Steganalysis of LSB Embedded Images Using Gray Level Co-Occurrence Matrix

H.B.Kekre hbkekre@yahoo.com
Department of Computer Science
Mukesh Patel School of Technology Management
& Engineering, NMIMS University
Mumbai, India

A.A. Athawale athawalearchana@gmail.com
Department of Computer Science
Mukesh Patel School of Technology Management
& Engineering, NMIMS University
Mumbai, India

S.A.Patki sayli_patki@rediffmail.com
Department of Information Technology
K.J.Somaiya Institute of Engineering
& Information Technology, Mumbai University
Mumbai, India

Abstract

This paper proposes a steganalysis technique for both grayscale and color images. It uses the feature vectors derived from gray level co-occurrence matrix (GLCM) in spatial domain, which is sensitive to data embedding process. This GLCM matrix is derived from an image. Several combinations of diagonal elements of GLCM are considered as features. There is difference between the features of stego and non-stego images and this characteristic is used for steganalysis. Distance measures like Absolute distance and Euclidean distance are used for classification. Experimental results demonstrate that the proposed scheme outperforms the existing steganalysis techniques in attacking LSB steganographic schemes applied to spatial domain.

Keywords: Steganography, Steganalysis, LSB Embedding, GLCM, Distance Measures.

1. INTRODUCTION

Steganography is the art of passing information through apparently innocent files in a manner that the very existence of the message is unknown. The term steganography in Greek literally means, "Covered Writing" [1]. 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. This gives indications that steganography can be used in criminal activities. The messages such as images, videos, sound files, text and other computer files can be hidden inside images or other digital objects which remains invisible to an ordinary observer. By embedding secret data into cover object, a stego object is obtained [2]. Steganalysis is the art of discovering and rendering useless the covert messages, hence breaking steganography. A steganalysis detector attempts to detect the presence or absence of an embedded message when presented with a stego signal. The basic rationale of steganalysis is that there should be differences between an original cover medium and its stego versions. Although the presence of embedded messages is often imperceptible to human eye, it may disturb the statistics of an image. Discovering the difference of some statistical characteristics between the cover and stego media becomes key issue in steganalysis [3].

In other view steganalysis techniques can be broadly divided as, (a). Passive steganalysis: Detect the presence or absence of a secret message in an observed media and (b). Active steganalysis: Extract an approximate version of the secret message or estimate some parameters such as embedding key, message length, etc. using a stego media [4].

2. RELATED WORK

In spatial domain, LSB-based steganography, in which the lowest bit plane of a bitmap image is used to convey the secret data, has long been used by steganographers. This is because the eye cannot detect the very small perturbations it introduces into an image and also because it is extremely simple to implement [5]. The tools used in this group include StegoDos, S – Tools, MandelSteg, Ezstego, Hide and Seek, Steganos [6] etc. LSB steganography methods can be divided into two classes, the LSB substitution [7] and LSB matching [7]. Several techniques for the steganalysis of the images for LSB embedding are present.

Fridrich J and Long M [8] proposed an algorithm for stego only attack. They analyzed the steganographic technique for the LSB embedding in 24-bit color images. The method is based on statistical analysis of the image colors in the RGB cube. Pfitzmann and Westfeld [9] introduced a method based on statistical analysis of Pairs of Values (PoVs) that are exchanged during message embedding. This method, which is the chi-square attack, is quite general and can be applied to many embedding paradigms besides the LSB embedding. Fridrich et al. [10] developed a steganographic method for detecting LSB embedding in 24-bit color images-the Raw Quick Pairs (RQP) method. This method is based on analyzing close pairs of colors created by LSB embedding. It works well if the number of unique colors in the cover image are less than 30 percent that of the total pixels. Sorina et .al [11], have introduced statistical sample pair approach to detect LSB steganography in digital signals such as images and audio. A quantitative steganalysis method to detect hidden information embedded by flipping pixels along the boundaries in binary images is presented in [12]. M. Abolghasemi et .al in [13] have proposed a method for detection of LSB data hiding based on Gray Level Co- Occurrence Matrix (GLCM). In [14] K.B.Raja et .al have proposed a method for LSB steganalysis to detect the embedded message length using pixel pair threshold. S. Mitra et.al [15], have described a detection theory based on statistical analysis of pixel pairs using their RGB components to detect the presence of hidden message in LSB steganography. They have used a fixed threshold method that resulted in poor detection rates. K.B.Raja et al [2], explain a LSB steganalysis method CPAVT based on variable threshold color pair analysis. They have employed "Color Density" as the measure to derive the variable threshold. S.Geetha et al [3], proposed another steganalysis method CCPASST based on variable threshold. Structural Similarity Index Measure is the measure used to obtain variable threshold.

3. GLCM (Gray Level Co-occurrence Matrix) and FEATURE EXTRACTION

A co-occurrence matrix is also referred to as co-occurrence distribution. It is defined over an image to be the distribution of co-occuring values at a given offset. Mathematically, a co-occurrence matrix C defined over an n x m image I, parametrized by an offset (Δx, Δy) is given as [13]:

$$C(i,j)=\sum_{p=1}^{n}\sum_{q=1}^{m}\begin{cases}1, & \text{if } I(p,q)=i \text{ and } I(p+\Delta x,q+\Delta y)=j\\ 0, & \text{otherwise}\end{cases}$$

Four different directions are selected for gray level co-occurrence matrix calculation, i.e. $\theta = 0°, 45°, 90°$ and $135°$ respectively. Thus four gray level co-occurrence matrixes: G1, G2, G3, G4 are obtained from these four directions respectively. From these four matrices the resultant co-occurrence matrix is generated as [13]:

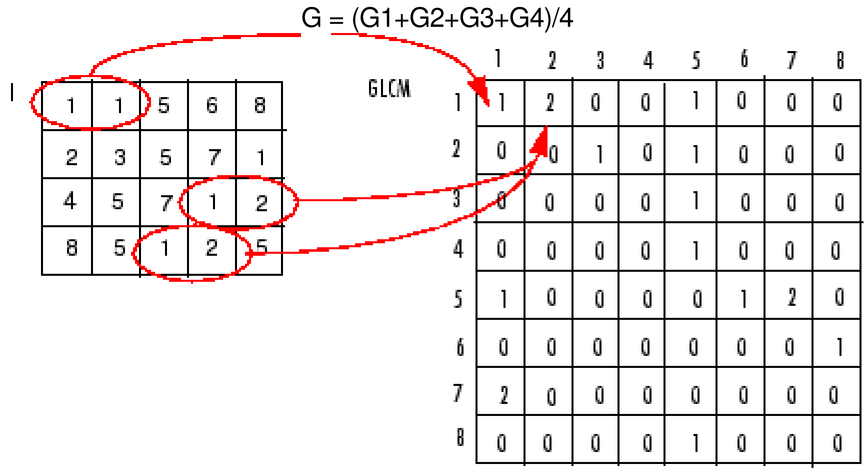

FIGURE 1: Matrix Elements and GLCM Matrix

The gray levels of neighboring pixels in natural images are often correlated, so the gray level co-occurrence matrix of the natural image tends to be diagonally distributed. However after data embedding the high concentration along the main diagonal of the matrix spreads as the high correlation between the pixels in the original image have been reduced as shown in Figure 1 and Figure 2.

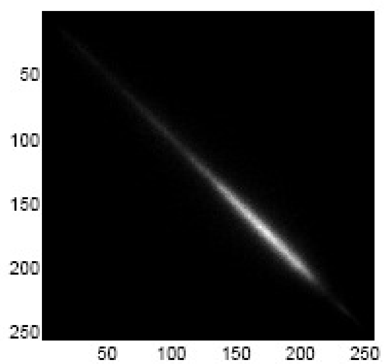

FIGURE 2: GLCM of Cover Image

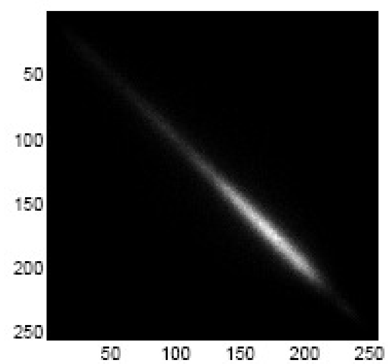

FIGURE 3: GLCM of Stego Image

Considering this asymmetry of the co-occurrence matrix, elements of the main diagonal (d0) and part of the upper (du1, du2) and lower (dl1, dl2) of main diagonal from GLCM are used to construct the feature vector [13]. Table 1 lists the 31 feature vectors used for the purpose of experiments. The 31 feature vectors are formed by considering the powerset of the five diagonals i.e. du2, du1, d0, dl1 and dl2

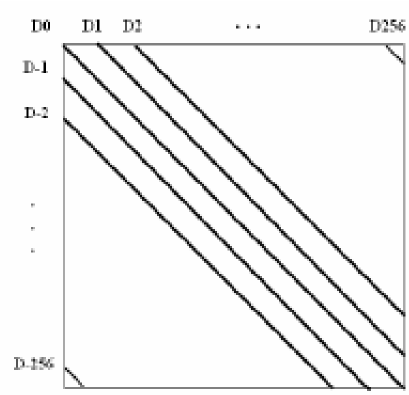

FIGURE 4: Diagonals of Co-occurrence Matrix as Features

F1(dl2,dl1,d0,du1,du2)	F17(d0,du2)
F2(dl1)	F18(dl2,du2)
F3(d0)	F19(dl1,d0,du1,du2)
F4(du1)	F20(dl2,dl1,d0,du1)
F5(du2)	F21(dl2,d0,du2)
F6(dl2,dl1)	F22(dl2,du1)
F7(du1,du2)	F23(dl1,du2)
F8(dl2,dl1,d0)	F24(dl2,dl1,d0,du2)
F9(d0,du1,du2)	F25(dl2,d0,du1,du2)
F10(dl1,d0,du1)	F26(dl2,du1,du2)
F11(dl2)	F27(dl1,du1,du2)
F12(dl1,du1)	F28(dl2,dl1,du1)
F13(dl2,dl1,du1,du2)	F29(dl2,dl1,du2)
F14(dl2,d0)	F30(dl2,d0,du1)
F15(dl1,d0)	F31(dl1,d0,du2)
F16(d0,du1)	

TABLE 1: List of Feature Vectors

Classification of images as stego or cover is done using two different distance measures: Absolute distance and Euclidean distance.

1. Feature vector of the test image is generated
2. Feature vector of the test image is compared with the feature vectors of all the images in the training database. Absolute distance measure and Euclidean distance measure is used to check the closeness of the test image and the training database images.
3. Distance values are sorted in ascending order and minimum of the values is considered
4. A threshold value is set to determine whether the image is stego image.

4. EXPERIMENTAL RESULTS

To test the performance a database of BMP images is used. It consists of 30 color images and 30 grayscale images of size 128 x 128. This database is augmented with the stego versions of these images using LSB steganography. Different payload strengths were used i.e. 25%, 45%, 50%, 90%, 100% of the size of the cover image. So the database consists of 180 color images (cover and stego) and 180 gray scale images (cover and stego). In the experiments, 30 randomly selected images (cover and stego) are taken as training images.

After prolonged testing with database of images, threshold is selected on trial and error basis. Using this threshold, stego images are identified from the database. We have considered four different threshold values 100, 150, 200 and 250. With increase in the threshold the results improve, so we have considered the results with maximum threshold i.e. 250. In case of grayscale images operations such as obtaining the GLCM and extracting the features from the same are performed on the image as a whole. On the other hand color images are first separated in to three planes (Red, Green and Blue), and operations of obtaining the GLCM and extracting the features from the same are performed on each of the planes separately. The maximum of the results of the three planes are taken into consideration.

Table 2 shows the percentage detection for 31 features using Absolute distance in grayscale images and Table 3 shows the percentage detection for 31 features using Absolute distance in color images. Table 4 shows the percentage detection for 31 features using Euclidean distance in grayscale images and Table 5 shows the percentage detection for 31 features using Euclidean distance in color images. Percentage detection indicates out of 30 test stego images, the number of images that are detected as stego.

Feature	Length of Embedding				
	25%	45%	50%	90%	100%
F1	23	37	23	17	23
F2	100	100	97	87	90
F3	70	73	67	47	50
F4	93	97	90	87	87
F5	83	90	80	70	77
F6	77	77	73	53	60
F7	77	77	73	50	60
F8	43	53	47	20	27
F9	40	50	47	20	27
F10	43	50	43	23	30
F11	83	87	77	70	83
F12	83	83	77	60	73
F13	47	53	50	20	27
F14	63	67	60	30	37
F15	63	67	63	33	37
F16	63	67	63	30	37
F17	63	67	57	30	37
F18	70	67	67	47	57
F19	30	40	27	20	27
F20	30	40	23	20	27
F21	37	50	50	20	27
F22	73	77	73	53	60
F23	80	77	73	50	60
F24	30	40	23	20	27
F25	30	40	23	20	27
F26	67	63	60	20	33
F27	67	67	63	27	43
F28	67	67	67	27	43
F29	63	67	60	20	37
F30	43	53	47	20	27
F31	40	50	47	20	27

TABLE 2: Detection accuracy comparison for 31 features: Absolute Distance and Grayscale images

Feature	Length of Embedding				
	25%	45%	50%	90%	100%
F1	33	33	33	20	20
F2	100	100	100	97	97
F3	80	83	87	63	57
F4	100	100	100	97	93
F5	93	90	93	87	87
F6	77	83	90	73	63
F7	77	83	90	73	63
F8	53	53	50	33	33
F9	47	50	50	33	33
F10	50	53	53	33	37
F11	90	90	93	90	87
F12	93	90	97	80	80
F13	50	53	53	43	30
F14	60	67	70	53	40
F15	63	70	83	53	43
F16	63	70	77	53	43
F17	60	70	67	47	40
F18	73	80	77	70	60
F19	40	43	37	27	30
F20	40	43	40	27	27
F21	47	50	53	30	33
F22	73	83	90	73	63
F23	77	83	90	73	63
F24	40	40	37	23	27
F25	37	40	37	23	27
F26	63	70	70	50	40
F27	70	73	80	53	40
F28	70	77	77	53	40
F29	63	70	70	47	40
F30	53	53	47	37	33
F31	47	53	53	33	33

TABLE 3: Detection accuracy comparison for 31 features: Absolute Distance and Color images

Feature	Length of Embedding				
	25%	45%	50%	90%	100%
F1	87	93	93	73	77
F2	100	100	100	93	93
F3	97	100	100	83	83
F4	100	100	100	93	93
F5	100	100	100	97	97
F6	100	100	100	87	93
F7	100	100	100	87	93
F8	87	97	97	80	77
F9	87	97	97	80	77
F10	87	100	100	80	83
F11	100	100	100	97	97
F12	100	100	100	90	93
F13	97	97	97	83	90
F14	87	97	97	80	80
F15	93	100	100	80	83
F16	93	100	100	80	83
F17	87	97	97	80	80
F18	100	100	100	93	97
F19	87	97	97	77	77
F20	87	97	97	73	77
F21	87	97	97	77	77
F22	100	100	100	87	93
F23	100	100	100	87	93
F24	87	97	97	73	77
F25	87	97	97	73	77
F26	97	97	97	83	90
F27	100	100	100	83	93
F28	100	100	100	83	93
F29	97	97	97	83	90
F30	87	97	97	77	77
F31	87	97	97	80	77

TABLE 4: Detection accuracy comparison for 31 features: Euclidean Distance and Grayscale images

Feature	Length of Embedding				
	25%	45%	50%	90%	100%
F1	90	93	97	80	73
F2	100	100	100	97	97
F3	90	97	100	87	83
F4	100	100	100	97	97
F5	100	100	100	100	97
F6	100	100	100	93	93
F7	100	100	100	93	93
F8	90	93	97	83	80
F9	90	93	97	83	80
F10	90	93	97	87	83
F11	100	100	100	100	97
F12	100	100	100	97	93
F13	97	97	97	90	90
F14	90	93	97	83	80
F15	90	93	97	87	83
F16	90	93	97	87	83
F17	90	93	97	83	80
F18	97	97	97	97	97
F19	90	93	97	80	80
F20	90	93	97	80	73
F21	90	93	97	83	80
F22	100	100	100	93	93
F23	100	100	100	93	93
F24	90	93	97	80	73
F25	90	93	97	80	73
F26	97	97	97	93	90
F27	100	97	97	93	93
F28	100	97	97	93	90
F29	97	97	97	93	90
F30	90	93	97	80	80
F31	90	93	97	83	80

TABLE 5: Detection accuracy comparison for 31 features: Euclidean Distance and Color images

5. CONSLUSION

This paper discusses a steganalysis method based on features that are extracted from co-occurrence matrix of an image. Two different distance measures: Absolute and Euclidean are used for the purpose of classification. This scheme outperforms previous works in steganalysis for LSB hiding. It works in case of both grayscale and color images. Euclidean distance gives the best results. It is observed that results obtained using Euclidean distances are better than Absolute distance by around 329% in grayscale images and by 265% in color images. Detection accuracy in case of color images is better than that of grayscale images by around 18% in Absolute distance and almost same in Euclidean distance. Superiority is observed for low embedding rates. The feature vectors which consist of the diagonal d0 exhibit poor results as compared to feature vectors that do not contain the diagonal d0.

6. REFERENCES

1. S. K. Jena, G.V.V. Krishna, “Blind Steganalysis: Estimation of Hidden Message Length”, *International Journal of Computers, Communications and Control*, Vol. II, No. 2: 149-158

2. K. B. Raja, Shankara N, Venugopal K. R., and L.M. Patnaik, “Steganalysis of LSB Embedded Images using Variable Threshold Color Pair Analysis”, *International Journal of Information Processing*, Vol. 1, 2007

3. S.Geetha, Siva S. Sivatha sindhu, R. Renganathan, P. Janaki Raman and Dr. N. Kamraj, “StegoHunter: Steganalysis of LSB Embedded Images based on Stego – Sensitive Threshold Close Color Pair Signature”, *Sixth Indian Conference on Computer Vision, Graphics and Image Processing*, 2008, pp. 281 – 288, 2008

4. M.Abolghasemi, H.aghainia, K.Faez, M.A.Mehrabi, "LSB data hiding detection based on gray level co-occurrence matrix", *International symposium on Telecommunications,*2008, pp. 656-659, 2008

5. J. Fridrich, M. Goljan, r. Du, “Reliable detection of LSB steganography in grayscale and color images”, *Proceeding of ACM, Special Session on Multimedia Security and Watermarking*, Ottawa, Canada, pp. 27-30, 2001.

6. Wayner P., “Disappearing Cryptography”, *Morgan Kaufmann Publisher*, 2002

7. S. Manoharan, “An Empirical Analysis of RS Steganalysis”, *International Conference on Internet Monitoring and Protection,* 2008, pp. 172-177.

8. Westfeld and A. Piftzmann, "Attacks on Steganographic Systems", Proc. 3rd International Workshop on Information Hiding*, LNCS 1768,* pp. 61-76, Springer-Verlag, 2000

9. J. Fridrich, M. Goljan and R. Du, "Detecting LSB steganography in color and gray-scale images", *IEEE Multimedia,* vol. 8, no. 4, pp. 22-28,2001

10. S. Dumitrescu, X. Wu, and Z. Wang, "Detection of LSBsteganography via sample pair analysis", *IEEE Trans. SignalProcessing,* vol. 51, no. 7, pp. 1995-2007, 2003

11. Q. Liu, A. H.Sung, Jianyun Xu and Bernardete M. Riberio, “Image Complexity and Feature Extraction for Steganalysis of LSB Matching Steganography”, *International Conference on Pattern Recognition*, 2006,

12. M.Abolghasemi, H.aghainia, K.Faez, M.A.Mehrabi, "steganalysis of LSB Matching based on co-occurrence matrix and removing most significant planes", *International conference on Intelligent Information Hiding and Multimedia Signal Processing,*2008, pp. 1527- 1530, 2008

13. A. D. Ker, “Steganalysis of LSB Matching in Grayscale Images”,*IEEE Signal Processing Letters*, Vol. 12, No. 6, june 2005

14. M.A.Mehrabi, H. Aghaeinia, M. Abolghasemi, “Steganalysis of LSB Matching Steganography by Removing Most Significant Bit Planes”, *International Symposium on Telecommunications*, 2008, pp. 731-734, 2008

15. F. Huang, B. Li, J. Huang, "Attack LSB Matching Steganography by counting alteration rate of the numnber of neighbourhood gray levels", *International Conferenece on Image Processing*, 2007, pp. 401 – 404, 2007