International Journal of Engineering, Applied and Management Sciences Paradigms, Vol. 12, Issue 01, February 2014
An Indexed and Referred Journal
ISSN (Online): 2320-6608
www.ijeam.com

Statistical Approaches for Digital Image Steganalysis

Amit Kumar[1]
Teaching Associate,
GJUS&T, Hisar ,Haryana
malikamit7@gmail.com

Poonam[2]
Mtech Scholar RIEM,
Rohtak ,Haryana

Gaurav Pruthi[3]
TeachingAssociate,
GCE ,Bilaspur,Gurgoan
pruthi.gaurav@gmail.com

Abstract
In this paper we presented three steganalysis techniques which are developed using statistical properties of an image. When secret data is hidden in an image, the statistical properties like variance, correlation, entropy, (Peak Signal to Noise Ratio) *PSNR*, and (Mean Square Error) *MSE* are changed due to the hidden secret data. We have used these quantitative measures to detect whether any secret data is present in the image or not. Using a statistical approach, we investigated the inherent detectability of several commonly used steganography techniques to check the performance of proposed steganalysis approaches.
Keywords: *Statistical Approaches, Image.*

1. Introduction

1.1 Steganography:

Steganography is the technique to hide secret information within cover objects like images, audio, video and text files. It is art of science that involves communicating secret data in an appropriate multimedia carrier i.e., image, audio, and video files. The word steganography is derived from Greek words which mean Covered Writing. The main purpose of steganography is to hide the message by embedding it into host carrier. It has been widely reported that there has been a surge in the use of steganography for criminal activities and therefore, implementing effective detection techniques is an essential task in digital forensics. Unfortunately, building a single effective detection technique still remains one of the biggest challenges. The proliferation of steganographic tools has created a demand for powerful means to detect hidden data. The host carrier is known as cover object such that it is not detected. The sender embeds a secret message *m* into the cover object *c* to obtain a stego object *s* using an embedding scheme and a secret key *K*. Steganography is different from cryptography where the main goal is to convert the message into a form that is not easily comprehensible or deciphered. The common point between steganography and cryptography is that, the security of underlying methods lie in the secrecy of the embedding and cryptographic keys respectively. We can say that without having access to the secret key, the attacker is not being able to detect the presence of the message in the former or be able to decipher the message in the latter. As in cryptography, we assume the details of the embedding algorithm are known to the attacker [6].

1.2 Steganalysis:

It is the process to decide if an image or other medium contains the hidden message. It is a way of distinguishing between a cover-object and stego-object. A steganalyst may be passive or active.

- A steganalyst is known as passive if his/her aim to detect the presence of a message. He/she may try to find out the embedding method used to hide the messages in the code medium.
- An active steganalyst tries to estimate the hidden message by him/her.

Criteria for Steganalysis:

The main goal of a steganalysis is to identify whether or not a suspected medium is embedded with secret data, in others words, to determine the testing

medium belongs to the cover or stego class. If a certain steganalytic method is used to steganalyze a suspicious medium, there are four possible resultant situations.

- True positive (*TP*): A stego image medium is correctly classified as stego.
- False negative (*FN*): A stego image medium is wrongly classified as cover.
- True negative (*TN*): A cover medium is correctly classified as cover.
- False positive (*FP*): A cover medium is wrongly classified as stego.

At least one of above the condition will occur during the setganalysis process [35].

Steganalysis Techniques [25]: There are two types of steganalysis as given below:

- **Universal Steganalysis Technique:** It attempts to detect the presence of embedded message independent of the embedded algorithm. This is also known as Blind Steganalysis Technique.
- **Embedded Algorithm Based Steganalysis Technique:** This approach takes the advantage of particular algorithmic detail of the embedding algorithm.

1.3 Statistical and Quality Parameters: In this section, we have discussed the image measure parameters which are used in the development steganalysis techniques.

1.3.1 Variance: It is a measure of how far a set of numbers is spread out. If a random variable X has the expected value (mean) $\mu = E[X]$, then the variance of X is given by equation 1:

$$\text{Variance}(X) = E[(X-\mu)^2] \quad \ldots..(1)$$

That is, the variance is the expected value of the squared difference between the variable's realization and the variable's mean.

Correlation: It is degree of linear relationship between two variables is called correlation. Correlation coefficient between two random variables X and Y, usually denoted by $r(X, Y)$ or $r_{xy,}$ is a numerical measure of linear relationship between them and is defined as:

$$r(X,Y) = \frac{Cov(X,Y)}{S.D.(X) \times S.D.(Y)} \quad \ldots (2)$$

where $Cov(X, Y)$ is the covariance between X and Y and $S.D.$ is the standard deviation.

XOR: It is a logical operation that implements an exclusive or, that is, a true output (1) results if one, and only one, of the inputs to the gate is true (1). If both inputs are false (0) and both are true (1), a false output (0) results. Its behavior is summarized in the truth table as given below:

Table 1.1 Truth table of XOR.

INPUT A	B	OUTPUT A XOR B
0	0	0
0	1	1
1	0	1
1	1	0

Mean Square Error (*MSE*): It measures the average of the square of the error. The error is the amount by which value implied by the estimator differ from the quantity to be estimated. The difference occurs because of randomness or because the estimator does not account for information that could produce a more accurate estimate. It is the second moment (about the origin) of the error.
MSE for two images *A* and *B*, each of size *x*× *y*, is defined as:

$$MSE = \sum_{m=1}^{x} \sum_{n=1}^{y} \frac{(A_{mn} - B_{mn})^2}{x \times y} \quad \ldots.(3)$$

where A_{mn} is the pixel of reconstructed image *A* and B_{mn} is the pixel of original image *B*, x and y are the height and width of the images, respectively.

Peak Signal–to-Noise Ratio (*PSNR*): It is used in the comparison between an original image and a coded/decoded image. It is measured in decibels (*dB*).The syntax for *PSNR* is given by

$$PSNR = 10\, log_{10} \frac{(2^B - 1)^2}{MSE} \quad ...(4)$$

where *B* is the bit depth of the image and *MSE* is the mean square error.

2. Literature Survey

Westfeld *et al.* [7] introduced a powerful statistical attack that can be applied to any steganography technique in which a set of Pairs of Values (*PoVs*) are used to detect the presence of secret message. Authors exploited the fact that any steganographic techniques change the frequency of pair of value during message embedding process. This method was effective in detecting Stego-images generated from variety of steganography algorithms. Westfeld [8] analyzed that many steganographic systems are weak against visual and statistical attacks. Systems without these weaknesses offer only a relatively small capacity for steganographic messages. The newly developed algorithm *F5* withstands visual and statistical attacks, yet it still offers a large steganographic capacity. *F5* implements matrix encoding to improve the efficiency of embedding. Thus it reduces the number of necessary changes. *F5* employs permutative straddling to uniformly spread out the changes over the whole steganogram. Avcibas *et al.* [12] proposed *LSB* detection scheme by using binary similarity between the 7th bit plane and 8th bit plane. It is assumed that there is a natural correlation between the bit planes that is disrupted by the *LSB* hiding. This scheme does not auto-calibrate on a per image basis and instead calibrate on a training set of cover and stego images. The scheme works better than a generic steganalysis scheme, but not as well state of the art *LSB* steganalysis. Farid [14] analyzed that techniques for information hiding have become increasingly more sophisticated and widespread. With high-resolution digital images as carriers, detecting hidden messages has become considerably more difficult. This paper describes a new approach to detecting hidden messages in images. The approach uses a wavelet-like decomposition to build high-order statistical models of natural images. A Fisher linear discriminate analysis is then used to discriminate between untouched and adulterated images.

2.1 Problem Statement: Steganalysis is used to detect, identify, and/or extract hidden information within various media sources. When secret data is embedded into cover media like image, audio, video and text files, the statistical properties of the cover file are changed. After going through the literature, we analyzed that there is the need to do statistical analyses of the digital images using some mathematical formulas and to develop approaches to detection the presence of hidden data on the basis of these statistical results.

3. Proposed Approach:

3.1 First Approach:

Variance Based Steganalysis Approach:

a) Read the cover image.
b) Read the suspicious image.
c) Find the variance between both images row wise.

d) Find the variance between both images column wise.
e) Find the difference of the variance between both images.
f) Draw a histogram between variance of both images.
g) Count the number of rows and number of columns in which histogram is not override.
h) Find the percentage of the pixel that has been changed with the total number of pixels.

If percentage is greater than 1, then image is stego

Else image is cover.

3.2 Second Approach:

Correlation Based Steganalysis Approach:

a) Read the cover image.
b) Read the suspicious image.
c) Find out the 8-neighbors of the pixel $P_c(x, y)$ of cover.
d) Find the mean of all neighbors.
e) Find the difference between mean and $P_c(x, y)$.
f) Apply the same process for the cover image.
g) Find the correlation between both images.
h) If correlation is equal to 1 then image is cover

Else image is stego.

3.3 Third Approach:

XOR Based Steganalysis Approach:

a) Read the cover image.
b) Read the suspicious image.
c) Convert the both image into bit stream
d) Find the sum of *XOR* between cover image and suspicious image's *LSB.*
e) If sum is greater than 0 then image is stego. Else image is cover.

Table 1: Three approaches applied on hundred images

Image name	***PSNR* (in *dB*)**	**Variance**		**Correlation**		***XOR***	
		Flag	**Time(sec)**	**Flag**	**Time(sec)**	**Flag**	**Time(sec)**
C1	61.50	1	0.1406	1	0.1969	1	0.8438
C2	60.84	1	0.1406	1	0.2500	1	0.7813
C3	61.58	1	0.1563	1	0.2188	1	0.7813
C4	61.08	1	0.1563	1	0.2344	1	0.8125
C5	52.80	0	0.1406	1	0.2344	1	0.9688
C6	61.10	0	0.1875	1	0.2344	1	0.8125
C7	62.46	1	0.1406	1	0.1875	1	0.8750
C8	60.95	1	0.1406	1	0.1875	1	0.8750
C9	60.92	0	0.1406	0	0.2031	1	0.8750
C10	60.54	1	0.1719	1	0.1875	1	0.8750
C11	60.63	1	0.1094	1	0.1563	1	0.9063
C12	60.77	1	0.1406	1	0.2188	1	0.8594
C13	60.55	0	0.1250	1	0.1563	1	1.2969
C14	60.58	1	0.1406	1	0.2188	1	0.8750

C15	60.35	1	0.1406	1	0.2969	1	0.9375
C16	60.91	0	0.1250	0	0.1875	1	0.9063
C17	61.77	0	0.1250	1	0.1875	1	0.7969
C18	63.63	0	0.1563	1	0.1563	1	0.9375
C19	63.19	0	0.1563	1	0.2344	1	0.9531
C20	62.61	0	0.1563	1	0.2031	1	1.2500
C21	47.66	0	0.1563	1	0.2344	1	0.8594
C22	63.48	0	0.1563	1	0.2344	1	1.1094
C23	53.34	0	0.1563	1	0.1719	1	1.0781
C24	52.95	0	0.1250	1	0.2031	1	0.9219
C25	55.39	0	0.1406	1	0.1719	1	1.3125
C26	53.26	0	0.1406	1	0.2188	1	0.9688
C27	55.39	0	0.1875	1	0.1875	1	0.9219
C28	56.27	0	0.1406	1	0.2344	1	0.9063
C29	64.26	0	0.1406	1	0.3125	1	0.7969
C30	55.49	0	0.1406	1	0.2969	1	0.9219
C31	54.56	0	0.1719	0	0.1406	1	1.4688
C32	55.85	0	0.1250	1	0.1875	1	1.1563
C33	59.37	0	0.1563	0	0.3281	1	0.8906
C34	46.32	1	0.1406	1	0.3281	1	0.9219
C35	46.82	0	0.1250	1	0.2344	1	1.0000
C36	47.33	0	0.1406	1	0.2500	1	0.9844
C37	47.89	1	0.1406	1	0.1875	1	0.9063
C38	54.55	0	0.1406	1	0.2188	1	1.3906
C39	55.34	0	0.1250	1	0.1563	1	1.4844
C40	55.42	0	0.1563	1	0.2344	1	1.2969
C41	51.28	0	0.1563	1	0.2656	1	1.1719
C42	48.22	0	0.1406	1	0.2031	1	0.8594
C43	47.63	0	0.1406	1	0.2031	1	1.0000
C44	50.52	0	0.1563	1	0.1250	1	1.4531
C45	51.27	1	0.1719	1	0.1563	1	1.4688
C46	55.26	0	0.1406	1	0.3281	1	0.8906
C47	51.92	0	0.1563	1	0.3281	1	1.1875
C48	51.92	0	0.1406	1	0.3281	1	0.7813
C49	48.66	1	0.1406	0	0.1875	1	0.9063
C50	47.79	0	0.1563	1	0.2344	1	1.0938
C51	47.79	0	0.1563	1	0.1563	1	1.4844
C52	46.00	0	0.1563	1	0.3438	1	1.2969
C53	61.58	0	0.1875	1	0.2188	1	1.0781
C54	60.87	0	0.1406	1	0.3125	1	0.9531
C55	60.87	0	0.1406	1	0.2188	1	1.2031
C56	60.82	0	0.1719	1	0.1250	1	1.2656
C57	61.02	0	0.1563	1	0.1875	1	1.2031
C58	60.88	0	0.1250	1	0.1563	1	1.4688
C59	60.95	0	0.1406	1	0.1719	1	1.4844
C60	60.92	0	0.1406	0	0.1719	1	1.1094

C61	60.79	1	0.1719	1	0.3125	1	0.8281
C62	60.88	0	0.1563	1	0.2656	1	0.8906
C63	60.71	0	0.1563	1	0.2188	1	1.1563
C64	60.97	0	0.1563	1	0.2656	1	0.9063
C65	61.69	0	0.1719	1	0.2656	1	0.7813
C66	63.14	0	0.1406	1	0.2656	1	0.9688
C67	60.79	0	0.1563	1	0.2969	1	0.7656
C68	61.71	0	0.1406	1	0.2969	1	0.7813
C69	61.80	0	0.1563	1	0.3594	1	0.7656
C70	60.73	0	0.1406	1	0.2500	1	1.0000
C71	60.54	1	0.1563	1	0.2813	1	0.8594
C72	61.66	0	0.1719	1	0.2813	1	0.8125
C73	61.67	0	0.1406	1	0.3125	1	1.3281
C74	61.75	0	0.1563	1	0.3281	1	0.9219
C75	61.98	0	0.15563	1	0.2344	1	1.0625
C76	61.98	0	0.1406	1	0.3281	1	1.1094
C77	61.09	0	0.1406	1	0.2344	1	1.0938
C78	61.90	0	0.1563	1	0.2344	1	1.1094
C79	48.94	1	0.1563	1	0.2344	1	1.0938
C80	47.06	0	0.1250	1	0.3281	1	0.9844
C81	48.93	0	0.1406	1	0.2344	1	1.0156
C82	47.83	0	0.1406	1	0.2656	1	1.0625
C83	48.87	0	0.1563	1	0.2344	1	1.0469
C84	49.10	0	0.1406	1	0.2656	1	1.3906
C85	50.61	0	0.1406	1	0.3438	1	1.4844
C86	49.82	0	0.1406	1	0.2500	1	0.9688
C87	49.17	0	0.1904	1	0.2969	1	0.9844
C88	49.34	0	0.1875	1	0.2500	1	1.0156
C89	49.26	0	0.1563	1	0.2969	1	1.0000
C90	49.12	0	0.1719	1	0.2656	1	1.3594
C91	49.26	0	0.1406	1	0.4844	1	1.3125
C92	59.37	0	0.1563	0	0.3594	1	1.0313
C93	56.52	0	0.1250	1	0.4688	1	0.7969
C94	55.82	0	0.1563	1	0.4844	1	0.8906
C95	54.77	0	0.1719	1	0.3750	1	0.8906
C96	56.86	0	0.1406	1	0.1969	1	0.9844
C97	55.34	0	0.1563	1	0.3438	1	0.9844
C98	55.49	0	0.1250	1	0.3281	1	0.9844
C99	54.00	0	0.1563	1	0.3594	1	0.8750
C100	55.37	0	0.1406	1	0.3594	1	0.7813

Table 1 shows the hundred images and their PSNR with corresponding Stego. Flag 0 shows the image has no data in the image and flag 1 shows there is some hidden data in the image. The First approach take minimum time (0.148937 seconds on an average) but show only 18 images has hidden data, second approach takes more time (0.251607 seconds on an average) rather first approach but shows 93 images have secrete data, while the third approach takes maximum time (1.029547 seconds on an average) among three but shows all the images have hidden data. The proliferation of steganographic

tools has created a demand for powerful means to detect hidden data.

4. Conclusion

The primary focus of this paper is to develop the steganalysis techniques using statistical properties of an image. Using a statistical approach, we investigated the inherent detectability of several commonly used data hiding techniques. Though our approaches are providing satisfying results in the study of steganalysis, we acknowledge that still there are some problems yet to be solved. We conclude with a look to future research directions, which we believe will advance the study of stealthy transmission of, and interception of, hidden data in the images. For future work, the statistical properties will be further investigated in order to achieve a blind steganalysis of digital images and videos of different formats.

References

[1] R. C. Gonzalez, R. E. Woods and S. L. Eddins, "Digital Image Processing Using MATLAB," 5th Edition, Pearson, 2009.

[2] G. J. Simmons, "The Prisoners' Problem and The Subliminal Channel," In Proceedings of Advances in Cryptology, pp. 51-67, 1983.

[3] T. M. Cover, and J. A. Thomas, "Elements of Information Theory," Wiley, 1991.

[4] E.T. Lin and E. T. Delp, "A Review of Data Hiding in Digital Images," In Proceedings of the Image Processing, Image Quality, Image Capture Systems Conference, pp. 274-278, 1999.

[5] http://www.freefoto.com

[6] L. Marvel, C. G. Boncelet Jr., C. T. Retter, "Spread Spectrum Image Steganography," In Proceedings of IEEE Transactions on Image Processing, Vol. 8, No. 8, pp. 1075-1083, 1999.

[7] A. Westfeld and A. Pfitzmann, "Attacks on Steganographic Systems," In Proceedings of Lecture Notes in Computer Science, Vol. 1768, pp. 61−75., 2000

[8] A. Westfeld, "High Capacity Despite Better Steganalysis (F5- S Steganographic Algorithm)," In Proceedings of Lecture Notes in Computer Science: 4th International Workshop on Information Hiding, Vol. 2137, pp. 289-302, 2001.

[9] B. Chen, and G. Wornell, "Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Theory and Information Embedding," In Proceedings of IEEE Transactions on Information Theory, Vol. 47, No. 4, pp. 1423-1443, 2001.

[10] N. Provos, "Defending Against Statistical Steganalysis," In Proceedings of 10th USENIX Security Synposium, 2001.

[11] J. Fridrich, M. Goljan, and R. Du. "Detecting LSB Steganography in Color and Grayscale Images," In Proceedings of Magazine of IEEE Multimedia, Special Issue on Security, Vol. 8, pp. 22-28, 2001.

[12] I. Avcibas, N. Menon, and B. Sankur, "Image Steganalysis with Binary Similarity Measures," In Proceedings of ICIP, 2002.

[13] S. Lyu, and H. Farid, "Detecting Hidden Messages Using Higher Order Statistics and Support Vector Machines," In Proceedins of Lecture Notes in Computer Science: 5th International Workshop on Information Hiding, Vol. 2578, 2002.

[14] H. Farid, "Detecting Hidden Messages Using Higher Order Statistical Models," In Proceedings of the IEEE International Conference on Image Processing, Vol. 2, pp. 905-908, 2002.

[15] I. Avcibas, N. Menon, and B. Sankur, "Steganalysis Using Image Quality Metrics", In Proceedings of IEEE Transactions on Image Processing, Vol. 12, No. 2, pp. 221-229, 2003.

[16] S. Dumitrescu, X. Wu, and Z. Wang, "Detection of LSB Steganography via Sample Pair Analysis," In Proceedings of IEEE Transactions on Signal Processing, Vol. 51, No. 7, pp. 1995-2007, 2003.

[17] J. J. Harmsen, and W.A Pearlman,

"Steganalysis of Additive Noise Modelable Information Hiding," In Proceedings of IST/SPIE's 15th Annual Symposium on Electronic Imaging Science and Technology, 2003.

[18] P. Sallee, "Model Based Methods for Steganography," In Proceedings of Seconds International Workshop on Digital Watermaking, pp. 154-167, 2003.

[19] M. U. Celik, G. Sharma and A. Tekalp, " Universal Image Steganalysis Using Rate Distortion Curves," In Proceedings of IST/SPIE's 16th Annual Symposium on Electronics Imaging Science and Technology, 2004.

[20] S. Lyu, and H. Farid, "Steganalysis Using Color Wavelet Statistics and One Class Support Vector Machines," In Proceedings of IST/SPIE's 16th Annual Symposium on Electronic Imaging Science and Technology, 2004.

[21] Y. Wang and P. Moulin, "Steganalysis of Block Structured Stegotext," In Proceedings of IST/SPIE's 16th Annual Symposium on Electronic Imaging Science and Technology, 2004.

[22] P. Moulin, and Y. Wang, "New Results on Steganographic Capacity," In Proceedings of Conference on Information Sciences and Systems, 2004.

[23] M. Kharrazi, H. T. Sencar, and N. Menon, "Benchmarking Steganographic and Steganalysis Techniques," In Proceedings of IST/SPIE's 17th Annual Symposium on Electronic Imaging Science and Technology, 2005.

[24] P. Sallee, "Model Based Methods for Steganography and Steganalysis", In Proceedings of International Journal of Image and Graphics, Vol. 5, No. 1, pp. 167-190, 2005.

[25] H. Farid, S. Lyu, "Steganalysis Using Higher-Order Image Statistics," In Proceedings of IEEE Transactions on Information Forensics and Security, Vol. 1, No. 1, pp. 1-10, 2006.

[26] K. Sullivan, U. Madhow, S. Chandrasekran, B.S. Manjunath, "Steganalysis for Markov Cover Data With Applications to Images," In Proceedings of IEEE Transactions on Information Forensics and Security, Vol. 1, No. 2, pp. 275-287, 2006.

[27] A. D. Ker, "Steganalysis of Embedding in Two Least-Significant Bits," In Proceedings of IEEE Transactions on Information Forensics and Security, Vol. 2, No. 1, pp. 46-54, 2007.

[28] H. Cai, S. S. Agaian, "JPEG Steganalysis Using Color Correlation and Training On Clean Images Only," In Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, pp. 3710-3713, 2008.

[29] A. G. H. Chamorro, A. E. Trujillo, J. L. Hernandez, M. N. Miyatake, H. P. Meana, "A Methodology of Steganaysis for Images," In Proceedings of International Conference on Electrical, Communications, and Computers", pp. 102-106, 2009.

[30] A. G. H. Chamorro, M. N. Miyatake, "A New Methodology of Image Steganalysis Including for JPEG Steganography," In Proceedings of Electronics, Robotics, and Automotive Mechanics Conference, pp. 434-438, 2010.

[31] S. Bera, M. Sharma, "Steganalysis of Real Time Image by Statistical Attacks," In Proceedings of International Journal of Engineering Science and Technology, Vol. 2, No. 9, 2010.

[32] V. Singhal, D. Yadav, D. K. Bandil, "Steganography and Steganalysis: Review," In Proceedings of International Journal of Electronics and Computer Science Engineering, Vol. 1, pp. 399-405, 2011.

[33] H. B. Kekre, A. A. Athawale, S. A. Patki, "Steganalysis of LSB Embedded Images Using Gray Level Co-Occurrence Matrix," In Proceedings of International Journal of Image Processing, Vol. 5, No. 1, pp. 36-45, 2011.

[34] A. S. Hashemi, M. M. Ghazi, S.

Ghaemmaghami, H. S. Zadeh, "Universal Steganalysis Based on Local Prediction Error in Wavelet Domain," In Proceedings of Seventh International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 165-168, 2011.

[35] B. Li, J. He, J. Huang, Y. Q. Shi, "A Survey on Image Steganography and Steganalysis," In proceedings of Journal of Information Hiding and Multimedia Signal Processing, Vol. 2, No. 2, 2011.