR (Section),
BSNL, Thane (E),
amitkhare@bsnl.co.in

[2]Neha Khare
Assistant Professor,
KCCEMSR, Thane (E),
nehakhare31jan@gmail.com

Abstract - Steganography is the art of hiding the fact that communication is taking place, by hiding information in other information. Many different carrier file formats can be used, but digital images are the most popular because of their frequency on the Internet. For hiding secret information in images, there exist a large variety of steganographic techniques some are more complex than others and all of them have respective strong and weak points. Different applications have different requirements of the steganography technique used. This is accomplished through hiding information in other information. This paper is based on OUTGUESS algorithm. Outguess is one of the embedding algorithm which embeds messages in the DCT domain. Outguess goes about the embedding process in two separate steps. First it identifies the redundant DCT and then depending on the information obtained in the first step, chooses bits in which it would embed the message. Digital Image Steganography system is a stand-alone application that combines steganography and encryption to enhance the confidentiality of intended message. The user's intended message is first encrypted to create unintelligible cipher text. Then the cipher text will be hidden within an image file in such a way as to minimize the perceived loss in quality. The recipient of the image is able to retrieve the hidden message back from the image with Digital Image Steganography system.

Keywords
Steganography, Digital Image, OUTGUESS algorithm, JPEG Compression

I. INTRODUCTION

A. Introduction

We are living in an age of science. Communication has become major part of our daily life today. With the increasing communication traffic demand, data security has become very important field. Lots of data security and data hiding algorithms have been developed in the last decade. In this project we are implementing a method of "Digital Stenography" for image data hiding.

B. Problem Definition

Digital Image Steganography system allows an average user to securely transfer text messages by hiding them in a digital image file. A combination of Steganography and encryption algorithms provides a strong backbone for its security. Digital Image Steganography system features innovative techniques for hiding text in a digital image file or even using it as a key to the encryption.

C. Goals & Objective

This paper is feasible for several reasons. First, there is a lot of research that has been done in the field of steganography and encryption, so we will be able to reference pre-existing models. Also, Java has an extensive API that will serve the purpose of this project. We will use Java to program the user interface, which will greatly simplify creation of an easy-to-use GUI. Java is a good choice also because it has built-in capabilities for encryption and decryption.

The project contains several challenges that make it interesting to develop. The central task is to research available steganography and encryption algorithms to pick the ones that offer the best combination of strong encryption, usability, and performance. Another important aspect of the assignment is to create a simple, compact, and powerful GUI because it will play a large role in the success of the application. The GUI will include a walkthrough, which will be similar to a wizard to help first-time users. The program window will be informative in general, since it will include image previews and a text box. The program should also gracefully handle errors, giving the user meaningful error messages. Another important goal is to make sure that the changes in the image are unnoticeable to human eyes and largely undetectable to programs designed to look for repetitive and obvious patterns.

Because of its user-friendly interface, the application can also be used by anyone who wants to securely transmit private information. The main advantage of this program for individuals is that they do not have to have any knowledge about steganography or encryption. The visual way to encode the text, plus the visual key makes it easy for average users to navigate within the program.

D. Literature Survey

a. Articles

- Quick Study: Steganography: Hidden Data at computerworld.com
- Steganography Articles, Links, and Whitepapers at Forensics.nl.
- Steganography & Digital Watermarking—papers and information related to steganography and steganalysis research by Neil F. Johnson from 1995 to the present.
- Detecting Steganographic Content on the Internet, 2001. Paper by Niels Provos and Peter Honeyman, Center for Information Technology Integration, University of Michigan.
- Steganography, Steganalysis, and Cryptanalysis BlackHat and DefCon presentations by Michael T. Raggo.

b. Books

- Disappearing Cryptography by Peter Wayner

- Information Hiding—Techniques for Steganography and Digital Watermarking edited by Stefan Katzenbeisser and Fabien, A.P. Petitcolas
- Information Hiding: Steganography and Watermarking—Attacks and Countermeasures by Neil F. Johnson, Zoran Duric, and Sushil Jajodia
- The Complete Reference Java 2 by Herbert Schildt

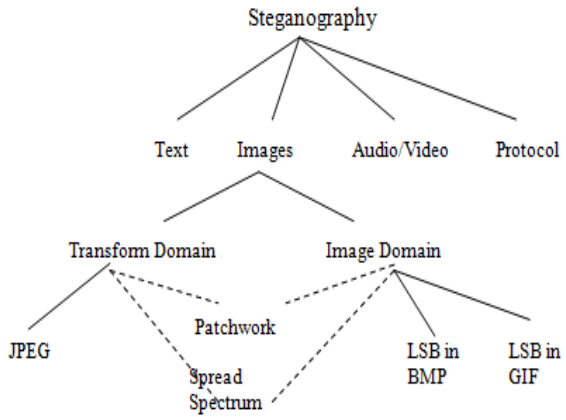
II. DIGITAL IMAGE STEGANOGRAPHY

The word steganography is of Greek origin and means "covered, or hidden writing".Generally, a steganographic message will appear to be something else: a picture, an article, a shopping list, or some other message. This apparent message is the cover text. For instance, a message may be hidden by using invisible ink between the visible lines of innocuous documents.

The advantage of steganography over cryptography alone is that messages do not attract attention to themselves, to messengers, or to recipients. An unhidden coded message, no matter how unbreakable it is, will arouse suspicion and may in itself be incriminating, as in countries where encryption is illegal. Often, steganography and cryptography are used together to ensure security of the covert message.

A steganographic message (the plaintext) is often first encrypted by some traditional means, producing a cipher text. Then, a cover text is modified in some way to contain the cipher text, resulting in stego text. For example, the letter size, spacing, typeface, or other characteristics of a cover text can be manipulated to carry the hidden message; only the recipient (who must know the technique used) can recover the message and then decrypt it. Francis Bacon is known to have suggested such a technique to hide messages.

A. Different kinds of steganography



Almost all digital file formats can be used for steganography, but the formats that are more suitable are those with a high degree of redundancy. Redundancy can be defined as the bits of an object that provide accuracy far greater than necessary for the object's use and display. The redundant bits of an object are those bits that can be altered without the alteration being detected easily. Image and audio files especially comply with this requirement, while research has also uncovered other file formats that can be used for information hiding.

a. Image steganography

As told earlier, images are the most popular cover objects used for steganography. In the domain of digital images many different image file formats exist, most of them for specific applications. For these different image file formats, different steganographic algorithms exist.

Image definition

To a computer, an image is a collection of numbers that constitute different light intensities in different areas of the image. This numeric representation forms a grid and the individual points are referred to as pixels. Most images on the Internet consists of a rectangular map of the image's pixels (represented as bits) where each pixel is located and its colour. These pixels are displayed horizontally row by row.

The number of bits in a colour scheme, called the bit depth, refers to the number of bits used for each pixel. The smallest bit depth in current colour schemes is 8, meaning that there are 8 bits used to describe the colour of each pixel. Monochrome and grey scale images use 8 bits for each pixel and are able to display 256 different colours or shades of grey. Digital colour images are typically stored in 24-bit files and use the RGB colour model, also known as true colour. All colour variations for the pixels of a 24 bit image are derived from three primary colours: red, green and blue, and each primary colour is represented by 8 bits. Thus in one given pixel, there can be 256 different quantities of red, green and blue, adding up to more than 16-million combinations, resulting in more than 16-million colours. Not surprisingly the larger amount of colours that can be displayed, the larger the file size.

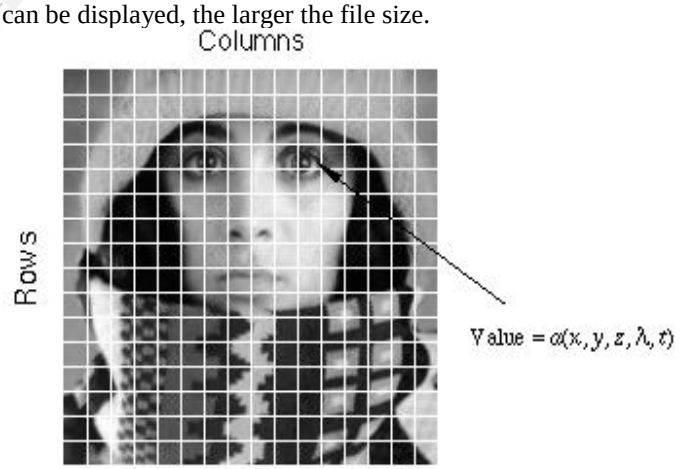


Figure 1 Digitization of a continuous image

B. Transform Domain

To understand the steganography algorithms that can be used when embedding data in the transform domain, one must first explain the type of file format connected with this domain. The JPEG file format is the most popular image file format on the Internet, because of the small size of the images.

JPEG compression

To compress an image into JPEG format, the RGB colour representation is first converted to a YUV representation. In this representation the Y component corresponds to the luminance (or brightness) and the U and V components stand for chrominance (or colour). According

to research the human eye is more sensitive to changes in the brightness (luminance) of a pixel than to changes in its color. This fact is exploited by the JPEG compression by down sampling the color data to reduce the size of the file. The colour components (U and V) are halved in horizontal and vertical directions, thus decreasing the file size by a factor of 2.

The next step is the actual transformation of the image. For JPEG, the Discrete Cosine Transform (DCT) is used, but similar transforms are for example the Discrete Fourier Transform (DFT). These mathematical transforms convert the pixels in such a way as to give the effect of "spreading" the location of the pixel values over part of the image. The DCT transforms a signal from an image representation into a frequency representation, by grouping the pixels into 8 × 8 pixel blocks and transforming the pixel blocks into 64 DCT coefficients each. A modification of a single DCT coefficient will affect all 64 image pixels in that block.

The next step is the quantization phase of the compression. Here another biological property of the human eye is exploited: The human eye is fairly good at spotting small differences in brightness over a relatively large area, but not so good as to distinguish between different strengths in high frequency brightness. This means that the strength of higher frequencies can be diminished, without changing the appearance of the image. JPEG does this by dividing all the values in a block by a quantization coefficient. The results are rounded to integer values and the coefficients are encoded using Huffman coding to further reduce the size.

JPEG steganography

Originally it was thought that steganography would not be possible to use with JPEG images, since they use lossy compression which results in parts of the image data being altered. One of the major characteristics of steganography is the fact that information is hidden in the redundant bits of an object and since redundant bits are left out when using JPEG it was feared that the hidden message would be destroyed. Even if one could somehow keep the message intact it would be difficult to embed the message without the changes being noticeable because of the harsh compression applied.

However, properties of the compression algorithm have been exploited in order to develop a steganographic algorithm for JPEGs. One of these properties of JPEG is exploited to make the changes to the image invisible to the human eye. During the DCT transformation phase of the compression algorithm, rounding errors occur in the coefficient data that are not noticeable. Although this property is what classifies the algorithm as being lossy, this property can also be used to hide messages.

It is neither feasible nor possible to embed information in an image that uses lossy compression, since the compression would destroy all information in the process. Thus it is important to recognize that the JPEG compression algorithm is actually divided into lossy and lossless stages. The DCT and the quantization phase form part of the lossy stage, while the Huffman encoding used to further compress the data is lossless. Steganography can take place between these two stages. Using the same principles of LSB insertion the message can be embedded into the least significant bits of the coefficients before applying the Huffman encoding. By embedding the information at this stage, in the transform domain, it is extremely difficult to detect, since it is not in the visual domain

III. IMAGE STEGANOGRAPHY ALGORITHM

A. Outguess

Proposed by Neils Provos in 2001 as a response to the ststistical chi –square attact by Andreas Westfeldin 1999.

Main feature of outguess

1. outguess hides messages in JPEG files
2. It embeds hides messages in bits in LSBs of quantized

Algorithm for Encoding

1) The input is a Buffered Image object, which contains a Color Model and a matrix representing the image with pointers aimed at indices of the Color Model. The RGB values of the uncompressed input image are converted into three components: one luminance component and two chrominance components (YUV). The luminance component is considered more important.

2) The image is separated into 8x8 pixel blocks starting from the upper left-hand corner.

3) The component signals for each 8x8 block are transformed into the frequency domain by using the two-dimensional discrete cosine transform (DCT). This transformation is similar to the two-dimensional fast Fourier transformation.

4) While the coefficients closest to 0 are eliminated, the remaining coefficients are quantized using various degrees of accuracy. This can be modified by changing the quantization tables. The DC luminance coefficients are the most important and are quantized with the most accuracy.

5) Since the DC luminance coefficients are quantized with the most accuracy, we will hide the information inside them

6) Finally, the quantized coefficients are compressed using a Huffman encoder.

3.2.2 Algorithm for Decoding

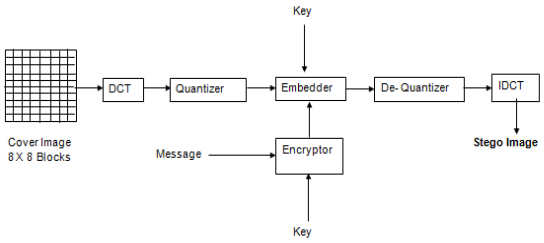
The decoding scheme is much simpler than encoding. The averages of the luminance of the 8x8 blocks just need to be calculated and converted back to bits.

Basically what we exactly do in decoding process is:-

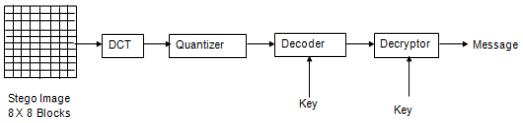
1) The user needs to provide a Source Image and any Keys used when the Source Image was generated.

2) If a non-default key was used in text hiding process, the receiving party must have prearranged knowledge of the key for use in retrieving the text.

3) After the Source Image and any required Keys are loaded into the application, the hidden text can be retrieved by selecting Extract Text from the Tools menu.



Block Diagram of Encoding



Block Diagram of Decoding

Figure 2 Block Diagram of Encoding & Decoding

IV. RESULT ANALYSIS

A. JPEG compression

The process of embedding information during JPEG compression results in a stego image with a high level of invisibility, since the embedding takes place in the transform domain. JPEG is the most popular image file format on the Internet and the image sizes are small because of the compression, thus making it the least suspicious algorithm to use. However, the process of the compression is a very mathematical process, making it more difficult to implement.

Table 1 Comparison of image Steganography algorithms

	LSB in BMP	LSB in GIF	JPEG compression	Patchwork	Spread spectrum
Invisibility	High*	Medium*	High	High	High
Payload capacity	High	Medium	Medium	Low	Medium
Robustness against statistical attacks	Low	Low	Medium	High	High
Robustness against image manipulation	Low	Low	Medium	High	Medium
Independent of file format	Low	Low	Low	High	High
Unsuspicious files	Low	Low	High	High	High

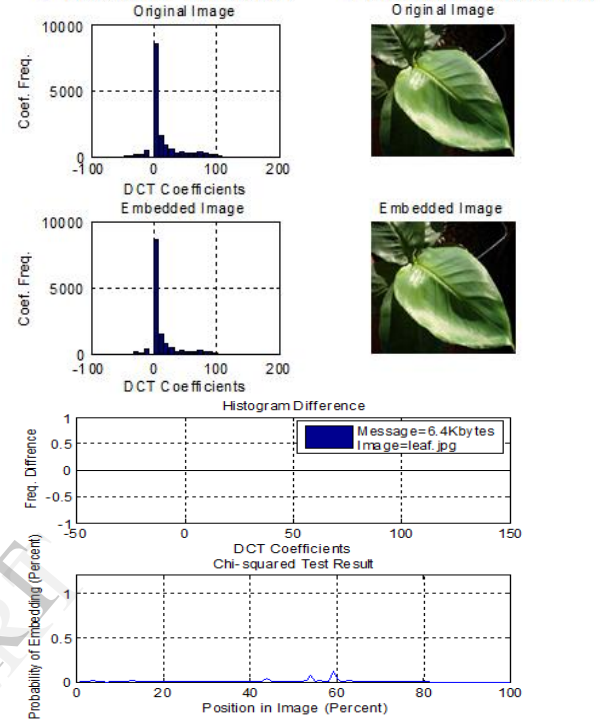
* - Depends on cover image used

Original
- Size : 244KB
- Dimension : 1152 X 864

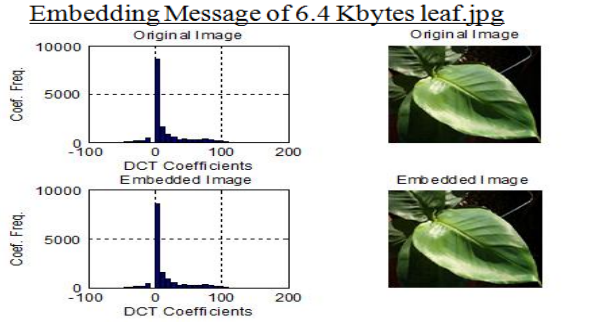
Stego Image
- Size : 246KB
- Dimension : 1152 X 864

Original
- Size : 196KB
- Dimension : 1152 X 864

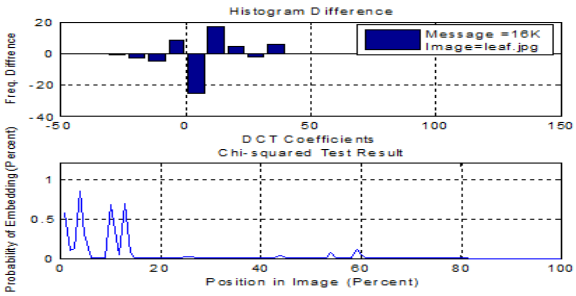
Stego Image
- Size : 208KB
- Dimension : 1152 X 864



Embedding Message of 6.4 Kbytes leaf.jpg



Embedding Message of 16 Kbytes leaf.jpg



V. CONCLUSIONS & FUTURE DEVELOPMENT

The meaning of Steganography is hiding information and the related technologies. There is a principal difference between Steganography and Encryption; however they can meet at some points too. They can be applied together, i.e.

encrypted information can be hidden in addition. To hide something a covering medium is always needed. (Picture, sound track, text or even the structure of a file system, etc.) The covering medium must be redundant; otherwise the hidden information could be detected easily. The technology of hiding should match the nature of the medium. The hidden information should not be lost, if the carrying medium is edited, modified, formatted, re-sized, compressed or printed. That's a difficult task to realize. It's an expectation as well, that the fact of hidden information should be impossible to detect by other then the addressee. On the other hand security services should have methods to detect such information. At least its existence. Realizing a wood trade-off there are different technologies. Nowadays the most popular application of Steganography is hiding copy rights and other commercial information. Such kind of hidden information is known as e-watermark. The e-watermark is not always invisible. There are cases when it is made deliberately strikingly visible. e.g. in case of trial versions of software.

The application is primarily intended to be used to inconspicuously hide confidential and proprietary information by anyone seeking to hide information. This software has an advantage over other information security systems because the hidden text is in the form of image, which is a not obvious text information carrier. Because of its user-friendly interface, the application can also be used by anyone who wants to securely transmit private information. The main advantage of this program for individuals is that they do not have to have any knowledge about steganography or encryption. The visual way to encode the text, plus the visual key makes it easy for average users to navigate within the program. Digital Image Steganography system allows an average user to securely transfer text messages by hiding them in a digital image file. A combination of Steganography and encryption algorithms provides a strong backbone for its security.

REFERENCES

[1] N Provos, "Defending Against Statistical Steganography," Proc 10th USENEX Security Symposium 2012.

[2] N . Provos and P. Honeyman, "Hide and Seek: An introduction to Steganography," IEEE Security & Privacy Journal 2013.

[3] S . Katzenbeisser and Petitcolas , "Information Hiding Techniques for Stenography and Digital watermaking" Artech House, Norwood, MA. 2010 .

[4] L. Reyzen And S. Russell , "More efficient provably secure Steganography" 2013.

[5] S.Lyu and H. Farid , "Steganography using higher order image statistics , " IEEE Trans. Inf. Forens. Secur. 2011.

[6] Venkatraman , s, Abraham , A . & Paprzycki M." Significance of Steganography on Data Security " , Proceedings of the International Conference on Information Technology : Coding and computing , 2012.

[7] Fridrich , J ., Goljan M., and Hogea , D ; New Methodology for Breaking stenographic Techniques for JPEGs. " Electronic Imaging 2009".

[8] http:/ aakash.ece.ucsb.edu./ data hiding / stegdemo.aspx.Ucsb data hiding online demonstration . Released on Mar .09,2010.

[9] Mitsugu Iwanmoto and Hirosuke Yamamoto, "The Optimal n-out-of-n Visual Secret Sharing Scheme for GrayScale Images", IEICE Trans. Fundamentals, vol.E85-A, No.10, October 2012, pp. 2238-2247.

[10] Doron Shaked, Nur Arad, Andrew Fitzhugh, Irwin Sobel, "Color Diffusion: Error Diffusion for Color Halftones", HP Laboratories Israel, May 2009.

[11] Z.Zhou, G.R.Arce, and G.Di Crescenzo, "Halftone Visual Cryptography", IEEE Tans. On Image Processing,vol.15, No.8, August 2011, pp. 2441-2453.

[12] M.Naor and A.Shamir, "Visual Cryptography", in Proceedings of Eurocrypt 1994, lecture notes in computer science, 2004, vol.950, pp. 1-12.

[13] Robert Ulichney, "The void-and-cluster method for dither array generation", IS&T/SPIE Symposium on Electronic Imaging and Science, San Jose, CA, 2003, vol.1913, pp.332-343.

[14] E.R.Verheul and H.C.A. Van Tilborg, "Constructions and properties of k out of n visual secret sharing scheme", Designs, Codes, and Cryptography, vol.1, no.2, 2007, pp.179-196.

[15] Daniel L.Lau, Robert Ulichney, Gonzalo R.Arce, "Fundamental Characteristics of Halftone Textures: Blue-Noise and Green-Noise", Image Systems Laboratory, HP Laboratories Cambridge, March 2013.

[16] C.Yang and C.Laih, "New colored visual secret sharing schemes", Designs, Codes and Cryptography, vol.20, 2010, pp.325-335.

[17] C.Chang, C.Tsai, and T.Chen, "A new scheme for sharing secret color images in computer network", in Proc. of International Conference on Parallel and Distributed Systems, 2010, pp. 21-27.

[18] R.L.Alder, B.P.Kitchens, M.Martens, "The mathematics of halftoning", IBM J. Res. & Dev. Vol.47 No.1, Jan. 2013, pp. 5-15.

[19] R.Lukac, K.N.Plantaniotis, B.Smolka, "A new approach to color image secret sharing", EUSIPCO 2010, pp.1493-1496.