IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 10, Issue.4, April 2022 E-ISSN: 2347-2693

A Comparative Study of LSB based Statistical Steganalysis and Gray Level Co-Occurrence Matrix based Blind Image Steganalysis

Bibek Ranjan Ghosh

Department of Computer Science, Ramakrishna Mission Residential College, Narendrapur, Kolkata-700103, India

Author's Mail Id: bibekghosh2019@gmail.com, Tel.: 6291992963

DOI: https://doi.org/10.26438/ijcse/v10i4.15 | Available online at: www.ijcseonline.org

Received: 20/Mar/2022, Accepted: 02/Apr/2022, Published: 30/Apr/2022

Abstract- Image steganography is used as a covert communication technique which hides secret data in cover image intelligently so that it is visually imperceptible. This is often used by individual or organization with bad intent to harm people, organization or society. Steganalysis technique is used to break these systems to extract the secret information, reveal such covert communication and thwart imminent threat. Steganalytic techniques can be broadly classified as targeted or blind. In the former the knowledge of steganographic system used should be known and the latter adopts a more general approach where no knowledge of the process used to hide data is required. This paper studies some well-established statistical methods of targeted steganalysis and gray level co-occurrence matrix based blind steganalysis and compare their performances.

Keywords- RS, Sample pair analysis, Chi-squared, Gray level co-occurrence matrix

I. INTRODUCTION

Image steganography and steganalysis try to defeat each other [1][2]. The former hides data and creates stego image from cover whereas latter does the opposite i.e. uncover secret data in cover image. Statistically hidden data can be perceived as noise addition to image signal resulting in alteration of statistical properties of the cover image. A good steganalytic system tries is to discover these patterns to distinguish between clean and stego image. So this job is more difficult and complex than steganography. In this cyber age it is quite essential for organization or government agencies to use it in forensics, tracking terrorism and criminal activities [3]. Steganalysis may be either targeted or blind. The former utilizes the knowledge of the steganographic algorithms used to hide data and the latter is independent of such knowledge. Numerous methods exist in each of the categories [4]. Among these methods statistical image steganalysis techniques such Chi-square (χ^2) attack, RS analysis, Sample Pair Analysis (SPA) and Least Square are widely accepted [5][6][7][8]. Blind image steganalysis takes a more general learning based approach to distinguish clean and stego image and dives deeper in stego image to discover patterns or features that are generated as a result of embedding secret data. Steganalytic features may be image quality measures (IQM), correlation based, moment based etc. Gray level (gray level) co-occurrence matrix (CM) represents spatial correlation among neighboring pixel pairs in an image. Haralick et al. introduced GLCM and identified fourteen texture features of image and further focused on seven most important feature for image classification[9][10]. Later Sullivan et al. utilizes GLCM and introduced Markov chain representation of GLCM [11]. The low dimensional representation of GLCM is also predominant [12]. Our study is distributed as follows. Section-II studies three well accepted statistical steganalysis techniques, Section-III discusses related work of these algorithms, Section-IV deals with GLCM based blind techniques, Section-V discussed related work of Section IV, Section VI compares their performances and Section-VII concludes the study with future direction.

II. STATISTICAL STEGANALYSIS

In this section the three widely accepted statistical steganalysis techniques are studied.

A) Chi-square (χ^2) attack

In 1999, A. Westfeld and A. Pfitzmann et al. coined this method to detect least significant bit (LSB) substitution [5]. Considering LSB substitution of an 8-bit grayscale image C, an even valued pixel 2z in C (where 0<=z<=127) may change to odd pixel 2z+1 in stego image (S) or remain as 2z . In the same way an odd pixel 2z+1 in C may change to even (2z) or may remain odd. This leads to statistical patterns for certain pairs of value (PoV) in S histogram. Histogram of S shows the co-occurrence of both PoV i.e. (2z, 2z+1) and (2z+1,2z) becomes closer as to bit flipping increases. Let the frequencies of such 2z and 2z+1 in C are, efre(z)=frequency(2z) and ofre(z)=frequency(2z+1) respectively, then average is

$$afre(z) = \frac{efre(z) + ofre(z)}{2} \quad (1)$$

The Chi-squared (χ^2) statistic computed with (n-1) degrees of freedom is,

$$\chi^{2}_{(n-1)} = \sum_{z=0}^{127} \frac{(efre(z) - ofre(z))^{2}}{afre(z)} \quad (2)$$

Message embedding probability can be derived by Equation-3.

$$p = 1 - \frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} \int_{0}^{\chi^{2}(n-1)} e^{-y/2} y^{\frac{n-1}{2}-1} dy \quad (3)$$

B) Regular Singular (RS) Analysis

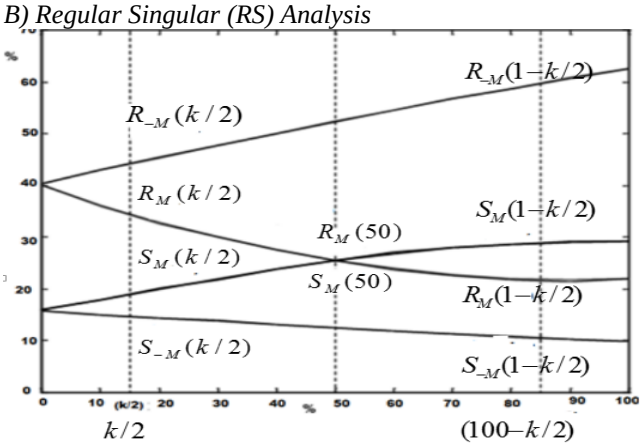

Figure. 1 RS analysis curve.

In 2001, Fridrich et al. introduced RS analysis [6]. At the outset the image under study is partitioned into disjoint pixel groups G. The group smoothness is computed by a discriminating function D(G). Higher D(G) represent nosier group. Another flipping function F can simulate the bit flipping in LSB substitution. A group is termed as regular (R) if D(F(G))>D(G), singular(S) if D(F(G))<D(G) and unusable(U) if D(F(G))=D(G). A mask M may be used to get independent flipping in a group. Representing R_M and S_M as percentage of number of R and S groups respectively using mask M, then the null hypothesis for RS analysis in C is,

$$R_M \approx R_{-M}, S_M \approx S_{-M} \quad (4)$$

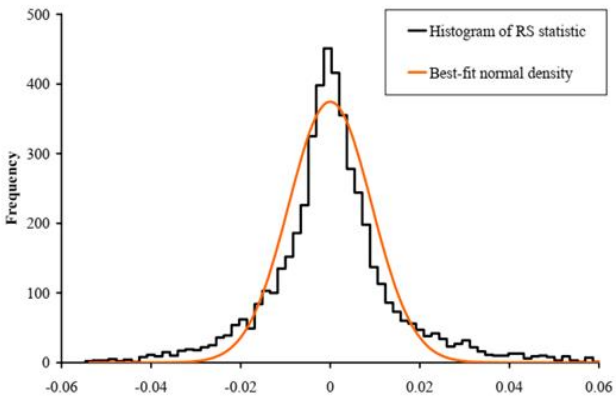

Figure. 2 Histogram of RS statistic.

Fridrich et al. tests equation (4) using numerous images and results is shown in Figure.1. If embedding rate (in pixel percent) is k then percentage of LSB flipped is k. (a) Difference of R_M and S_M diminishes with the rise of k and reaches zero when 50% pixels are flipped. On the contrary R_{-M} and S_{-M} difference increases as k rises. (b) The intersection point of (R_M , R_{-M}) and (S_M , S_{-M}) has same x axis value. The secret message length can be computed by observing (a) and (b). The accuracy of this estimation depends on starting cover bias, cover image noise and distribution of embedding locations. Histogram of RS statistic as experimented by Andrew D. Ker et al. on 5000 jpeg images is found as in Figure-2 [13]. It is not Gaussian and its estimated kurtosis is nearly 20.

C) Sample Pair Analysis (SPA)

In 2003, Dumitrisecu et al. pinned this method which is based on finite state machine (FSM) model [7]. The internal states of FSM are trace multi sets which is a multi-set of sample pair pixels. LSB flipping give rise to state transition between states represented as trace multi-sets. A quadratic equation derived from FSM shown in Figure.3 cab be used to estimate the hidden message length.

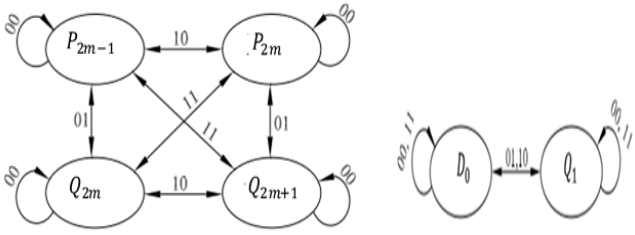

Figure. 3 a) The FSM whose states are trace multi sets of Cm b) FSM for C_o.

For a digital signal DS={v1,v2,..vN} of N samples, a possible sample pair is (vi,vj) where 1<=i,j<=N. Multi-set MS is all such pairs from DS. A subset D_n from MS is created where pairs differs by n where 0<=n<=2^b-1 where b= sample encoding bit length. Similarly C_m is created from MS in which pairs differ by m in (b-1) MSB bits. D_{2m+1} can be further partitioned into P_{2m+1} and Q_{2m+1} where even and odd pair member are larger respectively. D_{2m} creates similar partitions. C_m can be further partitioned into P_{2m}, Q_{2m}, P_{2m}-1 and Q_{2m+1}. Again, C0 is creates partition D_0 and Q_1. If 0 represent no bit change and 1 represent 1 bit change in LSB substitution then 4 possibilities are PI={00,01,10,11}. The probability of each such possibilities can be calculated in terms of k, the proportion of secret message length and length of DS. The quadratic equation derived from the FSM in Figure.3 can estimate the hidden message length. Andrew D. Ker has shown that error distribution in SPA is not normal has kurtosis 15.53 [13].

III. RELATED WORK ON STATISTICAL STAGNALYSIS

This section studies related work on targeted statistical steganalysis. Zhang et al. used LSB steganography to join

adjacent pixel pair LSB bits with modulo 2 addition [14]. RS and Chi-square on stego image for 50 images show poor result. Luo et al. used chaotic map and dynamic compensation with SPA for LSB substitution detection on 2000 images [15]. Highest mean p value is 0.11% which is less than threshold 1.8%. S. Manoharan et al. used 25 color images with 24 bit depth and 16 synthetic logo for LSB substitution and matching using with mask= [0 1 1 0] resulting in 6% and %5 detection accuracy respectively [16]. RS performs superior with LSB substitution than matching. Qian, T et al. verified RS, Histogram Characteristic Function (HCF) and Raw Quick Pair (RQP) technique [17]. ROC curve shows histogram of HCF is superior to basic one for grayscale and colour images. RS and RQP detect LSB substitution at low (10%) payload, but HCF perform better with full load in LSB substitution. N. Prokhozhev et al. evaluates performance on grayscale images with RS and SPA along with difference image histogram (DIH) and other [18]. Performance remain same for algorithms w.r.t. ROC but degrades above 5% payload. R. A. Solodukha et al. improved RS analysis with variable group size (VGS) for BMP images [19]. The method is tested with photorealistic images and results shows improvement of RS-VGS. N. Mewalal et al used a mixed model for LSB in png file with many algorithms [20]. Detection accuracy for long size data in RS, SPA and Chi-square, are 54%, 55% and 99% respectively. B. Lin et al. introduced chi-square fit function detection method for modified PVD (MPVD) on difference images [21]. DIH show step effect which can be reduced by dynamic interval. Using BOSS database detection accuracy observed 90.3% for 10% payload and 99.8% for full payload.

IV. GRAY LEVEL CO-OCCURRENCE MATRIX BASED BLIND STEGANALYSIS

A) Gray Level Co-occurrence Matrix (GLCM)

In 1973 Haralick et al. introduced GLCM to describe texture feature of an image for satellite image analysis [9]. Co-occurrence matrix CM of grayscale image IM, calculates the frequency of occurrence of a pixel pair with an offset (Δx,Δy) in IM having Δx= difference of vertical pixel position and Δy=difference of horizontal pixel position. For an 8 bit grayscale image the C has size 256x256. If IM has size M x N then, for a pair pixels of gray levels (i,j), co-occurrence matrix CM of IM at offset (Δx,Δy) is defined in Equation-5.

$$CM_{\Delta x,\Delta y}(i,j) = \sum_{x=1}^{M}\sum_{y=1}^{N} \begin{cases} 1, \text{if}(IM(x,y) = i \,\&\, IM(x+\Delta x, y+\Delta y) = j \\ 0, \text{otherwise} \end{cases} \quad (5)$$

Alternatively offset of a pixel pair can be expressed as (d,θ) where d= number of pixels between pixel pairs and θ =angle (e.g. 0, 45, 90, 135 etc.) between them. Figure-4 represent CM of a grayscale image having intensity values {0, 1, 2, 3}

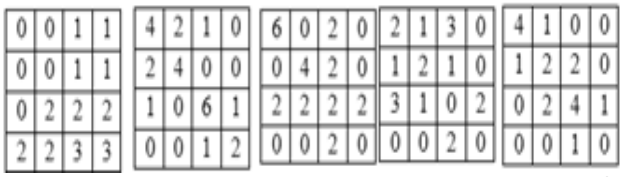

Figure-4. a) Image with 4 intensity levels b) CM for 0^o (horizontal) c) CM for 90^o (vertical) d) CM for 135^o (Right down) (e) CM for 45^o (Left down).

B) Blind Steganalysis based on GLCM

The Haralick features can be used as input to machine learning model like support vector machine (SVM) or neural network to classify stego or clean images. In natural images the nonzero values of GLCM are clustered around the diagonal. But the embedding spread the diagonal to some extent. This is a major cause to selection of diagonal and some off diagonal elements as features. Even low dimensional features are extracted by using principal component analysis [12]. Related techniques are discussed in next section.

V. RELATED WORK ON GLCM BASED STEGANALYSIS

This section studies some GLCM based steganalytic methods. H.B. Kekre, et al. implemented LSB steganography using a ratio R and R' of close color pair and unique color in an image before and after embedding [22]. For stego image R=R' and R'≥R for cover. Compute percentage change $m=(R-R')*100/R$ and threshold t=m/SSIM. If m<t then image is categorized as stego else cover with 83% accuracy. A.A. Athawale et al. computed average GLCM for four directions {0, 45, 90,135} for LSB steganalysis [23].31 features are generated taking five central GLCM diagonals.Accuracy in colour images is better than grayscale by18% using Manhattan distance. Z. Xia et al. observed smoothness of multi-order histogram post-LSB matching [24]. The GLCM used to extract features and SVM used for classification with 0.1 to 1 bits per pixel payload. The detection reliability ranges from 0.5621 to 0.8906 for BOSS dataset. A. Anjum et al. used neighbor pixel predictor (NPP) for edge and boundary pixels in weighted steganography (WS) method[25].A better detection rate is 0.47491 at 0.5 bpp payload with BOSS database. O. Juarez-Sandoval et al. used LSB matching with full payload using 12 feature derived from the GLCM of the difference image achieved 96.25 % and 90.96% detection accuracy with 100% payload for BOWS-2 and UCID dataset respectively[26]. S. Ziwen et al. computes the forward difference at four direction {0, 45, 90,135}[27]. Maximum difference are thresholded to remove redundancy and to reduce GLCM features with detection accuracy of 78%, 91.75% and 92.75% accuracy for spread spectrum, LSB matching and generic LSB respectively at 0.3 bpp. S. Ghanbari et al. implemented a neural network based method to identify the clean and stego images based on GLCM features and achieved 80% detection accuracy [28]. I.A. Khalifa et al. produce GLCM of the image followed by 3 level discrete wavelet

transform (DWT) generating 12 sub band and DCT with neural network based accuracy as 81.82% [29].

VI. COMPARISON OF METHODS

This section compares the techniques discussed in the previous sections based on the parameters a) Modelling Technique: The mathematical models and methods used to establish the technique. b) Quantitative Detection: The accuracy of the estimation of the hidden message length from stego image. c) Usage: Steganalysis class in which the method belong d) Error distribution: The error rate of detection e) Image types: The file type on which it can be applied. f) Compression: The effect of compression on detection rate. g) Maximum error rate: Maximum estimated error rate. h) Feature base: Important feature types

Table-1: Comparison among the techniques.

Parameters	Chi-Square	RS	SPA	GLCM
Modelling technique	χ^2 statistics of pairs of values[5]	Discriminating and flipping operation on pixel group[6]	Multi-sets of pairs as internal states of FSM.[7]	Frequency of co-occurrence of pairs of values[9]
Quantitative Detection	Message length estimation unreliable[30]	Message length estimate quite accurate[13]	More accurate than RS and Chi-square[13]	Classification is of importance.
Usage	Targeted	Targeted	Targeted	Blind
Error Distribution	Not Gaussian	Heavy tailed Kurtosis nearly 20. Follows Cauchy Distribution.	Heavy tailed Kurtosis nearly 15.	Diagonally clustered values for natural images. Depends on the Learning model used.
Image types	Color/Grayscale[30]	Color/Grayscale[30]	Better for Color JPEG[30]	Separate GLCM for 3 channels
Compression	Compressed image shows better detection[13]	Better detection with JPEG compression[13]	Slightly better performance with compression.[13]	Less affected by compression.
Maximum Error rate	12.9% [4]	11.38% [4]	8.6%[4]	Depends on the learning model.
Feature base	Frequency of pairs of values	Group type in pixel percent	Sample pair transition probability	Haralick, diagonal centric, low dimensional etc.

VII. CONCLUSION

This work aims to study the principle of targeted classical statistical steganalysis techniques along with GLCM based blind approach of steganalysis. This study also brings in the work that has been done so far in each of these categories. Statistical method are pretty older but reliable and computation need less time. They are still significant as found from the literature. GLCM based blind methods are comparatively new and need huge data set for training and need lot hardware support. Dimensionality reduction techniques are also used as a solution. But their detection accuracy is better. They can complement each other and can be applied in a blended mode when needed.

REFERNECE

[1] R.J. Anderson, F. A Petitcolas, "On the Limits of Steganography", IEEE Journal on Selected Areas in Communications, Vol.**16,** Issue.**4**,pp.**474-481,1998**.

[2] B. Li, J. He, J. Huang, Y. Q. Shi, "A Survey on Image Steganography and Steganalysis", Journal of Information Hiding and Multimedia Signal Processing., Vol. **2**, Issue.**2** ,pp.**142-172**, **2011.**

[3] A. Nissar, A.H. Mir, "Classification of Steganalysis Techniques. A Study", Digital Signal Processing, Vol. **20**, Issue.**6**, pp.**1758-1770, 2010.**

[4] S. Pathak, R. Roy, S. Changder, "Performance Analysis of Image Steganalysis Techniques and Future Research Directives", International Journal of Information and Computer Security. Vol.**10**, Issue.**1**, pp.**1-24**, **2018.**

[5] A. Westfeld, A. Pfitzmann, "Attacks on Steganographic Systems",In Lecture Notes in Computer Science, Vol.1768, Springer, **Berlin**, pp. **61-76, 2000**.

[6] J. Fridrich ,M. Goljan, R. Du, "Reliable Detection of LSB Steganography in Color and Grayscale Images", In the Proceedings of the Workshop on Multimedia and Security: New Challenges, ACM, **New York**, pp. **27-30**,**2001.**

[7] S. Dumitrescu, X. Wu, Z.Wang, "Detection of LSB Steganography via Sample Pair Analysis", IEEE Transactions on Signal Processing, Vol.**51,** Issue.**7,** pp. **1995-2007, 2003**.

[8] P. Lu, X. Luo, Q. Tang, L. Shen, "An Improved Sample Pairs Method for Detection of LSB Embedding.", In Lecture Notes in Computer Science, Vol. 3200, Springer, **Berlin,** pp.**116-127,2004**.

[9] R. M. Haralick, K. Shanmugam, I. H. Dinstein,"Textural Features for Image Classification", IEEE Transactions On Systems, Man, and Cybernetics Vol.**3**, Issue. **6** , pp. **610-621**, **1973**.

[10] R. M. Haralick, "Statistical And Structural Approaches To Texture," Proceedings of the IEEE, vol. **67**, no. **5**, pp. **786-804**, **1979**.

[11] K. Sullivan, U. Madhow, S. Chandrasekaran , B. S. Manjunath, "Steganalysis for Markov Cover Data With

Applications to Images", IEEE Transactions on Information Forensics and Security Vol.**1**, Issue. **2**, pp. **275-287**, **2006.**

[12] L. Nanni, S. Brahnam, S. Ghidoni, E. Menegatti, T. Barrier, "Different Approaches for Extracting Information from the Co-Occurrence Matrix.," PloS one Vol.**8**, Issue. **12**, pp. **1-9, 2013.**

[13] A.D. Ker,"Quantitative Evaluation of Pairs and RS Steganalysis". In the Proceddings of SPIE 5306. Security, Steganography, and Watermarking of Multimedia Contents VI, Vol. 5306,SPIE,**San Jose**, pp. **83-97,2004**.

[14] H. Zhang, H. Tang, "A Novel Image Steganography Algorithm Against Statistical Analysis." In the Proceedings of International Conference on Machine Learning and Cybernetics, IEEE, **Hong Kong,** pp**. 3884-3888, 2007.**

[15] X. Luo,F. Liu,P. Lu, "A LSB Steganography Approach Against Pixels Sample Pairs Steganalysis", International Journal of Innovative Computing, Information and Control, Vol.**3,** Issue.**3,** pp. **575-588,2007.**

[16] S. Manoharan, "An Empirical Analysis of RS Steganalysis" In the Proceedings of the Third International Conference on Internet Monitoring and Protection, IEEE, **Bucharest**, pp.**172-177**, **2008.**

[17] T. Qian, S. Manoharan, "A Comparative Review of Steganalysis Techniques", In the Proceedings of 2nd International Conference on Information Science and Security (ICISS), IEEE, **Seoul**,pp.**1-4**, **2015**.

[18] N. Prokhozhev, , O. Mikhailichenko, , A. Sivachev, , D. Bashmakov, A. Korobeynikov, "Passive Steganalysis Evaluation: Reliabilities of Modern Quantitative Steganalysis Algorithms", In the Proceedings of the First International Scientific Conference Intelligent Information Technologies for Industry (IITI'16). Advances in Intelligent Systems and Computing, Vol. 451. Springer, **Cham,** pp. **89-94**, **2016**.

[19] R. A. Solodukha, I. V. Atlasov, "Modification of RS-Steganalysis to Attacks Based on Known Stego-Program", In the Proceedings of Second Russia and Pacific Conference on Computer Technology and Applications (RPC), IEEE, **Vladivostok**, pp. **176-179**, **2017.**

[20] N. Mewalal, W. S. Leung "Improving Hidden Message Extraction Using LSB Steganalysis Techniques", In the Proceedings of International Conference on Information Science and Applications, Lecture Notes in Electrical Engineering. Vol. 514, Springer, **Singapore,** pp. **273-284, 2018**.

[21] W. Lin, T. Lai, C. Chou "Chi-Square-based Steganalysis Method Against Modified Pixel-Value Differencing Steganography", Arabian Journal for Science and Engineering, Vol.**46**, Issue.**9**,pp. **8525-8533,2021.**

[22] H. B. Kekre, A. A. Athawale, S. A. Patki, "Improved Steganalysis of LSB Embedded Color Images Based On Stego-Sensitive Threshold Close Color Pair Signature", International Journal of Engineering Science and Technology (IJEST), Vol. **3**, Issue. **2**, pp. **836-842**,**2011**.

[23] H. B. Kekre, A. A. Athawale and S. A. Patki., "Steganalysis of LSB Embedded Images Using Gray Level Co-Occurrence Matrix", International Journal of image processing (IJIP), Vol. **5**, Issue. **1**, pp. **36-45**, **2011**.

[24] Z. Xia, X. Wang, X. Sun,B. Wang, "Steganalysis of Least Significant Bit Matching using Multi-Order Difference," Security and Communication Networks, Vol. **7**, Issue. **8**, pp. **1283-1291**, **2014**.

[25] A. Anjum, S. Islam, "LSB Steganalysis Using Modified Weighted Stego-Image Method," In the Proceedings of the 3rd International Conference on Signal Processing and Integrated Networks (SPIN), IEEE,**Noida**, pp. **630-635,2016**.

[26] O. J. Sandoval, M. C. Hernandez, G. S. Perez, K. T. Medina, H. P. Meana, M. N. Miyatake, "Compact Image Steganalysis For LSB-Matching Steganography", In the Proceedings of the 5th International Workshop on Biometrics and Forensics (IWBF), IEEE, **Coventry**, pp. **1-6**,**2017**.

[27] Z. Sun, M. Hui, C. Guan, "Steganalysis Based on Co-occurrence Matrix of Differential Image", In the Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IEEE, **Harbin**, pp. **1097-1100, 2008**.

[28] S. Ghanbari, M. Keshtegary, N. Ghanbari,, "New Steganalysis Method using GLCM and Neural Network", International Journal of Computer Applications Vol.**42,** Issue. **2**, pp. **45-50**, **2012.**

[29] I. A. Khalifa, S. R. Zeebaree, M. Ataş, F. M. Khalifa, "Image Steganalysis In Frequency Domain Using Co-Occurrence Matrix and BPNN" Science Journal of University of Zakho Vol. **7**, Issue. **1**, pp. **27-32, 2019**.

[30] J. Fridrich, M. Goljan, D. Soukal, "Higher-Order Statistical Steganalysis of Palette Images", In the Proceedings of SPIE 5020, Security and Watermarking of Multimedia Contents V, SPIE , **Santa Clara,** pp. **178 - 190,2003**.

AUTHORS PROFILES

Bibek Ranjan Ghosh (MCA), University of North Bengal, is actively teaching as an Assistant Professor in the Department of Computer Science (CS), Ramakrishna Mission Residential College, (RKMRC) , Narendrapur Kolkata, India since 2008. He has published and presented many research articles in national and international journal and conferences including Scopus index journals. His research areas are computer vision, steganography &data science.

