# Statistical Approaches for Digital Image Steganalysis

*Thesis submitted in partial fulfillment of the requirements for the award of degree of*

**Master of Technology**
in
**Computer Science and Applications**

*Submitted By*

**Balkar Singh**
**(Roll No. 601003004)**

Under the supervision of
**Dr. Jitender Kumar**

SCHOOL OF MATHEMATICS AND COMPUTER APPLICATIONS
THAPAR UNIVERSITY
PATIALA – 147004

June 2012

# Certificate

I hereby certify that the work which is being presented in the thesis entitled, **"Statistical Approaches for Digital Image Steganalysis"**, in partial fulfillment of the requirement for the award of degree of Master of Technology (Computer Science and Applications) submitted in School of Mathematics and Computer Applications, Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of Dr. Jitender Kumar and refers other researcher's work which are duly listed in the reference section. The matter presented in the thesis has not been submitted for award of any degree of this or other University.

(Balkar Singh)

Roll No. 601003004

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

14/06/2012

(Dr. Jitender Kumar)

School of Mathematics and Computer Applications

**Countersigned by**

**(Dr. S. S. Bhatia)**
Professor & Head
School of Mathematics and Computer Applications
Thapar University
Patiala

**(Dr. S. K. Mohapatra)**
Dean (Academic Affairs)
Thapar University
Patiala

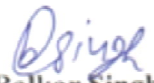
## Acknowledgement

First of all, I am thankful to God for his blessings and showing me the right direction. With His mercy, it has been made possible for me to reach so far.

It gives me great pleasure to express my gratitude towards my guide Dr. Jitender Kumar. I am thankful for his continual support, encouragement, and invaluable suggestions. He not only provided me help whenever needed, but also the resources required to complete this thesis report on time.

I am also thankful to Dr. S. S. Bhatia, Professor and Head, School of Mathematics and Computer Applications (*SMCA*) for his kind help and cooperation. I express my gratitude to all the staff members of *SMCA* for providing me all the facilities required for the completion of my thesis work.

Last but not the least I am highly grateful to all my family members for their inspiration and ever encouraging moral support, which enables me to purse my studies.



**(Balkar Singh)**

Roll No. 601003004

M.Tech. (CSA)

# Abstract

Steganography is the technique to hide secret information within cover objects like images, audio, video and text files. 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. This thesis presents 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, *PSNR*, and *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.

## List of abbreviations

| | |
|---|---|
| AWGM | Additive White Gaussian Noise |
| B | Bit depth |
| COV | Covariance |
| COM | Center of Mass |
| DCQIM | Distortion Compensated Quantization Index Modulation |
| DCT | Discrete Cosine Transform |
| DFT | Discrete Fourier Transform |
| FN | False Negative |
| FP | False Positive |
| HCF | Histogram Characteristic Function |
| IID | Independent and Identically Distributed |
| IQM | Image Quality Metrics |
| JPEG | Joint Photographic Experts Group |
| LSB | Least Significant Bit |
| MSE | Mean Square Error |
| O | Order |
| P | Probability |
| PMF | Probability Mass Function |
| PDF | Probability Distributed Function |
| PSNR | Peak Signal-to-Noise Ratio |
| QIM | Quantization Index Modulation |
| RGB | Red Green Blue |
| RS | Regular/Singular |
| SD | Standard Deviation |
| SS | Spread Spectrum |
| SVM | Support Vector Machines |
| TP | True Positive |
| TN | True Negative |
| WAM | Wavelet Absolute Moment |

# List of tables

# List of figures

# Table of Contents

# Chapter 1
# Introduction

With the development of digital multimedia and network technology, the communication using digital devices is increasing day by day. There is the need to secure this information. One of the methods discussed in the information security is the exchange of hidden information using some cover media. Information hiding is to hide the data in the cover media imperceptibly. Steganography, and watermarking are the techniques which are used to hide the information [1].

**1.1 Steganography:** 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. 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].

In the $20^{th}$ century, invisible ink was a widely used technique. In the Second World War, people used milk, vinegar, fruit juices and urine to write messages. When heated, these fluids becomes darker the message cold be read. After that, the Germans developed a technique called microdot. Microdots are photographs with the size of a printed period but have the clarity of a standard typewritten page. The microdots where then printed in a letter or on an envelope and being so small, they could be sent unnoticed [2].

We can say, steganography is not restricted to mediums like text, audio, video, images etc. A form of text steganography used by German spies in the World War 2nd is shown below. The following message was sent:

*"Apparently neutral's protest is thoroughly discounted and protested. Isman hard hit. Blockade issue affects pretext for embargo on byproducts, ejecting suets and vegetable oils"*. Taking the second letter from each word the sentence reads, *"Pershing sails from NY June 1"*.

The modern form of steganography is explained in terms of prisoner's problem [2], where Alice and Bob are two inmates, who wish to communicate in order to hatch an escape plan. However, all communication between them is examined by the warden, Wendy, who will put them in solitary confinement at the slightest suspicion of covert communication. Specifically, in the general model for steganography, we have Alice wishing to send a secret message *m* to Bob. In order to do so, she embeds *m* into cover object *c* (object not containing any secret message), to obtain the stego object *s* (object containing any secret message). The stego object is then sent through the public channel. In pure steganography framework, the technique for embedding the message is unknown to Wendy and shared as a secret between Alice and Bob. However, it is generally not considered as good practice to rely on the secrecy of the algorithm itself. In private key steganography, Alice and Bob share a secret key, which is used to embed the message. The secret key, for example, can be a password used to seed a pseudorandom number generator to select pixel locations in an image cover object for embedding the secret message (possibly encrypted). Wendy has no knowledge about the secret key that Alice and Bob share, although she is aware of the algorithm that they could be employing for embedding messages. In public key steganography, Alice and Bob have private-public key pairs and know each other's public key.

As long as people have been able to communicate with one another, there has been a desire to do so secretly. Two general approaches to covert exchanges of information have been:

i. Communicate in a way understandable by the intended parties, but unintelligible to eavesdrops;
ii. Communicate innocuously, so no extra party bothers to eavesdrop.

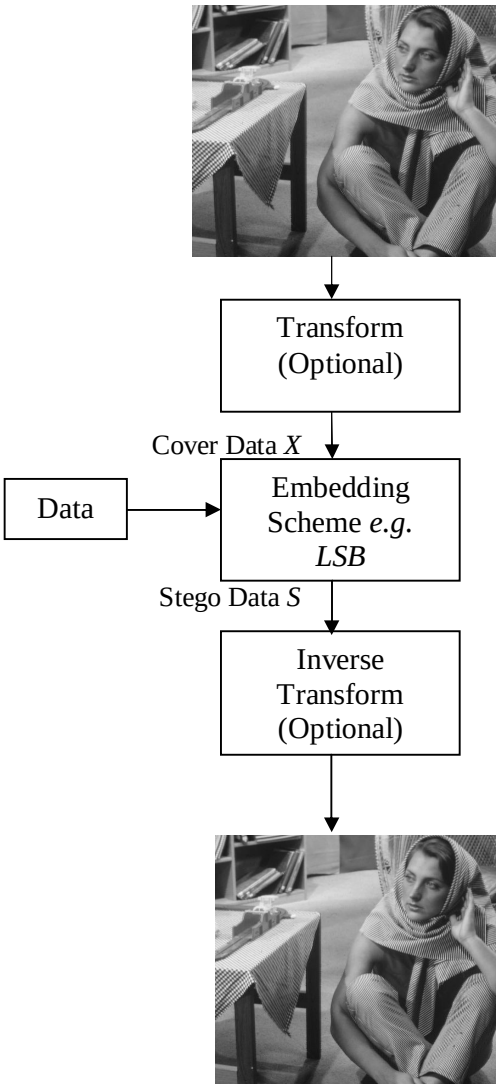



Fig. 1.1: Hiding data within an image.

Naturally both of these methods can be used concurrently to enhance privacy. The formal study of these methods, cryptography and steganography, have evolved and become increasingly more sophisticated over the centuries to the modern digital age. Methods for hiding data into cover or host media such as audio, images, and video, were developed about a decade ago. Although the original motivation for the early development of data hiding was to provide a means of watermarking media for copyright protection, data hiding methods were quickly adapted to steganography [4].

A typical image steganography system is described in Fig. 1.1. Although watermarking and steganography both imperceptibly hide data into images, they have

slightly different goals and so approaches differ. Watermarking has modest rate requirements, only enough data to identify the owner is required, but the watermark must be able to withstand strong attacks designed to strip it out. Steganograhpy generally is subjected to less vicious attacks, however as much data as possible is to be inserted. Additionally whereas in some cases it may actually serve a watermarker to advertise the existence of hidden data, it is of paramount importance for a steganographer's data to remain hidden.

Fig. 1.2: Original image to be stego.

Fig.1.2 shows an image which is used as a cover image. Fig.1.4 shows a stego image which is created using *F5* stego tool. This image contains the image shown in Fig. 1.3.

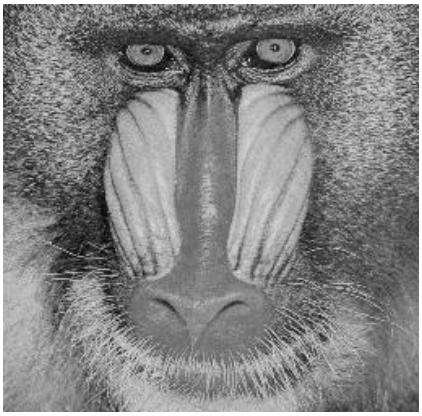


Fig.1.3: Image to be hidden.

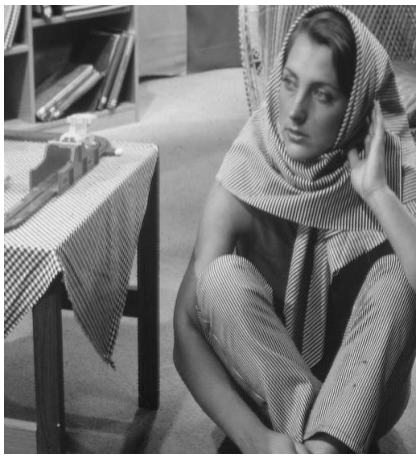


Fig. 1.4: Stego image.

For the steganographer, however, there are many scenarios with which the image remains untouched, and the covered image can be disposable, As such, Least Significant Bit (*LSB)* hiding is very popular today; a perusal of tools readily available online reveals numerous *LSB* embedding software packages.

Steganography represents a threat to the safeguarding of sensitive information and the gathering of intelligence. While it is possible to hide messages within a variety of data file types, image data is likely to be the medium of choice for the following reasons.

- First, because of the high level of redundancy in the images, it is possible to embed a great deal of hidden information.
- Second, innocuous-looking images are commonplace on Internet web sites and arouse little suspicion. By contrast, audio or video files posted on web sites are prone to be examined for copyright infringement.
- Third, the shear volume of image data available online makes it difficult to identify suspicious content.

After embedding a secret message into the cover (original) image, stego image is obtained. The less information, we embed into the cover image, the smaller the probability of introducing detectable artifacts by the embedding process. Another important factor is the choice of the cover image. The selection is at the discretion of the person who sends the message. Images with a low number of colors, computer art, and images with unique semantic content (such as font) should be avoided as cover images. Some steganographic experts recommend grayscale

images as the best cover images. They recommend uncompressed scans of photographs or images obtained with a digital camera containing a high number of colors and consider them safe for steganography.

There are three important aspects in information hiding systems are given below:

- Capacity refers to the amount of information that can be hidden in the cover medium.
- Robustness refers to the amount of modification the stego medium can withstand before an adversary can destroy hidden information.
- Security refers to an eavesdropper's inability to detect hidden information.

**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.

To find out true message may be impossible due to secure encryption methods, he/she may try to find out the location or length of the hidden message or estimate the parameters used in the embedding process. The statistical changes in the cover medium due to embedding help to design a passive steganalysis system. The changes can be quantified and compared to a threshold or to a known database to arrive at a decision. An attacker or steganalyst obtains a copy of the host signal from communication channel. After processing it, he/she measures the statistical change in the host signal due to embedding. The quantified change is compared with a threshold thresh to arrive at a decision of whether there is something is hidden or not.

It is developed for the urgent demands of network securities to block the covert communication with illegal information. It is an inherently difficult problem and requires a thorough investigation. The hider, who demands privacy, must carefully examine a means to guarantee stealth.

Naturally however, there are those who wish to detect this data. On the heels of developments in steganography come advances in steganalysis, the detection of images carrying hidden data, as explained in Fig.1.5.

The art of steganalysis is becoming increasingly more important in computer forensic, for screening and tracking documents that are suspect of criminal activities and for information security to prevent leakage of unauthorized data. It's important that the stego image does not contain any detectable artifacts due to message embedding. A third party could use such artifacts as an indication that a secret message is present. Once a third party can reliably identify which images contain secret messages, the steganographic tools become useless. A rigorous framework for analysis is required, both from the point of the steganalyst and the steganographer. In addition to preventing secret communication, steganalysis serves a way to judge the security performance of steganography techniques.

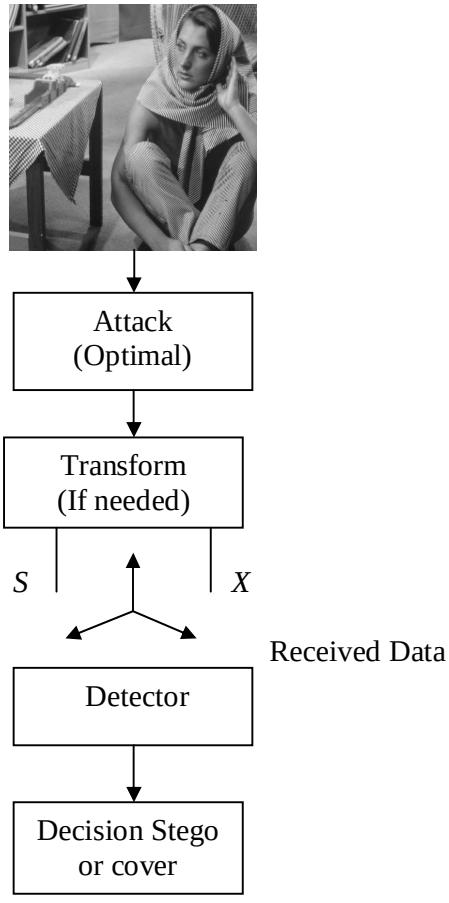



Fig.1.5: Steganalysis flow graph.

In the past few years, great expansions in the fields of steganography and steganalysis have witnessed. Several new steganography methods are being proposed each year, most of which are followed by new and improved steganalysis techniques for their detection.

**1.2.1 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].

**1.2.2 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:

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

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

**1.3.2 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.

**1.3.3 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 | | OUTPUT |
|---|---|---|
| A | B | A XOR B |
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

**1.3.4 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 $x \times y$, 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.

**1.3.5 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

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

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

# Chapter 2
# Literature Survey

In this chapter, detailed literature survey on Steganography Image tools and Steganalysis algorithms has surveyed and it is discussed in this section.

**2.1 Stego Image Tools:** There are various steganalytic tools available in market like 2Mosaic, StirMark Benchmark and PhotoTitle etc. these three steganalytic tools can remove steganographic content from any image. This is achieved by destroying secret message by two techniques break apart and resample. StegDetect, StegBreak, StegSpy identify information embedded via the following tools - Jsteg-shell, JPhide, and Outguess 0.13b, Invisible Secrets, F5, appendX Camouflage, Hiderman, JPHIde and Seek, Masker, JPegX, Steganography Analyzer Real-Time Scanner is the best available steganalysis software in the market at the moment, which can analyze all the network traffic to look for traces of steganographic communication [32].

**2.2 Steganalysis Methods:** There are many steganalysis methods reported in the literature survey. Research and development of steganography preceded steganalysis and steganalysis has been forced to catch up. Most recently, steganalysis had some success and steganographers have had to more carefully stealthiness of their hiding methods.

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.

Chen *et al.* [9] analyzed the problem of embedding one signal (e.g., a digital watermark), within another host signal to form a third, composite signal. The embedding is designed to achieve efficient tradeoffs among the three conflicting goals of maximizing information-embedding rate, minimizing distortion between the host signal and composite signal, and maximizing the robustness of the embedding. Authors introduce new classes of embedding methods, termed Quantization Index Modulation (*QIM*) and Distortion-Compensated *QIM* (*DC-QIM*), and develop convenient realizations in the form of what we refer to as dither modulation. Using deterministic models to evaluate digital watermarking methods, they show that *QIM* is provably good against arbitrary bounded and fully informed attacks, which arise in several copyright applications, and in particular, it achieves provably better rate distortion–robustness tradeoffs than currently popular spread-spectrum and low-bit(*s*) modulation methods. Furthermore, authors show that for some important classes of probabilistic models, *DC-QIM* is optimal (capacity-achieving) and regular *QIM* is near-optimal. These include both Additive White Gaussian Noise (*AWGN*) channels, which may be good models for hybrid transmission applications such as digital audio broadcasting, and mean-square-error-constrained attack channels that model private-key watermarking applications.

Another steganographic scheme based on *LSB* hiding, but designed to evade the chi square test is Provos Outguess [10]. Here *LSB* hiding is done as usual, but only half the available coefficients are used. The remaining coefficients are used to compensate for the hiding, by repairing the histogram. Although the rate is lower than *F5* hiding, since half the coefficients, we would expect this to not only be detectable by chi square, but by Fridrich's *F5* detector and in fact any detector using histogram statistics. However, because the embedding is done in the block wise transform domain, there are changes in the spatial domain at the block borders. Specifically, the change to the spatial joint statistics, i.e. the dependencies between the pixels, is different than for standard Joint Photographic Experts Group (*JPEG)* compression.
Fridrich *et al*. [11] discussed a reliable and accurate method for detecting *LSB* nonsequential embedding in digital images. The secret message length is derived by

inspecting the lossless capacity in the LSB and shifted LSB plane. An upper bound of 0.005 bits per pixel was experimentally determined for safe LSB embedding.

Avcibas *et al.* [12] proposed *LSB* detection scheme by using binary similarity between the $7^{th}$ bit plane and $8^{th}$ 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.

Lyu *et al.* [13] analyzed that techniques for information hiding have become more sophisticated and widespread. With high resolution digital images as carriers, detecting hidden messages has become considerably more difficult. Authors describes an approach to detect hidden messages in images that uses a wavelet-like decomposition to build higher order statistical models of natural images. Support Vector Machines (*SVM)* are then used to discriminate between untouched and adulterated images. Messages can be embedded into digital images in ways that are imperceptible to the human eye, and yet, these manipulations can fundamentally alter the under-lying statistics of an image. To detect the presence of hidden messages we have employed a model based on statistics taken from a multi-scale decomposition. This model includes basic coefficient statistics as well as error statistics from an optimal linear predictor of coefficient magnitude. These higher-order statistics appear to capture certain properties of natural images, and more importantly, these statistics are significantly altered when a message is embedded within an image.

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.

Avcibas *et al.* [15] proposed an approach to detect arbitrary hiding schemes. Author designed a feature set based on Image Quality Metrics (*IQM*), metrics designed to

mimic the human visual system. In particular they measure the difference between a received image and a filtered (weighted sum of 3x3 neighborhood) version of the image. This is very similar in spirit to the work done in [19], except with filtering instead of compression. The key observation is that filtering an image without hidden data changes the *IQMs* differently than an image with hidden data. The reasoning here is that the embedding is done here locally (either pixel wise or block wise), causing localized discrepancies. We see these discrepancies exploited in many stegnalysis schemes. Although their framework is for arbitrary hiding, they also attempted to fine-tune the choice of *IQMs* for two classes of embedding schemes: those designed to withstand malicious attack, and those not. A multivariate regression classifier is trained with examples of images with and without data. This work is an early example of supervised learning in steganalysis. Supervised learning is used to overcome the steganalyst's lack of knowledge of cover statistics. From experiments performed, we note that there is a cost of generality: the detection performance is not powerful as schemes designed for one hiding scheme. The results however are better than random guessing, reinforcing the hypothesis of the inherent unnaturalness of data hiding.

Dumitrescu *et al.* [16] analyzed that sample pair analysis is more rigorous analysis due to the basis of the Regular/Singular (*RS)* method, explaining why and when it works. The sample pairs are any pair of value (not necessary consecutive) in a received sequence. These pairs are partitioned into subsets depending on the relation on the two values to one another. It is assumed that in a cover image the numbers of pairs in each subset are roughly equal. It is shown that *LSB* hiding performs a different function on each subset, and so the numbers of pairs in the subsets are not equal. The amount of disruption can be measured and related to the known effect of *LSB* hiding to estimate the rate of hiding. Although the initial assumption does not require inter pixel dependencies, it can be shown that correlated data provides stronger estimates that the uncorrelated data. The *RS* scheme, the practical detector of *LSB* hiding, uses the same basic principal as sample pair analysis. As in sample pair analysis, the *RS* scheme counts the number of occurrences of pairs in given sets. The relevant sets, regular and singular are related to but slightly different the sets used in the sample pair analysis. Also as in sample pair analysis, equations are derived to estimate the length of hidden message. Since *RS* employs the same principle as sample pair analysis, we would expect it to also work better for correlated cover data.

Indeed the *RS* scheme focuses on spatially adjacent image pixels, which are known to be highly correlated. In practice *RS* and sample pair analysis perform comparably.

Harmsen *et al.* [17] analyzed the steganalysis of additive hiding scheme such as spread spectrum. Authors' detective statistic is based initially on a Probability Mass Function (*PMF)* estimate i.e. histogram. Since additive hiding is the addition of two random variables: the cover and the message sequence, the *PMF* of the cover and message sequences are convolved. In the Fourier it is equivalent to multiplication. Therefore the Discrete Fourier Transform (*DFT)* of the histogram, termed the Histogram Characteristic Function (*HCF*) is taken. It is shown for typical cover distributions that the expected value or Center Of Mass (*COM*), of the *HCF* does not increase after the data hiding and in practice typically decreases. The author chooses the *COM* as a feature to train a Bayesian multivariate classifier to discriminate between cover and stego. They perform tests on *RGB* images, using a combined *COM* of each color plane, with reasonable success in detecting additive hiding.

Sallee [18] proposed a means of evading optimal detection by using the detection theory from the steganographer's viewpoint. The basic idea is to create stego data with the same distribution model as the cover data. That is rather than attempting to mimic a parameterized model. The justification for this is that the steganalyst does not have access to the original cover distribution, but must instead use a model. As long as the steganographer matches the model the steganalyst is using, the hidden data does not look suspicious. The degree with which the model can be approximated with hidden data can be described as $\varepsilon$ secure with respect to that model. A specific method for hiding in *JPEG* coefficients using a Cauchy distribution model is proposed.

Celik *et al.* [19] proposed rate distortion curves for detecting the *LSB* hiding and Fridrich's content independent stochastic modulation which as studied, is statistically identical to spread spectrum. Authors observe that data embedding typically increases the entropy of the image, while attempting to avoid introducing perceptual distortion to the image. On the other hand, compression is designed to reduce the entropy while also not introducing any perceptual changes. It is expected therefore that the difference between a stego image and its compressed version is greater than the difference between cover and its compressed form. Distortion metrics such as *MSE*,

mean absolute error and weighted *MSE* are used to measure the difference between an image and compressed version of the image. A feature vector consisting of these distortion metrics for several compression rates (using *JPEG*2000) is used to train the classifier. False alarm and missed detection rates are about 18 %.

Lyu *et al*. [20] analyzed that steganographic messages can be embedded into digital images in ways that are imperceptible to the human eye. These messages, however, alter the underlying statistics of an image. Author previously built statistical models using first and higher-order wavelet statistics, and employed a non-linear *SVM* to detect steganographic messages. In this paper author extend these results to exploit color statistics, and show how a one-class *SVM* greatly simplifies the training stage of the classifier.

A thorough detection theoretic analysis of steganography was recently presented by Wang *et al.* [21]. Although the emphasis is on steganalysis of block-based schemes, authors make general observations of the detectability of Spread Spectrum (*SS)* and *QIM*. It is shown for Gaussian covers that spread spectrum hiding can be made to have zero divergence ($\varepsilon$ = 0). However this is not clear if this extends to arbitrary distributions, and additionally requires the receiver to know the cover distribution, which is not typically assumed for steganography. It is shown that *QIM* generally is not secure. Authors suggest alternatively hiding schemes that can achieve zero divergence under certain assumptions, though the effect on the rate of hiding and robustness is not immediately transparent.

Moulin *et al.* [22] address the secure hiding rate and derive an information theoretic capacity for secure hiding for a specified cover distribution and distortion constraints on hider and attacker. The capacity is explicitly derived for a Bernoulli (1/2) (coin toss) cover distribution and Hamming distance distortion constraint, and capacity achieving codes are derived. However for more complex cover distributions and distortion constraints, the derivation of capacity is not at all trivial.

Kharrazi *et al.* [23] analyzed that there have been a number of steganography embedding techniques proposed over the past few years. In turn the development of these techniques has led to an increased interest in steganalysis techniques. More specifically Universal steganalysis techniques have become more attractive since they

work independently of the embedding technique. Authors compare a number of universal steganalysis techniques proposed in the literature which include techniques based on binary similarity measures, wavelet coefficients' statistics, and Discrete Cosine Transform (*DCT)* based image features. These universal steganalysis techniques are tested against a number of well know embedding techniques, including Outguess, *F5*, Model based and perturbed quantization. Experiments are done using a large dataset of *JPEG* images, obtained by randomly crawling a set of publicly available websites. The image dataset is categorized with respect to the size and quality. They benchmark embedding rate versus detectability performances of several widely used embedding as well as universal steganalysis techniques.

Sallee [24] presents methods for performing steganography and steganalysis using a statistical model of the cover medium. The methodology is general, and can be applied to virtually any type of media. It provides answers for some fundamental questions that have not been fully addressed by previous steganographic methods, such as how large a message can be hidden without risking detection by certain statistical methods, and how to achieve this maximum capacity. Current steganographic methods have been shown to be insecure against simple statistical attacks. Using the model-based methodology, an example steganography method is proposed for *JPEG* images that achieves a higher embedding efficiency and message capacity than previous methods while remaining secure against first order statistical attacks. A method is also described for defending against blockiness steganalysis attacks. Finally, a model-based steganalysis method is presented for estimating the length of messages hidden with *Jsteg* in *JPEG* images.

Farid *et al*. [25] analyzed that techniques for information hiding are becoming increasingly more sophisticated and widespread. With high-resolution digital images as carriers, detecting hidden massages is also becoming considerably more difficult. Authors describe a universal approach to steganalysis for detecting the presence of hidden messages embedded within digital images. Authors show that, within multi-scale, multi-orientation image decomposition (i.e. wavelets), first and higher order magnitude and phase statistics are relatively consistent across a broad range of images, but are distributed by the presence of embedded hidden messages.

Sullivan *et al*. [26] analyze that the difficult task of steganalysis, or the detection of the presence of hidden data, can be greatly aided by exploiting the correlations inherent in typical host or cover signals. In particular, several effective image steganalysis techniques are based on the strong inter pixel dependencies exhibited by natural images. Thus, existing theoretical benchmarks based on independent and identically distributed (*i.i.d*.) models for the cover data underestimate attainable steganalysis performance and hence, overestimate the security of the steganography technique used for hiding the data. Authors investigate detection-theoretic performance benchmarks for steganalysis when the cover data are modeled as a Markov chain. The main application explored here is the steganalysis of the data hidden in the images.

Ker [27] proposed that steganalysis methods for extensions of *LSB* overwriting to both of the two lowest bit planes in digital images: there are two distinct embedding paradigms. The author investigates how detectors for standard *LSB* replacement can be adapted to such embedding, and how the methods of structural steganalysis , which gives the most sensitive detectors for *LSB* replacement, may be extended and applied to make more sensitive purpose-built for detectors for two bit plane steganography. The literature contains one other detector specialized to detect replacement multiple bits, and those presented here substantially more sensitive. The author also compares the delectability of standard *LSB* embedding with the two methods of embedding in lower two bit planes: although the novel detectors have a high accuracy from the steganographer's point of view, the empirical results indicate that embedding in the two lowest bit planes is preferable to embedding one.

Cai *et al*. [28] analyzed that steganalysis has becoming an emerging technique for detecting secret messages that are embedded in a clean-image. Universal steganalysis is mainly, useful due to its independence of prior knowledge of the embedding procedure. However, the detection results from majority of universal methods are largely determined by the training procedure on a mixture of clean-image and stego-images, and therefore not practically feasible. Moreover, many color steganalysis methods do not take color coefficients into special consideration and thus they can be viewed as a simple extension of the analysis for grayscale images, authors proposed a novel predictor based on the intra-color and inter-color correlations of wavelet

coefficients. This method achieves higher detection rates, under a blind condition that involves clean images at the training stage.

Chamorro *et al.* [29] proposed a global steganalysis methodology by comparing the steganalysis methods proposed in the literature. The secret messages detection capacities of these steganalysis methods are evaluated using stego images generated by typical data hiding algorithms. The evaluation of steganalysis methods is realized in terms of false negative false positive error rates using 100 images. There is not any steganalysis that can detect presence of secret message in all types of stego images. Therefore, to realize a reliable analysis about a suspicious image, several methods must be efficiently combined.

Chamorro *et al.* [30] analyzed that there are many image steganalysis methods that detect presence of hidden data into stego images generated by *LSB* steganography. However if the stego images is generated by *JPEG* steganography, these methods shows inefficiency of the detection of hidden data. The *JPEG* steganography is a data hiding technique that performs data hiding in *DCT* domains. In the steganalysis side, any information about the stego image is not available, therefore many steganalysis methods for different types of stego images must be combined to generate an efficient steganalysis methodology. The evaluation of steganalysis methods is realized in terms of false negative and false positive error rates.

Bera *et al.* [31] proposed two techniques for the detection of the hidden data in the cover image. Firstly, detection done by comparing the histogram of cover and stego image in which attacker knows about the cover image without the knowledge of the coding algorithm of the stego image and secondly in the image smoothening technique, the Probability Distribution Function *(PDF)* is used for the detection. Based on the difference in statistical parameter of the stego with cover image detection is done. Thus universal steganalysis techniques seem to be the real solution since they should be able to detect stego images even when a new embedding technique is being employed. Statistical techniques are based on the detection of the hidden data knowing the statistical parameter of the cover image. The variations in the statistical parameter are the basis of the detection of the secret data.

Kekre *et al.* [33] proposed a steganalysis technique for both grayscale and color images. This technique derived from grayscale level co-occurrence matrix in spatial domain, which is sensitive to data embedding process. This matrix is derived from an image. Several combinations of diagonal elements of matrix are considered as features. There is the difference between stego and non-stego images and this characteristic is used for steganalysis. Distance measures like Absolute distance and Euclidean distance are used for classification.

Hashemi *et al.* [34] proposed a method that is shown to be of higher detection accuracy than existing truly blind steganalysis methods including Farid's and Wavelet Absolute Moment (*WAM*). This is achieved by improving some weakness of Farid's steganalysis scheme in feature extraction; that is, instead of deriving an over-determined equation for each subband of the wavelet decomposition, the subbands are divided into overlapping blocks and over-determined equation is constructed for each block. To guarantee the existence of infinite answers, the over-estimated equation is solved in a way different from Farid's by using Moore-Penrose pseudo-inverse concept. Further improvement to the performance is achieved by adding diagonal directions and increasing the number of moments.

Not surprisingly, detecting many steganalysis schemes at once is more difficult than detecting one method at a time. We used a general framework, but approach each hiding one at a time. *LSB* hiding is a natural starting point, and we begin our study of steganalysis there. Other hiding methods have received less attention hence we continue our study with *QIM*, *SS* and a version of *QIM* adapted to reduce detectability.

**2.3 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.

# Chapter 3

# Proposed Approaches and Experimental Results

In this chapter, the solution of the problem stated in previous chapter is discussed. We have proposed three approaches for checking the image, whether is it stego or not? Proposed approaches are based on the statistical properties of the image. In this, all the three approaches have their own merits and demerits. Proposed approaches are applied according to their time complexity. The approach having lowest time complexity is applied first on the input image. We apply these three approaches one by one, first apply variance, if it gives desired result then accept it. Otherwise apply correlation, if it gives result according to expectation then accept result else apply *XOR* approach.

## 3.1 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.

**Merits:**

a) Execution time is very small.
b) Time complexity is $O(n)$.

**Demerits:**

a) Poor performance (18% successful results).

Fig. 3.1: Cover image.

Fig. 3.2: Stego image.

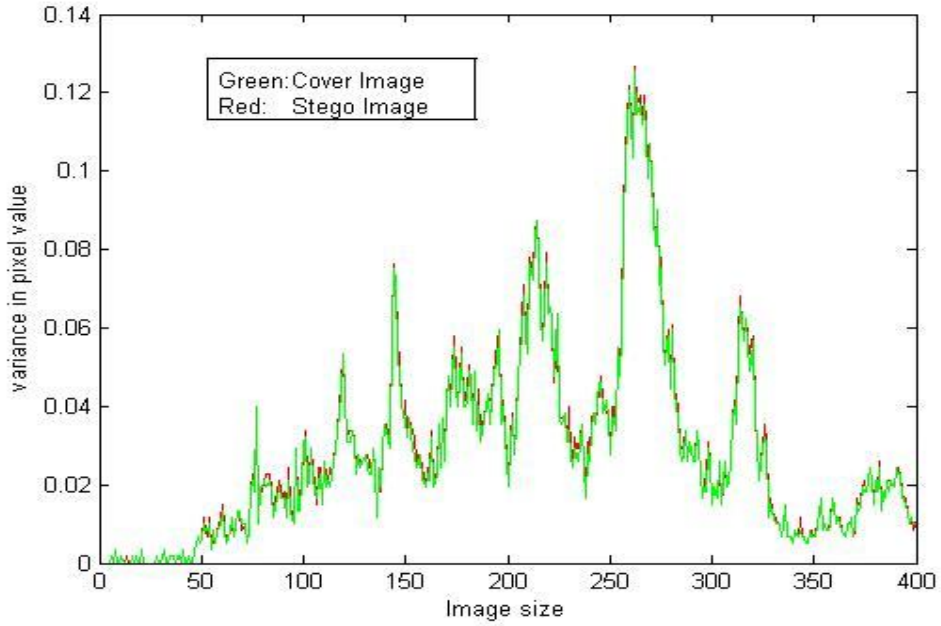



Fig. 3.3: Histogram of the variance of cover and stego images.

The histogram was taken between Fig.3.1 image and Fig.3.2 image. The histogram in fig.3.3 shows that there are two colors, green for cover image and red for stego image. So that Fig.3.2 is a stego image.

Fig. 3.4: Cover image.

Fig. 3.5: Cover image.

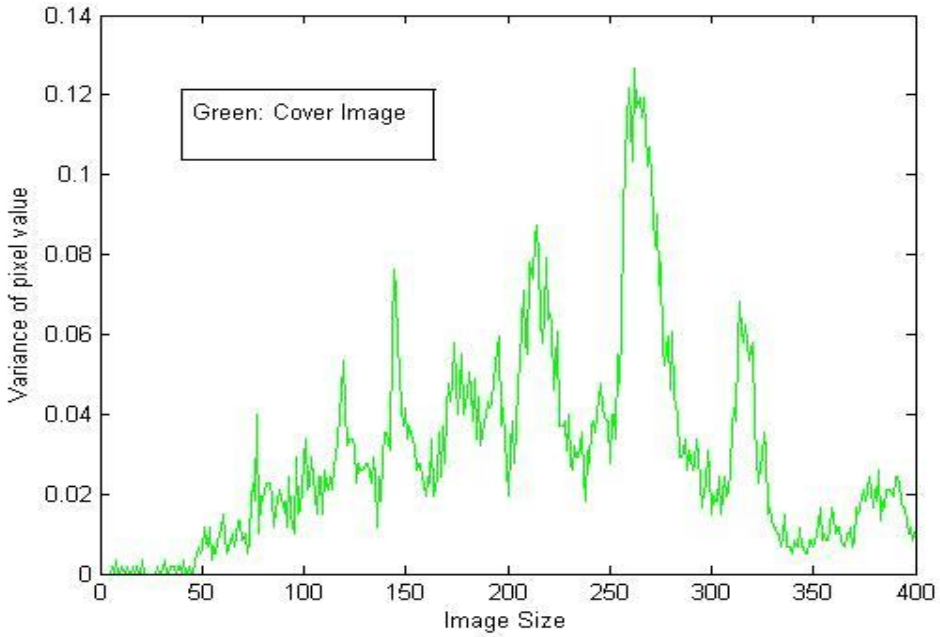



Fig. 3.6: Histogram of the variance of cover image.

Images shown in Fig. 3.4 and Fig. 3.5 are same. To check our approach, the histogram is taken between these two images. This histogram in Fig.3.6 shows that there is only single green color, which shows that no data is hidden in the image shown in Fig 3.4.

### 3.2 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.

**Merits:**

a) Execution time is less than *XOR* based approach.
b) Best performance.
c) Time complexity is $O(n^2)$.

**Demerits:**

a) Time variation is larger than the first approach.

b) Execution time is more than variance based approach.

### 3.3 *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.

**Merits:**

a) The best performance among three approaches.

b) Time complexity is $O(n^2)$.

**Demerits:**

a) Execution time is maximum.

b) Execution time variation is maximum.

To check the performance of the proposed approaches, we collected hundred images from www.freefoto.com web site [5]. Different types of images are hidden into these images using *F5*, Jsteg, Stegohide stego tool etc. Then these stego images are processed by the process of these proposed approaches. The result is summarized into Table 3.1. This table contains the execution time; and one flag attribute. If flag value is one, it shows that the approach is successfully finds out that the input image is a stego image; and zero value of flag shows that no data is hidden in the input image.

Table 3.1 *PSNR*, execution time of correlation, variance and *XOR* based approaches.

| **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 |


Fig. 3.7: Cover image.

Fig. 3.8: Secret image.

With the help of *F5* tool, secret image is hidden into cover image. The commands to run the *F5* are shown step wise.

(i) Firstly compile *F5* Main class, using the javac compiler.
javac Main.java

(ii) Then run the Main class by using java Main command

(iii) Now we run the command to embed secret data into cover image.

java main e -e bird.jpg -p 1234 -q tree.jpg stego.jpg

where

- e is used to embed data.
- bird.jpg is secret data.
- *p* is used for password which is 1234 in our case.
- *q* define quality factor for jpeg compression.
- tree.jpg is cover image.
- stego.jpg is stego image

After execution of the above command, the stego image is created, which is shown in Fig.3.9.


Fig. 3.9: Stego image.

From the table 3.1, one can analyze that variance based approach is poor in performance but it takes less execution time. Correlation based approach is better in performance but slightly time consuming and *XOR* based approach is the best among three in performance but execution time of this approach is very high as compared to the other approaches. This analysis is plotted in the following figures.

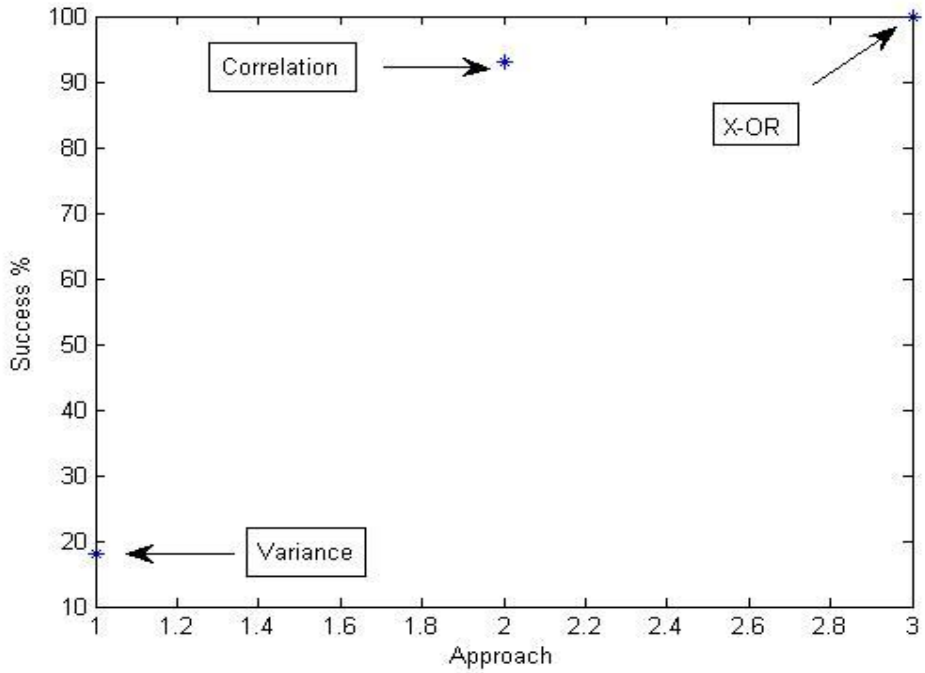



Fig. 3.10: Successful result ratio of the proposed approaches.

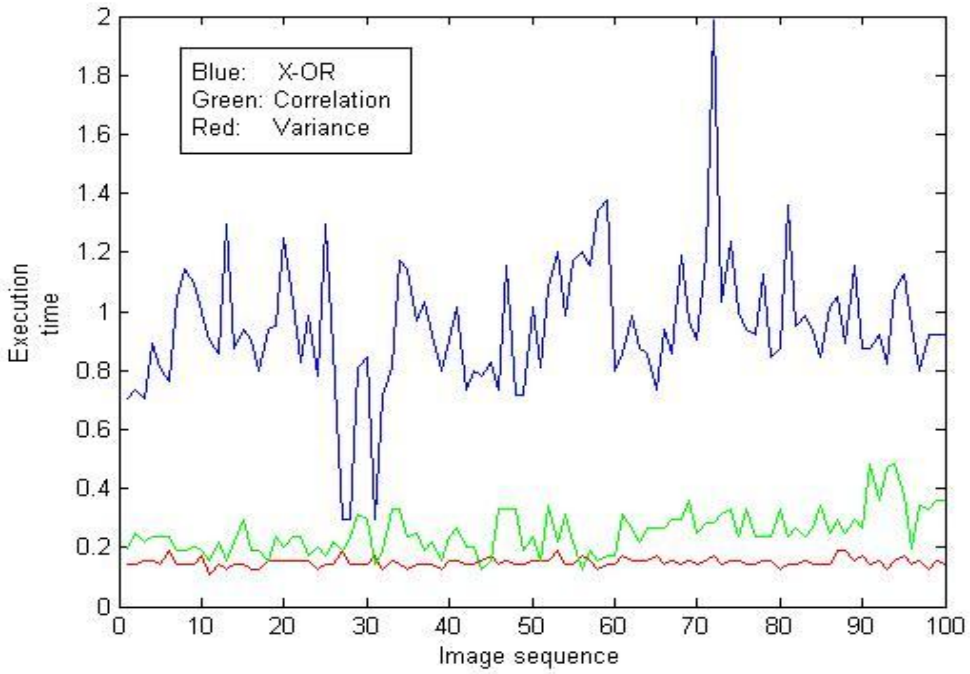



Fig. 3.11: Execution time of different images for the different approaches.

Fig. 3.10 shows the successful percentage of all the approaches. It is clear that variance based approach has minimum successful rate (18%) and correlation based approach (93%) is higher than this, and *XOR* based approach (100%) is the superior

among three. Fig. 3.11 shows the variation of execution time (in sec) of all the images taken from dataset. Red color is for variance based approach, green color is for correlation based approach and blue color is for *XOR* based approach. *XOR* based approach shows maximum time variation that implies, it must be used when the other two cases are failed.

We have compared our approaches with Farid's method. For this purpose, we have taken the average of success cases of our second and third approach. This comparison is shown in Table 3.2.

Table 3.2: Comparison of the proposed approaches with Farid's method

| Steganalysis / Steganography | Proposed (averages of second and third approaches) | Farid's Method [25] (average of blind and non blind steganalysis) |
|---|---|---|
| *F5* tool | 96% | 80% |

From this Table, one can observe that the correct classification rate achieved by our propsed system is 96% as compared to the Farid's approach.

# Chapter 4
# Conclusions and Future Work

The proliferation of steganographic tools has created a demand for powerful means to detect hidden data. The primary focus of this thesis 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.

## Bibliography


[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.