# A New Steganography Based on $\chi^2$ Technic

Zainab Famili[1], Karim Faez[2], and Abbas Fadavi[3]

[1] Department of Electrical, Computer and IT Eng., Azad University, Qazvin, Iran
z_electron590@yahoo.com
[2] Department of Electrical Eng., Amirkabir University of Tech, Tehran, Iran
kfaez@aut.ac.ir
[3] Department of Electrical Eng., Azad University, Garmsar, Iran
abbas_fadavi@yahoo.com

**Abstract.** In this paper, we proposed a new method for Steganography based on deceiving $\chi^2$ algorithm. Since the cover image coefficients and stego image coefficients histograms have sensible difference for purposes of statistical properties, statistical analysis of $\chi^2$-test reveals the existence of hidden messages inside stego image .We are introducing the idea for hiding messages in the cover image. It causes that DCT (Discrete Cosine Transforms) coefficient histogram not having remarkable modification before and after embedding message. As a result, the identifying of hidden message inside an image is impossible for an eavesdropper through $\chi^2$-test. In fact, the proposed method increases the Steganography security against $\chi^2$-test, but the capacity of embedding messages decreases to half.

**Keywords:** Steganography, cover image, stego image, $\chi^2$-test.

## 1 Introduction

Information hiding is a recently developed technique in the information security field and has received significant attention from both industry and academia [1]. Steganography is one technique for hiding information with heavy application in military, diplomatic, and personal area [2]. In the past, people used hidden tattoos or invisible ink to convey Steganographic contents. Today, computer and network technologies provide easy-to-use communication channels for Steganography [3,4]. In today's digital world, invisible ink and paper have been replaced by much more versatile and practical covers for hiding messages inside media such as digital documents, images ,video, and audio files. The digital image is one of the most popular digital mediums for carrying covert messages. There are two main branches Steganography and digital watermarking [1]; the modifications are in spatial domain for the watermarking, and in the frequency domain for the Steganography. The information-hiding process in a Steganographic system begins by identifying a cover medium's redundant bits (bits which can be modified without destroying that medium's integrity) [5].Then this redundant bits are replaced with the data by the hidden messages. In space-hiding systems, one simple method is that of least significant bit Steganography or LSB

E. Bayro-Corrochano and J.-O. Eklundh (Eds.): CIARP 2009, LNCS 5856, pp. 1062–1069, 2009.

embedding. LSB embedding has the merit of simplicity, but suffers from a lack of robustness, and it is easily detectable [6,7]. Steganography goal is to keep hidden message inside an image undetectable, but Steganographic systems for the reason of their invasive nature leave detectable traces in the statistical properties cover medium. Modifying the cover medium changes its statistical properties, so eavesdroppers can detect the distortions in the resulting stego medium's statistical properties. To accommodate a secret message, the original image, also called the cover-image, is slightly modified by the embedding algorithm. As a result, the stego-image is obtained [8,9]. Each Steganographic communication system consists of an embedding algorithm and an extraction algorithm. The system is secure if the stego images do not contain any detectable artifacts due to message embedding. It means the stego images should have the same statistical properties as the cover images [10]. Three different aspects in information-hiding systems contend with each other: capacity, security, and robustness [11]. Capacity refers to the amount of information which can be hidden in the cover medium, security refers to an eavesdropper's inability to detect hidden information, and robustness refers to the amount of modification the stego medium can withstand before an adversary can destroy the hidden information [3]. Steganography strives for high security and capacity, which often entails that the hidden information is fragile. While digital watermarking is mainly used for copyright protection of electronic products [12,13,14] and its primary goal is to achieve a high level of robustness. For Steganography to remain undetected, the unmodified cover medium must be kept secret, because if it is exposed, a comparison between the cover and stego media immediately reveals the changes. The plan of this paper is given by a brief review of JPEG image format in Section2. In Section 3, we present Steganographic systems. After reviewing statistical analysis in Section4, we present out proposed method in Section 5. In Section 6 we summarized the results.

## 2 JPEG Image Format

The format of cover-image is important because it significantly influences the design of the stego system and its security. There are many advantages using images in JPEG format as carrier-image in steganographic applications. JPEG [15] is a popular and widely-used image file format and has become a de facto standard for network image transmission. If we apply JPEG (Joint Photographic Experts Group) images for data hiding; the stego-image will draw less attention of suspect than that with most other formats. JPEG format operates in a DCT transform space and is not affected by visual attacks [16]. The JPEG image format uses a discrete cosine transform (DCT) to transform successive 8×8 pixel blocks of the image into 64 DCT coefficients each. The DCT coefficients $F(u, v)$ of an 8×8 block of image pixels $f(x, y)$ are given by equation (1):

$$F(u,v) = \alpha(u) \cdot \alpha(v) \sum_{x=0}^{7} \sum_{y=0}^{7} f(x,y) \cos[\frac{(2x+1)u\pi}{16}] \cos[\frac{(2y+1)v\pi}{16}]. \qquad (1)$$

Afterwards, the equation (2) quantizes the coefficients:

$$F^Q(u,v) = round\frac{F(u,v)}{Q(u,v)}. \quad (2)$$

$Q(u,v)$, is a 64-element quantization table. (This table is given in reference [18]). We can use the least-significant bits of the quantized DCT coefficients, for which $F^Q(u,v) \neq 0$ and $\neq 1$, are used as redundant bits into which the hidden message is embedded [17]. For more information about JPEG, the reader is referred to [18].

## 3 Steganographic Systems

There are five popular Steganographic algorithms which hide information in JPEG images [19]:

- JSteg: Its embedding algorithm sequentially replaces the least-significant bit of DCT coefficients with the message's data. The algorithm does not require a shared secret; as a result, anyone who knows the Steganographic system can retrieve the message hidden by JSteg [1].
- JSteg-Shell: It compresses the image contents before embedding the data with JSteg. JSteg-Shell uses the stream cipher RC4 (Ron's code #4 or Rivets) for encryption [20].
- JPHide: Before the content is embedded, it is Blowfish [21], encrypted with a user-supplied pass phrase.
- Outguess: Outguess 0.1 is a Steganographic system which improves the encoding step by using a PRNG to select DCT coefficients at random, and Outguess 0.2, which includes the ability to preserve statistical properties [22].
- F5 algorithm: In F5 instead of replacing the least-significant bit of a DCT coefficient with message data, it decrements the absolute value of DCT coefficients in a process called matrix encoding. As a result, there is no coupling of any fixed pair of DCT coefficients [10,23].

## 4 Statistical Analysis

Statistical tests can reveal if an image has been modified by Steganography by testing whether an image's statistical properties deviate from a norm. Westfield and Pfitzmann observe that for a given image, the embedding of encrypted data changes the histogram of its color frequencies [19]. In the following, we clarify their approach and show how it applies to the JPEG format. In their case, the embedding process changes the least signification bits of the colors in an image. The colors are addressed by their indices in the color table. If $n_i$ and $n_i^*$ are the frequencies of the color indices before and after the embedding respectively, then the following relation is likely to hold

$$|n_{2i} - n_{2i+1}| \geq |n_{2i}^* - n_{2i+1}^*|. \quad (3)$$

In other words, the embedding algorithm reduces the frequency difference between adjacent colors. In an encrypted message, zeros and ones are equally distributed. Given uniformly distributed message bits, if $n_{2i} > n_{2i+1}$ , then pixels with color 2i are changed more frequently to color 2i + 1 than the pixels with color 2i + 1 are changed to color 2i. The same is true in the case of the JPEG data format. Instead of measuring the color frequencies, we observe differences in the frequency of the DCT coefficient. Figure (1) displays the histogram before and after a hidden message has been embedded in a JPEG image [3,22]. A $\chi^2$ - test used to determine whether the observed frequency distribution $y_i$ in the image matches a distribution which shows distortion from embedding hidden data [4]. Although we do not know the cover image, we know that the sum of adjacent DCT coefficients remains invariant, which lets us compute the expected distribution $y_i^*$ from the stego image .We then take the arithmetic mean,

$$y_i^* = \frac{n_{2i} + n_{2i+1}}{2}. \tag{4}$$

To determine the expected distribution and compare it against the observed distribution

$$y_i = n_{2i}. \tag{5}$$

The $\chi^2$ value for the difference between the distributions is given as

$$\chi^2 = \sum_{i=1}^{\nu+1} [\frac{(y_i - y_i^*)^2}{y_i^*}]. \tag{6}$$

Where $\nu$ are the degrees of freedom that is, one less than the number of different categories in the histogram [3].

## 5 Proposed Method

Indeed, after doing Steganography inside an image, $\chi^2$ algorithm is operated on the basis of sensible modification which will increase the difference between $\Delta n$, and $\Delta n^*$. $n_{2i}$ is DCT coefficient frequency in $2i$ before of the embedding messages,and $n_{2i}^*$ is DCT coefficient frequency in $2i$ after of the embedding messages.

$$\Delta n = n_{2i} - n_{2i+1}, \quad \Delta n^* = n_{2i}^* - n_{2i+1}^* \ . \tag{7}$$

We propose a new method in this article to endure that the differences between $n_{2i}$ and $n_{2i+1}$ don't have remarkable changes before and after embedding message. In this approach, two sequential DCT coefficients hide only one message bit. In this method, the hiding capacity is reduced to half as compared with JSteg. At first, we arrange DCT coefficients to (2i, 2i+1) groups in terms of $i$ (see table 1).

In study of each group, we will realize that with LSB modification, each members of group, changes its own group and no member of one group conveys to

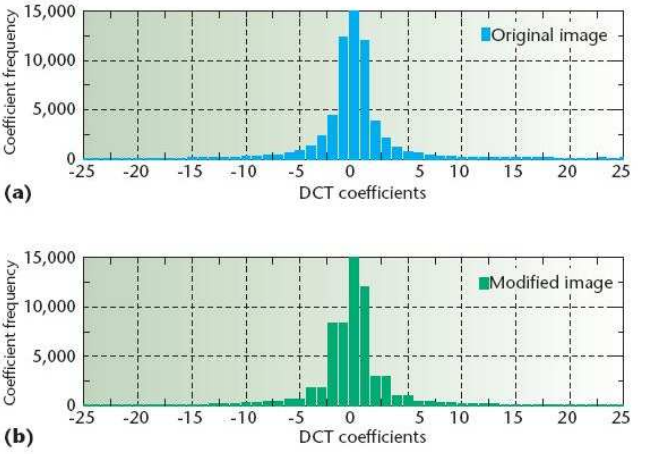


**Fig. 1.** Frequency histograms. Sequential changes to the (a) original and (b) modified image's least-sequential bit of discrete cosine transform coefficients tend to equalize the frequency of adjacent DCT coefficients in the histograms [3].

**Table 1.** Grouping the of two adjacent DCT coefficients according to $i$

| | Group1 | Group2 | Group3 | Group4 | ..... |
|---|---|---|---|---|---|
| 2i | 2 | 4 | 6 | 8 | ..... |
| 2i+1 | 3 | 5 | 7 | 9 | ..... |

**Table 2.** This table indicate the new coefficient replacement according to the message bit to be 0 or 1

| Message Bit | New Coefficient |
|---|---|
| 0 | (2i,2i) or (2i+1,2i) |
| 1 | (2i,2i+1) or (2i+1,2i+1) |

other group. In JSteg method, applying Steganography algorithm and transferring the coefficients in each group changes the difference between $\Delta n$, and $\Delta n^*$in group. So eavesdroppers will be able to recognize the existence of message. To avoid of this subject, we introduce a new approach to minimize the difference between $\Delta n$, $\Delta n^*$. First of all, we obtain in each group, and we examine the coefficients of each group separately two by two. According to the value of two coefficients and the hiding message, $\Delta n$ and $\Delta n^*$ two coefficients are replaced by two new coefficients. With due attention to a message bit to be 0, or 1, we create the table (2) for new coefficients. These two coefficients are chosen optional for hiding message.

For example, we assume $i$ be equal to 4 ( $i = 4$), and two sequential coefficients equal to (9, 8). For the purpose of a hiding message bit with zero value, we can replace (8, 8) or (9, 8) according to table (2). There are three cases in choosing (8, 8) or (9, 8) in the example.

**Table 3.** General convert method

| Old Coefficient1 | Old Coefficient2 | message | $\Delta n$ and $\Delta n^*$ | New Coefficient1 | New Coefficient2 |
|---|---|---|---|---|---|
| 2i | 2i | 0 | $\Delta n > \Delta n^*$ | 2i | 2i |
| 2i | 2i | 0 | $\Delta n = \Delta n^*$ | 2i | 2i |
| 2i | 2i | 0 | $\Delta n < \Delta n^*$ | 2i+1 | 2i |
| 2i | 2i | 1 | $\Delta n > \Delta n^*$ | 2i | 2i+1 |
| 2i | 2i | 1 | $\Delta n = \Delta n^*$ | 2i | 2i+1 |
| 2i | 2i | 1 | $\Delta n < \Delta n^*$ | 2i | 2i+1 |
| 2i | 2i+1 | 0 | $\Delta n > \Delta n^*$ | 2i | 2i |
| 2i | 2i+1 | 0 | $\Delta n = \Delta n^*$ | 2i+1 | 2i |
| 2i | 2i+1 | 0 | $\Delta n < \Delta n^*$ | 2i+1 | 2i |
| 2i | 2i+1 | 1 | $\Delta n > \Delta n^*$ | 2i | 2i+1 |
| 2i | 2i+1 | 1 | $\Delta n = \Delta n^*$ | 2i | 2i+1 |
| 2i | 2i+1 | 1 | $\Delta n < \Delta n^*$ | 2i+1 | 2i+1 |
| 2i+1 | 2i | 0 | $\Delta n > \Delta n^*$ | 2i | 2i |
| 2i+1 | 2i | 0 | $\Delta n = \Delta n^*$ | 2i+1 | 2i |
| 2i+1 | 2i | 0 | $\Delta n < \Delta n^*$ | 2i+1 | 2i |
| 2i+1 | 2i | 1 | $\Delta n > \Delta n^*$ | 2i | 2i+1 |
| 2i+1 | 2i | 1 | $\Delta n = \Delta n^*$ | 2i | 2i+1 |
| 2i+1 | 2i | 1 | $\Delta n < \Delta n^*$ | 2i+1 | 2i+1 |
| 2i+1 | 2i+1 | 0 | $\Delta n > \Delta n^*$ | 2i+1 | 2i |
| 2i+1 | 2i+1 | 0 | $\Delta n = \Delta n^*$ | 2i+1 | 2i |
| 2i+1 | 2i+1 | 0 | $\Delta n < \Delta n^*$ | 2i+1 | 2i |
| 2i+1 | 2i+1 | 1 | $\Delta n > \Delta n^*$ | 2i | 2i+1 |
| 2i+1 | 2i+1 | 1 | $\Delta n = \Delta n^*$ | 2i+1 | 2i+1 |
| 2i+1 | 2i+1 | 1 | $\Delta n < \Delta n^*$ | 2i+1 | 2i+1 |

**Table 4.** Result of running $\chi^2$-test over cover image, stego image (JSteg method), Stego image (proposed method)

| | $\chi^2$for cover image | $\chi^2$for stego image(JSteg method) | $\chi^2$ for stego image(proposed method) |
|---|---|---|---|
| 1 | 140 | 10 | 136 |
| 2 | 329 | 10 | 316 |
| 3 | 611 | 11 | 508 |
| 4 | 360 | 11 | 340 |
| 5 | 511 | 14 | 463 |
| 6 | 227 | 14 | 197 |
| 7 | 245 | 14 | 223 |

1-If $\Delta n > \Delta n^*$ , it means that the frequency of eights is less than its frequency in the cover image. So that, it should be convert one 9 to one 8.There for, we use (8, 8). As a result, one occurrence of the nines is decreased and it is increased to 8.

2 - If $\Delta n$ is $\Delta n^*$, it means that we don't need to change the number of eights and nines, then new coefficients are same (9, 8).

3 - If $\Delta n < \Delta n^*$ ,it means that the frequency of eights is more than its frequency in the cover image. Thus, it should convert one 8 to one 9. Because

in transmitting this message, there isn't any (9, 9) then we use (9, 8).Using this step stop the difference between $\Delta n$ , and $\Delta n^*$ to get higher. We show general convert method (approach) in table (3).

## 6 Conclusion and Result

Our proposed algorithm is applied on 16 different images. We present their results in table (4). In second column, we obtain $\chi^2$ value for cover the image. In third column we calculated $\chi^2$ value for stego image using the JSteg method [3,16]. As we observe a high difference between the second and third columns. As a result, $\chi^2$ -test can recognize the existence of hidden message in stego image. In forth column, $\chi^2$ values was calculated with our new method. Comparison with the second column show the sensible modification and it means that $\chi^2$ values remain nearly fixed and shows deceit of $\chi^2$ algorithm with this method. The existence of message is not revealed in stego image and indeed, this method support the system security against $\chi^2$-test.

## References

1. Zhang, T., Ping, X.: A Fast and Effective Steganalytic Technique against JSteg-like Algorithms. In: Proc. 8th ACM Symp. Applied Computing. ACM Press, New York (2003)
2. Judge, J.C.: Steganography: Past, Present, Future, SANS white paper November 30 (2001), `http://www.Sans.org/rr/papers/`
3. Provos, N., Honeyman, P.: Hide and seek: an introduction to steganography. IEEE Security and Privacy 1(3), 32–44 (2003)
4. Johnson, N.F., Jajodia, S.: Exploring Steganography: Seeing the Unseen. Computer 31(2), 26–34 (1998)
5. Anderson, R.J., Petitcolas, F.A.P.: On the Limits of Steganography. J. Selected Areas in Comm. 16(4), 474–481 (1998)
6. Liu, Q., Sung, A.H., Ribeiro, B., Wei, M., Chen, Z., Xu, J.: Image complexity and feature mining for steganalysis of least significant bit matching steganography. Information Sciences 178(1), 21–36 (2008)
7. Kurak, C., McHugh, J.: A cautionary note on image downgrading. In: Proceedings of the 8th Computer Security Application Conference, pp. 153–159 (1992)
8. Katzenbeisser, S., Petitcolas, F.A.P.: On Defining Security in Steganographic Systems. In: Proceedings of SPIE: Electronic Imaging 2002, Security and Watermarking of Multimedia Contents, San Jose, California, vol. 4675 (2002)
9. Cachin, C.: An Information-Theoretic Model for Steganography. In: Aucsmith, D. (ed.) IH 1998. LNCS, vol. 1525, pp. 306–318. Springer, Heidelberg (1998)
10. Fridrich, J., Goljan, M., Hogea, D.: Steganalysis of JPEG Images: Breaking the F5 Algorithm. In: Petitcolas, F.A.P. (ed.) IH 2002. LNCS, vol. 2578, pp. 310–323. Springer, Heidelberg (2003)
11. Chen, B., Wornell, G.W.: Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding. IEEE Trans. Information Theory 47(4), 1423–1443 (2001)

12. Petitcolas, F.A.P., Anderson, R.J., Kuhn, M.G.: Information Hiding – A Survey. Proceeding of IEEE 87(7), 1062–1078 (1999)
13. Fridrich, J., Goljan, M.: Practical Steganalysis-State of the Art. In: Proc. SPIE Photonics Imaging 2002, Security and Watermarking of Multimedia Contents, vol. 4675, pp. 1–13. SPIE Press (2002)
14. Chen, B., Wornell, G.W.: Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding. IEEE Trans. Information Theory 47(4), 1423–1443 (2001)
15. Wallace, G.W.: The JPEG Still Picture Compression Standard. Communications of the ACM 34(4), 30–44 (1991)
16. Westfeld, A., Pfitzmann, A.: Attacks on Steganographic Systems. In: Pfitzmann, A. (ed.) IH 1999. LNCS, vol. 1768, pp. 61–76. Springer, Heidelberg (2000)
17. Provos, N.: Defending Against Statistical Steganalysis. In: Proc. 10th Usenix Security Symp., pp. 323–335. Usenix Assoc. (2001)
18. Gonzalez, Woods: Digital Image Processing
19. Provos, N., Honeyman, P.: Detecting Steganographic Content on the Internet. In: Proc. 2002 Network and Distributed System Security Symp. Internet Soc. (2002)
20. RSA Data Security. The RC4 Encryption Algorithm (March 1992)
21. Schneier, B.: Description of a New Variable-Length Key, 64-Bit Block Cipher (Blowfish). In: Anderson, R. (ed.) FSE 1993. LNCS, vol. 809, pp. 191–204. Springer, Heidelberg (1994)
22. Provos, N.: Defending Against Statistical Steganalysis. In: Proceedings of the 10th USENIX Security Symposium, August 2001, pp. 323–335 (2001)
23. Westfeld, A.: F5-A Steganographic Algorithm: High Capacity Despite Better Steganalysis. In: Moskowitz, I.S. (ed.) IH 2001. LNCS, vol. 2137, pp. 289–302. Springer, Heidelberg (2001)