

New Steganalysis Method using Glcm and Neural Network

Sedighe Ghanbari
M.Sc of Information technology
Engineering
Shiraz University, Iran

Manije Keshtegary
Ph.D of Computer
Engineering
Dept. of Computer Eng. & IT
Shiraz University of
technology
Shiraz, Iran

Najme ghanbari
M.Sc of Electrical Engineering
, Zahedan Branch ,Islamic Azad
University, Zahedan,,Iran

ABSTRACT

Steganography is the art of hidden writing and secret communication. The goal of Steganography is to hide a message in a multimedia object such as image. Steganalysis is the art and science of detecting such the hidden messages. The Gray level Co-occurrence matrix (GLCM) is the matrix containing information about the relationship between values of adjacent pixel in an image. In this paper, we extract features from GLCM that are different between cover image (image without hidden information) and stego image (image with hidden information).

In the proposed algorithm, first, we use a combined method of steganography based on both location and conversion to hide the information in the original image and call it image-steg1 image. Then, we hide the information in imagesteg1 again and call it image-steg2. Using GLCM matrix properties, we investigate some different features in the GLCM of the original image and stego images. We can extract features that are different between these images. Features are used for training neural network and the classification step was accomplished using four layers Multi Layer Perceptron (MLP) neural network. We tested our algorithm on 800 standard image databases and we detected 80% of stego images. Therefore, our proposed algorithm efficiency is 80%.

Keywords

Steganography, Steganalysis, GLCM, Multi Layer Perceptron Neural Network

1. INTRODUCTION

Steganography is the art and science of hiding information. It uses the digital media such as text, image, audio, video and multimedia as a carrier (cover) for hiding private information in such a way that the third party cannot detect or even notice the presence of the communication [1]. Steganography is different from cryptography. The goal of cryptography is to make data unreadable by a third party, while the goal of steganography is to hide the data from a third party [2, 3].

There are two kinds of image Steganography techniques, spatial-domain and transform domain based methods. Spatial-domain based methods [4] embed messages in the intensity of pixels of images directly. For transform domain based [5, 6],

images are first transformed to another domain (such as frequency domain), and then messages are embedded in the transform coefficients. Steganalysis is the art and science of detecting hidden messages that are embedded using steganography. Discovering of information depends on several factors such as length of message, embedding percent, type of cover media (sound, video, image, and text), format of cover media, method of Steganography, etc. The goal of Steganalysis is to identify suspected packages and determine whether or not they have a payload encoded into them [7].

Because of limitations of human vision, images are the most common media that are used in Steganography, especially in the internet [8]. LSB-based steganography, in which the lowest bit plane of an image is used to convey the secret data, has been used by many Steganographer, because the eye cannot detect the very small perturbations this method introduces into an image and also it is extremely simple to implement [9]. In this paper, we use LSB to hide message in the image. Then, we use statistical properties from the GLCM and four layers neural network to detect the presence of embedded messages in the image blocks.

This paper is organized as follows. In section 2, we will review some steganography and steganalysis algorithms. Section 3 describes our feature selection algorithm. Then, in section 4 we will present specification of our neural network. Section 5 will conclude the paper.

2. RELATED WORK

Pfritzmann and Westfeld [10] proposed a method based on statistical analysis of Pairs of Values (PoVs) that are exchanged during message embedding. This method provides good results when the message placement is known. However, randomly scattered messages can only be reliably detected with this method when the message length becomes comparable with the number of pixels in the image.

Fridrich et al [11] have shown that images stored in the JPEG format are a very poor choice for cover images. This is due to the quantization introduced by JPEG compression. It can serve as a "watermark" or a unique fingerprint, and one can detect even very small modifications of the cover image by inspecting the compatibility of the Stego image with the JPEG format.

Fridrich et al [12] developed a Steganography method for detecting LSB embedding in 24-bit color images which they call it Raw Quick Pairs (RQP) method. This method is based on analyzing close pairs of colors created by LSB embedding.

It works well if the numbers of unique color in the cover image is less than 30 percent of the total pixels.

Zhang et al, [13] proposed a Steganalysis method which is based on a physical quantity derived from the transition coefficients between difference image histograms of an image and its processed version produced by setting all bits in the LSB plane to zero. They claimed that this quantity is a good measure of the weak correlation between successive bit planes and can be used to discriminate stego-images from cover images. They also indicate that there exists a functional relationship between this quantity and the embedded message length.

Sun et al [14] proposed the steganalysis method based on co-occurrence matrix. In this method, the forward difference is calculated in three directions, horizontal, vertical and diagonal, towards adjacent pixels to get three-directional differential images for a natural image. Then the differential images are threshold with a pre-set threshold to remove the redundant information. The co-occurrence matrixes of threshold differential images are used as feature selection for steganalysis.

Also, Kekre et al in [15] proposed a steganalysis method that uses different features between cover and stego images to detect LSB steganography. In the next section, we will extract new features from GLCM and will propose a new steganalysis method

3. FEATURE EXTRACTION FROM GLCM

Co-occurrence matrix of image created based on solidarity and values of pixels in image. Co-occurrence matrix for an image with Dimensions 5*5 is shown in figure 1. Dimension of co-occurrence matrix are the same and equal to number of color levels, by default, this number is 8. In figure 1, image has 8 color levels. Each element (i,j) in the resultant GLCM is simply the sum of the number of times that the pixel with value i occurred in the specified spatial relationship to a pixel with value j in the input image.

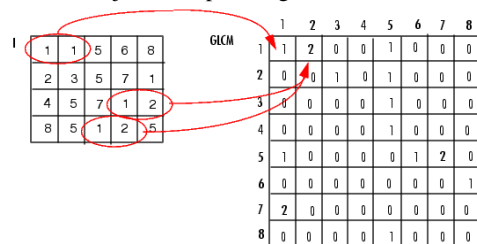


Figure 1 - create co-occurrence matrix

Steganography and embedding information in images changes pixel values and also co-occurrence matrix values. We used these changes to distinguish stego image from cover image.

Since the difference between adjacent pixel values is high, we focused on major diameter elements of co-occurrence matrix.

First, we obtained the GLCM from a cover image, calculated the sum of the major diameter values and named it S-image. Then, we used steganography method based on LSB in DCT coefficients to embed data in the original image and obtained imageSteg1. Then, we obtain GLCM from imageSteg1 and calculate the sum of the major diameter values and named it S-imageSteg1. Finally again, we embed data in imageSteg1 and obtained imageSteg2, get GLCM from imageSteg2,

calculated sum of major diameter values and named it S-imageSteg2.

In the next stage, we calculate difference between S-image and S-imageSteg1 as feature1 and also difference between S-imageSteg1 and S-imageSteg2 as feature2. These values are shown in table 1 for the Lena standard image.



Figure 1- Lena

Table 1 - Result for Lena from GLCM with 8 color levels

S-image	219989
S-imageSteg1	218142
S-imageSteg2	217821
Feature1 (Difference between S-image and S-imageSteg1)	1847
Feature2(Difference between S-imageSteg1 and S-imageSteg2)	321

As shown in Table 1, difference between S-image and S-imageSteg1 is much more than difference between S-imageSteg1 and S-imageSteg2. We created this table for 800 images that are selected from two standard databases, the USC-SIPI [16], the BSD [17] and also internet. By comparing values of table for each images, we found that the difference between a cover image and stego image is much more than difference between two stego images.

LSB mechanism uses the least significant bits to hide data. The least significant bits that are 1 or 0, convert to 0 or 1 if the bit of data and the least significant bit in pixel are not equal, and remains constant if the bit of data and the least significant bit in pixel are the same. So, after using steganography method, we have changes in (i, i), (i, i+1) and (i+1, i) elements of co-occurrence matrix. In this paper, we use symmetric GLCM. So, to improve our features, we replaced "sum of major diameter elements" with difference between "sum of major diameter elements" and "sum of the diameter elements above the major diameter". Also, we increased the color levels from 8 to 256 and extract features for 800 images. Results obtained after these changes for the Lena image and are shown in the table2.

Table 2- obtained result for Lena from GLCM with 256 color levels

S-image	37529
S-imageSteg1	32475
S-imageSteg2	32030
Feature1(Difference between(S-image , S-imageSteg1))	5054
Feature2(Difference between(S-imageSteg1 , S-imageSteg2))	445

Since the spectrum color in image is varied in different points of image, we found that it is better to divided image into smaller pieces, so we divided each image to some images with 64*64 dimensions (an image[256*256] convert to 16 images

of [64*64]) and extracted the features from each 64*64 image. Finally, features extraction for each image is the average of obtained features from each piece. The results of this change for the Lena image is shown in the table3.

Table 3 - obtained results from Lena after fragmentation to 64*64 pieces

Avg (S-image)	571.7813
Avg (S-imageSteg1)	518.2384
Avg (S-imageSteg2)	516.4384
Feature1(Avg (Difference between (S-image , S-imageSteg1)))	53.5429
Feature2(Avg (Difference between (S-imageSteg1 , S-imageSteg2)))	1.8

The algorithm for feature selection of feature1 and feature 2 are shown in figure 2.

Feature1 selection from GLCM

For (each pieces of Image)

Calculate (Sum of element of Major diameter in GLCM in Image) AS **A1-Image**

Calculate (Sum of element of Major diameter in GLCM of ImageSteg1) AS **A1-ImageSteg1**

Calculate (Sum of element of above of Major diameter in GLCM in Image) AS **A2-Image**

Calculate (Sum of element of above of Major diameter in GLCM in ImageSteg1) AS **A2-ImageSteg1**

Calculate (Different ((A1-Image – A2-Image), (A1-ImageSteg1 - A2-ImageSteg1)) As **D (Image, ImageSteg1)**

Feature1 is Avg (D (Image, ImageSteg1))

Feature2 selection from GLCM

For (each pieces of ImageSteg1)

Calculate (Sum of element of Major diameter in GLCM in ImageSteg1) AS **A1-ImageSteg1**

Calculate (Sum of element of Major diameter in GLCM of ImageSteg2) AS **A1-ImageSteg2**

Calculate (Sum of element of above of Major diameter in GLCM in ImageSteg1) AS **A2-ImageSteg1**



Calculate (Sum of element of above of Major diameter in GLCM in ImageSteg2) AS **A2-ImageSteg2**

Calculate (Different ((A1-ImageSteg1 – A2-ImageSteg1), (A1-ImageSteg2 - A2-ImageSteg2)) As **D (ImageSteg1, ImageSteg2)**

Feature2 is Avg (D (ImageSteg1, ImageSteg2))

Figure 2: Features 1 and 2 selection algorithm

Table 4 – Comparison of results from Simian and Pepper

Pepper[16]	Simian[16]
	
If initial image is a cover image	
Different between image and imageSteg1	
39.3143	187.7714
Different between imageSteg1 and imageSteg2	
8.1714	30.7429
If initial image is a stego image	
Different between image and imageSteg1	
19.6571	57.5429
Different between imageSteg1 and imageSteg2	
4.0857	28.8

Various images are created from different color Spectrum and domain of features that extracted is very wide and various. Some values of extracted features are small while some of them are large numbers. The wide and various range, make the separation of cover image from stego image be difficult. In table 4, a sample different range is shown.

In table4, Different between image and imageSteg is 39.3143 for the pepper and Different between imageSteg1 and imageSteg2 is 8.1714. Therefore, Pepper is a cover image. But, Different between image and imageSteg1 is 57.5429 for Simian, And Different between imageSteg1 and imageSteg2 is 28.8. Therefore, Simian is a stego image.

For resolving this problem, we extracted two new features from GLCM as detection factors. In fact, these factors are the ratio of the elements that are on major diameter to the elements that aren't on the major diameter. Detection factor that obtained from the image as feature3 and detection factor that obtained from imagesteg1 as feature4. So, if feature3 is large then we expect that difference between "image and imageSteg1" would be large, and if this factor is a small number, then difference between "image and imageSteg1" would be small. Also if feature4 is large then we expect the difference between "imageSteg1 and imageSteg2" be large, and if this factor is a small number, then difference between "image and imageSteg1" would be small. Detection factors that obtained from simian and pepper are shown in table5.

Table5- detection factor

Pepper[16]	Simian[16]
If initial image is a cover image	
Feature3	
44%	86%
Feature4	
33%	63%
If initial image is a stego image	
Featurer3	
33%	63%
Feature4	
31%	57%

The algorithm for feature selection of feature3 and feature 4 are shown in figure 3.

For (each pieces of Image)

Calculate (Sum of Major diameter elements in GLCM Image) AS **Sum-Image**

For (each pieces of ImageSteg1)

Calculate (Sum of Major diameter elements in GLCM in ImageSteg1) AS **Sum-ImageSteg1**

Then

$C_Image = Avg (Sum-Image)$

$C_ImageSteg1 = Avg (Sum-ImageSteg1)$

Feature 3 is $\%(C_Image / GLCM \text{ of Image})$

Feature 4 is $\%(C_ImageSteg1 / GLCM \text{ of ImageSteg1})$

Figure 3: Features 3 and 4 selection algorithm

4. CLASSIFICATION USING MULTI LAYER PERCEPTION NEURAL NETWORK

In this paper, we used Multi Layer **Perceptron** (MLP) neural network as classifier. A MLP neural network is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate output. The experiments were performed using MatLab programming version 7.1. Our ANN has 4 layers, one input layer, two hidden layers and one output layer. The number of input layer neurons is 4, equal with features. The number of output layer neurons is 1 that determines if image is stego or non-stego (cover). The number of first and second hidden layer neurons selected 2 by try and error method. The neural network epochs are 50. Our neural network training is done with 400 images . We tested our algorithm for 800 images and could determine stego and non-stego (cover) images with 80 percent success

5. CONCLUSION

In this paper, we present a new algorithm for Steganalysis. Using GLCM matrix properties, we investigate some different values in the GLCM of the cover and stego images. We can extract features that are different between these images. Features are used for training neural network. The classification step was accomplished using four layers Multi Layer Perceptron (MLP) neural network. We tested our algorithm on 800 standard image databases and we detected 80% of stego images. Therefore, our proposed algorithm efficiency is 80%.

6. REFERENCES

[1] Dumitrescu S., Wu X. and Wang X., "Detection of LSB steganography via sample pair analysis", IEEE Transactions on Signal Processing, Vol. 51, No. 7, pp. 1995-2007, 2003.

- [2] Shieh C.-S, Huang H.-C, Wang F.-H and Pan J.-S, "Genetic Watermarking Based On Transform-Domain Techniques", Pattern Recognition, Vol. 37, pp: 555-565, 2004.
- [3] Hopper N., "Toward a theory of steganography", Ph.D. Thesis, School of Computer Science, Carnegie Mellon University, July 2004.
- [4] Swanson M., Kobayashi M., and Tewfik A., "Multimedia data embedding and watermarking technologies", Proceedings of the IEEE, Vol. 86, No. 6, pp. 1064-1087, 1998.
- [5] Cox I., Kilian J., Leighton T. and Shamoon T., "Secure spread spectrum watermarking for multimedia", IEEE Transactions on Image Processing, Vol. 6, No. 12, pp. 1673-1687, 1997.
- [6] [6] Provos N., "Defending against statistical Steganalysis", Proceedings of 10th Usenix Security Symposium, pp. 323-335, 2001.
- [7] Westfeld A., "F5 A steganographic algorithm: High capacity despite better Steganalysis", Proceedings of 4th International Information Hiding Workshop, Springer-Verlag, Vol. 2137, pp. 289-302, 2001.
- [8] [8] Mahdavi M., Samavi Sh., Zaker N. and Modarres-Hashemi M., "Steganalysis Method for LSB Replacement Based on Local Gradient of Image Histogram", Iranian Journal of Electrical & Electronic Engineering, 2008.
- [9] Fridrich, J., Goljan, M. and Du,R. "Reliable detection of LSB steganography in grayscale and color Images", Proceeding of ACM, Special Session on Multimedia Security and Watermarking, Otawwa, Canada, pp. 27-30, 2001.
- [10] Westfeld A. and Pfitzmann A., "Attacks on steganographic systems", roceedings of 3rd International Information Hiding Workshop, Springer-Verlag, pp. 61-76, 1999.

- [11] Fridrich J., Goljan M., and Du R., "Steganalysis based on JPEG compatibility", Proceedings of Special Digital Watermarking Data Hiding, pp. 275-280, 2001.
- [12] Fridrich J., Du R. and Meng L., "Steganalysis of LSB encoding in color images", Proceedings of IEEE International Conference on Multimedia, Vol. 3, pp. 1279-1282, 2000.
- [13] Zhang T. and Ping X., "A new approach to reliable detection of LSB steganography in natural images", Elsevier Journal of Signal Processing, Vol. 83, pp. 2085–2093, 2003.
- [14] Sun Z, Hui M, Guan C," Steganalysis Based on Co-occurrence Matrix of Differential Image ", International Conference on Intelligent Information Hiding and Multimedia Signal Processing, PP.1097-1100, 2008.
- [15] Kekre H.B, Athawale A.A, PatkiS.A," Steganalysis of LSB Embedded Images Using Gray Level Co-Occurrence Matrix ", International Journal of Image Processing, (IJIP), Vol.5, PP.711-720, 2011.
- [16] USC-SIPI available at <http://sipi.usc.edu/database/index.php>.
- [17] BSD available at <http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/fg>.