International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 1 (2018) pp. 634-640

Analysis of Effect of Spatial Domain Steganography Technique on DCT Domain using Statistical Features for Digital Images

Govind R. Suryawanshi
Research Scholar, Department of Computer Engineering,
Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, Maharastra, India.
Orcid Id: 0000-0003-4491-3533

Dr. Suresh N. Mali
Principal, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, Maharastra, India.
Orcid Id: 0000-0001-6540-1604

Abstract

Steganography is an art of hiding secret data behind the digital images for covert communication. This could be down by different Steganographic techniques like frequency domain, spatial domain. While hiding the data in digital images may changes its statistical properties through which the analyzer gets the track for reveal the covert communication. In this paper, we have studied the different spatial domain Steganography techniques and their artifacts that leaved after data hiding. We have observed DCT domain statistical features of digital images that could be affected after spatial domain embedding. Our work shows that there will be significant changes in DCT domain statistical features after spatial domain embedding w. r. t. embedding algorithms

Keywords: steganography, Steganalysis, Feature Extraction

INTRODUCTION

Steganography is a technique to conceal the data in cover media like image, text etc such manner should not be detected through human eyes. The covert communication done by embedding secret data bits in digital media in a manner that no one rather than the sender knows that data is present in that digital media [1]. Steganography means "covered writing" in Greek. Now days, internet is widely used to exchange secret information securely without any interpretation of any Steganography analyzer techniques. Lot of Steganography techniques exist in literatures which are used for covert communication in different domain for hiding the data in carrier images like frequency, Spatial. The counter measures for malicious activity is Steganalysis whose ultimate aim is to detect the presence of secret communication. There are different challenges in Information Forensic out that most important challenge is to improve the performance of Steganalysis techniques for detection covert communication.

The Information Security Forensic professionals, Law Enforcement, Intelligence Professionals require the potential not only for detecting application of digital Steganography to hide information also to want an ability to detect hidden information in cover media. For steganography, digital media is used to carry secret information but image is more famous for the same as it has redundancy of high degree. There are number of techniques used to hide data in images in spatial or frequency domain like LSB, F5 Outguess etc.

For covert communication media and hiding strategies plays important role. If the embedding technique is known then it's very easy to identify covert communication through images. The technique which is used analyze the hide data in images is exactly reverse of embedding technique, it is called as targeted steganalysis. The process of analysis of cover media which results in separation or detection of secrete information from cover media is called as Steganalysis. Steganalysis is tool to strengthen steganography. The detailed categorization of steganalysis is as shown in figure 1.

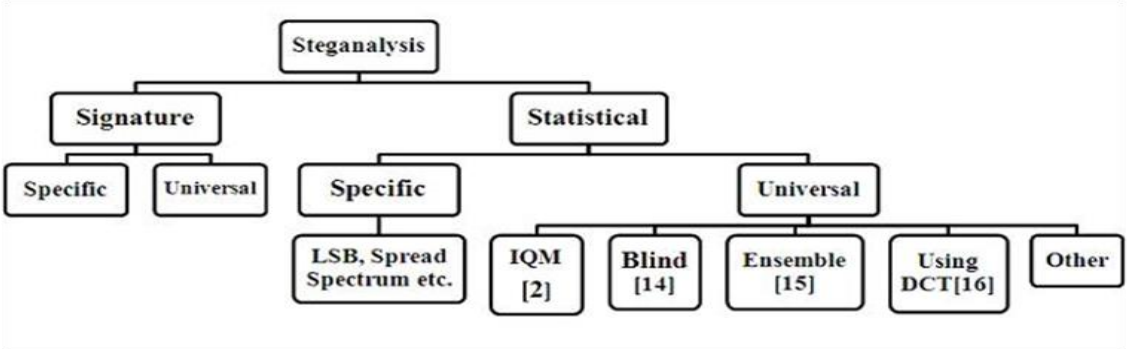


Figure 1: Categorization of Steganalysis

Basically steganalysis is classified in two main classes as signature based and statistical analysis approach [7]. Signature analysis used the common artifacts that are left by the respective steganography through which the embedding carried out. It is also referred as targeted steganalysis. Whereas in statistical based approach used the statistical clues that left by embedding algorithm. In this approach the algorithm search for statistical modifications done by the steganography algorithm.[8]. In statistical analysis cover media analyzed and its statistical values are used as reference for detection of covert communication. Most of researcher used binary classifier for treating steganalysis and it is highly successful [1-9]. However the analysis carried out by assuming that steganalyst has information about cover media and embedding steganographic method. Most of the targeted steganalysis use this hypothecation. However there are two different approaches for steganalysis. First one is with all information of steganographic method and second on without having information about cover and source. In this case, analyst finds some prominent features and train system only on sampled data set of cover images. The system may experience form cover source mismatch as system is trained on one type but tested with different type. This issue can be resolved by using ensemble classifier which train and steganalyzer on different samples of images [11]-[13]. One more method uses training a group of steganalyzer for different possible cover media types and this set of analyzers able to detect defect, pattern or content of images or source identification for recognize the cover media type The real challenge in this is selection of features through which analyzer will be able to identify the cover and source of communication. In this paper we have focused on the statistical features of images that are modified by the embedding algorithm. As explained in being, there are different techniques in market available for embedding the secret information in images in different domain. While designing the analyzer it is required to find the embedding domain and effect of it on statistical features. We have focused this issue and tracked the effect of spatial domain embedding algorithms on frequency domain using statistical features of images.

IMAGE FORMAT

From the digitization digital image are produced. It's a process of conversion of analogue information into digital information. A digital image is the representation of an original image by discrete sets of points called pixel. These are ordered in a two-dimensional lattice analogous to spatial coordinates in the original image. The number of Bits per Pixel (bpp) decides the distinct colors in digital image. Hence, numbers of bits per pixel is used to classify the digital images. Following are the common types of digital image.

Binary Image, only one bit per pixel allocated. Therefore a bit has only two possible values or states (on or off). In this format, each pixel has one of two colors i.e. black or white. It is also called as a bi-level image.

Gray scale image is a digital image which has only shades of grey colors. The darkest possible shade is black, whereas the lightest possible shade is white. Eight bits are used to represent a pixel value of gray colour. Possible permutation will be 256 for different shades of grey color.

Color image in this format, pixel is represented by three different primary colors (Red, Blue and Green). A pixel has three different values that are represented by eight bits. 24 bit are required to represent the color pixel. It produces near about 16.7 million distinct color combinations.

LITERATURE SURVEY

There are many different techniques proposed in literature for hiding data in images. The main aim of all the techniques is to hide the secret data in digital images in spatial or frequency domain. So that no one can knows about it except sender. While hiding the data in images may change its statistical properties [2]. In literature, the statistical properties are grouped in to 26 categories according to respective domains as described by Avcibas [13] that are Co-occurrence Matrix, Statistical Moments, Wavelet Sub bands and Pixel Difference.

We have visible many instances of a new steganographic scheme created to steer clear of current steganalysis. In flip this new scheme can detected through an stepped forward detector, and steganographer tries to thwart the improved detector. Preferably, in place of iterating on this way, the inherent detectability of a steganographic scheme to any detector, now or inside the future, could be pre-decided. An approach that yields wish of determining this is to version an image as a attention of a random procedure, and leverage detection idea to decide most reliable solutions and estimate performance. The important thing advantage of this version for steganalysis is the provision of results prescribing ideal (mistakes minimizing) detection methods as well as imparting estimates of the effects of gold standard detection. Moreover, the idealized detection regularly shows a technique for realistic realizations. There has been a few steganographic techniques with this approach, in particular within the closing couple of years. An early example of a detection-theoretic method to steganalysis is Cachin's work [14]. The steganalysis problem is framed as a hypothesis check between cowl and stego hypotheses. Cachin indicates certain on the Kullback-Leibler (KL) divergence (relative entropy) among the quilt and stego distributions as a degree of the security between cover and stego. This safety measure is denoted - secure, where is the bound on the K-L divergence. If is 0, the device is described as perfectly secure. Under this assumption, through Stein's Lemma [15] that is equal to bounds on the error prices of an premier detector. Any other data theoretic derivation is finished for a slightly one of a kind model via Z¨olner et al [16]. They first count on that the steganalyst has get admission to the exact cover, and show the instinct that this will in no way be made secure. They modify the version so that the detector has some, but now not entire, data on the duvet. From this version they locate constraints on conditional entropy similar to Cachin's, even though greater abstract and as a result more tough to evaluate in exercise. Chandramouli and Memon [17] use a detection-theoretic framework to analyze LSB detection. However, though the analysis is correct, the model is not accurate enough to provide practical

results. The cover is assumed to be a zero mean white Gaussian, a common approach. Since LSB hiding effectively either adds one, subtracts one, or does nothing, they frame LSB hiding as additive noise. If it seems likely that the data came from a zero mean Gaussian, it is declared cover. If it seems likely to have come from a Gaussian with mean of one or minus one, it is declared stego. However, the hypothesis source distribution depends on the current value. For example, the probability that a four is generated by LSB hiding is the probability the message data was zero and the cover was either four or five; so the stego likelihood is half the probability of either a four or five occurring from a zero mean Gaussian. Under their model however, if a four is received, the stego hypothesis distributions are a one mean Gaussian and a negative one mean Gaussian. Guillon et al [18] analyze the detectability of QIM steganography, and observe that QIM hiding in a uniformly allotted cover does now not trade the statistics. this is, the stego distribution is also uniform, and the device has considering the fact that standard cowl facts is not in fact uniformly allotted, they endorse the usage of a non-linear "compressor" to convert the quilt information to a uniformly disbursed intermediate cover. The information is hidden into this intermediate cover with general QIM, after which the inverse of the characteristic is used to convert to final stego facts. However Wang and Moulin [19] factor out that such processing may be unrealizable. The usage of detection idea from the steganographer view factor, Sallee [20] proposed a means of evading top-rated detection. The basic idea is to create stego facts with the identical distribution model as the duvet information. That is, rather than trying to mimic the precise cover distribution, mimic a parameterized model. The justification for that is that the steganalyst does not have get entry to the original cover distribution, but need to as a substitute use a version. As long as the steganographer matches the version the steganalyst is using, the hidden records does not appearance suspicious. The degree with which the model may be approximated with hidden facts may be described as –secure with appreciate to that version. a specific approach for hiding in JPEG coefficients using a Cauchy distribution model is proposed. Even though this particular approach is found to be inclined by using B¨ohme and Westfeld [21], the authors strain their a success detection is due to a weak point within the version, in place of the overall framework. Recently Sallee has protected [22] a protection towards the blockiness detector, through explicitly compensating the blockiness degree after hiding with unused coefficients, much like OutGuess' histogram repayment. The author concedes an superior solution might require a technique of matching the entire joint distribution within the pixel area, and leaves the improvement of this technique to future work. Form literature it is discovered that steganography and steganalysis are constantly developing generation and it's miles very tough to use equal generation to crack maximum of steganography strategies. Therefore we've got targeted to examine modifications in extraordinary domain after embedding mystery facts in virtual pictures.

PROPOSED SYSTEM

In proposed a system secret data can embedded into digital using images spatial domain steganography techniques and observed the effect of same on frequency domain through statistical properties like IQM proposed by Avibas. Figure 2 shows architecture of proposed model. Pre-processing of cover images removes unnecessary data so that the high accuracy of analysis can be achieved. Features extracted from given input. The set of that features used to classify the category of input image. Domain classifier classify the category of image to which is belongs to. There are different classifiers are available which has their own advantages like FLD, SVM, NN, Multivariate Regression. SVM is not scalable with respect to dimensionality of feature set.

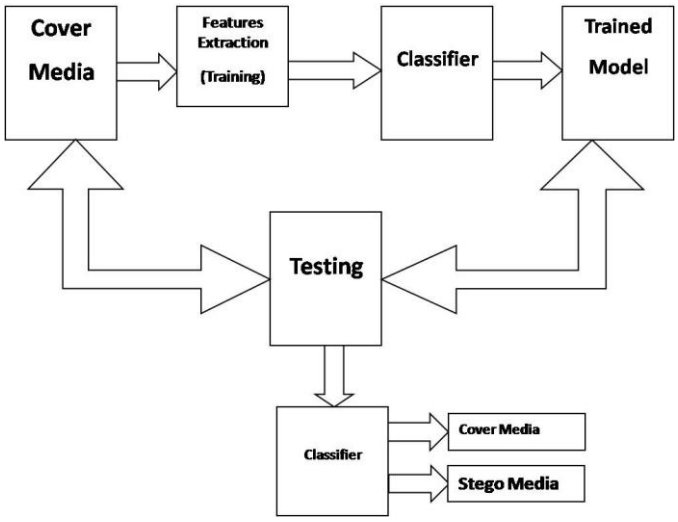


Figure 2: Proposed Framework

MATHEMATICAL MODEL

Co-occurrence of DCT coefficient with high dimensional features set are used for Steganalysis purpose, if there is any difference any it that implies intra and inter block dependencies. Multispectral components of image are identified by the pixel at position i, j and in band k as C_k (i, j), where k=1…3 for color images. Let C (i, j), C^ (i, j) are used to represent multispectral pixel vectors of position (i, j).

C and C^ are used for representing multiband image matrix

Let Mi, i= 1….10 represent IMQ features which are used in the detector.

1. Minkowsky Measures: It is average of the pixel differences in spatial and then chromatical (that is over the bands)

$$Mr = 1/k_0 \sum_{k=1}^{K} \left(1/_{n2} \sum_{i,j=1}^{N} | C_K(i,j) - \hat{C_k}(i,j) |\gamma\right)^{1\backslash\gamma} \quad (1)$$

2. Correlation Measures Czekanowski distance is used to compare vectors (Non negative components) as in the case of images, is the

$$M_3 = \frac{1}{N^2}\sum_{i,j=0}^{N-1}\left(1 - 2\sum_{k=1}^{K}\mathbf{Min}(C_K(i,j) - C_k^{\wedge}(i,j)) \Big/ \sum_{k=1}^{K}(C_K(i,j) - C_k^{\wedge}(i,j))\right) \quad (2)$$

The angular correlation is used to calculate difference between two vectors which is as follows

$$M_4 = 1 - \frac{1}{N^2}\sum_{i,j=1}^{N}\frac{2}{\pi}\cos^{-1}\frac{(C(i,j),C^{\wedge}(i,j))}{||C(i,j)||,||C^{\wedge}(i,j)||} \quad (3)$$

In proposed work image will be divided into smaller blocks for better results of Steganalysis. Secret message hides in such area which has high energy blocks. Generally Steganalysis algorithm takes entire images for processing. Work focused on to find the difference between the blocks whose energy difference is high. SS-SPAM l Features of block are calculated. The correlation of that block is stored for further calculations

Algorithm 1: Image Feature Extraction

Step1: Read an Image (I)

Step 2: Preprocessing of I if necessary

Step 3: Upload the database of BOSS images

Step 4: Upload input Image

Step 5: Divide the input image into blocks of Size 8x8

B= {b1, b2, b3,….}

Step 6: Extract Features from image through each blocks

Step 7: Stored extracted features in vector

F= { Set of features w. r. t B}

Step 8: F= SPAM(I)

Step 9 : Read a Stego Image(I')

Step 10: Repeat step 2 to 8 for I' and F'

Step 11: Compare F and F'

Step 12: If (F=F')

Original Image -> I'

Step 13: Else

Stego image -> I'

End

Algorithm 2: Analysis of Image

Step 1 : Read Image (X and Y)

Let a two dimension array of pixel values in an n1 x n2 cover and Stego image. ⟦X,Y{0,……255}⟧^(n_1 x⟦ n⟧_2) .

Step 2: Preprocessing of I and I' Noise components are estimated using SPAM different pixel predictors [F].

Step 3: Calculate Residual Z= (zkl) with same dimension as X

Z= K * X – X

Step 4 : Calculate

$$\boldsymbol{Z^h = Z_{kl}^h}\ , \boldsymbol{Z^v = Z_{kl}^v}$$

$$\boldsymbol{Z_{kl}^h = x_{k,l+1} - x_{kl}}$$

$$\boldsymbol{Z_{kl}^v = x_{k+1,l} - x_{kl}}$$

$$\boldsymbol{z_{kl}^{min} = \mathbf{min}\{z_{kl}^h, z_{kl}^v\}}$$

$$\boldsymbol{z_{kl}^{max} = \mathbf{max}\{z_{kl}^h, z_{kl}^v\}}$$

Step 6: Generate Random matrices $\boldsymbol{\pi^i \in R^{r\,x\,s}, i \in \{1, \ldots v\}}$. where r, s uniformly randomly selected from {1,…..,S}. For each $\boldsymbol{i \in \{1, \ldots s\}}$

Step: 7 compute the projections$\boldsymbol{P^i \equiv z * \pi^i}$.

Step 8: Compute v separate histogram of the quantized values

$$\boldsymbol{h_j^i = |\{k,l\}|\left|p_{kl}^i\right| = j + 1/2|,}$$

$$\boldsymbol{j\epsilon\{0 \ldots . T-1\}, i\epsilon\{1, \ldots v\}\mathbf{forlinearResiduals}}$$

EXPERIMENTAL SETUP

For this experiment, Break Our Steganography Schema (BOSS) base image database is used along with different types like JPEG,PGM and BMP each of 481x381, 512x512, 256x256 resolution respectively. Dataset is used for training, testing and validation propose. 1000 images used in dataset out of that 800 images are used for training propose. Remaining are used as testing dataset. Out of the 200 testing images, 100 images are used as clean images and the rest images are used as Stego images. Stego images are generated through spatial domain steganography algorithms domain like LSB, MB, UNWARD etc. for different payload l. Similarly from the training images, out of 800 images, 400 images were used as clean. Remaining 400 images are used as Stego generated through well known spatial domain Steganography algorithms. Four statistical features are calculated from 400

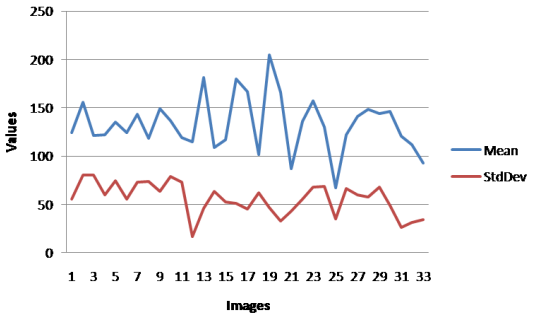
Stego and 400 cover images of different type. Calculated features are skewness, kurtosis, mean and standard deviation. Table 1 shows the details of dataset used

Table 1: Details of Dataset

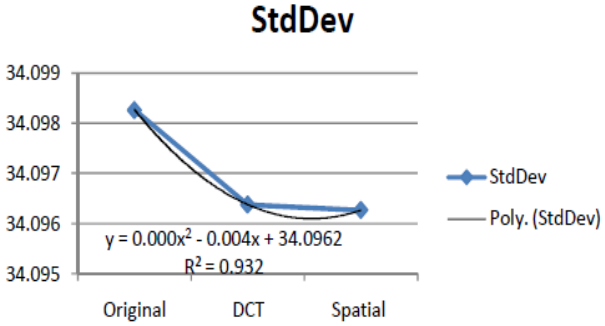
Sr No	Type	Resolution	Size	Quantity	PSNR
1	Gray	512x512	256k	1000	75
2	Color	481x321	603K	1000	75
3	Color	256x256	256k	100 0	75

RESULTS

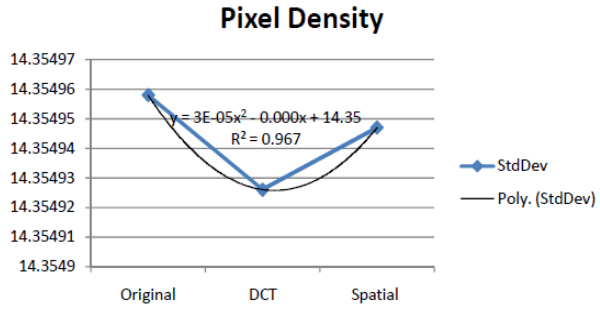
Observed statistical features of cover and Stego images show significant changes in their values which show the effect of spatial domain steganography on DCT domain. In spatial domain steganography, secret data is covered under least significant value of pixel. So that overall LSB bits may be affected which revert the changes in AC or DC coefficient of images and that can be traced using frequency features of images. The experiment shows effect of change of Mean, Skewness, kurtosis and Standard Deviation. Graph 1 to Graph 6 shows the relationship between mean and standard deviation and variation in pixel density and slandered deviation. Value of mean of pixel is must be less than the standard deviation of any image for different embedding algorithms.



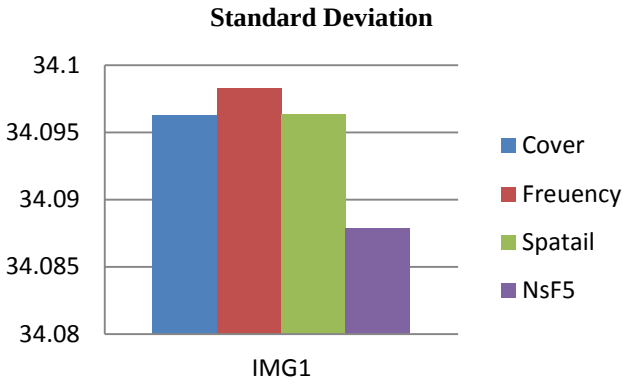
Graph 1:. Mean vs Standard Deviation



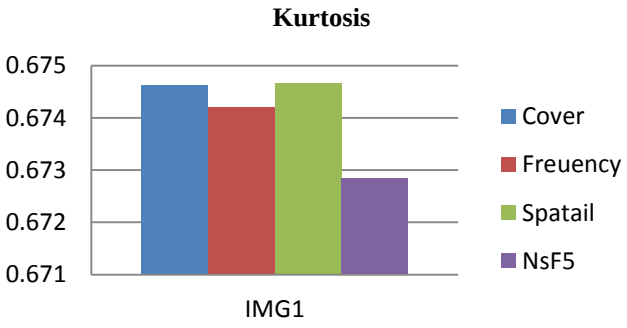
Graph 2: Variation in StdDev



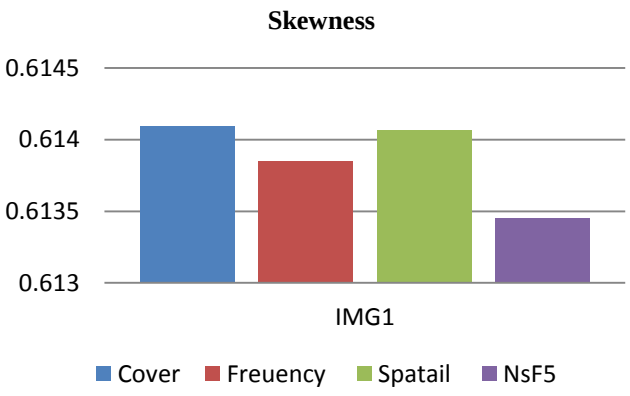
Graph 3: Variation in Pixel Density



Graph 4: Standard Deviation



Graph 5: Kurtosis



Graph 6: Skewness

From above Graphs, it is observed that the effect of changes in pixel can be captured or observed by statistical features of image as seen in above graphs. The feature selection is important for the steganalysis of digital images. In this experiment only four features are observed i.e standard deviation, Mean, kurtosis, and skewness . spatial changes also affect the DCT domain features that can be monitored by the same feature set. Spatial embedding algorithms show significant changes the DCT coefficients of images with respect to DCT value of image. The linear equation for standard deviation and pixel density obtained with R value 0.932 and 0-967 respectively.

CONCLUSION

Spatial domain Steganography algorithms embed the data in pixel values of digital images which will change not only spatial values but also affect the DCT coefficients of images. Embedding changes can be monitored by the spatial and DCT domain statistical features of digital images. Work will be useful for analysis of spatial domain steganography and its effect on DCT domain through which novel spatial algorithm can be innovated which will not leave an artifact behind to trace the embedding technique

REFERENCES

[1] Arooj Nissar , A.H. Mir, "Classification of steganalysis techniques: A study," Digital Signal Processing 20 (2010) Elsevier, pp.1758–1770 ,2010.

[2] G R Suryawanshi, Dr. S N Mali.(2015).Study of Effects of DCT Domian Steganography Techniques in Spatial Doamain for JPEG Images Steganalysis. Internation Journal of Computer Applciaiton, 127(6), 16-20

[3] G R Suryawanshi, Dr. S N Mali.(2015)." Universal steganalysis using IQM and multiclass discriminator for digital images" DOI:10.1109/SCOPES.2016.7955568

[4] Jessica Fridrich, *Member, IEEE*, And Jan Kodovský, "Rich Models For Steganalysis Of Digital Images," IEEE Transactions On Information Forensics And Security, Vol. 7, No. 3, pp no. 868-882, 2012

[5] Min Wu, *Member, IEEE*, and Bede Liu, *Fellow, IEEE* ,"Data Hiding in Image and Video:Part I—Fundamental Issues and Solutions," IEEE Transactions on Image Processing, Vol. 12, No. 6, June 2003"

[6] Chunfang Yang, Fenlin Liu, Xiangyang Luo, and Bin Liu, "Steganalysis Frameworks of Embedding in Multiple Least-Significant Bits," IEEE Transactions on Information Forensics and Security, Vol. 3, No. 4, December 2008

[7] Yun Cao, Xianfeng Zhao, and Dengguo Feng, " Video Steganalysis Exploiting Motion Vector Reversion Based Features," IEEE Signal Processing Letters, Vol 19, No. 1, pp no 35-38, 2012

[8] Mengyu Qiao, Andrew H. Sung , Qingzhong Liu, "MP3 audio steganalysis," Information Sciences 231 Elsevier 123–134, 2013

[9] R. Sridevi, A. Damodaram and S.V.L. Narasimham, "Efficient Method of Audio Steganography by Modified LSB Algorithm and Strong Encryption Key with Enhanced Security," Journal of Theoretical and Applied Information Technology, Vol. 5, No. 6, pp. no 768 – 771, June 2009.

[10] D. Kirovski and H. Malvar, "Spread spectrum Watermarking of Audio Signals," IEEE Transactions on Signal Processing, vol. 51, no. 4, pp. 1020 – 1033, April 2003.

[11] Tomáš Pevný, Jessica Fridrich, And Andrew D. Ker, "From Blind To Quantitative Steganalysis," IEEE Transactions On Information Forensics And Security, Vol. 7, No. 2, pp. no 445-454, 2012

[12] Jan Kodovský, Jessica Fridrich, "Ensemble Classifiers For Steganalysis Of Digital Media," IEEE Transactions On Information Forensics And Security, Vol. 7, No. 2, pp no. 432-444, 2012

[13] I. Avcibas, M. Nasir, and B. Sankur., "Steganalysis Based on Image Quality Metrics," *IEEE 4th Workshop on Multimedia Signal Processing*, pages 517–522, 2001

[14] K. Sullivan, U. Madhow, S. Chandrasekaran, and B. Manjunath, "Steganalysis for Markov Cover Data With Applications to Images," *IEEE Transactions on Information Forensics and Security*, 1(2):275–287, 2006

[15] R. J. Anderson, "Stretching the Limits of Steganography," *1st International Workshop on Information Hiding*, 1174:39–48, 1996.

[16] R. J. Anderson and F. A. P. Petitcolas, "On the limits of Steganography," *IEEE Journal of Selected Areas in Communications*, 16(4):474–481, 1998

[17] S. Badura and S. Rymaszewski, "Transform Domain Steganography in DVD Video and Audio Content," *IEEE International Workshop on Imaging Systems and Techniques*, pages 1–5, 2007

[18] X.Chen, Y. Wang, T. Tan, and L. Guo, "Blind Image Steganalysis Based on Statistical Analysis of Empirical Matrix," *International Conference on Pattern Recognition*, 3:1107–1110, 2006

[19] G. Xuan, Y. Q. Shi, J. Gao, D. Zou, C. Yang, Z. Zhang, P. Chai, C. Chen, and W. Chen,"Steganalysis Based onMultiple Features Formed by Statistical Moments of Wavelet Characteristic Functions," *7th International Workshop on Information Hiding*, 3727:262–277, 2005

[20] Y. Q. Shi, G. Xuan, D. Zou, J. Gao, C. Yang, Z. Zhang, P. Chai, W. Chen, and C. Chen, "Image Steganalysis Based on Moments of Characteristic Functions Using Wavelet Decomposition,

Prediction-Error Image, and Neural," Network. *IEEE International Conference on Multimedia and Expo*, pages 269–272, 2005

[21] C. Chen, Y. Q. Shi, W. Chen, and G. Xuan, " Statistical Moments Based Universal Steganalysis using JPEG 2-D Array and 2-D Characteristic Function," *IEEE International Conference on Image Processing*, pages 105–108, 2006

[22] H. Farid, "Detecting Steganographic Messages in Digital Images," *TR2001-412, Department of Computer Science, Dartmouth College*, 2001