Republic of Iraq
Ministry of Higher Education and Scientific Research
University of Babylon
College of Information Technology
Software Department

DETECTION OF STEGO IMAGE BASED ON CORRELATED ANALYSIS

A Dissertation

Submitted to the Council of the College of Information Technology, University of Babylon in Partial Fulfillment of the Requirements for the Doctor of Philosophy Degree in Information Technology/Software

By

Natiq Mutashar Abdali Hussain

Supervised by

Prof. Dr. Zahir M. Hussain

2022A.D. **1443 A.H.**

بسم الله الرحمن الرحيم

اللهُ نُورُ السَّمَاوَاتِ وَالْأَرْضِ مَثَلُ نُورِهِ كَمِشْكَاةٍ فِيهَا مِصْبَاحٌ الْمِصْبَاحُ فِي زُجَاجَةٍ الزُّجَاجَةُ كَأَنَّهَا كَوْكَبٌ دُرِّيٌّ يُوقَدُ مِن شَجَرَةٍ مُّبَارَكَةٍ زَيْتُونِةٍ لَّا شَرْقِيَّةٍ وَلَا غَرْبِيَّةٍ يَكَادُ زَيْتُهَا يُضِيءُ وَلَوْ لَمْ تَمْسَسْهُ نَارٌ نُورٌ عَلَى نُورٍ يَهْدِي اللهُ لِنُورِهِ مَن يَشَاء وَيَضْرِبُ اللهُ الْأَمْثَالَ لِلنَّاسِ وَاللهُ بِكُلِّ شَيْءٍ عَلِيمٌ (35)

صدق الله العلي العظيم

سورة النور– اية 35

Declaration

I hereby declare that this dissertation, submitted to the University of Babylon as fulfillment of requirements for the degree of doctor of Philosophy in Information Technology\ Software has not been submitted as an exercise for a similar degree at any other university. I also certify that the work described here is entirely my own.

Signature:

Name: Natiq Mutashar Abdali Hussain

Date: \ 6 \ 2022

Supervisor Certification

I certify that the dissertation entitled **(Detection of Stego Image Based on Correlated Analysis)** was prepared under my supervision at the department of Software/ College of Information Technology/ University of Babylon as partial fulfillment of the requirements of the degree of Doctor of Philosophy in Information Technology-Software.

Signature:
Supervisor Name: Prof. Dr. Zahir M. Hussain
Date: / 6 / 2022

Head of the Department Certification

In view of the available recommendations, I forward the thesis entitled **"Detection of Stego Image Based on Correlated Analysis"** for debate by the examination committee.

Signature:
Asst. Prof. Dr. Ahmed Saleam
Head of Software Department
Date: / 6 / 2022

Certification of the Examination Committee

We, the undersigned, certify that (Natiq Mutashar Abdali Hussain) candidate for the degree of Doctor of Philosophy in Information Technology-Software, has presented his dissertation of the following title (**Detection of Stego Image Based on Correlated Analysis**) as it appears on the title page and front cover of the dissertation that the said dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on: (22 / 5 /2022).

Signature:
Name: Dr. Sattar B. Sadkhan
Title: Professor
Date: / / 2022
(**Chairman**)

Signature:
Name: Dr. Suhad Ahmed Ali
Title: Professor
Date: / / 2022
(**Member**)

Signature:
Name: Dr. Kadhim Mahdi Hashim
Title: Professor
Date: / / 2022
(**Member**)

Signature:
Name: Dr. Mohammed Najm Abdullah
Title: Assistant Professor
Date: / / 2022
(**Member**)

Signature:
Name: Dr. Loay E. George
Title: Assistant Professor
Date: / / 2022
(**Member**)

Signature:
Name: Dr. Zahir M. Hussain
Title: Professor
Date: / / 2022
(**Member and Supervisor**)

Approved by the Dean of the College of Information Technology, University of Babylon.

Signature:
Name: Dr. Hussein Atiya Lafta
Title: Professor
Date: / / 2022
(Dean of Collage of Information Technology)

Dedication

To the sake of Allah, my Creator, and my Master,

to my parents and my kids,

to my wife, my pillar of support. Thanks for holding me when I fell and helping me rise above and beyond,

to my brothers and sisters, who supported me until the completion of this research,

and to my friends, colleagues, and relatives,

I dedicate this work.

Acknowledgement

First and foremost, I would like to thank my God, Allah Almighty, for giving me endless graces. My deep sense of gratitude goes to the beacon of science, to the master of creatures, to the greatest Prophet, Mohammed (Peace be upon Him and His Family).

I take this opportunity to express my sincere gratitude and greatest appreciation to my supervisor **Prof. Dr. Zahir M. Hussain** for his continuous support for my Ph.D. study, his patience, motivation, enthusiasm, and immense knowledge. The words are inadequate and I can't say thank you enough for his tremendous support and help. I feel motivated and encouraged every time I attend his meeting. His tireless guidance has helped me immensely in researching and writing this dissertation.

Also, I would like to show my gratitude to who weaved my happiness from strings woven from her heart to my dear Mother. My thanks and appreciations also go to my family, for their encouragement, support and patience.

Moreover, I wish to express my love and gratitude to my beloved wife for her understanding and endless love through the duration of my study.

Finally, I would also like to express my thanks and gratitude to all those who contributed to making this dissertation done, and foremost among them all the teaching and staff members at the College of Information Technology at the University of Babylon.

Abstract

Steganography and Steganalysis are two sides of the same coin. The techniques of steganography and steganalysis are developed at the same time. Steganography is a technique for concealing secreted data and making it invisible by embedding it in a multimedia range. Image steganalysis is a technique for determining whether or not images contain concealed information. Because developed steganographic methods can payload tiny secret messages into covers, state-of-the-art image steganalysis techniques with the improvement of steganography technology cannot distinguish between the cover and stego images or are detected with lower accuracy performance, making it challenging.

Recent research has demonstrated the effectiveness of utilizing neural networks to detect stego images. However, because accessing a database is complex, which is needed in the classification process to detect whether image is cover or stego image. The current study presents a technique to address these problems for detecting stego image in the spatial domain.

The proposed system depends on the statistical properties of the input image. The system depends on deriving the auto-correlation function of the image histogram, then applying a high-pass filter as a threshold to evaluate it. This technique can be used to decide which image is a cover or a stego without adopting the original image. Although this study has focused on least-significant-bit (LSB) steganography, it is found that the proposed approach could be applied successfully on sequentially and randomly LSB steganography with different orders of histogram-correlation derivatives. Also, the ratio stego-image to cover limits is considered, where tiny ratios can escape this detection method unless

modified. We further this strategy has been examined for other image steganographic ways. The results have eventually revealed the validity of this universal steganalysis system. We use a different threshold to evaluate the proposed system. These thresholds depend on the format of the input image. The results found that the high-pass filter is more suitable than other thresholds when using the grayscale as an input image and the entropy correlation derivatives of the image histogram are appropriate for a color image. We found the proposed system is better when comparing the performance of our proposed method with other specific methods in terms of recognizing sequentially or random embedded messages and the capacity of detecting messages. The proposed method is universal but doesn’t use a dataset in the processing phase to classify images into cover or stego images.

The suggested method's outcomes are evaluated using five steganographic approaches. The method capable to detect secret message if the message size is small (Rm > 0.01). The programming language of the proposed system is MATLAB R2021a (64-bit).

Declaration Associated with this Dissertation

Some of the works presented in this dissertation have been published as listed below.

1. Natiq. M. Abdali and Zahir. M. Hussain, "Reference-free Detection of LSB Steganography Using Histogram Analysis," 2020 30th International Telecommunication Networks and Applications Conference (ITNAC), 25-27 Nov, IEEE Xplore Press, Melbourne, VIC, Australia, pp.1-7, 2020, doi: 10.1109/ITNAC50341.2020.9315037.

2. Natiq. M. Abdali and Zahir. M. Hussain, " Reference-free differential histogram-correlative detection of steganography: performance analysis," Indonesian Journal of Electrical Engineering and Computer Science, vol 25, no.1, pp. 329-338, Jan. 2022, doi: 10.11591/ijeecs.v25.i1.pp329-338.

Table of Contents

List of Tables

List of Figures

.

List of Algorithms

List of Abbreviations

Abbreviation	Description
abs	Absolute
ACF	autocorrelation function
AES	advanced encryption standard
BMP	bitmap image file
bpp	bits per pixel
CWT	Complex Wavelet Transform
DCT	Discrete Cosine Transform
DES	Data Encryption Standard
DFT	Discrete Fourier Transform
DNA	Deoxyribonucleic acid
DWT	Discrete Wavelet Transform
EMD	Exploiting Modification Direction
FIR	finite impulse response
FT	Fourier Transform
FTP	file transfer protocol
GIF	Graphics Interchange Format
hex	Hexadecimal system
HP	high-pass
HPF	high-pass filtered
HSV	hue, saturation, value
IDEA	International Data Encryption Algorithm

IIR	infinite impulse response
IWT	Integer Wavelet Transform
JPEG	Joint Photographic Experts Group
Km	number of different hidden message sizes
LP	Low-pass
LSBs	Least Significant Bits
max	Maximize
mH	Maximize high-pass region
mL	Maximize low-pass region
MPVD	Modified Pixel-value differencing
MSE	mean squared error
PCX	Picture Exchange
PDH	pixel difference histogram
POVs	Pairs of Values
PSNR	Peak signal-to-noise ratio
PVD	Pixel-value differencing
RGB	Red Green Blue
Rm ratio	message-to-cover ratio
RQP	Raw Quick Pair
RS	Regular and Singular Groups
STD	Standard Deviation
SVM	Support Vector Machine
WNS	White Noise Storm
WOW	Wavelet Obtained Weights

List of Symbols

Symbol	Description
C	Cover
D	Decoding process
d1	First derivative
d2	Second derivative
d3	Third derivative
E	Embedding process
K	Secret Key
M	Message

Chapter One

General Introduction

Chapter One
General Introduction

1.1 Introduction

Data Hiding pervades practically every part of our lives, whether for good or malicious reasons. According to David Kahn and several historians, it started from hidden writing hundreds of years ago. It began in Egyptian civilization as hieroglyphs, which was supposed to represent past timeframes for certain lords. Other cultures at that period, such as the Chinese, used a more physical method to concealing communications by handwriting them on silk or paper, rolling them into a ball, and coating them with wax to transmit political or military truths. The ball was even swallowed during transportation as an extra safety step. Hidden transmissions got more complex as civilization progressed, and cryptograms and anagrams advanced.

Digital music, podcasts, online and recorded webinars, video chats, and movies have altered the way of interaction today and are found in practically every company. These techniques are used to communicate ideas, develop personnel, build strong customer relationships, and, of course, amuse. Is it true that digital media poses a threat? Could these pathways be used to exchange information discreetly, remove copyrights, provide insider knowledge, transmit command and control information, or offer the technology needed to combat advanced threats [1].

This scientific danger to community safety is a genuine one. As a result, researchers have investigated and invented clever image steganalysis approaches to avoid and assess this danger. However, few articles were done about the difficulty of receiving a database. Image steganography is a

famous confidential communication technique., which involves concealing data in an image [2].

Steganalysis is the discipline of finding hidden data or messages in an electronic cover and discriminating between stego item and cover product with no or little knowledge of steganography techniques, similar to cryptanalysis, which focuses on cryptography. The goal of steganalysis is to collect evidence that points to the presence of a secret message [3].

1.2 Related Works

In this section, a number of researchers' works are discussed which are related to steganalysis system techniques.

In 2000, Westfeld and Pfitzmann [4] presented the first statistical steganalysis approach. This method identifies Pairs of Values (POVs) transferred during secret data covering. Values of the pixel, quantized coefficients of DCT, and palette indices that vary in the least significant bit are examples of POVs. According to the Chi-squared method, the frequencies of each of the two-pixel values in each POV tip are distant from a POV's mean. It is approach finds close-equal POVs in images and, as a result, embedded data. This approach consistently recognizes successively embedded information, but it fails miserably if the information is random.

In 2000, The Raw Quick Pair (RQP) approach, presented by Fridrich and Long [5], was developed to identify the Least Significant Bit encoding in true-color photos. RQP examines LSB embedding-created near pairings of colors. Close color couples demonstrate that the only difference between two colors is at the LSB. The frequency of nearby color couples rises as the process of embedding messages into images increases. As a result, it may classify an image as stego or not by calculating a numeral of nearby color schemes. The authors demonstrated that a significant degree of detection

reliability might be achieved even with encrypted data capacities of 0.1–0.3 bits per pixel. This approach has the limitation of being limited to color images.

In 2001, As a result, Fridrich and Goljan introduced RS steganalysis [6], a novel technique for detecting the Least Significant Bit encoding in grayscale and color photos. This method separates the image into groups and then measures the noise in each group. The LSBs of a defined set of pixels within each group are flipped (by employing a mask, i.e. the scheme of pixels to flip). Each group is categorized as regular or singular depending on whether the pixel clutter inside it is decreased or increased. For a dual kind of flipping, the categorization is repeated. As a result, the RS steganalysis approach to be additional trustworthy from the Chi-square technique.

In 2003, Zhang and Ping [7] suggested another approach for noncolored photos. The system is based on the histogram of different images. Translation coefficients through difference photo histograms were used to detect the poor correlation between the least significant bit (LSB) plane and the other bit planes. This metric was then utilized to build a classifier that could distinguish between the stego and the clear images. The encoding capacity ranged from 0% to 100% in 10% increments, with the highest detection ratio at 96.03 %. The suggested approach performs well for random and sequential LSB substitution, with improved performance, computation speed compared to RS analysis.

In 2004, Celik et al. [8] proposed a different strategy. They created a feature set based on the rate-distortion features of images built on the view that secret data encoding enlargements photo entropy and other concealing methods generate tiny undetectable distortions. This feature set was used to train a Bayesian classifier, then used to describe images as evident or stego after a Karhunen–Loeve transform. The embedding rates were 0.1, 0.2, 0.4, 0.6, 0.8, and 1.0 bits per pixel (bpp), and 27% of the cover images were

mislabeled as stego-images, with the miss rate dropping as the embedding rate went up.

In 2005, Andrew D. Ker [9] establishes a comprehensive framework for structural analysis of LSB substitution finding and message length assessment. They combined several previously known structural detectors and demonstrated a more precise approach. They also used various images in JPEG and BMP formats to test other recommended tactics and compare them to other ways; they also discovered that the new algorithm performs well in a more significant number of circumstances where message length exceeds 2%.

In 2009, Malekmohamadi and Ghaemmaghami [10] provided a grayscale image steganalysis approach based on spatial and Gabor features. They looked at spatially correlations across pixels in unclean and clear images for feature selection. Gabor filter coefficients were used to produce input data for a training model, and those characteristics were used to train an SVM classifier. The entire image's first and higher-order statistics, as well as its DCT transform, were used. After that, the trained model was applied to unseen changes and clear photos. According to the findings, the algorithm had an extremely accurate detection rate of 93 % for changed pictures and 96 % for pure photos, with an encoding ratio of 14.1 %.

In 2010, Zhang et al. [11] suggested an LSB matching steganalysis approach based on the statistical model of pixel differences distributions. The number of non-zero difference values from stego-images compared to the number of zero-difference values. As a result, the classifying characteristic estimates the relative error between the expected and actual values of the number of the zero-difference value.

In 2013, Cho et al., [12] proposed the method to identify stego from its clean image; researchers employed a technique of deconstructed photo blocks. The authors divided the stego into groups, organised the groups into numerous classes, and then a classifier. The entire picture was then classified as stego or clean by combining the findings blocks for the whole image with the judgment phase. The authors show that the execution of this method is minimally susceptible to judgment fusion approaches and additional sensitivity to classifier selection.

In 2014, The histogram of multi-order differences was smoothed using LSB matching. Xia et al., employed the co-occurrence array to depict the discrepancies with tiny actual values to extract features by this discovery. SVM classifiers were trained using these parameters to identify stego photos from the originals. Results of this method, with encoding ratio ranging from 0.1 to 1.0 bpp [13].

In 2015, Goljan et al., [14], a close edition of the spatial rich model for steganalysis of color images, was introduced. Three-dimensional co-occurrences of residuals calculated from all three-color channels were utilized to extract the additional features. These features can capture cross-color channel interdependence. Tests were done on different datasets: various color versions of BOSSBase v1.01-with an encoding ratio of 0.1 bpp for LSB Matching, 0.4 bpp for WOW-with an embedding rate of 0.1 bpp for LSB Matching and 0.4 bpp for WOW. In these experiments, the suggested 18,157 characteristics were quite effective at detecting LSB Matching steganography in images. The detection error for one payload is 0.0297–0.1790, whereas the detection error for other loads is similar (0.05–0.5 bpc).

In 2016, Tang et al., [15] Proposed a pixel embedding probabilities-based adaptive steganalysis technique on the BOSSBase dataset that was subjected to various encoding ratios (0.05–0.5 bpp). Experiments with

different adaptive approaches have revealed that the proposed system is effective, especially at low insertion rates and less than 0.20 bpp.

In 2017, Nouri and Mansouri, [16] created a method for constructing the steganalysis feature vector by modifying the singular value curve; two spatial feature vectors were recovered in the statistical exploitation indicated. According to testing findings on photos encoded with relative payloads (0.05-0.4 bpp), the recommended feature vectors work adequately for both universal and JPEG-based steganalysis algorithms.

In 2018, Li et al., [17] ReSt-Net is a steganalysis approach based on multiple activation modules and parallel subnet-based CNN, which was suggested. Diverse activation modules (DAMs) are used in their design to activate the convolution outputs in various ways before aggregating the results for subsequent layers. Rather than increasing the number of filters in the pre-processing layers, the network employed more sub-nets with fewer filters. Pre-trained the subnets individually to speed up the training process. S-UNIWARD, HILL, and CMD-HILL were used to evaluate the findings, which ranged from 65.67 % to 85.44 %, 62.38 % to 81.66 %, and 58.92 % to 79.16 %, respectively.

In 2019, Liu et al. [18] suggested a new steganalysis approach based on the binary bat Algorithm and employing a nature-inspired FS technique. To improve detection accuracy, this technique selects the most compelling feature subset from the features extracted by the Subtractive Pixel Adjacency Matrix (SPAM) [19] method. The suggested technique improves detection while reducing redundant features, according to the results of the experiments. This strategy was tested using HILL, WOW, and HUGO steganographic approaches at a 40% bpp embedding rate. The detection accuracies of the findings were 64.11 %, 68.08 %, and 64.07 %, respectively.

In 2020, Xiang et al. [20] suggested a CNN-based steganalysis approach that contains two contributions: First, by adding additional convolutional

layers to the lower parts of the network, it provides a new structure of convolutional and pooling layers to analyse local information better than other CNN-based models in steganalysis. Second, the (GAV) pooling layer is placed before the softmax layer, rather than before the fully connected layer, which is the best position for steganalysis. The trials revealed that when compared to previous CNN-based steganalysis approaches, the suggested strategy achieves the best results. Satisfying outcomes for S-UNIWARD and WOW steganographic methods ranged from 57.7% to 81.79 % and 63.91 % to 86.17 % at 10%, 20%, 30%, and 40% bpp embedding rates, respectively.

In 2021, Lin et al. [21] suggested a method for recognizing MPVD steganography successfully. They presented a way for MPVD steganalysis based on the chi-square fit of the model since relevant works on PVD steganography have mostly highlighted embedding capacity and image quality and concentrated on preventing attacks by RS and PDH study. They tested the suggested technique with 1,000 images and discovered that it outperformed existing state-of-the-art techniques across various image datasets, embedding ratios, and classification methods. In the trials, this technique performed substantially better in terms of accurate results at low insertion ratios. As a result, this approach may be utilized to do steganalysis of MPVD steganography and is a viable alternative to the regularly used RS and PDH analysis methods.

Table (1.1) presents a summary of all the related works and illustrates the algorithm of detecting, datasets, and measurement of evaluation.

Table (1.1): Summary of the Mentioned Related Works

Ref. No.	Technique of steganalysis	Dataset	Accuracy – Mistake ratio Detection ratio –
[4]	Chi-squared detects of POVs	No database used	Various tests depend on the steganography tool (Steganos, S-Tools, Jsteg, EZStego) and the size of the embedding message
[5]	Raw Quick Pairs method. Statistical analysis of the image colors in the RGB cube	Color images, 350×250 pixels, stored as JPEGs	Several tests showing threshold and error probability for numerous different test message sizes and different secret message sizes.
[6]	RS steganalysis	No database used	Multiple tests and outcomes are dependent on the starting bias, software of technique (Steganos, Hide4PGP, S-Tools), photo utilized.
[7]	Coefficients of the Gabor filter and statistics of the gray-scale class co-occurrence array of photos as characteristics. As a classifier SVM	Images from USC-SIPI in grayscale	For cover and stego images, the medium detection ratio is 94.50 %. The embedding rate is 0.141 bits per second.
[8]	The distributions of pixel differences are statistically simulated.	NRCS image database	50–100% embedding rate: 68.4 8–98.27 % True Positive and False Positive
[9]	Analysis of LSB substitution.	UCID image database	Several experiments and outcomes depend on the system and classifier.
[10]	To extract features, a co-occurrence matrix was utilized to represent the differences with a modest absolute value. As a classifier, SVM is used.	NRCS, BOSSBase v0.92	Various tests for detecting HUGO were assessed using the "detection reliability" p (p = 2A-1, where A is the area beneath the ROC curve).
[11]	Three-dimensional co-occurrences of residuals calculated from all	BOSSBase v1.01	On variants of the BossBase dataset and its grayscale versions, several experiments with varying encoding ratios

	three-color channels were used to extract additional features.		(0.05–0.5bpp) with medium detection mistake as a measure.
[12]	Fisher criterion-based feature selection approach	BOSSBase v1.02	Different outcomes are obtained based on the encoding rate (bpp) and steganography technique.
[13]	The steganalysis feature vector was created by altering the singular value curve.	UCID	Embedding rates of 0.05, 0.1, 0.2, and 0.4 bits per second were used. Many steganographic algorithms yielded various outcomes. Several relevant feature extraction techniques are compared.
[14]	Additional features extracted by three-dimensional co-occurrences of residuals computed from all three-color channels.	BOSSBase v1.01 (7)	Various tests for different embedding rates (0.05–0.5bpp) with average detection error as metric.
[15]	Feature selection method based on the Fisher criterion, in which the separability of single-dimension and multiple dimension features, combined with measurement of the Euclidean distance, is analyzed.	BOSSBase v1.02 (7)	Numerous results dependent on embedding ratio (bpp) and embedding technique.
[16]	Variation of singular value curve was used to construct the steganalysis feature vector.	UCID (11)	Embedding rates of 0.05, 0.1, 0.2 and 0.4 bpp. Numerous outcomes on different steganographic algorithms.
[17]	ReSt-Net, is a steganalysis technique based on multiple activation modules and parallel subnet-based CNN.	BOSSbase ver. 1.01	The detection accuracies of the findings ranged from 65.67 % to 85.44 %, 62.38 % to 81.66 %, and 58.92 % to 79.16 %
[18]	Binary Bat Algorithm (BBA)	BOSSbase ver. 1.01	Approach evaluated using HILL, WOW, and HUGO steganographic methods at 40% bpp embedding rate. The results satisfied 64.11%, 68.08% and 64.07% detection accuracies.

[20]	CNN-based steganalysis approach	BOSSbase ver. 1.01	Satisfying outcomes for S-UNIWARD and WOW steganographic methods ranged from 57.7% to 81.79 % and 63.91 % to 86.17 % at 10%, 20%, 30%, and 40% bpp embedding rates, respectively.
[21]	Method for recognizing MPVD steganography	BOSSbase ver. 1.01	Approach evaluated using MPVD steganographic methods at 10%, 20%, 30%, and 40% bpp embedding rate. The results satisfied 90.03%, 95.09% and 99.06% detection accuracies.

1.3 Problem Statement

Currently, there are billions of photos found online. Individuals utilize these photographs to record real-life events, express their emotions and pursue other desires. Unfortunately, due to notable advances in image application and its related software, terrorist gangs may now communicate messages using normal image communication with great efficiency. Since this steganographic software aims to disguise the payload as a random image-noise process produced by the electronic systems of the camera, identification of images related to illegal activity is difficult using the human eye alone.

In addition, most previous detecting methods depend on the knowledge of original images by taking the features from the availability of datasets and depend on the format of tested images. Then, classified the image as stego or cover.

As a consequence, developing a detecting method, which are updated and independent of any steganographic method and without any previous hint about the concealed content or the embedding process has become essential.

1.4 Dissertation Objectives

Steganalysis aims to build a technique that can assess whether an image includes a secret message while remaining completely blind. The following goals are established to achieve this overall aim.

1. Design an image steganalysis approach capable of identifying the existence of secret information from a variety of steganographic algorithms, even with low embedding rates.
2. Prepare image datasets for steganalysis, involving images embedded at various embedding rates and using different steganographic algorithms.
3. Propose a method to detect a stego image without adopting the original image (dataset).
4. Prove that the proposed method for derivative histogram autocorrelation techniques and the proposed method as image classifiers, whether an image if stego or cover, can be effective.

1.5 Dissertation Contribution

The main contribution of this dissertation is the designing and implementation of a trusted approach to detect the secret message in an image. The following are the main contributions: -

1. Build image datasets for detecting image steganography purposes, including images embedded at different insertion rates. It uses K_LSB, enhanced_LSB, Chaotic_LSB, MPVD, and ICA steganographic methods.
2. The system is used to detect different image steganography that uses different image formats like BMP, RGB, BMP palette, and grayscale PNG.

3. Prove the proposed system does not rely on the availability of original images (dataset).

1.6 Dissertation Outline

This dissertation contains several chapters. It is organized as follows:-

Chapter Two: an overview of information hiding, the steganalysis techniques related to digital images, and the general concept of correlation coefficients.

Chapter Three: provides the details of the proposed system, including the block diagram and the algorithms to implement the system.

Chapter Four: discusses the experimental results of the proposed system and compares it with other works.

Chapter Five: explains the conclusions of the dissertation, and suggestions for future works.

Chapter Two

THEORETICAL BACKGROUND

CHAPTER TWO
THEORETICAL BACKGROUND

2.1 Introduction

The art of hiding secret data under a digital cover for covert communication is known as data hiding. It refers to approaches for embedding secret data into a digital object (usually referred to as a cover), such as a digital image, video, or audio, by slightly altering the cover item's noise-like component. The resultant object (titled stego) holding the secret data will not introduce visible artifacts. It should be sent to the data receiver via an insecure channel such as a social network. The data receiver can reconstruct secret data from the stego object by utilizing a secret key to drive the data extraction operation [1].

Even when a piece of evidence is found, successful criminals are capable of committing crimes and then getting away with them. Many of these accomplished criminals are well-known in the news and are being investigated by local and federal law enforcement on a regular basis. Even if the crimes are attributed to them, their effectiveness depends on not producing evidence against them.

The best criminals are almost unknown. They aren't in the news. They don't brag about their wrongdoings. They are not being investigated by investigators. In reality, because their illegal conduct is hidden from view, no one is aware of their involvement. These thieves are the most difficult to track down because their conversations are hidden behind layers of technology and physical protection. Not only they do conceal their involvement in crimes, but they also conceal the conversations that assisted the crimes.

Hiding communications is such a powerful tool for facilitating covert operations that it has been utilized for millennia for criminal purposes, national goals, and aberrant conduct. When investigators investigate in the appropriate areas, at the right time, and with the proper mindset, they might occasionally uncover these secret messages. Rarely will an investigator discover concealed data without previously searching for it or being informed where it is located [22].

One of the essential characteristics of an information society is information communication. As a critical component of information communication, secure communication safeguards important state data, critical data in trade, and individual information privacy, which are essential for the country, community, and individuals. Transmission protection and secrecy are utilized with martial objectives and for real-life, as communication for mobile and online communication. [23].

This chapter is organized as follows: The Characterization of Information Hiding is introduced in section 2. In section 3, Image steganalysis is explained. Correlation effects are introduced in section 4. In section 5 autocorrelation function is described. Image processing in the spatial domain is described in section 6. First and second-order derivatives operations are introduced in section 7. In section 8, the Histogram of an image is explained. The entropy concept is described in section 9. In section 10, the Butterworth filter concept is explained. The Fourier transforms is described in section 11. Then, Measures of Evaluation are introduced in section 12.

2.2 Information Hiding

Data embedding is a fundamental concept built on age-old steganography, known as data covering. Data obscuring is sometimes known as "information concealment" or "information veiling." In general,

steganography tries to conceal the existence of helpful or significant messages by embedding them in other data [24,25].

As the name indicates, information hiding technology embeds private information in cover data with a degree of integrity that appears standard. Any digital medium can transmit personal letters and public records (also information).

Hostile attackers cannot know if the sent data contains sensitive information. The host data, including sensitive data, do not draw notice or raise worry. The risk of harm is reduced, identical to natural disguise, which allows animals to evade discoveries by their adversaries. The essence of information hiding technology is this.

Information hiding (also known as data embedding) is a communication problem [25] with two major components: digital media sources and transmission channels. An idea of data embedding is to hide crucial private information in general data. Petitcolas et al. [24] studied current information hiding based on applications, classifying it into numerous research categories, such as watermarking, fingerprinting, and steganography. For the purpose of copyright protection of digital media, watermarking and fingerprinting are used for authentication and authority, which are related to signal sources of communications in order to safeguard the copyright of digital media. Steganography is the art of covering or concealing secret messages, which is a covert transmission for private data to hide their presence from hostile attackers listening in on a transmission Instrument. Most information [26] handles copyright keep for digital media. Information hiding is a broad phrase that covers a wide range of topics. As illustrated in Figure (2.1), it is one of the most significant areas of study [24].

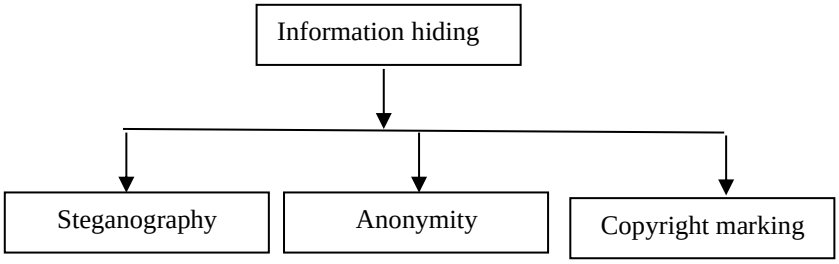

Figure (2.1): Information hiding techniques [24].

2.2.1 steganography system

A general steganography system is shown in Figure (2.2).

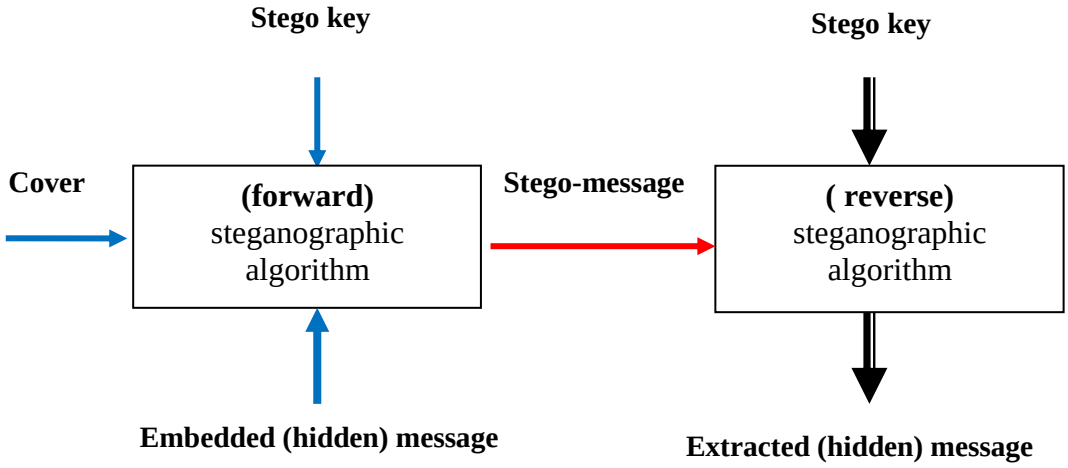

Figure (2.2) General Steganography System [27].

It is assumed that the sender aims to send a message to a receiver using a steganographic transmission. The sender begins with a cover message, which is an input to the stego-system that conceals the embedded message. The embedded message is the name of the secret message.

A steganographic key (Stego key), which is extra secret data that may be required in the hidden process, may or may not be used by the hidden method. To retrieve the encoded message again, the same key (or a similar

one) is generally required. The stego message is the result of the steganographic method. The cover and stego messages must both be of the same data type, while the embedded message can be of any data type. To retrieve the encoded message, the receiver reverses the embedding process [27].

In another word, the most popular medium that attracted steganographers is the image [28]. The availability of digital photos online, the image's sufficient redundancy to manage steganography, and the Human Visual System's HVS attributes encourage investigators to use these features in information embedding techniques are all reasons for its popularity. Though images are commonly used in steganography, the text is also a viable option [29]. There are several data-hiding technologies related to the text. Text watermarking, for example, can be used for copy-right protection. In contrast, steganography can add a "hash" to a document to prevent unauthorized use. Unfortunately, text steganography is more complex than digital images steganography due to the absence of redundancy in texts [30].

2.2.1.1 Characterization of Steganography System

Various aspects describe the advantages and disadvantages of steganographic approaches, which embed a message inside a cover. Each feature's relative relevance is determined by the application [31].

1. Robustness

If the embedded information can be reliably recognized after the image has been changed but not destroyed beyond recognition, it is robust. Linear and nonlinear filters (blurring, sharpening, median filtering), lossy compression, scaling, rotation, noise addition, color quantization (as in palette images), and other techniques are examples.

It has highlighted that attacks on the encoding scheme dependent on the embedding algorithm's knowledge or the availability of the detector function are not considered robust. Robustness refers to the ability to withstand "blind" or non-targeted changes, as well as standard image processes [31].

2. Undetectability

In most cases, this feature is necessary for secure covert communication. If the image with secret encoded data is compatible with the source model from which images are derived, we say the embedded information is undetected. For example, if a steganographic approach embeds a hidden message in the noise component of digital photos, it should do so without causing statistically significant changes in the carrier noise. The idea of Undetectability is inextricably linked to the image source's statistical model. A more comprehensive representation of the original may allow an attacker to discover the presence of personal information. Note: The capability to identify the presence does not indicate that the hidden message can be read [31].

3. Invisibility (Perceptual Transparency)

Noise modification or distortion of the cover image is required to hide the message in the over. It's critical that the embedding occurs without causing the cover's perceived quality to suffer. Even if the attacker cannot extract the message, the failure of steganographic encoding in secret communications arouses suspicion of the presence of concealed data in a stego-image. Because the integrity of the original work must be preserved, maintaining perceptual transparency in an embedded watermark for copyright protection is equally critical.

Allowing more significant distortion in the stego-image can enhance hiding capacity, robustness, or both in cases where perceptual transparency of embedded data is not critical [31].

4. Security

The encoding method is safe if the embedded data cannot be destroyed behind accurate detection by focused assaults based on the complete understanding of the steganography method and detector (excluding the encryption key) and knowing of at minimum one transport with a concealed payload [31].

5. Capacity

The needs listed above are in direct competition with one another and cannot be clearly optimized at the same time. If we wish to hide a huge message inside an image, we can't expect total undetectability and high resilience at the same time. It is always necessary to reach an acceptable compromise. If robustness to significant distortion is a concern, on the other hand, the message that can be securely hidden cannot be too lengthy [31].

2.2.1.2 Images Steganography Applications

Generally, steganography may be used whenever data has to be hidden. There are numerous reasons for hiding data, but they all boil down to preventing unauthorized people from accessing the data or learning of the message's existence. Steganography may be valuable for the automatic monitoring of a radio commercial or music. It is possible to build up an automated system to monitor a specific stego message [32].

Because of recent computer and networking approaches, individuals can use the fundamental uses of steganography linked to covert communications. People, organizations, and businesses can host a web page

that may include sensitive data intended for a third person. Someone may download the web page; nevertheless, the concealed data is invisible and goes unnoticed [33]. Some recent steganography uses are employed in medical imaging systems [34], where a separation between patients' image data or DNA sequences and captions such as a physician, patient's name, address, and other particulars is considered necessary for confidentiality. Using steganography to protect patient's personal information from falling into the wrong hands might benefit.

Bucerzan and Raţiu may create a technique to embed data in a printed photo invisible to the human eye and can subsequently be recovered by a mobile phone with a camera, inspired by the idea that steganography may be embedded as part of the usual printing process [35]. Because the embedded data is 12 bytes long, the procedure takes less than a second. As a result, people may collect encoded data using their cell phones. The core idea is to convert the image color scheme to hue, saturation, and value components (HSV) before printing and then embed it in the Hue domain, which human eyes are not sensitive to it. Mobile cameras may see and decode the coded data [36].

There are a variety of different applications that may utilize steganography to support their transmissions unknown [32], such as:

• Cleverness or intelligent property services.

• Ensuring the safety of multimodal biometric data.

• Businesses have confidential information to safeguard.

• Governments say that criminals can communicate via steganography. As a result, it may be restricted by legislation.

• Military-to-military and defense-to-military communication.

Steganography may be used in the business world to conceal a secret chemical formula or a strategy for innovations. It can also be used for business espionage by disseminating private information. It can also be utilized in the nonprofit sector to conceal confidential information.

2.2.1.3 Steganography difficulties

The statistics of the cover image are utilized in steganography techniques to hide secret knowledge within it without altering its characteristics [37]. A stego image is the product of this process. The stego image must be free of visible alterations so that a third party cannot detect these changes and treat the cover as if it were a regular image, while the secret data transferred via it remains safe. The following significant problems confront any image steganographic system, as demonstrated in figure (2.3).

- Payload size: how can the highest encoding capability be performed? Steganography targets a high level of embedding capability. More high payload requirements and safe transmission requirements are frequently in opposition [36].
- Visible photo quality: how similar is the tampered image to its untampered image perceptually? As a result, image steganography algorithms should create a stego image with a high level of imperceptibility [38].
- Robustness: how can a stego image withstand the various steganalysis detection attacks?

The stego image should be robust to image processing techniques such as compression, cropping, scaling, and so on; that is, secret information should not be lost when any of these steganalysis procedures are applied to the stego image [39].

As a result, the optimal steganographic approach must concurrently achieve the aforementioned goals of great capacity, acceptable visual image quality, and undetectability. High payload steganographic methods, on the other hand, frequently create distortion problems in stego images that are sensitive to steganalysis. The tiny payload concerns steganographic systems with excellent visual image quality. As a consequence of the conflicts between them, achieving a large payload, acceptable visual quality, and undetectability at the same time is a major research challenge [40].

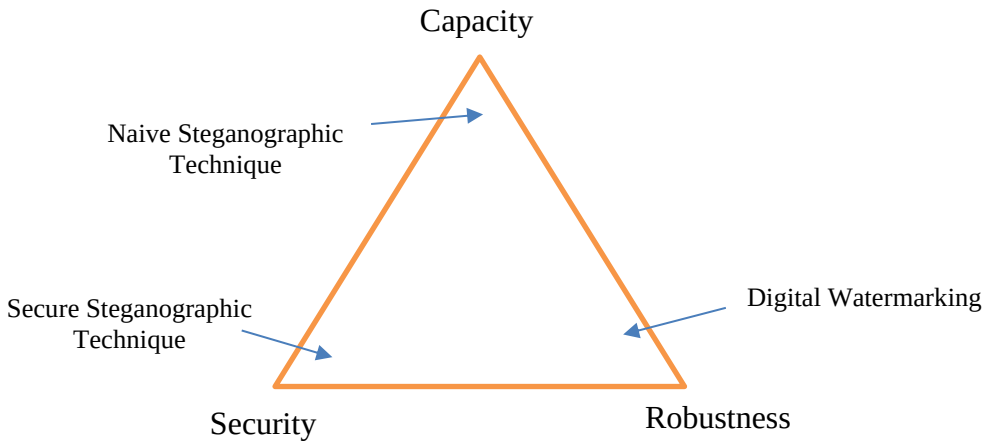

Figure (2.3) Balance -off among capacity, security, and robustness [41].

2.2.1.4 Steganography types

There are three major forms of steganography. These forms are:

1. ***Pure Steganography***

Pure steganography is a system that does not need the preceding sharing of some secret information (such as a stego-key). The embedding process is formalized as a mapping E: C× M⟶ C, where C denotes the set of potential covers and M is the set of possible messages. A mapping D: C× ⟶ M is used to extract the secret message during the extraction procedure. Indeed, it is important that /C/ ≥ /M/. The embedding and

extraction procedure must be accessible to both the sender and the receiver, but it should not be made public [34].

2. ***Secret Key Steganography***

Asymmetric cipher is comparable to a secret key steganography method. The sender selects a cover C and uses a secret key K to insert the secret message inside C. If the recipient knows the key used in the embedding procedure, he may reverse the process and retrieve the hidden message. Anyone who does not have access to the secret key should be unable to get proof of the encoded data. Perceptually, again, a cover C, and a stego-object can all be identical. The embedding process is a mapping E_k: $C\times M\times K \rightarrow C$, and the extraction process is a mapping D_k:$C\times K \rightarrow M$, where K is the set of all potential secret keys [42].

3. ***Public Key Steganography***

Two keys are required in a public key steganography system: private and public. A public database stores the public key. The secret key is used to reconstruct the hidden message, whereas the public key is utilized in the embedding process. A public-key cryptosystem is one approach to creating a public key steganography system. The decoding function D in a steganography system may be applied to any cover C (recall that D is a function of the whole set C) is used in public key steganography. In the latter scenario, there will be random components of M as outcomes, which will be referred to as the cover's "natural randomness." Suppose this natural randomness is statistically indistinguishable from the cipher text generated by a public-key cryptosystem. In that case, safe steganography may be built by embedding cipher text rather than the unencrypted secret message [34].

2.2.1.5 Steganography approaches

This section provides an overview of the numerous famous techniques that employ photos as carriers. There are multiple steganographic approaches accessible under digital photo carrier media. In this technique, embedded information is multiplied by the pseudo-noise sequence, then modulated before being embedded in a carrier media in the spread spectrum. The transform domain is defined as the transformation of an image into its frequency representation followed by modification of the spectral components of the image. The spatial domain approaches employ bitwise techniques that use simple mechanisms for bit injection, including noise management. In contrast, a transform domain is described as transforming a photo into its frequency form and then altering the photo's spectral components. Choosing the image pixels concerned, the type of alteration to be performed, and the number of bits encoded in a pixel are all examples of how adaptive nature may be included in data embedding systems [41]. Figure 2.4 depicts the classification of image steganography approaches.

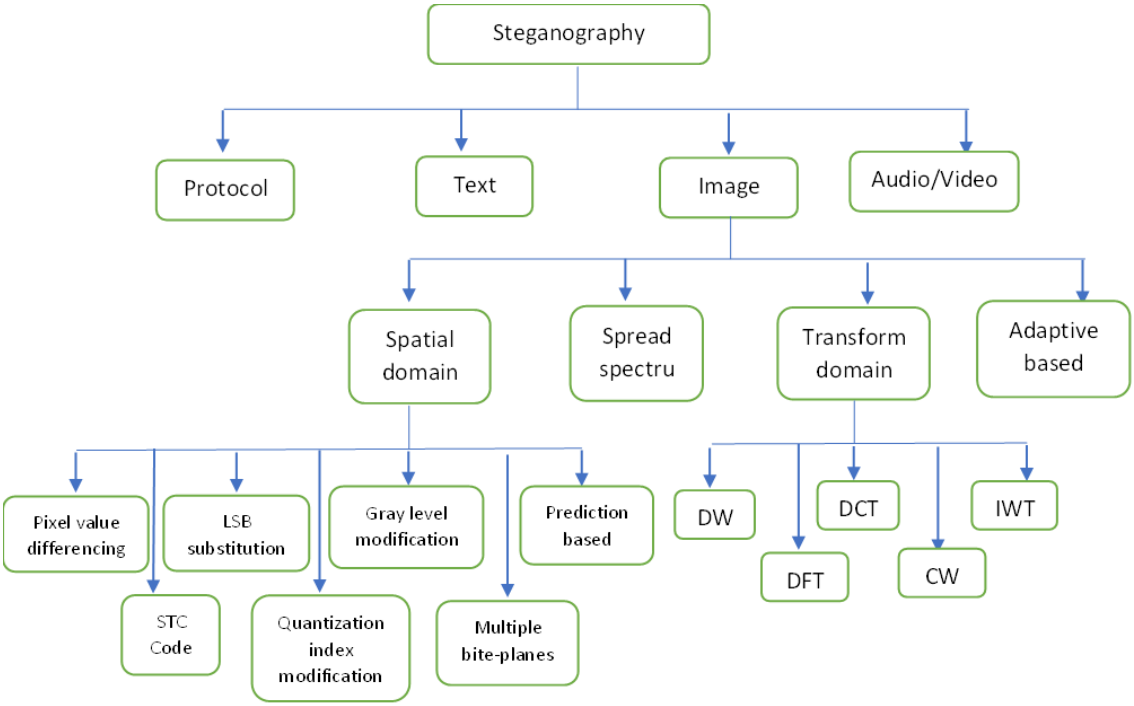

Figure (2.4): Types of digital steganography techniques [43].

1. **Spread spectrum techniques**

These approaches conceal and retrieve the long-secret data within digital images while preserving the image's original size and dynamic range. The encoded secret message may be retrieved using the proper keys with no knowledge of the original image. Image restoration, error-control coding, and spread spectrum methods are all demonstrated. Some applications for such techniques are in-band captioning, secret contact, photo stego-proofing, authentication, encoded control, and revision tracking [44].

2. **Spatial domain approaches**

Spatial domain approaches are better suited to the human visual system (HVS) and give higher hiding capacity than transform domain methods when suitable image quality [45]. It is the most basic type of data encoding in photos, in which the secret message bits are encoded directly by changing pixel values [46]. The following sections describe a number of these approaches:

A. ***Gray level modification***

The grey-level pixels values checked to the bit stream utilized to map the photo. First, the gray level values of the selected pixels (odd pixels) are made even by altering the gray value by one unit. In contrast, a bitstream must always be mapped when all chosen pixels have the same gray level. The bit stream's principal bit is compared to the pixel that was first selected. When a primary bit is even, it is unaffected since the gray level values of all selected pixels are the same. When the bit is odd, the gray level value of the pixel is reduced by one unit to make the esteem odd, resulting in an odd bit mapping. All bits are in the bitstream, and each bit is mapped due to modifying the gray level values. Compared to previous approaches, these strategies enable us to offer higher-quality stego images [47].

B. ***Pixel value differencing (PVD)***

This approach divides a cover image into non-overlapping blocks of two connected pixels. By changing the difference between these two pixels, it hides the data. The size of the pixel determines the concealing capability of this method. For example, if the edge area is picked, the difference between connected pixels is significant, but the difference between connected pixels is minimal in smooth regions. As a result, the optimum option is to embed the secret message in edge areas with higher embedding capacity [48], as illustrated in Figure (2.5).

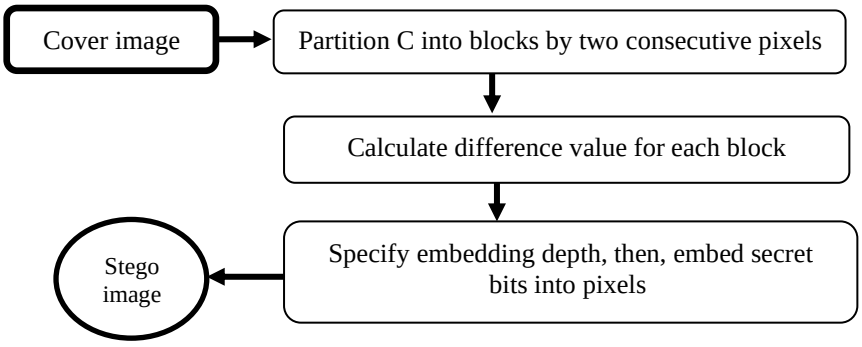

Figure (2.5) The pixel value differencing method's steps [43].

C. *Least significant bit substitution (LSB)*

It is a fundamental and well-known method for concealing massive amounts of covert data within the cover image [49]. In this approach, all Least Significant Bits of pixels in the carrier image are replaced with secret bits. As illustrated in Figure (2.6), this technique embeds the fixed-length covert bits in the identical fixed-length LSBs of pixels. Despite its simplicity, this method produces considerable distortion if the number of embedded bits per pixel exceeds four [50].

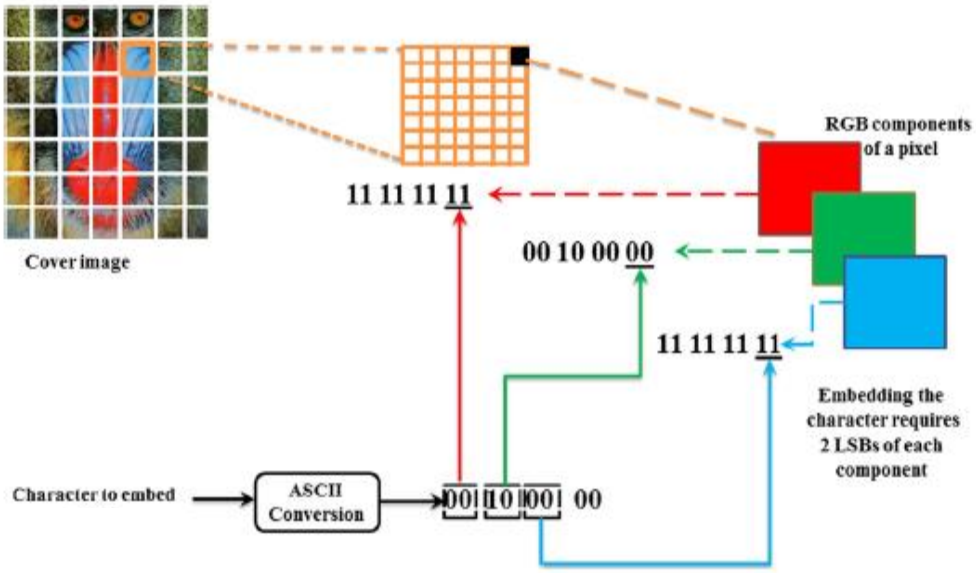

Figure (2.6) Using the LSB technique, data is hidden in images [43].

D. *Enhanced least significant bit substitution (ELSB)*

It used the final pair of bits of the cover-image pixels for embedding and retrieval of a covert message. In the process of embedding, when the final pair of bits at the value of the pixel are 00 or 10, it embeds the message bit 0. If the final pair of bits of the pixel value are neither 00 or 10, try to equal them to 00 or 10 by subtracting or adding 1 from the value of the pixel to embed 0. It also embeds the message bit 1 at a pixel value if the final pair of bits of the pixel value are 01 or 11. If the final pair of bits are neither 01

or 11, we attempt to convert them to 01 or 11 by subtracting or adding 1 from that pixel value to embed 1. The message bit is 0 if the final pair of bits of a pixel in the retrieval process are 00 or 10. The opposite of that is 1[51].

E. *Chaotic least significant bit substitution (CLSB)*

It used a chaotic map to disguise a hidden message (text or image) inside the cover image by shuffling address bits. The proposed scheme employs a chaotic map to produce integer chaotic series (secret key), which are then used to select pixel addresses of the cover image for embedding the secret message. The chaotic map's parameters are secret keys known only to the sender and receiver. Without recognizing these secret keys, the observer cannot sense the presence of a hidden message [52].

F. Method based on independent component analysis

It is a safe system for concealing a single or more extra image within a cover image of equal size. It employed an independent component analysis strategy to produce the secret key and employed a genetic algorithm strategy for the best values for the mixing matrix. The key would be exchanged between the sender and the recipient to ensure that special data remains secure and difficult for the untrustworthy to discover. Furthermore, the suggested technique improves the effectiveness of hiding capability, security level, and resistance to specific attacks [53].

G. *Exploiting modification direction (EMD)*

Using n pixels as a group, Exploiting Modification Direction (EMD) establishes hidden digits to 2n + 1 array notational system to reduce stego image distortion. In addition, embedding requires a decrease or rise in a single-pixel value within the collection. It is required to compute the value

of n before embedding for this approach. When the number of n is equal to 2, and the embedding is represented by only one secret digit within each two pixels, the image quality is the best [54].

H. ***Quantization-based approaches***

This type of steganographic system employs any type of encoding scheme to conceal secret data bits. Any common compression codec, such as JPEG, vector quantization, and so on, can be used as the encoding system. The secret data is split into little data bits, which are then embedded with an encoded cover image. These technologies are implemented to improve capacity while limiting tamper photo distortion. Unfortunately, these methods are insufficient to withstand geometrical attacks and steganalysis [41].

I. ***Multiple bit-planes-based approaches***

These approaches are presented as an extension to the LSB replacement technique, in which secret data bits are hidden via bit planes [55]. Bit plane stego techniques are frequently used in conjunction with other methods to improve the overall system's performance. An extended bit-plane encoding has two benefits over the 8-bit LSB techniques: it can host more hidden bits and the degree of randomness of embedding is high [56].

3. **Adaptive-based approaches**

"Statistics-aware embedding" or "Masking" are terms used to describe adaptive steganography. To put it another way, the cover image's statistics are utilized to embed covert data without affecting its characteristics. This embedding may be accomplished by selecting pixels in a block with a large local STD (Standard Deviation) and performing a random adaptive selection of pixels based on the cover image [57]. The secret bit-carrying pixels are chosen adaptively based on the content of the cover image [58].

4. **Transform domain approaches**

In the area of steganography, several transform domain methods are used, including the Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT), IntegerWavelet Transform (IWT), and Complex Wavelet Transform (CWT) [59]. In general, this approach is more resistant to typical image processing procedures, including lossy compression. The Joint Photographic Experts Group (JPEG) compression standards accommodate DCT-based steganography, whereas the Joint Photographic Experts Group 2000 (JPEG2000) compression standards accommodate DWT-based steganography.

2.2.1.6 Steganography Instruments

Several shareware and freeware applications may be downloaded via the Internet using the file transfer protocol (FTP). This section gives a quick overview of how they work.

A. ***Hide and seek***

This freeware is an MS-DOS software that embeds and extracts data from image files in GIF format. It can fit up to 19,000 bytes into a GIF file with a total size of 320×480 pixels. It encodes using Least Significant Bit (LSB) replacement, which removes the image byte's LSB and replaces it with a bit from the message file contents. IDEA is also used to encrypt program-specific header information [60].

B. ***Stego Dos***

This set of applications, commonly known as Black Wolf's Picture Encoder, is also available for free. It can encode data files up to 8kb in a picture file with a maximum size of 320 x 480 pixels and 256 colors. It doesn't specify an image format. Moreover, to use this application, the user

needs to have access to graphics file display and screen capture software. It employs the LSB replacement method to encode the screen capture file with data, and because it uses screen capture, it does not alter the original image [60].

C. ***White Noise Storm***

The White Noise Strom tool is a DOS package of software. The written message was easily included in the cover images, and there was no noticeable deterioration. By extracting the LSBs from the cover image and saving them in a file, WNS applies steganography to the LSBs of PCX files. To produce a new set of LSBs, the message is encrypted and applied to these bits. To produce the new stego-image, the changed bits are injected into the cover image. The White Noise Strom tools use Spread Spectrum Technology and frequency hopping to disperse the data over the image (similar to DES block encryption) [61].

D. ***S-Tools***

It not only encodes and extracts, utilizing LSB replacement BMP and GIF files, but also audio and WAV files. It also includes a tool for steganographically encoding unused floppy drive space. It supports 24 bit BMP color and encourages encryption of the input message file utilizing (IDEA, DES) [62].

2.3 Image steganalysis

The art and science of identifying secret communications hidden via steganography is steganalysis [63]. The objective of steganalysis is to gather enough evidence to prove the existence of hidden data and to compromise a carrier's safety. As a result, the goal of steganography is defeated. Steganalysis methods that can consistently identify the existence of embedded data in photos are becoming increasingly important. Steganalysis

is used in computer forensics, cyber warfare, tracking illegal activity on the internet, and gathering evidence for investigations, especially in cases involving anti-social elements [63,64]. Steganalysis has a peaceful use in addition to its law enforcement and anti-social importance—enhancing steganography technologies' safety by assessing and detecting their flaws.

Currently, billions of images may be found online. Images are used to capture real-life events, communicate emotions, and pursue additional events. Regrettably, with the advances in image steganography techniques, terrorist gangs may now communicate messages using regular image communication with excellent efficiency. Since these techniques aim to disguise security data as random image noise built with camera electronics, identifying images related to illegal activity is difficult using the human eye alone. This technological threat to public security is a real-world problem. As a result, studies explored and created novel image steganalysis methods to prevent plus analyze this danger. One of the most popular secret communication approaches is image steganography, which involves concealing data in an image [2].

As a result, in a manner similar to cryptanalysis, it is focused on cryptography; steganalysis is the practice to detect secret material or data in a cover media and distinguishing between stego object and cover object with little or no understanding of the steganography techniques. Steganalysis aims to gather some evidence indicating the existence of an encoded message [3,65].

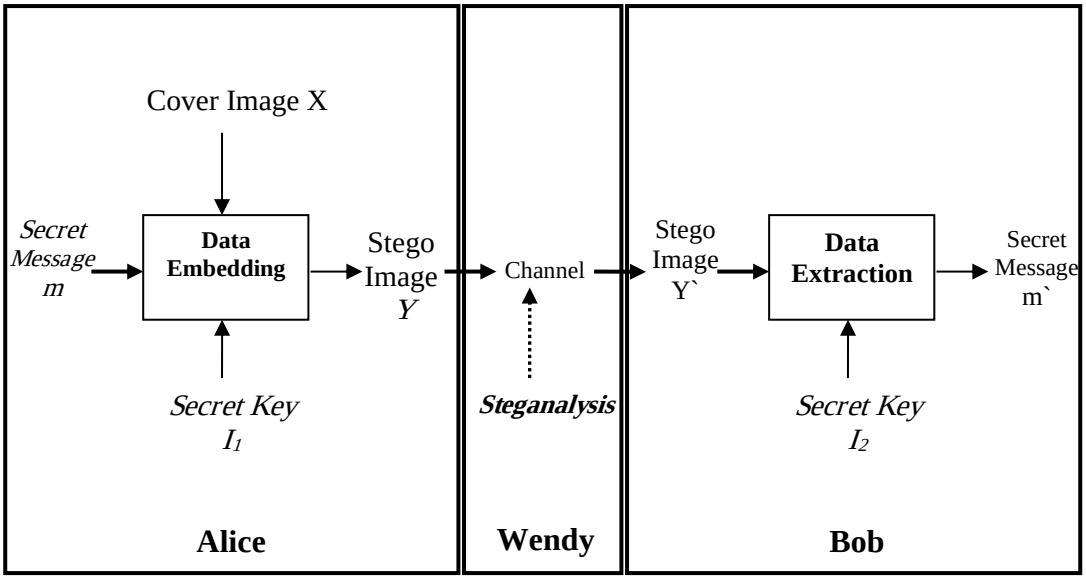

Figure (2.7) Steganalysis and steganography model [3].

2.3.1 Techniques of steganalysis

Signature steganalysis and statistical steganalysis are the two types of steganalysis that are often used. The split is dependent on whether the steganography technique's signature or image statistics are utilized to detect hidden messages in steganography-enabled images. It may be split into specific and universal techniques depending on their application sectors. A specific steganalytic approach makes use of the knowledge of a certain steganographic methodology and may be proprietary to that steganography. To detect various types of steganography, a universal steganalytic technique is utilized. Typically, universal techniques do not require an understanding of embedding processes. As a result, it's also known as the blind technique.

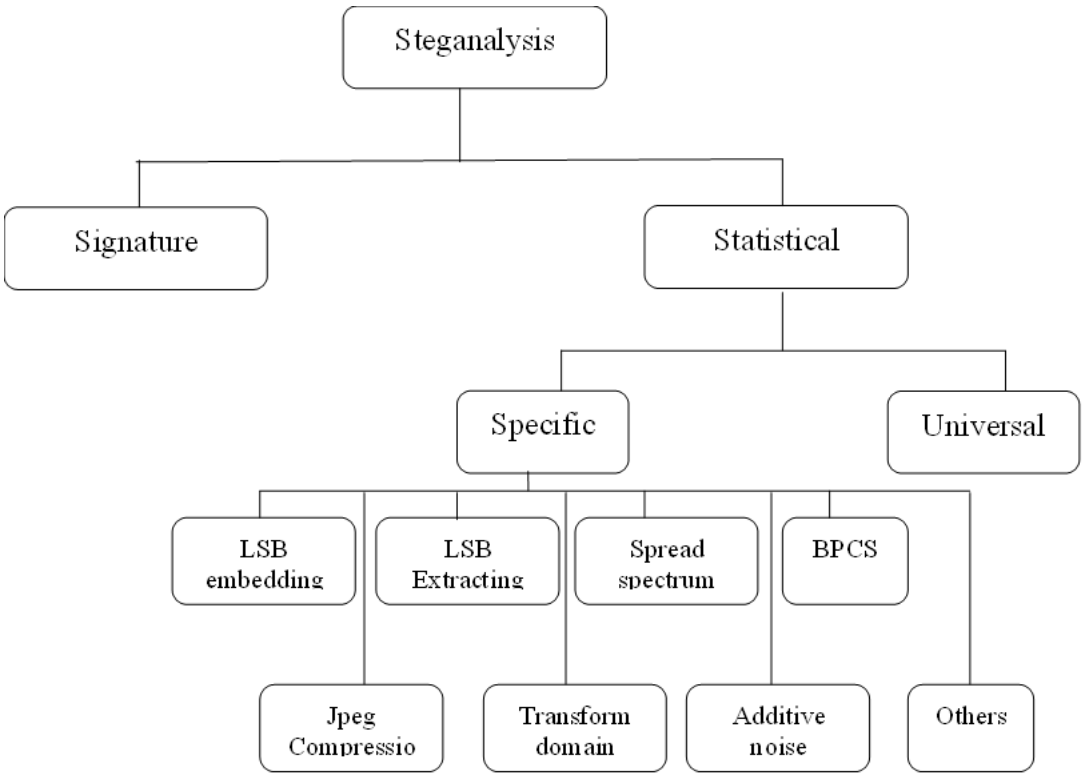

Figure (2.8) Steganalysis methods classification [66].

1. ***Signature steganalysis****:* Methods of steganography conceal secret information and modify images and other digital material in undetectable ways to the naked eye [7]. Steganography modifies the media characteristics by inserting payload bits in a deterioration state or repeating patterns, which serve as signatures indicating a presence of hidden data [4]. Look for these repeated patterns as signatures of a steganography tool to identify the existence of a concealed message in a suspicious image. These signatures automatically make use of the tool used to embed the data. Palette tables in GIF images are examined for any abnormalities generated by standard stego tools. When the message is encoded in sequential order, such assaults produce

encouraging outcomes, but they are difficult to automate, and their trustworthiness is questioned.

2. ***Statistical steganalysis****:* The statistics of an image are altered because of the data hiding. Statistical steganalysis examines an image's underlying statistics to uncover hidden information. It is more commanding than signature steganalysis because mathematical approaches are better sensitive than optical perception [4].

   1) *Specific statistical steganalysis:* These approaches are developed by examining an encoding process then calculating specific photo statistics. The methods require a thorough understanding of the hiding strategy. When applied to a target steganography approach, these techniques produce exact results. For detecting hidden messages from stego-images encoded via least significant bit encoding, least significant bit matching, spread spectrum, JPEG compression, and other transform domains, statistical steganalytic tools are applied [4].

   2) *Universal statistical steganalysis:* The statistical steganalysis approach that is not specialized for a specific steganography encoding approach is known as universal statistical steganalysis. It needs little or no prior knowledge of the steganographic technologies under assault to detect the hidden message. It employed a learning-based method that included cover and stego-image training. The detection model is built from the experimental data using neural networks, clustering algorithms, and other soft computing tools. These methods are independent of how embedding algorithms behave [66].

2.4 Correlation effects

Correlation as a general concept is a measure of an association between variables. In correlated data, the change in the magnitude of one variable is related to an adjustment in the extent of another variable, either

in the equivalent (positive correlation) or in the inverse (negative correlation) direction. Frequently, the term correlation is used in the context of a linear relationship between 2 continuous variables and expressed as Pearson product-moment correlation.

The Pearson correlation coefficient is commonly used for jointly normally distributed data (data that follow a bivariate normal distribution). For nonnormally appropriated continuous data, for ordinal data, or for data with significant outliers, a Spearman rank correlation can be utilized as a measure of a monotonic association. Both correlation coefficients are scaled with the end goal that they range from – 1 to +1, where 0 demonstrates that there is no linear or monotonic association, and the relationship gets stronger and at last methods, a straight line (Pearson correlation) or an always increasing or decreasing curve (Spearman correlation) as the coefficient approaches an absolute value of 1. Hypothesis tests and certainty intervals can be utilized to address the statistical significance of the outcomes and to estimate the strength of the relationship in the population from which the data were sampled [67].

The correlation analysis or cross-correlation is calculated according to the Equation (2.1) below, and the range of the coefficient is [-1,1]:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2}} \qquad (2.1)$$

r always has a value between –1 and 1, i.e.$-1 \leq r \leq 1$. We say there is a positive or direct connection between Y and X if Y rises as X increases. If, on the other hand, Y falls as X rises (or vice versa), It can be said they are negatively or inversely linked. Direct and inverse are words employed in the context of variation or proportionality, as the reader has already noted.

The maximum values of r, i.e. when r = ±1, imply that X and Y are perfectly correlated (positive or negative). If r is 0, however, it can be concluded that there is no correlation. Pearson's correlation coefficient is only used to assess linear relationships. It should not be utilized with non-linear relations since it will inevitably result in an incorrect interpretation.

The remaining numbers, which lie inside the [–1, 1] subinterval, reflect the relationship's strength. The values of r obtained after computation can be used as a guideline to choose which adjective should be used to explain the relationship, as shown in Figure (2.9).

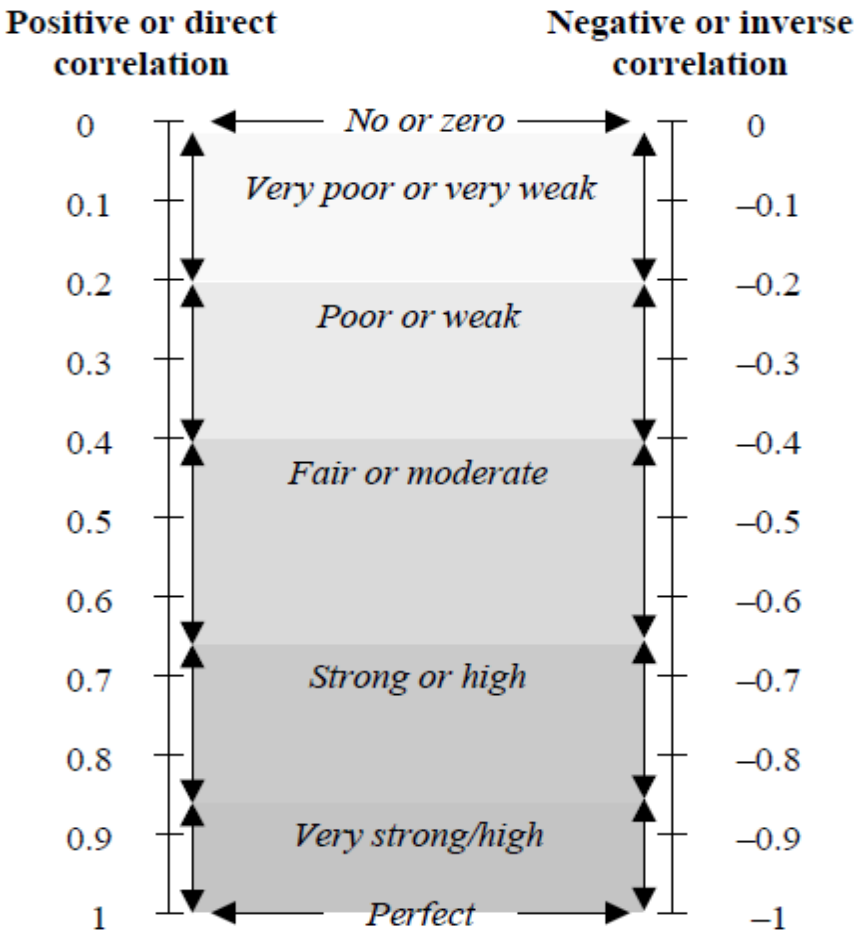

Figure (2.9) Correlation coefficient interpretation [68].

It is noticed that the quality of the correlation between X and Y is the same whether r = 0.85 or – 0.85. The only distinction is that there is a direct correlation in the first scenario and an inverse correlation in the second. It's important to remember that r stands for linear correlation coefficient and that its value, as previously stated, might be misinterpreted when the relationship between X and Y is non-linear. That is why there is a need to examine a scatter plot of points (x, y) and determine whether the relationship is quadratic, logarithmic, exponential, or trigonometric (in other words, non-linear).

If r = 0, we shouldn't assume there's no relationship between X and Y. Consider the scenario when the two variables are interrelated by the equation $Y= X^2$ and there is a perfect (but unexpected) non-linear correlation between them (see Figure 2.10 below). The reader may readily check that both $\sum x$ and $\sum xy$ are equal to zero by starting with the points (–3, 10), (–2, 5), (–1, 2), (0, 0), (1, 2), (2, 5) and (3, 10). As a result, r = 0 (check the formula for r in Equation (2.1)). The linear product-moment correlation coefficient can not be utilized to determine the strength of a non-linear relationship, we conclude.

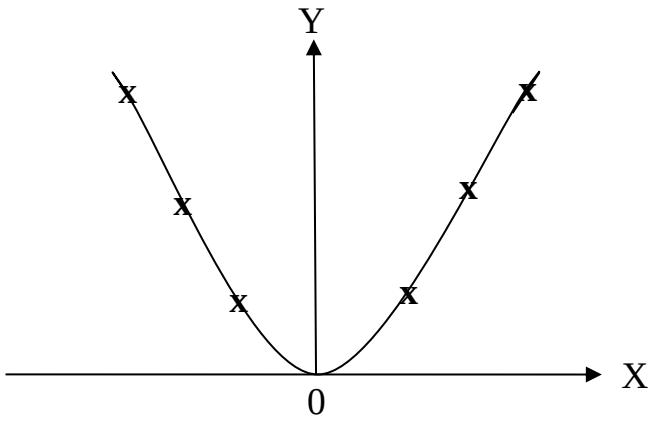

Figure (2.10) Relationship of non-linear [68].

With enough effort and skill, you can even estimate the value of r by looking at a scatter diagram. The strength of the link between the variables

is determined by the position (amount of scattering) of the points in reference to the least-squares regression line. The weaker the relationship and the nearer the correlation of r is to zero, the more scattered the points are.

In the regression equation $Y`=b + cX$, the sign of r is always the same as the sign of (the gradient) *c*. The scatter diagram in Figure (2.11) demonstrates how one may determine the value of r to a certain degree of precision [68].

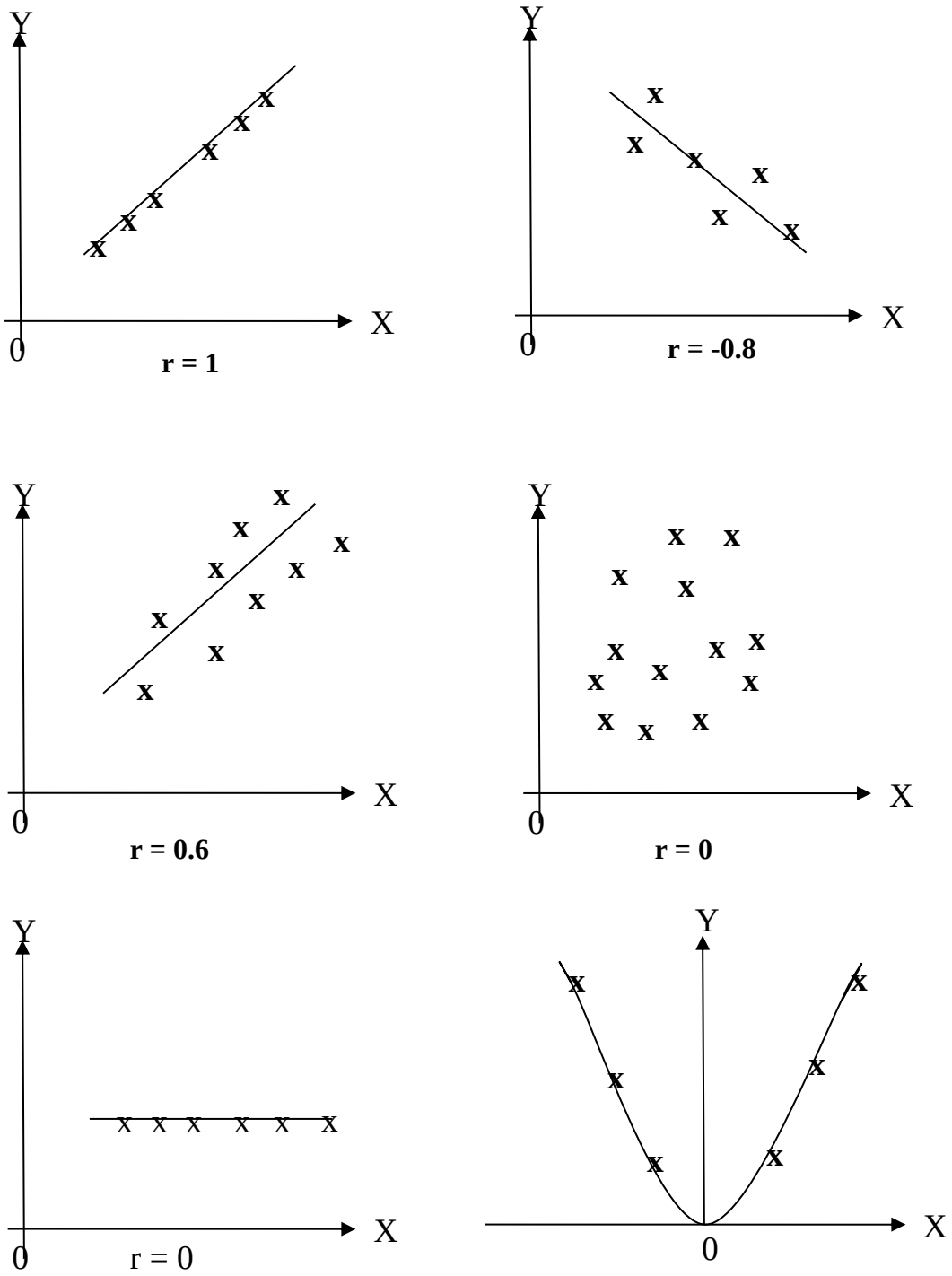

Figure (2.11) Using scattering charts to discover out about r [68].

2.5 Autocorrelation function

The cross-correlation of two waveforms was used to determine their similarity as a function of a time-lag applied to one of them. The cross-correlation of a signal with itself is known as autocorrelation. It is a time domain analysis that can be used to determine a signal's periodicity or repeated patterns. The autocorrelation function (ACF) is a popular metric for detecting whether or not there is a serial relationship. Because it provides a more detailed representation of the basic process, the autocorrelation function is more useful than the cross-coefficient test. x_t, $t = 1,..., N$, is a time series of length N. A scatterplot of the latest N-k observations versus the initial N-k observations is a lagged scatterplot for lag k [69]. Equation(2.2) may be refined to yield a correlation between observations split by k time steps. The autocorrelation coefficient at lag k is denoted by the value R_K.

$$R_k = \frac{\sum_{t=1}^{N-k} (x_t - x`)(x_{t+k} - x`)}{\sum_{t=1}^{N-k} (x_t - x`)^2} \qquad (2.2)$$

where $x` = \sum_{i=1}^{N} (x_t)$ is the mean.

The autocorrelation function (ACF) is a useful diagnostic tool for time series analysis in the time domain [70]. When analyzing stationarity and picking from a variety of non-stationary models, the ACF comes in handy. Lag is a time interval that separates the required data and computes the coefficients in autocorrelation. When it is calculated, the range resultant numeral can be from +1 to-1. The autocorrelation of +1,-1 means an excellent positive and negative correlation, and it is symmetric about the x=0 line, as you can see.

The likelihood of a link between data values split by a certain number of time stages is used in autocorrelation plots, also known as correlograms, to supply a more helpful understanding of the development of a strategy via time (lags). The correlogram shows autocorrelation coefficients on the vertical axis and the horizontal axis, lag values. The correlogram illustrates the time series' key characteristics, such as randomization, rising or falling trend, oscillation, and so on [70]. Because the mathematics used to compute the correlation coefficient looks the same in both situations, positive and negative lags will be identical:

lag 2

abcdefghijklmnopqrstuvwxyz

abcdefghijklmnopqrstuvwxyz

lag -2

abcdefghijklmnopqrstuvwxyz

abcdefghijklmnopqrstuvwxyz

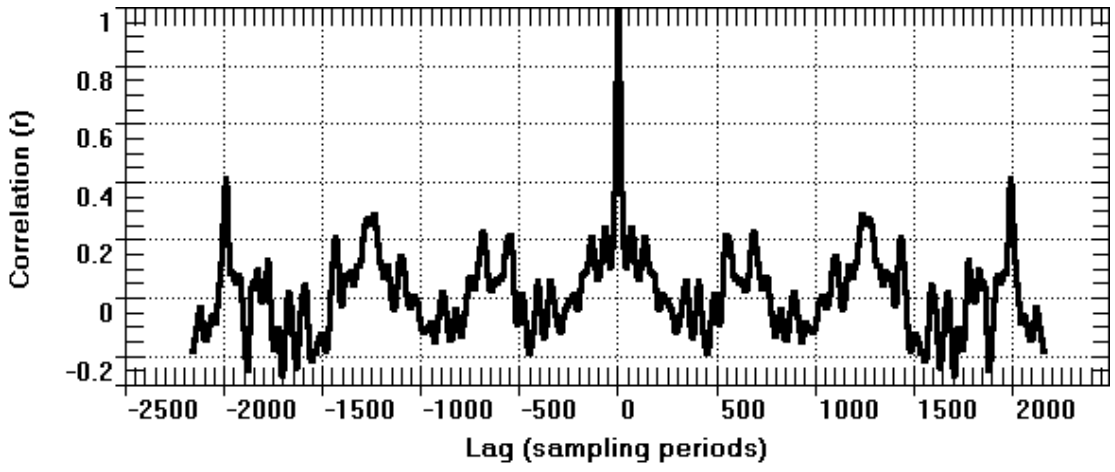

Figure (2.12) The time series k has a correlogram [71].

2.6 Image processing in spatial domain

The spatial domain refers to the collection of pixels that make up an image. Strategies that work straight on these pixels are known as spatial domain techniques. The expression that employed to represent spatial domain processes

$$g(x,y) = T[\,f(x,y)] \tag{2.3}$$

Where $f(x,y)$ denotes the input image, $g(x,y)$ is the processed image, and T denotes an operator on f , defined across some neighborhood of (x,y). T may also perform operations on a collection of input images, such as doing a pixel-by-pixel sum of K images to reduce noise.

As shown in Figure (2.13), the most common method for creating a neighborhood around a point (x, y) is to utilize a square or rectangular subimage region centered at (x, y). Starting in the top left corner, the center of the subimage is shifted from pixel to pixel. At each point (x, y), the operator T is used to produce the output, g, at that position. The pixels in the region of interest spanned by the neighborhood are used in the procedure.

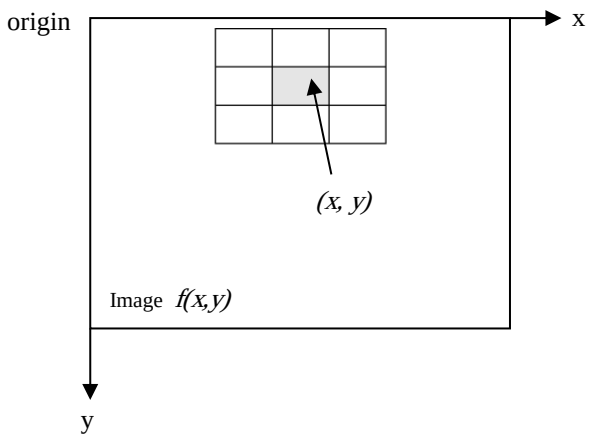

Figure (2.13) In an image f (x, y), the 3*3 neighborhood around a pixel (x, y) [72].

The subimage is referred to as a filter, mask, kernel, template, or window, with the first three words resonating the most. Instead of pixels, the values of a filter subimage are referred to as coefficients [72].

2.7 First and second-order derivatives operations

Some filter operations are based on first-order and second-order derivatives, respectively. In a digital environment, we pause to consider some of the essential characteristics of these derivatives. To make things easier, we'll concentrate on one-dimensional derivatives. The behavior of these derivatives in regions of gray level (flat parts), at the onset and end of discontinuities (step and ramp discontinuities), and along gray-level ramps is of special interest to us. In an image, these kinds of discontinuities may be utilized to simulate noise spots, lines, and edges. Also of interest is the behavior of derivatives during transitions into and out of these image features.

Differences are used to determine the derivatives of a digital function. These differences can be defined in a variety of ways. The greatest potential gray-level change is finite, since we're dealing with digital variables with finite values and the smallest distance across which that change may occur is between neighboring pixels.

The first-order derivative of a one-dimensional function *f(x)* is defined as the difference

$$\frac{\partial f}{\partial x} = f(x+1) - f(x) \qquad (2.4)$$

We utilized a partial derivative here to keep the notation identical with when we examine an image function of two variables, *f (x, y),* in which case partial derivatives along the two spatial axes will be involved. The use of a partial derivative in this discussion has no bearing on the nature of what we're trying to do.

A second-order derivative is defined as the difference

$$\frac{\partial^2 f}{\partial x^2} = f(x+1) - f(x-1) - 2f(x) \qquad (2.5)$$

It is simple to verify that these two definitions meet the previously stated requirements for first and second order derivatives. Consider the example in Figure (2.14) to demonstrate this, as well as to emphasize the essential similarities and differences between first and second-order derivatives in the context of image processing [73].

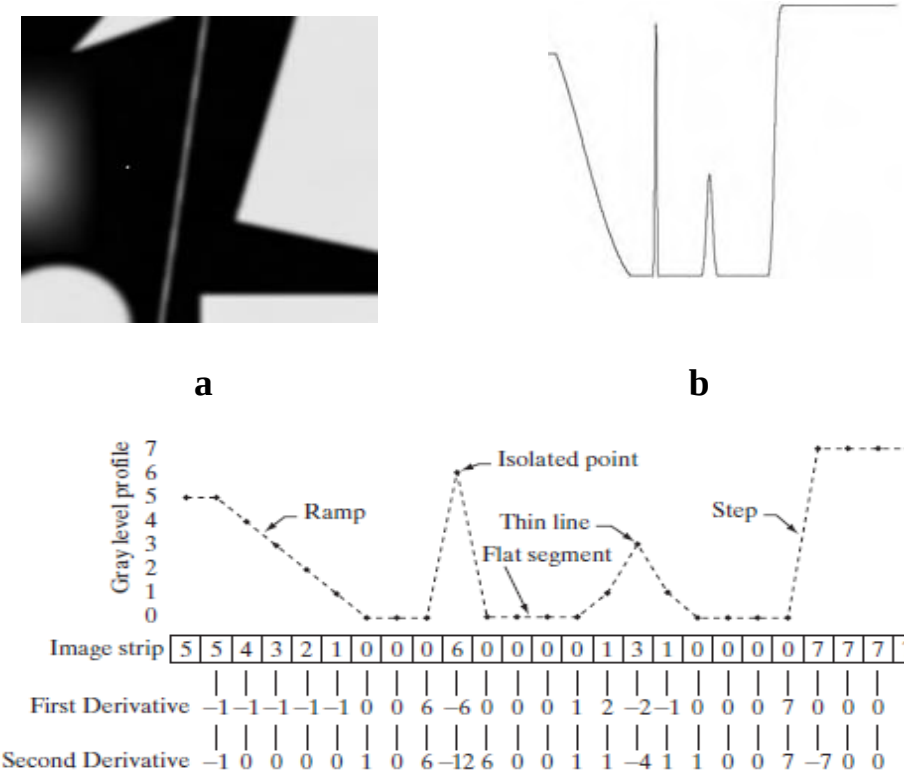

Figure 2.14 (a) An image. (b) A one-dimensional horizontal gray level profile running through the image's center and containing the isolated noise point. (c) Simplified profile (dashed lines connect the spots to make understanding easier) [72].

A basic image with several solid objects, a line, and a single noise point is shown in Figure (2.14 (a)). The image's horizontal gray-level profile (scan line) at the center, including the noise point, is shown in Figure (2.14 (b)). This profile is the one-dimensional function we'll utilize for this figure's

illustrations. Figure (2.14 (c)) depicts a simplified version of the profile, with enough data to demonstrate how the 1st and 2nd derivatives act as they approach a noise point, a line, and finally, an object's edge. The transition in the ramp spans four pixels in our simplified diagram. The noise point is one pixel, the line is three pixels thick, and the transition into the gray-level step occurs between neighboring pixels. The number of gray levels has been reduced to only eight.

In conclusion, the following findings after comparing the responses of first and second-order derivatives. (1) In general, first-order derivatives result in thicker image edges. (2) Fine details, such as thin lines and isolated points, are more responsive to second-order derivatives. (3) 1st derivatives show a higher answer to a gray-level step than 2nd derivatives. (4) At step changes in gray level, second-order derivatives generate a double reaction. Second-order derivatives similarly have a higher reaction to a line than to a step, and to a point than to a line, for identical changes in gray-level values in an image [72].

2.8 Histogram of an image

An image's histogram, like other histograms, displays frequency. On the other hand, an image histogram displays the frequency of pixel intensity values. The gray level intensities are shown on the x-axis, and the frequency of these intensities is shown on the y axis in an image histogram. Color images may also be histogrammed, either as separate red, green, and blue channel histograms or as a 3-D histogram. The three axes represent the red, blue, and green channels, and the brightness at each point indicates the pixel count. The precise output of the operation is dependent on the implementation — it may simply be a picture of the needed histogram in a suitable image format, or it could be a data file of some type, including the histogram statistics [72].

2.9 Entropy

The digital world has been developing practically daily. In particular, the capabilities of technological devices and their use in almost all aspects of life have grown dramatically during the last ten years (i.e., mobile, digital players, robot machines, digital readers, etc.). These technologies demonstrate how modern technologies' computational and storage capabilities quickly increase. Numerous cryptographic techniques, like cryptography (DES), the Blowfish cipher, the Twofish cipher, and the advanced encryption standard (AES), despite the fast growth of electronic and computational machines abilities. Although flaws in these approaches have been identified for bulk data such as digital images and electronic media, the ancient algorithms predominate cryptographic algorithms at all classes (people, institutions, businesses, and governments) [74].

Histogram analysis, global Shannon entropy measure, and neighboring pixel correlations are all ways for determining image randomness. The Shannon entropy, named after Claude Shannon, was developed for the first time in 1948, and Shannon entropy has been primarily applied in the information sciences since then. The Shannon entropy of a random variable measures its uncertainty, and Shannon entropy, in particular, measures the anticipated value of the information in a letter. Equation (2.6) defines the Shannon entropy of a random variable X [75].

$$H(X) = -\sum_{i=1}^{n} p_i \, log_2 p_i \qquad (2.6)$$

Where $p_i = \Pr(X = x_i)$

2.10 Butterworth filter

Digital signal processing techniques are utilized to solve the issues arising in the field of digital communication system architecture. Analog filters are continuous-time systems with continuous-time signals at both the input and output. Digital filters are discrete-time systems with discrete-time signals as input and output. A filter is a mechanism or network that modifies the waveform, amplitude-frequency, and phase frequency properties of a signal selectively. Filtering goals include improving signal quality (for example, removing or reducing noise), extracting information from signals, and separating two or many signals [76].

Digital filters are widely classified into two types: infinite impulse response (IIR) and finite impulse response (FIR). The convolution sum connects the filter's input and output signals. The equation of the IIR filtering

$$y(n) = \sum_{k=0}^{N} b_k\, x(n-k) + \sum_{k=1}^{M} a_k\, y(n-k) \qquad (2.7)$$

a_k , b_k filter coefficient

For N Butterworth filter with a passband edge ω_p, the magnitude equation is:

$$/\mathrm{H}(j\omega)/ = \frac{1}{\sqrt{1+\epsilon^2\left(\frac{\omega}{\omega_p}\right)^{2N}}} \qquad (2.8)$$

At $\omega=\omega_p$

$$/\mathrm{H}(j\omega_p)/ = \frac{1}{\sqrt{1+\epsilon^2}} \qquad (2.9)$$

the parameter ϵ specifies the greatest variance in passband transmission.

$$\text{Amax} = 20\log\sqrt{1+\epsilon^2} \qquad (2.10)$$

The value ϵ determine by equation

$$\epsilon = \sqrt{10^{\text{Amax}/10}} - 1 \qquad (2.11)$$

Remember that the highest deviation in passband transmission happens only at the passband edge in the Butterworth reply [77].

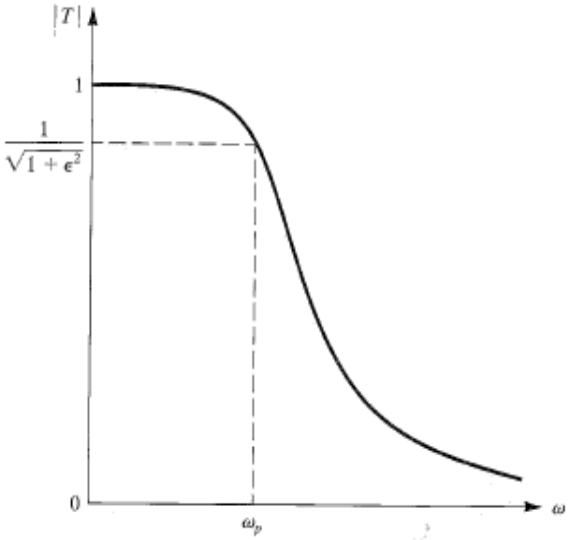

Figure (2.15): The Butterworth filter's magnitude response [78].

As we all know, some noise is added to signals during processing. As a result, removing the noise is required for signal transmission. Different ways can be used to eliminate the noise. The Butterworth IIR filter strategy employs denoise the incoming signal for several of these approaches [79].

2.11 Fourier Transforms

Signal and system analysis relies heavily on frequency domain analysis and Fourier transforms. These concepts are also one of the foundations of electrical engineering. These concepts are so fundamental that they are employed in a broad range of domains, including electrical engineering and almost every department of engineering and science and multiple mathematics places.

The Fourier Transform is a charming mathematical technique. Using the Fourier Transform, any function may be decomposed into a sum of sinusoidal basis functions. Each of these essential functions is a different frequency complex exponential. As a result, the Fourier Transform provides us with a unique perspective on every function: as the sum of simple sinusoids.

While the Fourier Transform is a wonderful mathematical tool, it is widely used in research and engineering because of its practical applications. It isn't easy to comprehend why the Fourier Transform is so crucial. But it simplifies the answer to complicated difficulties. Furthermore, the Fourier Transform provides us with a new way of experiencing the world, which is ideal for gaining a more intuitive understanding of our environment [80]. The Fourier Transform is a mathematical technique that transforms a function of time, x(t), to a function of frequency, X(ω). Define the Fourier transform of a function g(t) by Equation (2.12)

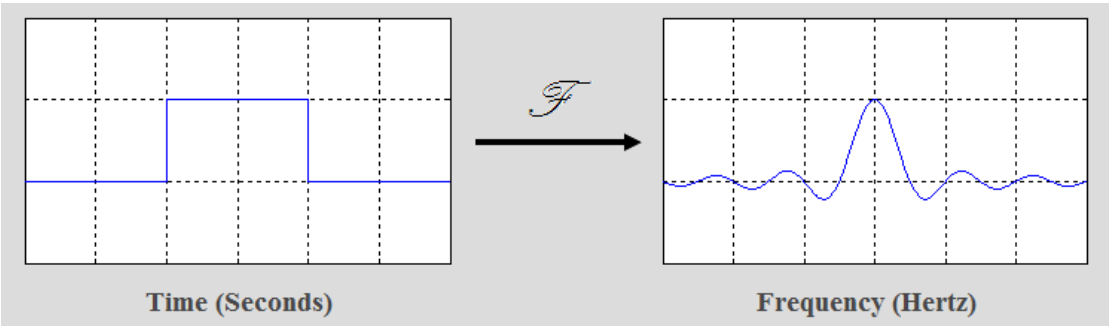

Figure (2.16): The Fourier Transform (ℑ) transforms a function of time, x(t), to a function of frequency, X(ω) [80].

$$\mathcal{F}\{g(t)\}=G(f) = \int_{-\infty}^{\infty} g(t)e^{-2\pi ift}\, dt \qquad (2.12)$$

The outcome is a function of f, or frequency. Consequently, G(f) indicates how much power g(t) has at frequency *f*. G(f) is also known as the g

spectrum. Furthermore, the inverse Fourier Transform may be used to extract g from G:

$$\mathcal{F}^{-1}\{G(f)\}=g(t)=\int_{-\infty}^{\infty} G(f)\,e^{2\pi ift}\,df \qquad (2.13)$$

2.12 Measures of Evaluation

On different sizes of the secret image (cell.tif secret message image 5*5, 10*10, 20*20, 30*30), the size ratio Rm between message size and cover size may be computed by the following Equation (2.14):

$$R_m = \frac{MessageSize}{Cover\ ImageSize} \qquad (2.14)$$

Then, Various indicators were used to assess the efficiency of the blind proposed steganalysis approach. There are four possible results if suspicious images are embedded:

1. True positive (TP): a stego image is accurately categorised as a stego image.
2. False negative (FN): a stego image is incorrectly classified as a cover image.
3. True negative (TN): a cover image is correctly organised as a cover image.
4. False positive (FP): a cover image is wrongly classified as a stego image.

Metrics such as accuracy are used to evaluate the success of the suggested strategy. The fraction of all accurately predicted classes are referred to as accuracy. The accuracy should be as high as possible [21]. Equation (2.15) is used to calculate accuracy.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \qquad (2.15)$$

Chapter Three

The Proposed System

Chapter Three

The Proposed System

3.1 Introduction

The design and implementation of the suggested system for detecting the existence of a secret message in an image are described in this chapter. The proposed method includes several stages.

The first stage is the correlation pixel and correlation of the tested image histogram, the autocorrelation function (ACF) is a popular metric for detecting whether or not there is a serial relationship. Then the first, second, and third derivatives of the histogram autocorrelation function will have significant ripples when steganography exists. Finally, using the power of a high pass filter, Fourier transform or the entropy correlation derivative as a threshold would imply that it is used to produce a decision. It is used to detect whether an image is clean or stego.

3.2 General System Architecture

The general architecture of the proposed system for stego detection is shown in Figure (3.1). This system presents a new idea to detect stego image without a dataset. According to the correlative analysis explained in the previous chapter, it is suggested to use the proposed system of steganalysis method. Steganalysis is a method for detecting the presence of a secret message in the cover image. The proposed system computes the autocorrelation function of the image histogram, then takes the first, second, and third derivatives of the autocorrelation function to test whether any steganography exists in the image. If steganography exists, then the derivatives of the autocorrelation function will have significant ripples. A high-pass filter can extract the high-pass content, Fourier transforms correlation derivatives or entropy correlation derivatives where a threshold could be used for a decision.

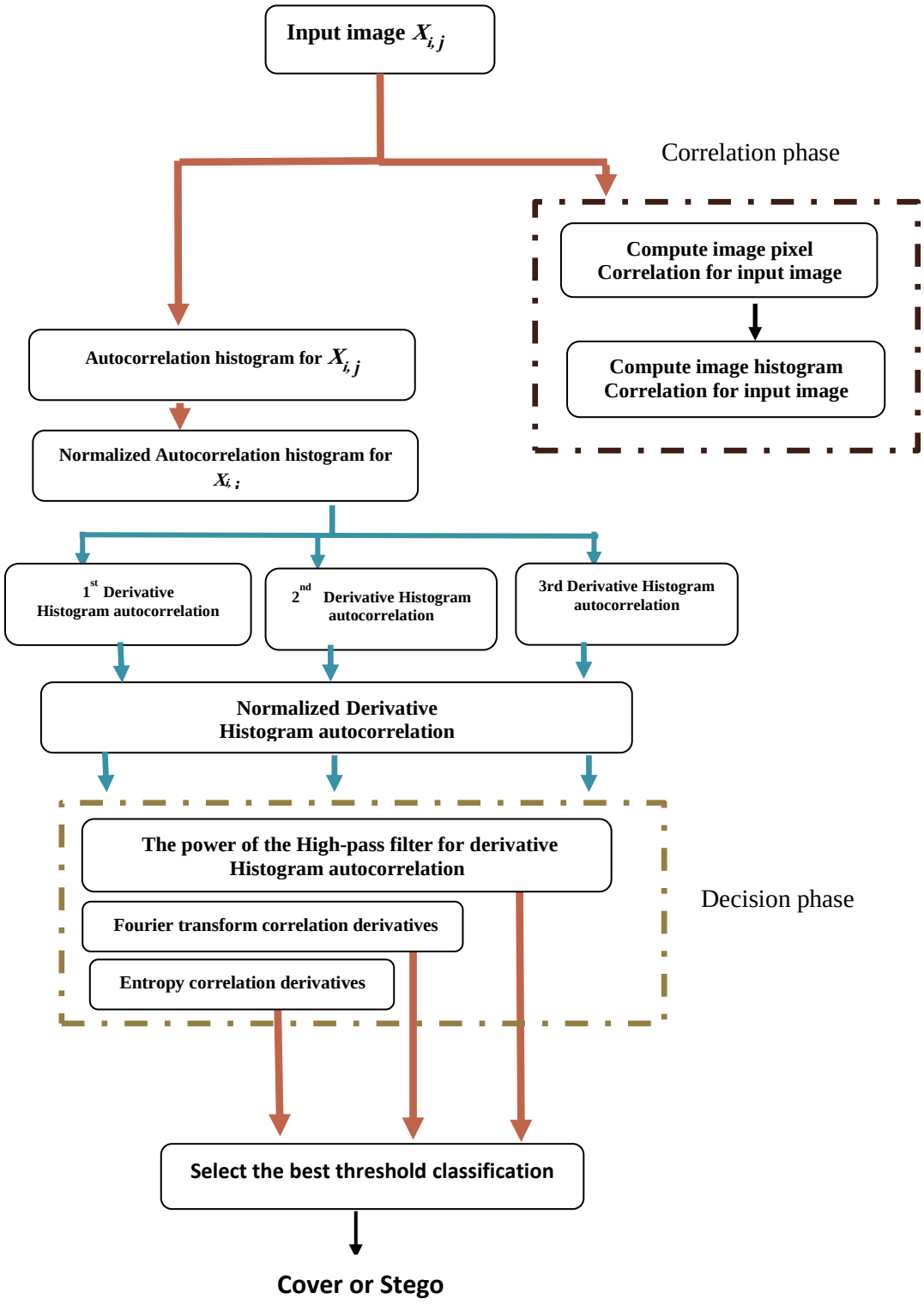

Figure 3.1: Block diagram of the proposed system.

The elements of the proposed work are described in detail in the subsections below.

3.2.1 Image histogram

We compute an image's histogram. The image histogram displays the frequency of pixel intensity values. It is the first step of the proposed system.

3.2.2 Correlation statistics

In this step, we compute the Pearson (linear) image pixel correlation and the Pearson (linear) image histogram correlation for the tested image. Then the comparison between them to decide on what to do in the next stage.

I. Image pixel correlation

Before applying for the derivative order, the Pearson (linear) correlation for the tested image was utilized to discover the correlated value between image pixels. Algorithm (3.1) show how to compute the correlation (Pearson correlation) using equation (2.1) in Chapter Two. Two attributes (*X*, *Y*) from the tested image values are selected randomly, and the Pearson (linear) correlation r is applied between these two attributes. The values of ***r*** fall into the range [-1, 1]. If *X* and *Y* are completely correlated, then ***r*** is equal or close to 1 or −1. If *X* and *Y* are independent, then ***r*** is equal or closely to 0 as shown in Figure (2.9).

Algorithm (3.1): Image pixel Correlated Feature

Input: g **Tested_ Image**

Output: ***r*** image pixel correlated feature

1. ***Begin***

2. *g←read image (Tested_Image)*

3. *b_ c ← size input image*

4.	$x_1 \leftarrow$ *input image (1_to_c-1)*
5.	$t_1 \leftarrow$ *input image (2_to_c)*
6.	$s_1 \leftarrow b*(c-1)$ // *size* x_1
7.	$I_r \leftarrow$ *rand* (s_1) // *random index between 1 and* s_1
8.	$K_p \leftarrow$ *round toward positive _s1*
9.	$I \leftarrow I_r(1\_ to\_K_p)$ // *choose* K_p *points only*
10.	$x \leftarrow x_1(I)$ // K_p *random pixel-pairs*
11.	$y \leftarrow t_1(I)$ // K_p *random pixel-pairs*
12.	*Apply Pearson correlation* (*Equation 2.1*)
13.	***return r***
14.	***End Algorithm.***

Steps (4-5) in the algorithm represented the range of input image as x_1 is the vector of the input image from 1 to c-1, and t_1 is the vector of the input image from 2 to c, where c is the final value in this image. Also, steps (10-11) in the algorithm represented the random pixel pairs between 1 to c. The output of this algorithm is the value of ***r*** represent the image pixel correlation value of input image.

II. Image histogram correlation

The Pearson (linear) image histogram correlation was utilized to discover the correlated value between image histogram. Algorithm (3.2) show how to compute the correlation (Pearson correlation) using equation (2.1) in Chapter Two.

Algorithm (3.2): Image histogram Correlated Feature

Input: **g Tested_ Image**

Output: **r** image histogram correlated feature

1. ***Begin***

2. *g← read image(Tested_Image)*

3. *b_ c ← size input image*

4. *V← 255*

5. *v← 0_to V*

6. *gh← Convert_tested image to one column*

7. *H← histogram of the Tested_ image_ (gh ,v)*

8. *x← H(1_to_v-1)*

9. *y← H(2_to_v)*

12. *Apply Pearson correlation*（*Equation 2.1*）

13. return r

17. End Algorithm.

Step (5) in the algorithm represented the range of image histogram as *v* is the vector of the image histogram from 0 to *V*, *x* is the vector of the image histogram from 1 to *v-1*, and *y* is the vector of the input image histogram from 2 to *v*, where *v* is the final value in this image histogram. Also, step (7) in the algorithm represented the histogram of the Tested image between 0 to *v*. The output of this algorithm is the value of ***r*** represent the image histogram correlation value of input image.

3.2.3 Autocorrelation of an image histogram

We compute the autocorrelation value of the histogram image that will be utilized to detect the autocorrelated values between the tested image histogram at this stage. Algorithm (3.3) demonstrated how to use equation (2.2) in Chapter Two, to calculate the autocorrelation function. The autocorrelation function at lag *k* is represented by the value R_K and is applied between two attributes (X_t, X_{t+k}), where *k* is the time step. After normalizing autocorrelation, the resulting number can range from 0 to 1, both of which are symmetric about the x=0 line when calculated.

Algorithm (3.3): Autocorrelated Feature

Input: **g Tested_ Image**

Output: R_K normalize the autocorrelated feature

1. ***Begin***

2. *g ← read image (Tested_Image)*

3. *V ← 255*

4. *v ← 0_to V*

5. *gh ← Convert_tested image to one column*

6. *H ← histogram of the Tested_ image_ (gh ,v)*

7. *Apply Autocorrelation of histogram of the Tested Image using* equation (2.2)

8. *N ← length_ histogram of the Tested_ image*

9. *T ← -(N-1)_to_N-1*

10. *Lc ← length_T* // *T is the lag variable for* R_K

11. *Mx ← max_*R_K // *Mx is max value of* R_K

12. $R_K \leftarrow R_K/Mx$ // *normalize the autocorrelation*

13. return the vector R_K

19. End Algorithm.

Step (4) in the algorithm represented the Autocorrelation of the histogram of the tested image as a vector from 511 values. Also, step (10) is the lag variable for Autocorrelation. The output of this algorithm is the vector values of R_K. It values in the range [0,1] both of which are symmetric about the T= 0 line and the large values at T is equal to 0.

3.2.4 Derivative histogram autocorrelation

In the current work, we compute the first, second and third derivatives of the histogram-autocorrelation of the input image. Algorithm (3.4) shows how to compute the derivative order using equations (2.4) and (2.5) in Chapter Two. The derivative order value at lag T is represented by the value d. The resulting number can range from [+1,-1] when calculated.

Algorithm (3.4): Derivatives-order of the Histogram-autocorrelation

Input : ***vector R_K***

Output : **Normalized derivative-order d1,d2,d3**

1. ***Begin***

2. *vector R_k from algorithm 3.3*

3. *compute first-order derivative of histogram autocorrelation using* Equation (2.4)

4. *$d1 \leftarrow diff(R_k)$* *// first-order derivative function*

5. *normalize the first derivative*

6. *$T1 \leftarrow T(1\_to\_Lc\text{-}1)$* *// T1 is the lag variable for d1*

7. *compute second-order derivative of histogram autocorrelation using Equation (2.5)*

8. *d2← diff(d1)* *// second-order derivative function*

9. *normalize the second derivative*

10. *T2← T1(2_to_Lc-1)* *// T2 is the lag variable for d2*

11. *compute third-order derivative of histogram autocorrelation*

12. *d3← diff(d2)* *// third-order derivative function*

13. *normalize the third derivative*

14. *T3← T2(2_to_Lc-2)* *// T3is the lag variable for d3*

15. ***return the vectors d1, d2, d3***

19. ***End Algorithm.***

Steps (4,8,12) in the algorithm represented the first, second, and third derivative of histogram autocorrelation of the input image. Then normalized the derivative. Also, *T1*, *T2*, and *T3* is the lag variable for derivatives of histogram autocorrelation of the input image. The output of this algorithm is the values of three vectors, *d1, d2, d3*. It values in the range of [-1,1] at lag *T1, T2, T3*.

3.2.5 Decision threshold of the system

The final step of the system is deciding (classification) whether the input image is a cover or a stego, without adopting the original image. This decision can you taking with one of three methods:

A. The power of high-pass filtered derivatives

The high pass filter can extract the high pass content, where a threshold could be used for a decision. Algorithm (3.5) show how the power of high-pass filtered derivatives can be used as a decision.

Algorithm (3.5): The power of the high-Pass Filter

Input: ***The vectors d1, d2, d3***

Output: The values of the decision threshold sz0, su0, sw0

1. ***Begin***
2. ***The vectors d1, d2, d3 from above algorithm 3.4***
3. ***Design of the High-Pass Filter***
4. $n \leftarrow 10$ *// the number of values for the filter 0 to 10 values*
5. $wn \leftarrow 0.5$ *// cutoff frequency*
6. $b\_a \leftarrow butter(n,wn,'high')$ *// the values of high pass filter*
8. ***Compute z0,u0,w0 for the input image***
9. $z0 \leftarrow filter(b,a,d1)$ *// high pass filtered 1st derivative*
10. $u0 \leftarrow filter(b,a,d2)$ *// high pass filtered 2nd derivative*
11. $w0 \leftarrow filter(b,a,d3)$ *// high pass filtered 3rd derivative*
12. ***change in power of derivatives***
13. $sz0 \leftarrow \sum(z0)^2$ *// sum the high pass filtered 1st derivative*
14. $su0 \leftarrow \sum(u0)^2$ *// sum the high pass filtered 2nd derivative*
15. $sw0 \leftarrow \sum(w0)^2$ *// sum the high pass filtered 3rd derivative*
16. ***End Algorithm.***

Steps (9,10,11) in the algorithm represented high pass filtered of the first, second, and third derivative of histogram autocorrelation of the input image. Also, Steps (13,14,15) in the algorithm represented the sum of the high pass filter of the first, second, and third derivative of histogram autocorrelation of the input image. The output of this algorithm is the values *sz0*, *su0* and *sw0*. These values represent the decision of the proposed system. For natural images, the power of its filtered derivative of histogram autocorrelation was low, while the power of the stego one was high; hence a threshold of 1 or higher would be sufficient to detect a stego image.

B. Fourier transform correlation derivatives

It is the second method of making a threshold that could be used for a decision. Algorithm (3.6) demonstrated how to use Fourier transform correlation derivatives as a threshold of the proposed system.

Algorithm (3.6): Fourier transform correlation derivatives

Input: ***The vectors d1, d2, d3***

Output: The values of the decision threshold r1o, 12o, r3o

1. ***Begin***

2. ***The vectors d1, d2, d3 from above algorithm 3.4***

3. ***compute Fourier transform correlation derivatives***

4. *Apply function of Fourier transform correlation 1st derivative using equation (2.12)*

5 *normalize of Fourier transform correlation 1st derivative*

6. *Apply function of Fourier transform correlation 2nd derivative using equation (2.12)*

7. *normalize of Fourier transform correlation 2nd derivative*

8. *Apply function of Fourier transform correlation 3rd derivative using equation (2.12)*

9. *normalize of Fourier transform correlation 3rd derivative*

10. ***Define HP-region and LP-region of FT correlation derivatives***

11. *fo ← 0.1* *// border of LP/HP*

12. *Ln←256* *// +ve range, including f=0*

13. *no ← (fo*Ln)* *// index of fo*

14. *mL1 ← max(ft1n(1_to_no))* *// the max LP-region of FT(hist-corr 1st derv's)*

15. *mH1 ← max(ft1n(no+1_to_Ln))* *// the max HP-region of FT(hist-corr 1st derv's)*

16. *mL2 ← max(ft2n(1_to_no))* *// the max LP-region of FT(hist-corr 2nd derv's)*

17. *mH2 ← max(ft2n(no+1_to_Ln))* *// the max HP-region of FT(hist-corr 2nd derv's)*

18. *mL3← max(ft3n(1_to_no))* *// the max LP-region of FT(hist-corr 3rd derv's)*

19. *mH3← max(ft3n(no+1_to_Ln))* *// the max HP-region of FT(hist-corr 3rd derv's)*

20. *r1o ← mH1/mL1* *// Max of HP/Max of LP(FT(der1))*

21. *r2o ← mH2/mL2* *// Max of HP/Max of LP(FT(der2))*

22. *r3o ← mH3/mL3* *// Max of HP/Max of LP(FT(der3))*

23. End Algorithm.

Steps (4,6,8) in the algorithm represented the Fourier transform correlation 1st, 2nd, and 3rd derivative of histogram autocorrelation of the input image. Then normalize of Fourier transform correlation 1st, 2nd, and 3rd derivative. Step (10) in the algorithm defines the High pass region and Low pass region of Fourier transform correlation derivatives. Step (12) represented a positive range, including *f*=0. Also, Steps (14,16,18) represented the max Low pass region of Fourier transform (histogram

correlation 1st, 2nd, and 3rd derivatives). Steps (15,17,19) represented the max High pass region of Fourier transform (histogram correlation 1st, 2nd, and 3rd derivatives). Then the values r1o, r2o, and r3o are the results of this algorithm. These values indicate the proposed system's decision. For natural images, most information is LP (Low pass) region. Hence, max (HP)/max (LP) is less than 1. Powerful HP (High pass) components appear after steganography, making max (HP) bigger than max (LP). hence a threshold of 1 or higher would be sufficient to detect a stego image.

C. Entropy correlation derivatives

It is the third method of making a threshold that could be used for a decision. Algorithm (3.7) demonstrated how to use Entropy correlation derivatives of the image histogram as a decision of the proposed system.

Algorithm (3.7): Entropy correlation derivatives

Input: ***The vectors d1, d2, d3***

Output: The values of the decision threshold *sz11*, *su11*, *sw11*

1. ***Begin***

2. ***The vectors d1, d2, d3 from above algorithm 3.4***

3. ***compute Entropy correlation derivatives***

4. *eeo1 ← wavelet_entropy(d1,'shannon')* *// function of Entropy correlation 1st derivative*

5. *eeo2 ← wavelet_entropy(d2,'shannon')* *// function of Entropy correlation 2nd derivative*

6. *eeo3 ← wavelet_entropy(d3,'shannon')* *// function of Entropy correlation 3rd derivative*

7. ***change in power of Entropy derivatives***

8. $sz11 \leftarrow (eeo1)^2$ *// power the Entropy 1st derivative*

9.	$su11 \leftarrow (eeo2)^2$	*// power the Entropy 2nd derivative*
10.	$sw11 \leftarrow (eeo3)^2$	*// power the Entropy 3rd derivative*
11.	***End Algorithm.***	

Steps (4,5,6) in the algorithm represented the wavelet_entropy correlation 1st, 2nd, and 3rd derivative of histogram autocorrelation of the input image. The output of this algorithm is the values *sz11*, *su11*, and *sw11*. These values represent the decision of the proposed system. The wavelet entropy of the autocorrelation derivative of the image histogram will be low since the image is original. It will be high if the tested image is stego.

Chapter Four

Implementation and Results

Chapter Four

Implementation and Results

4.1 Introduction

This chapter explains the implementation and the results of the proposed system. Grey-scale and color images are used to test the proposed system. The proposed system is used for testing and attack detection. The dataset is grey-scale images created by K-LSB [50], Chaotic-LSB [52], enhanced-LSB stego images [51], and stego color images created by A Hybrid Image Steganography Method based on a Genetic Algorithm [53] to test and evaluate the results of the proposed system. Then, the blind proposed system is tested and evaluated by different images steganography formats. The measures of evaluation are used to evaluate the proposed system.

4.2 System Requirement

Hardware: Processor Intel ® Core ™ i7-4510 CPU, Ram 8 GB, Storage 500 GB, Freq. 2.2 GHz.
Operating System**:** Windows 10 Home 64bit.

Programming Language**:** MATLAB R2021a (64-bit).

4.3 Datasets Preparation

Five steganographic methods are used to evaluate the proposed method for detecting the stego image. These steganographic methods are K-LSB, Chaotic-LSB, enhanced-LSB stego images, and stego color images created by A Hybrid Image Steganography Method Based on a Genetic Algorithm and MPVD, which were prepared using the BOSSbase v1.01 dataset for the

embedding in the spatial domain. BOSSbase dataset contains 10,000 uncompressed natural images.

Then using a dataset containing 10,000 grayscale images embedding by MPVD steganography that were 512 × 512 pixels in size [81] and Mendeley Data containing 1500 RGB images that were 512 × 512 pixels in size [83] to evaluate the proposed system.

4.4 Correlation statistics

In this section, the Pearson (linear) correlation for the tested image was used to measure the correlated value between image pixels and image histogram. Comparison of the results to decide what to do in the next step.

4.4.1 Image pixel correlation

The algorithm (3.1) is applied to the grey-scale and color images, the Pearson (linear) correlation concept is a good statistical indicator for testing and discovering the correlated value of the tested image pixel. Table (4.1) shows the adjacent pixels correlation decrease when the size message ratio(Rm) and K-LSB are growing up. It decreases slowly and stays positive very strong correlation for the cover and stego images Figure (4.2) confirms this fact. To find the correlation between two adjacent pixels, pairs of adjacent pixels have been selected randomly from the original and stego images. Figure (4.2) shows the correlation of adjacent pixels in the cover and k-LSB-stego images for different values of k (the number of least significant bits to be encoded with the message). It is clear that introducing a message in the cover image will reduce

the natural correlation between its pixels. The correlation damage will be more if we used more LSBs (i.e., larger k). Hence, for better security of the message, the message size should be selected as small as possible when compared to the cover image.

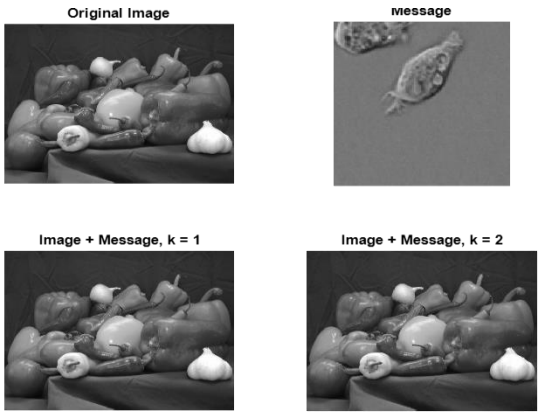

Figure 4.1: The grey-scale cover and stego MATLAB images.

Table (4.1): Pixel correlation to images in figure (4.1).

No.	Pearson correlation (Cover image)	Pearson correlation (Stego image) (k=1)	Pearson correlation (Stego image) (k=2)	Pearson correlation (Stego image) (k=4)	size ratio between message size and cover size (Rm)
1	0.9913272	0.9913265	0.9911616	0.9908681	0.001
2		0.9913259	0.9911604	0.9908567	0.005
3		0.9913227	0.9911527	0.9907994	0.020
4		0.9913184	0.9911456	0.9907484	0.046
5		0.9913045	0.9911182	0.9905693	0.127
6		0.9912515	0.9910314	0.9898853	0.509
7		0.9911621	0.9908691	0.9888058	0.114

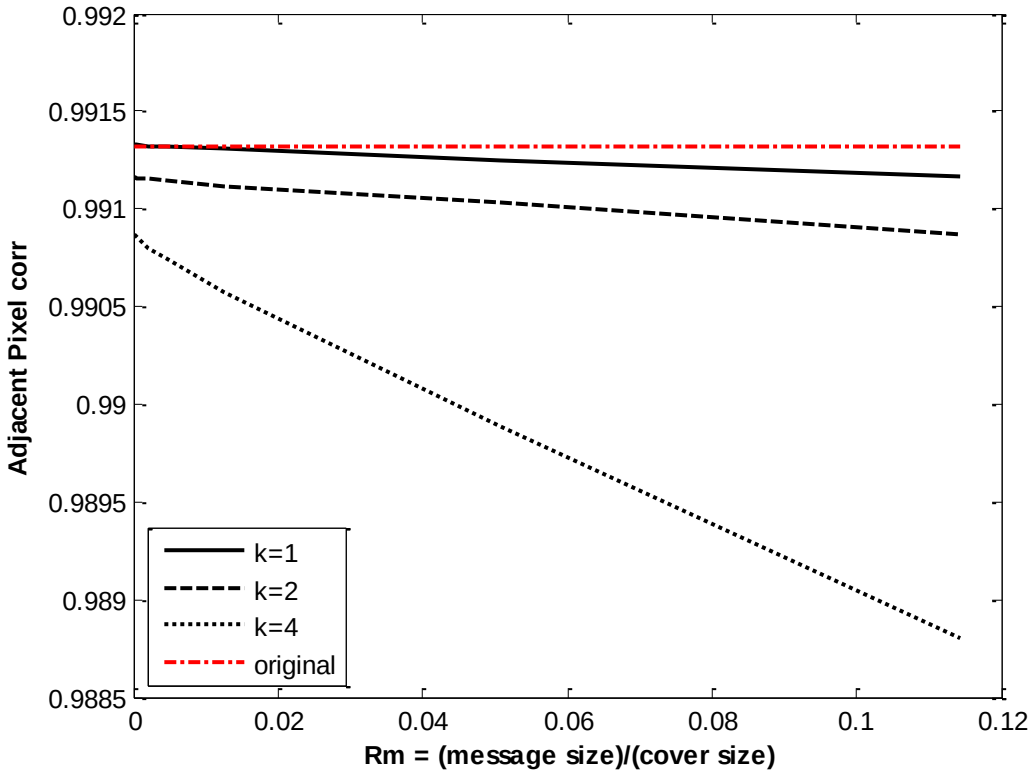

Figure (4.2): Correlation of adjacent pixels in the cover and k-LSB-stego images for different k.

Figure 4.3 depicts the pixel correlation between the cover and stego images. If we utilize more LSB's, the correlation damage will be greater (i.e., larger k). Figure (4.4) demonstrates the scatter plot for the cover and the stego images and it depicts the correlation value between adjacent pixels. It is a very strong correlation even when the secret message in the image exists.

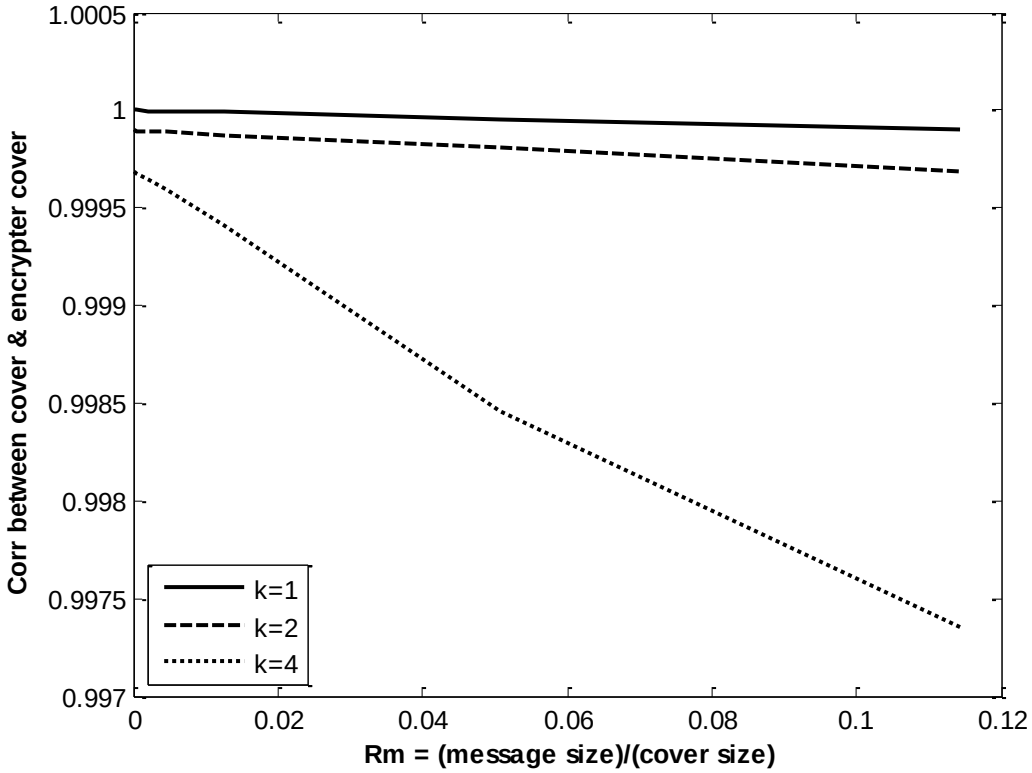

Figure (4.3): pixel correlation between the cover and stego images.

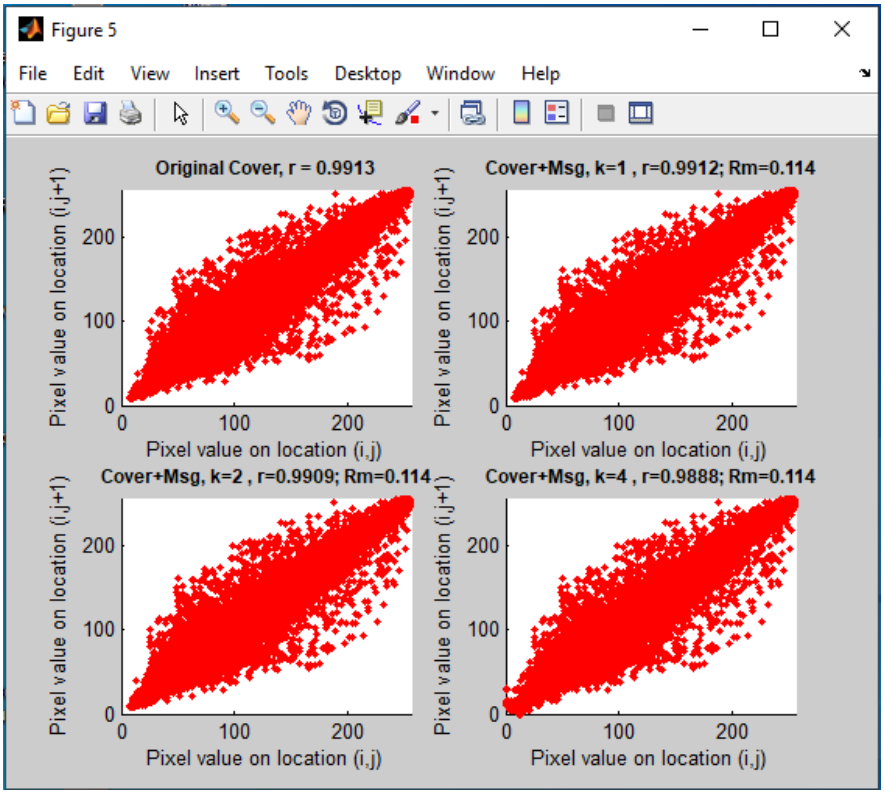

Figure (4.4): The scatter plot of adjacent pixels in the cover and k-LSB-stego images for different k.

4.4.2 Image histogram correlation

The Pearson (linear) correlation idea is a suitable statistical indicator for testing and discovering the correlated value of the tested image histogram. The algorithm (3.2) is implemented for grey-scale and color images. Table (4.2) shows the adjacent histogram correlation decreases when the size message ratio (Rm) and K-LSB grow up. It is clear that introducing a message in the cover image will reduce the natural correlation between its histogram and that the truth becomes clearer in histogram correlation than in pixel correlation. Figure (4.5) confirms this fact clearly.

Table (4.2): Histogram correlation to images in the cover and k-LSB-stego images for different k.

No.	Histogram correlation (Cover image)	Histogram correlation (Stego image) (k=1)	Histogram correlation (Stego image) (k=2)	Histogram correlation (Stego image) (k=4)	size ratio between message size and cover size (Rm)
1	0.9885112	0.9884820	0.8804239	0.91964235	0.001
2		0.9885485	0.8796743	0.9200170	0.005
3		0.9882802	0.8789658	0.9203808	0.020
4		0.9875907	0.8751177	0.9209426	0.046
5		0.9819610	0.8586152	0.9197193	0.127
6		0.9615548	0.8583395	0.8961505	0.509
7		0.8800120	0.9195122	0.9280765	0.114

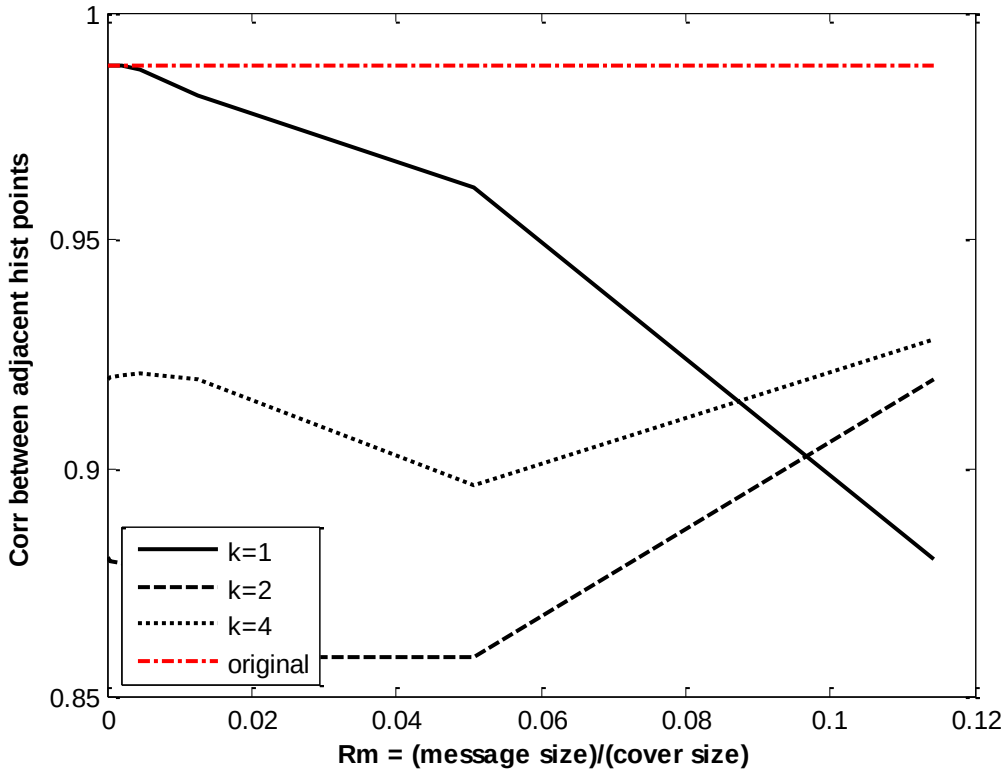

Figure (4.5): Correlation of adjacent histogram in the cover and k-LSB-stego images for different k.

4.5 Autocorrelation of an image histogram

The algorithm (3.3) is applied to the tested image histogram. Figure (4.6) shows the normalized autocorrelation of the histograms of the cover and stego images using the k-LSBs steganography system (i.e., the stego image is the cover image loaded with a hidden message as another image whose pixels have been distributed over k-LSB's of the cover image).

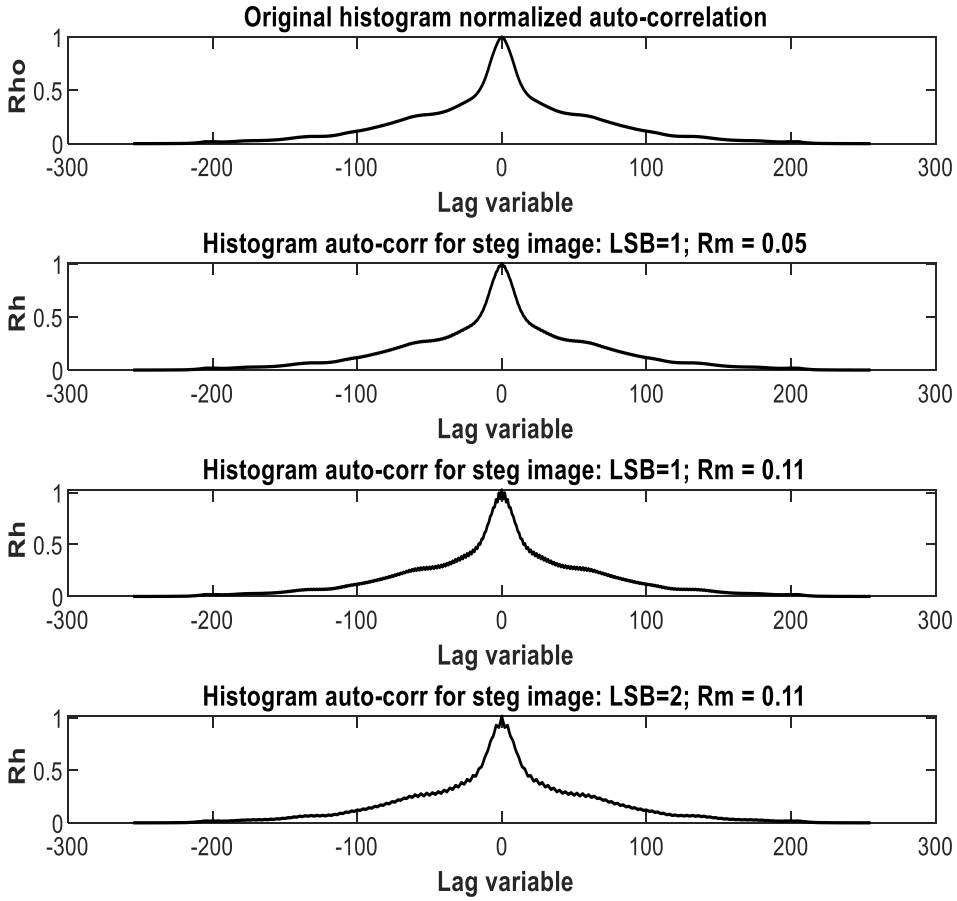

Figure (4.6): Normalized autocorrelation of the histograms of the cover and stego-image (cover image loaded with a message) using k-LSBs steganography system.

4.6 Derivative histogram autocorrelation

The algorithm (3.4) is applied to the histogram autocorrelation of the tested image. Figure (4.7) shows the effect of the first derivative on the histogram autocorrelation of the cover and LSB-stego images. It can be noticed that LSB-stego will introduce significant ripples when the autocorrelation function of the image histogram is subjected to derivation. On the other hand, it is found that using higher derivatives may not be necessary (only in this case),

as they do not introduce a significant difference from the first derivative in revealing stego image when the ripples are clear in the first derivative. Figures (4.8) and (4.9) show the ripples introduced by the second and third derivatives of the histogram autocorrelation functions. Hence, the second and third derivatives can be used as a support for the decision which can be based on the first derivative only.

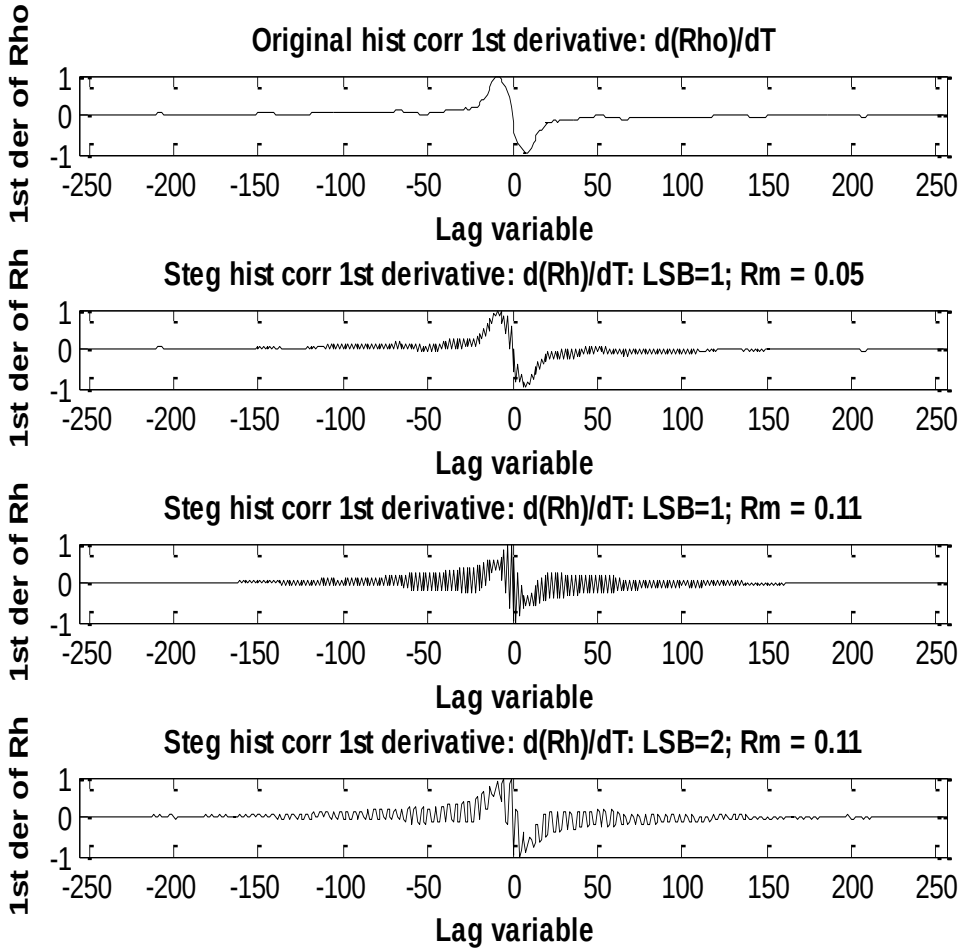

Figure (4.7): First derivative of the histogram autocorrelation function of the cover and LSB-stego images for different LSB levels and a different message to cover size ratios.

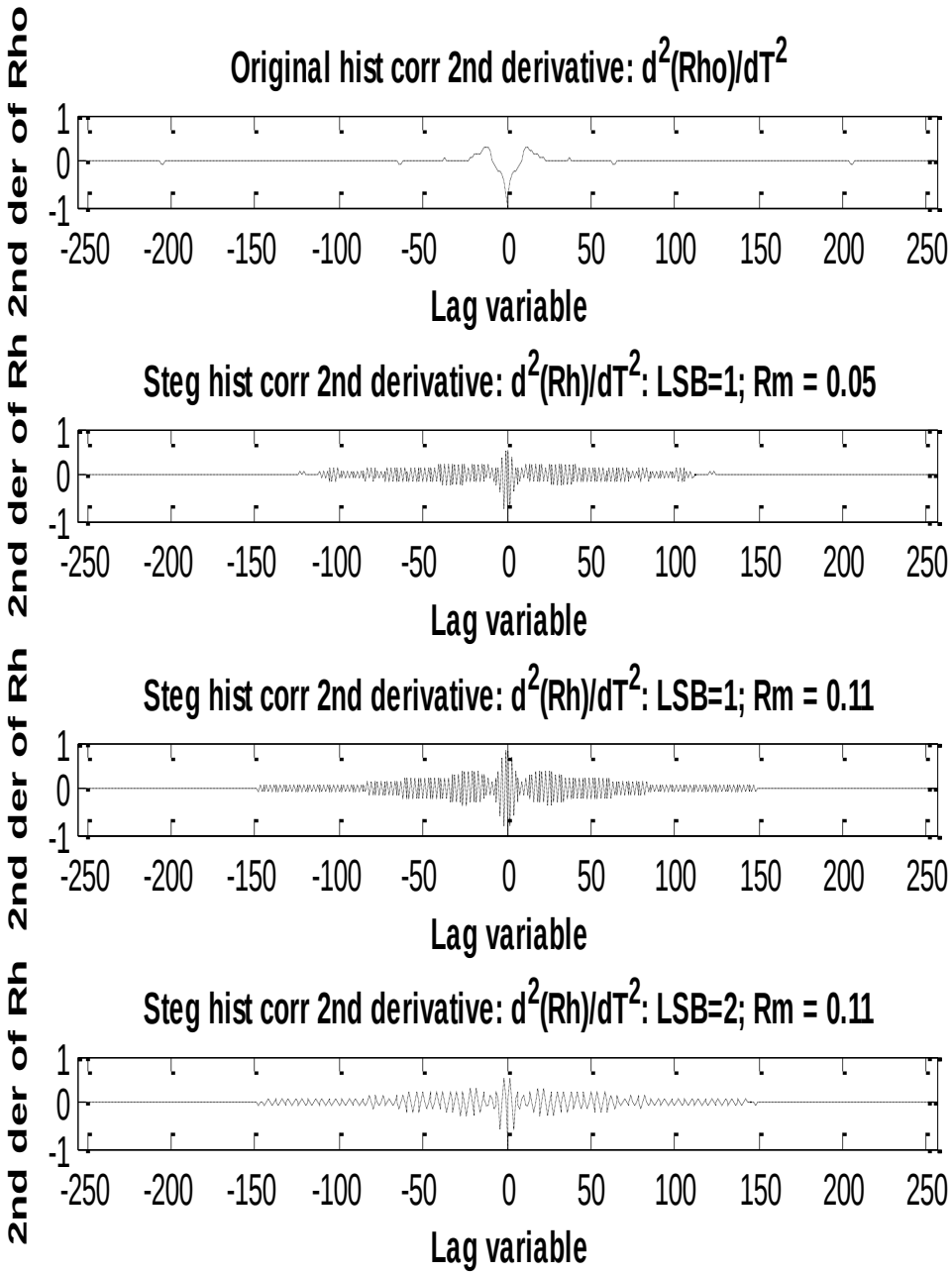

Figure (4.8): The second derivative of the histogram autocorrelation function of the cover and LSB-stego images for different LSB levels and a different message to cover size ratios.

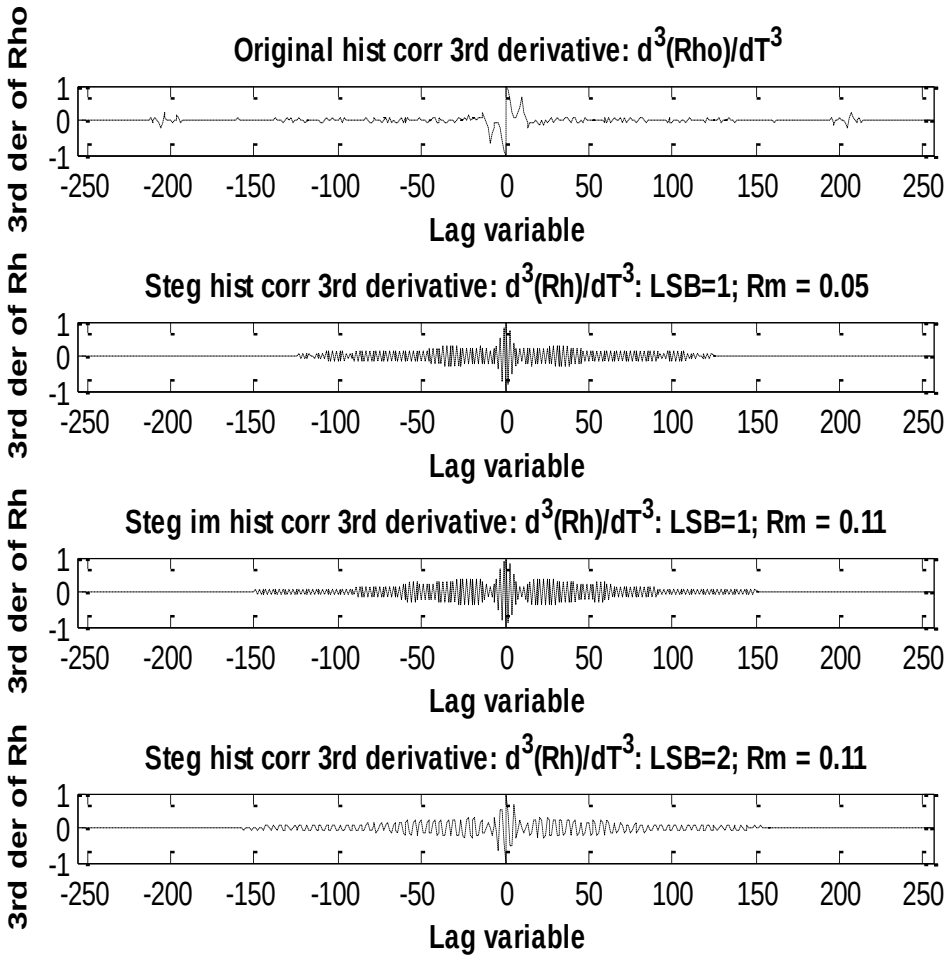

Figure (4.9): The third derivative of the histogram autocorrelation function of the cover and LSB-stego images for different LSB levels and a different message to cover size ratios.

4.7 Decision of the system

In the final step, the proposed system provides significant tools to discover these ripples, which represent the decision on the tested image whether it is stego or clear.

4.7.1 The power of high-pass filtered derivatives

The algorithm (3.5) is implemented to the derivatives histogram autocorrelation of the tested image. The high-pass content may be extracted

using a high-pass filter, as well as a threshold can be utilized to make a decision. Hence, high-pass filtering of this derivative will be a good tool to discover these ripples, where a threshold should be used as will be explained below. The high-pass filter used in this dissertation is the digital Butterworth filter of order 10 and whose frequency response is shown in Figure (4.10).

After the HPF of the derivative of the autocorrelation function of the histogram, the power of the filtered signal is calculated. As the original image is smoothly correlated with smooth histogram autocorrelation, the filtered version of the derivative of histogram autocorrelation will have a very low.

However, if the image is stego, then the power of the filtered derivative of the histogram autocorrelation will be high.

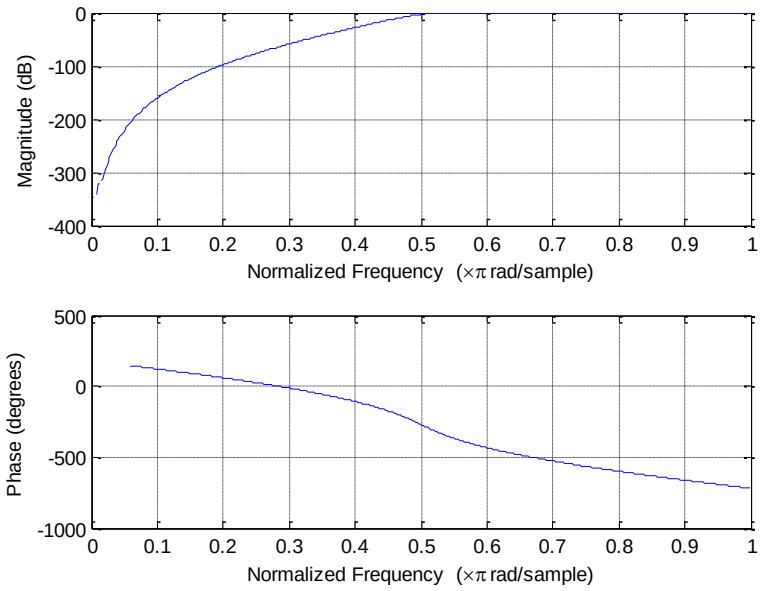

Figure (4.10): The frequency response of the high-pass filter used to determine the threshold of stego detection.

In the following experiments, we found that the ratio of the stego power to the original power is more evident if we use the first derivative, where this ratio exceeds 1000, figure (4.11) confirms this fact. For untampered natural images, the power of its filtered derivative of histogram autocorrelation was as

low as 0.0015, while the power of the stego one was 2.3733; hence a threshold of 1 or higher would be sufficient to detect stego, although the threshold could be image-dependent.

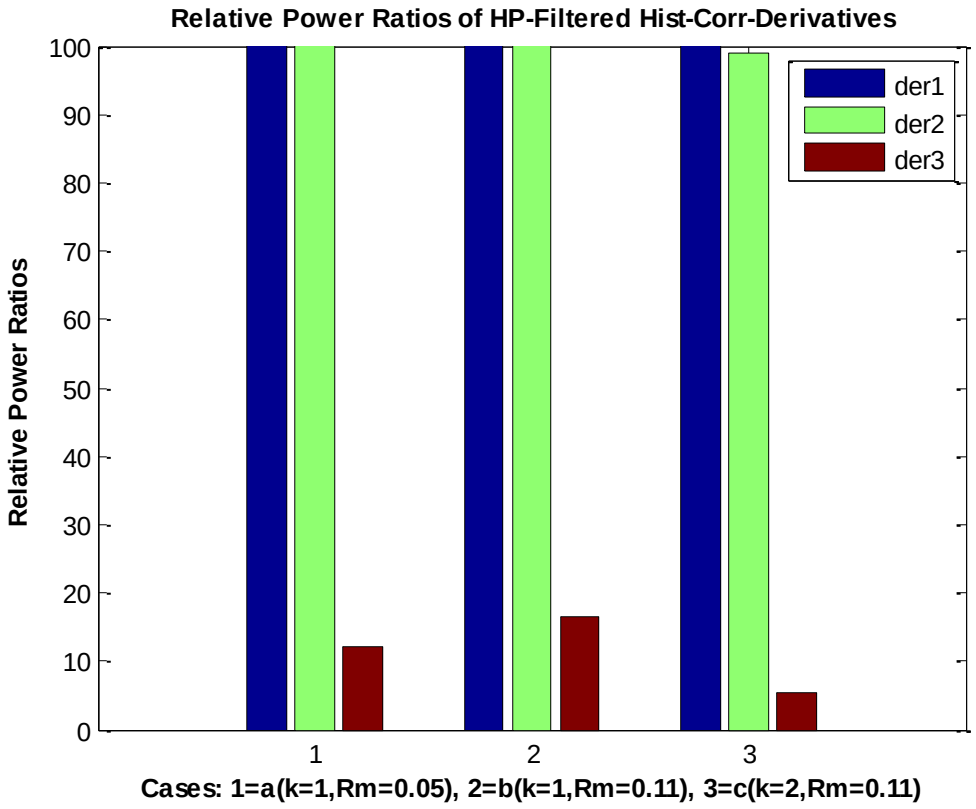

Figure (4.11): Relative Power Ratios of HP-Filtered Hist-Corr-Derivatives.

4.7.2 Fourier transform correlation derivatives

It is the second way of determining a threshold that may be used in a decision for the gray-scale image. Algorithm (3.6) is applied on the tested image's derivatives histogram autocorrelation. In Figure (4.12), we found that for natural images, most information is LP (Low pass) region. Hence, max (HP)/max (LP) is less than 1. Powerful HP (High pass) components appear after steganography, making max (HP) is bigger than max (LP). In figures (4.12), (4.13), case 1 is the original image, and cases 2,3, and 4 are the stego images with different embedding ratios.

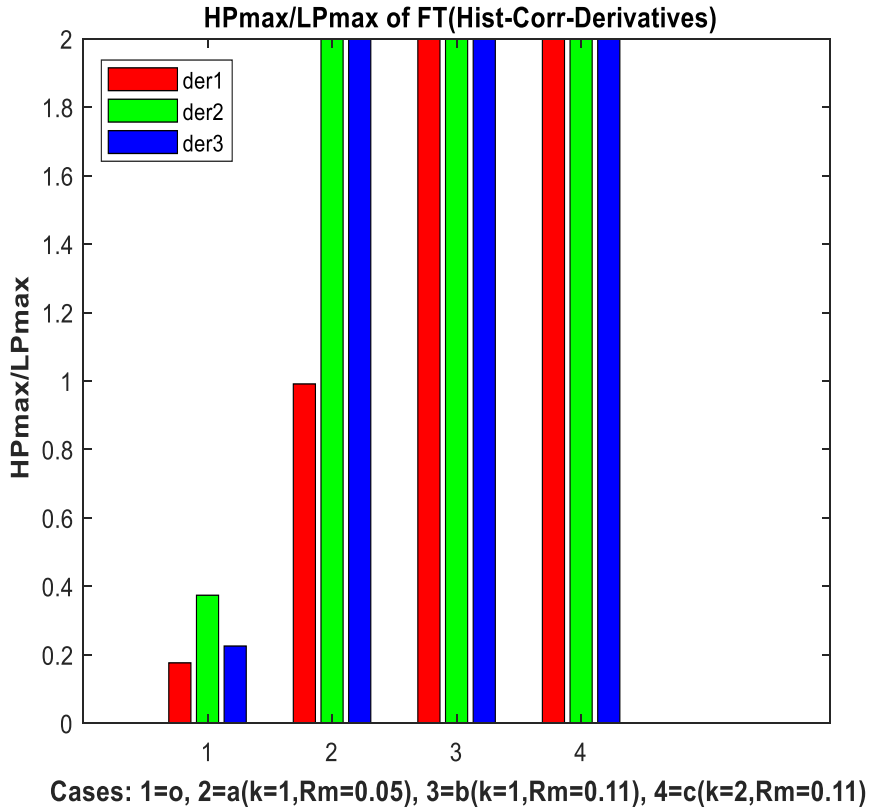

Figure (4.12): HP max / LP max of FT(Hist-corr-Derivatives).

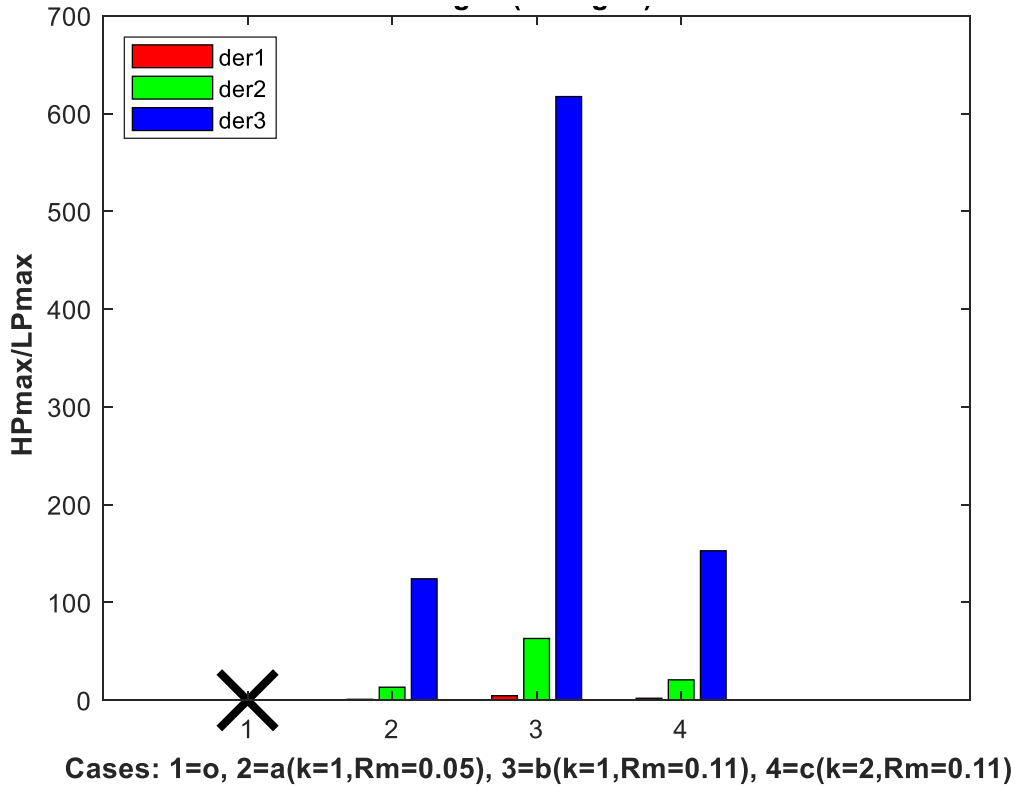

Figure (4.13) : HP max / LP max of FT(Hist-corr-Derivatives) (enlarged).

Figure (4.13) is the same as Figure (4.12), but Figure (4.13) is enlarged. Where the x-axis is the same in the two figures, but the y-axis in the Figure (4.13) is enlarged.

I. Expansion of experiments

In this experiment, we will show that higher-order derivatives could be necessary for other types of steganography. The suggested system was tested on three different LSB steganography techniques to conceal the message in the gray-scale image, and it tested the suggested system on another different steganography approach with a color image. The proposed method should be able to detect tampered images. Figure (4.14) describes the gray-scale images used in these experiments, with the message modified to achieve various message-to-cover ratios (Rm ratio). As the Histogram-Correlative of the image is a derivation, LSB-stego introduces major ripples. As a result, high-pass

filtering of this derivative would be a useful method for detecting these ripples, where a decision can be used.

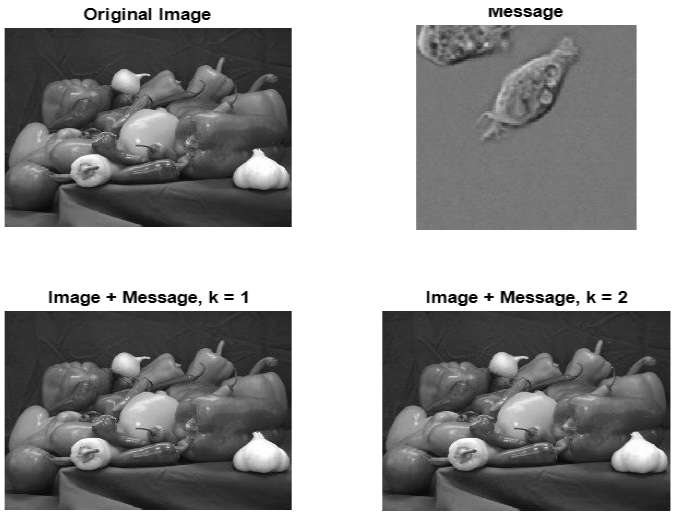

Figure (4.14): The grey-scale cover and stego MATLAB images.

Figures (4.15), (4.16), and (4.17) depict the effect of the first, second, and third derivatives on the histogram-correlative of the cover and K-LSB, Chaotic-LSB stego images. If the message is small compared to the cover (R_m), we should use the second and third derivatives since the first is unclear.

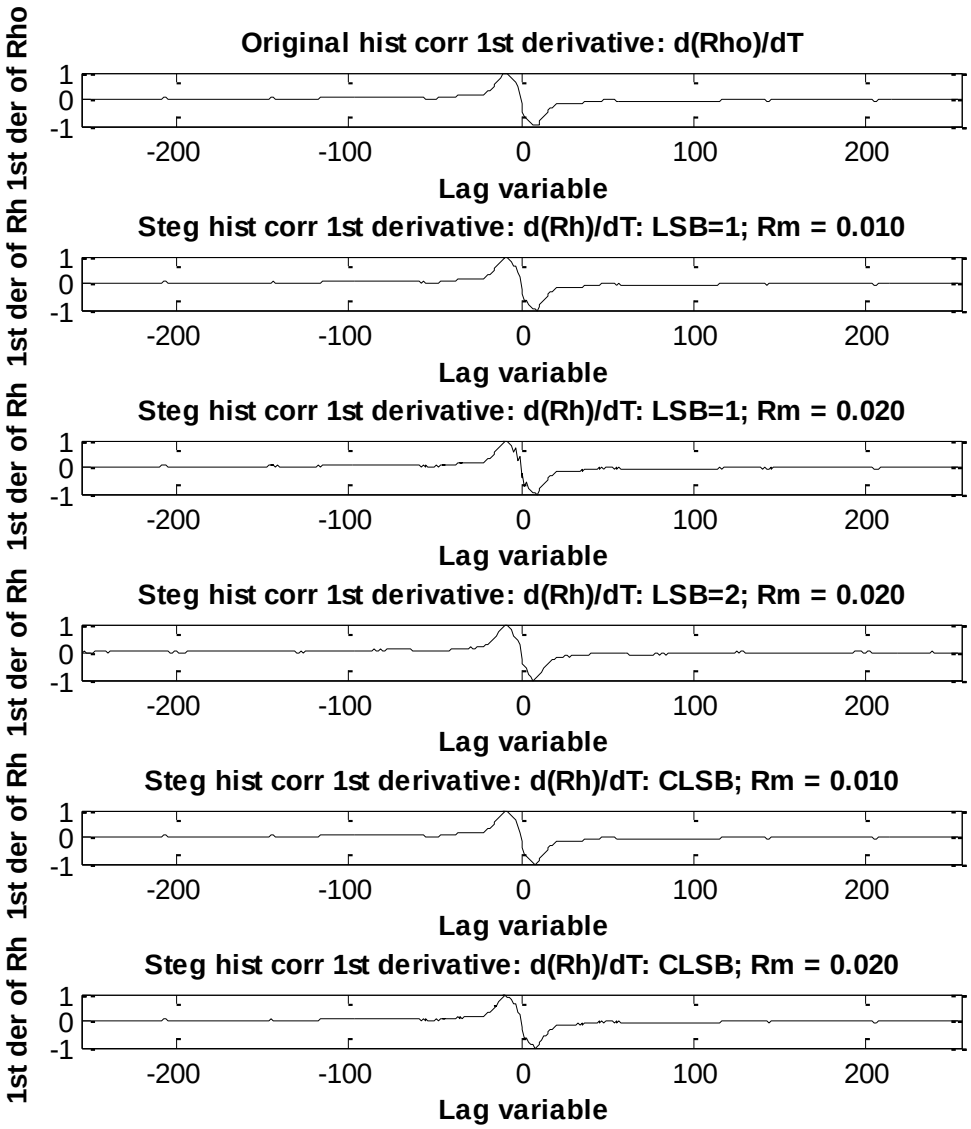

Figure (4.15): The 1st derivative of the histogram-correlative for Chaotic_LSB, K_LSB stego, and cover images for several LSB levels and several messages to cover size ratios.

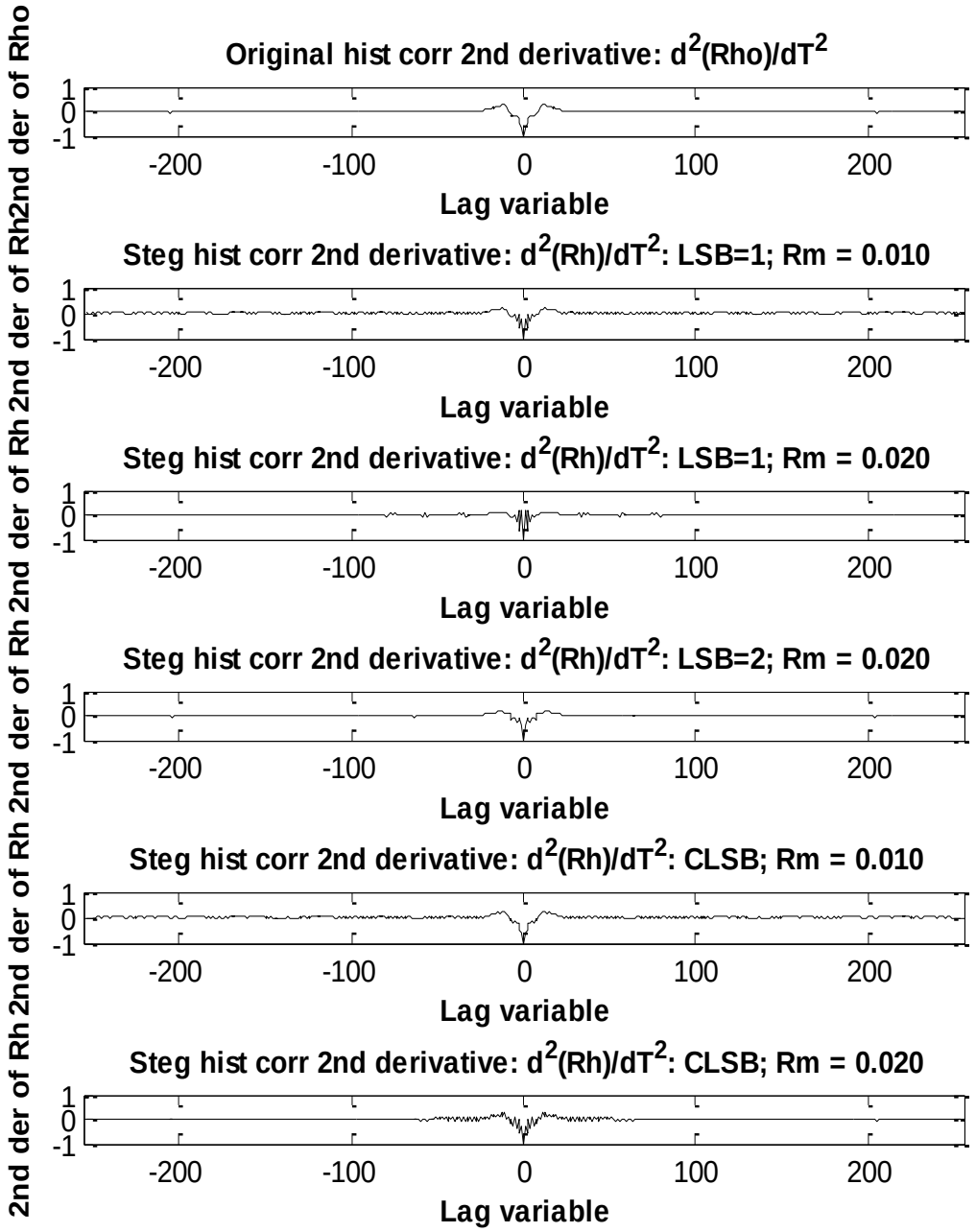

Figure (4.16): The 2nd derivative of the histogram-correlative for Chaotic_LSB, K_LSB stego, and cover images for several LSB levels and several messages to cover size ratios.

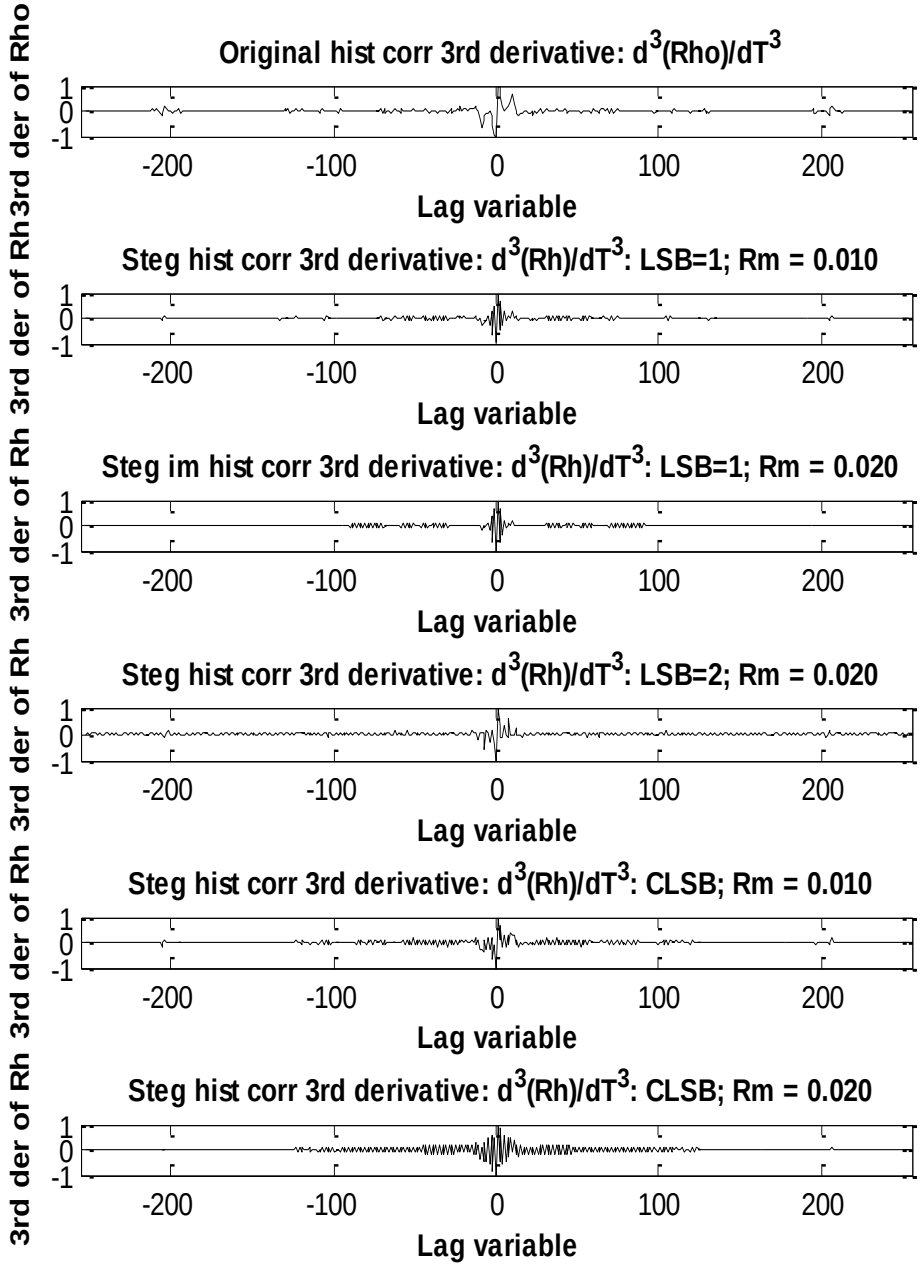

Figure (4.17): The 3rd derivative of the histogram-correlative for Chaotic_LSB, K_LSB stego, and cover images for several LSB levels and several messages to cover size ratios.

The effect of the first, second, and third derivatives on the histogram-correlative of the cover and K-LSB, enhanced-LSB stego images, is seen in Figures (4.18) a, b and c. The method fails if the message size is very small ($R_m < 0.01$). Higher derivatives are almost certainly required (future work).

K_m represents the number of different hidden message sizes in the system. K_m=[5 10 20 30 50 100 150]. As a result, Chaotic-LBS is not appropriate for large $R_m > 0.1$ and $K_m > 65$, so only K-LSB and enhanced-LSB are compared.

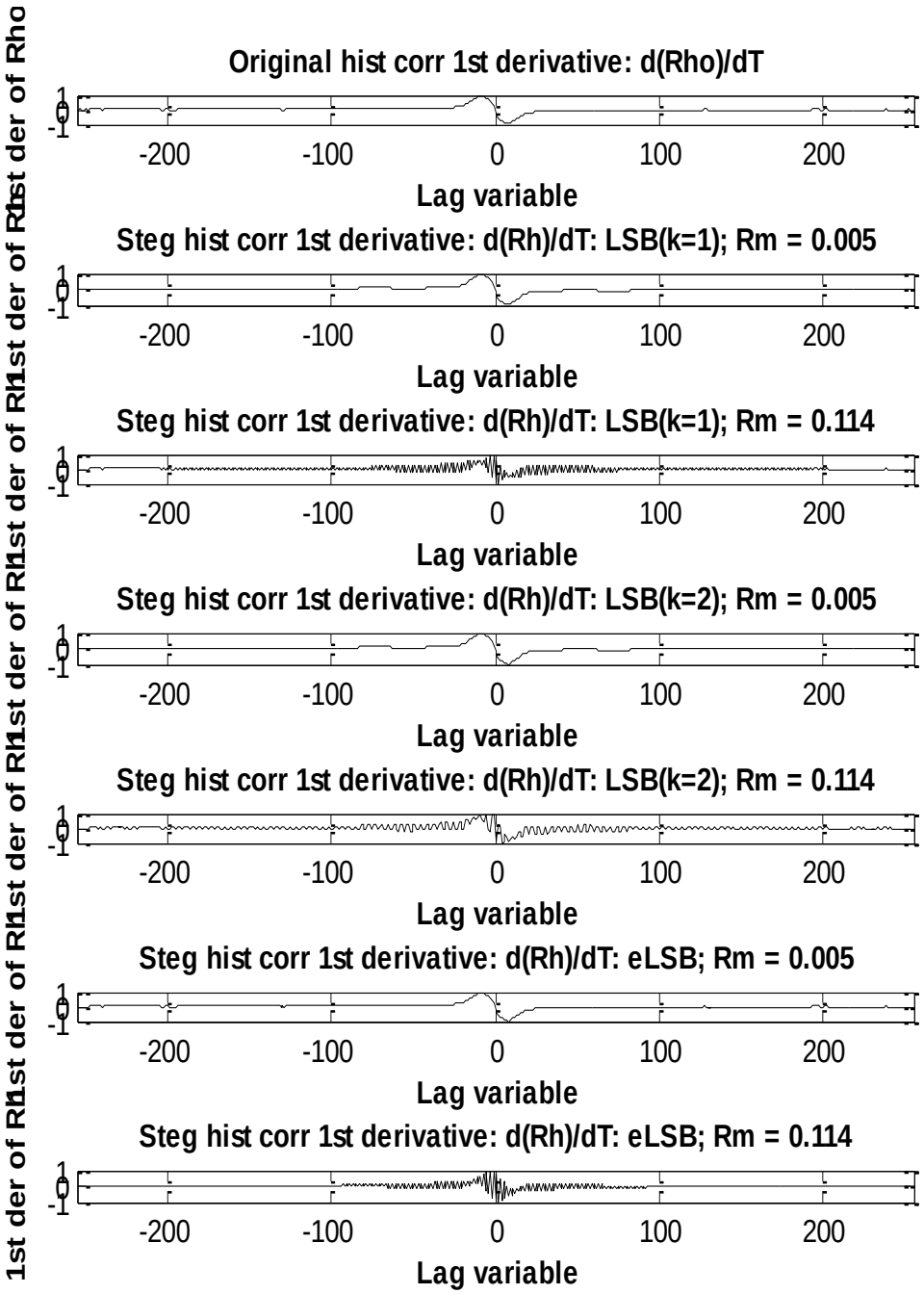
1st der of Rho
Original hist corr 1st derivative: d(Rho)/dT
Steg hist corr 1st derivative: d(Rh)/dT: LSB(k=1); Rm = 0.005
Steg hist corr 1st derivative: d(Rh)/dT: LSB(k=1); Rm = 0.114
Steg hist corr 1st derivative: d(Rh)/dT: LSB(k=2); Rm = 0.005
Steg hist corr 1st derivative: d(Rh)/dT: LSB(k=2); Rm = 0.114
Steg hist corr 1st derivative: d(Rh)/dT: eLSB; Rm = 0.005
Steg hist corr 1st derivative: d(Rh)/dT: eLSB; Rm = 0.114
1st der of Rh
Lag variable
-200
-100
0
100
200
1
0
-1

Figure (4.18) (a)

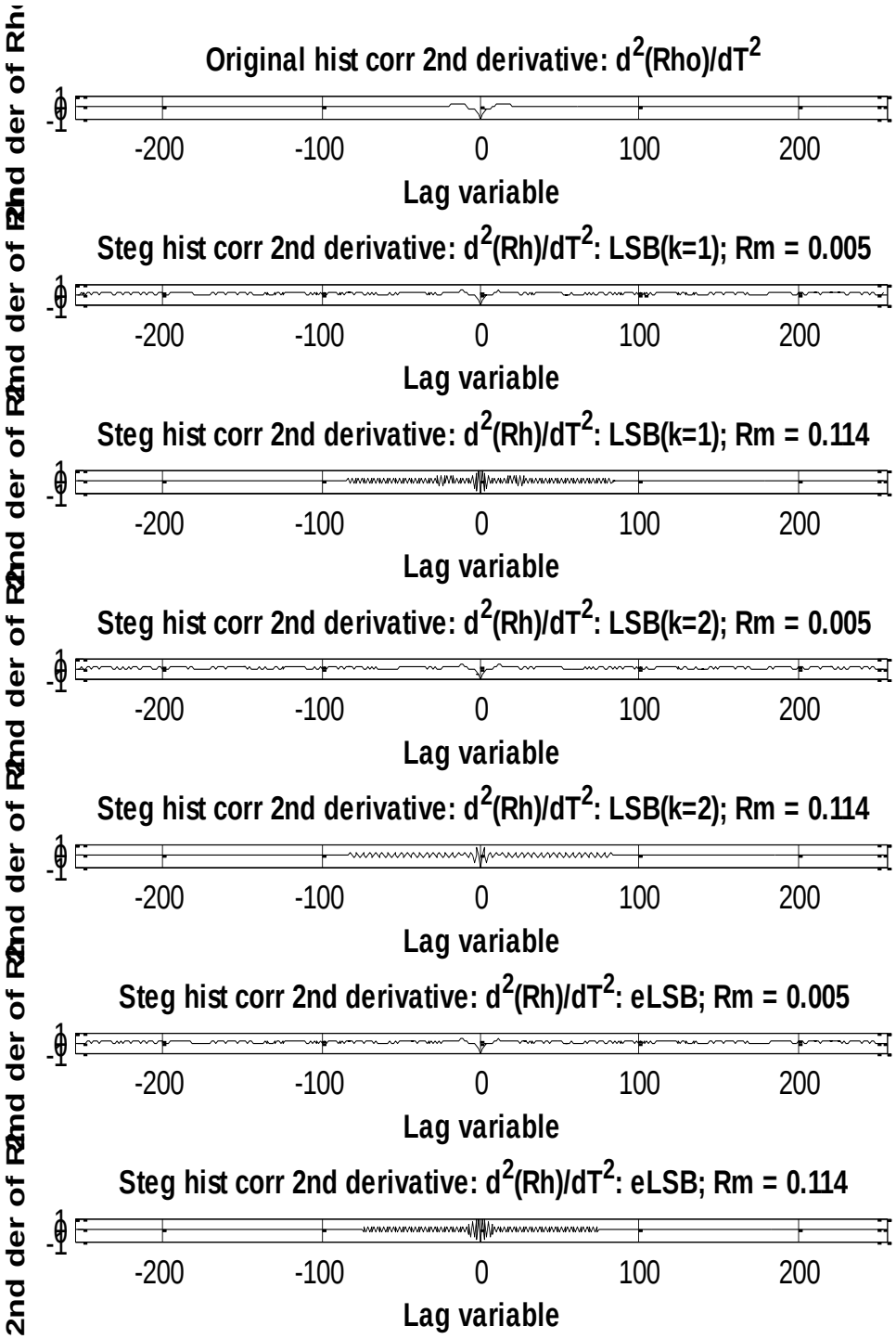
Original hist corr 2nd derivative: d2(Rho)/dT2
Steg hist corr 2nd derivative: d2(Rh)/dT2: LSB(k=1); Rm = 0.005
Steg hist corr 2nd derivative: d2(Rh)/dT2: LSB(k=1); Rm = 0.114
Steg hist corr 2nd derivative: d2(Rh)/dT2: LSB(k=2); Rm = 0.005
Steg hist corr 2nd derivative: d2(Rh)/dT2: LSB(k=2); Rm = 0.114
Steg hist corr 2nd derivative: d2(Rh)/dT2: eLSB; Rm = 0.005
Steg hist corr 2nd derivative: d2(Rh)/dT2: eLSB; Rm = 0.114
2nd der of Rho
-200
-100
0
100
200
Lag variable

Figure (4.18) (b)

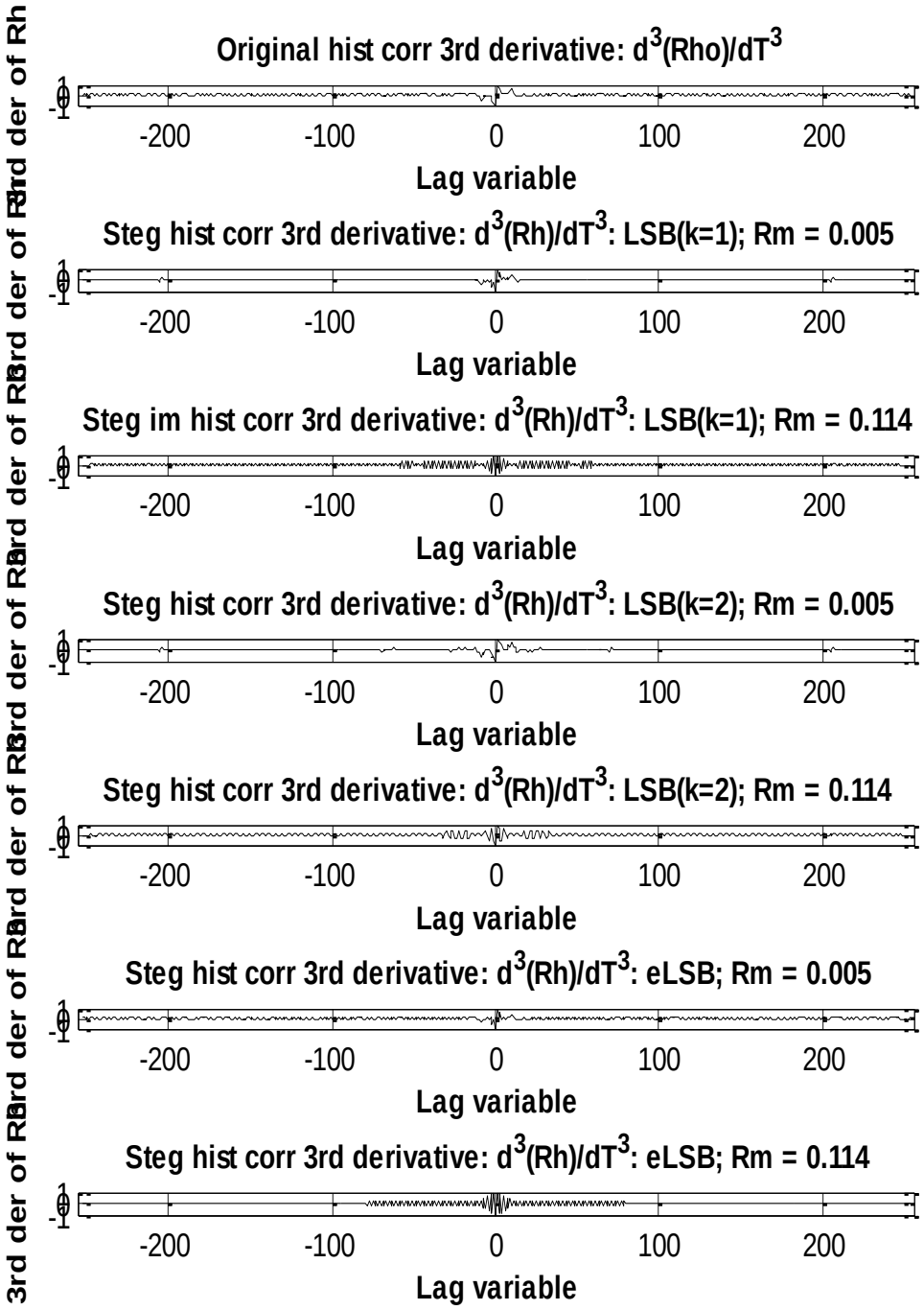

Figure (4.18) (c)

Figure (4.18): The derivative of the histogram-correlative for enhanced_LSB, K_LSB stego, and cover images for several LSB levels and several messages to cover size ratios. (a) the First derivative. (b) the second derivative. (c) Third derivative.

The proposed system in the following experiment tested a different image steganography approach, this time using a color image format (BMP)[53].

Figures (4.20) and (4.22) show the influence of the first derivative of the histogram-correlative for cover and stego color images. According to Figures (4.20), (4.22) and Tables (4.3), (4.4), the first derivative is sufficient to detect image tampering, as evidenced by the influence of the PSNR value, image type for the cover and message images, and whether they are 256 or true color. If the PSNR is high, we should probably use the second and third derivatives instead of the first because the first is vague.

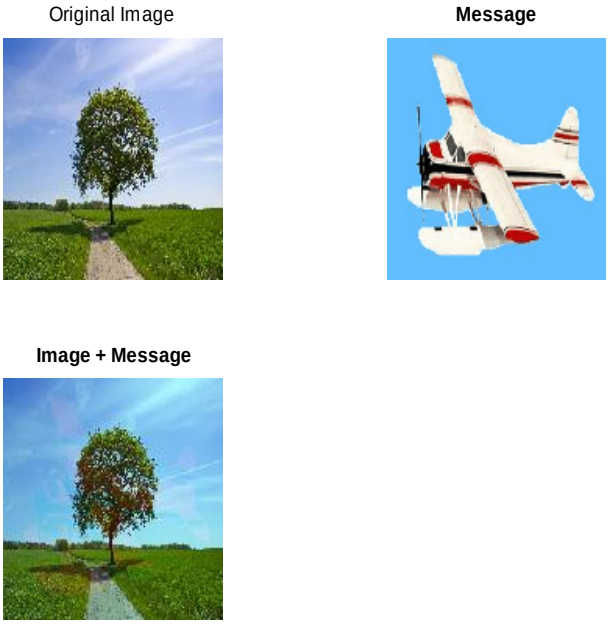

Figure (4.19): The images of the cover (true color), message (256 color), and stego (true color).

Table (4.3): Details of hiding an image in a cover image as in Figure (4.19).

Mixing Matrix	PSNR	MSE	Cover & message images
$\begin{bmatrix} 0.002 & 0.9 \\ 0.7 & 0.2 \end{bmatrix}$	52.603	0.598	256×256 pixels

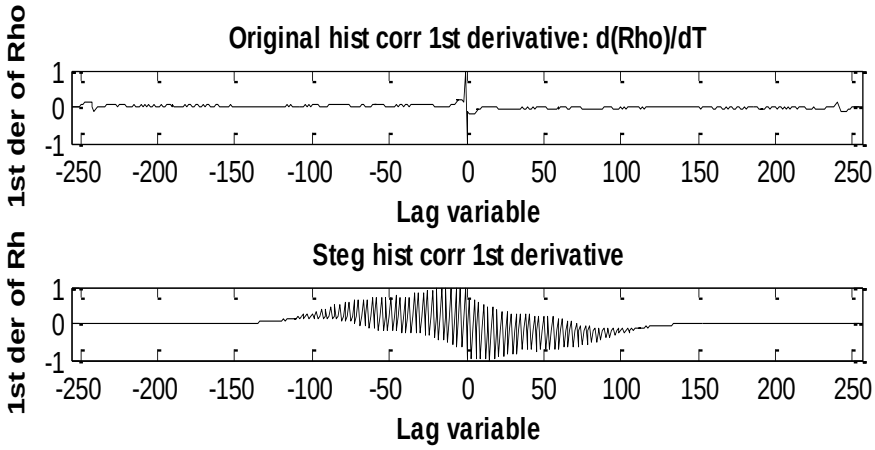

Figure (4.20): The 1st derivative of the histogram-correlative for cover and stego color images.

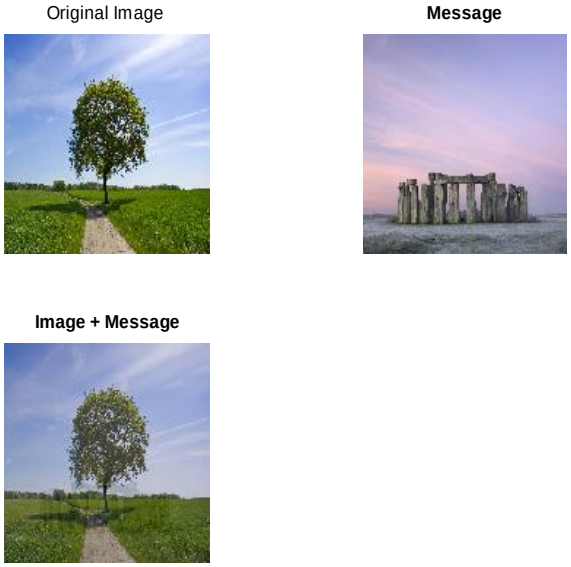

Figure (4.21): The cover, the message and the stego images are all in true color.

Table (4.4): Details of hiding an image in a cover image as in Figure (4.21).

Mixing Matrix	PSNR	MSE	Cover &message images
$\begin{bmatrix} 0.001 & 0.9 \\ 0.6 & 0.2 \end{bmatrix}$	58.038	0.320	256×256 pixels

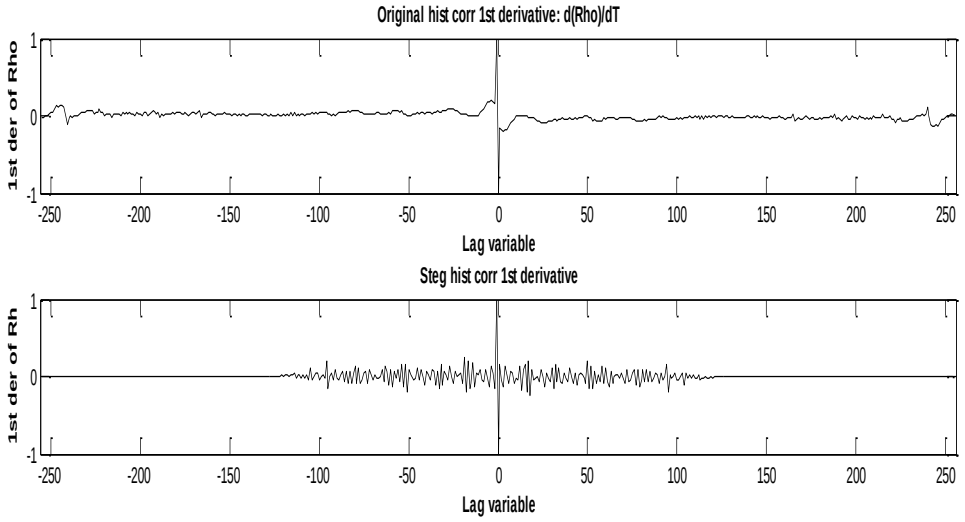

Figure (4.22): The 1st derivative of the histogram-correlative for cover and stego color images.

4.7.3 Entropy correlation derivatives

The wavelet entropy of autocorrelation derivative of the image histogram utilized as the decision suitable for the color images was created using MATLAB command:

threshold = wentropy(d1,'shannon');

The wavelet entropy of the autocorrelation derivative of the image histogram will be low since the image is original. It will be high if the tested image is stego. However, the criterion is more accurate for color images than other decisions classifier till now to discover stego image. The algorithm (3.7) is applied on the tested image's derivatives histogram autocorrelation. In the

following experiments, the proposed method was tested using several image steganography techniques to embedded the message in grayscale and color BMP (true, 256) images. The suggested approach should be capable of detecting manipulated images. As a result, using the wavelet entropy measure on this derivative would be an excellent way to identify these vibrations, where a criterion may be utilized.

In the next experiment, from Figure (4.23) to Figure (4.27) the suggested system evaluated for detecting the embedding image by the modified pixel-value differencing steganography method (MPVD) [21]. The cover images, which were grayscale images (PNG), were taken from the BOSS database. The test was performed randomly on 1,000 images selected from 10,000 images that were 512 × 512 pixels in size. It is found that if the image is clear, then the value of the wavelet entropy of the first, second, and third derivatives of histogram correlative is small. When the value of the wavelet entropy of derivatives of histogram correlative is high, then the tested image is stego.

Figure (4.23): The grey-scale cover and stego images.

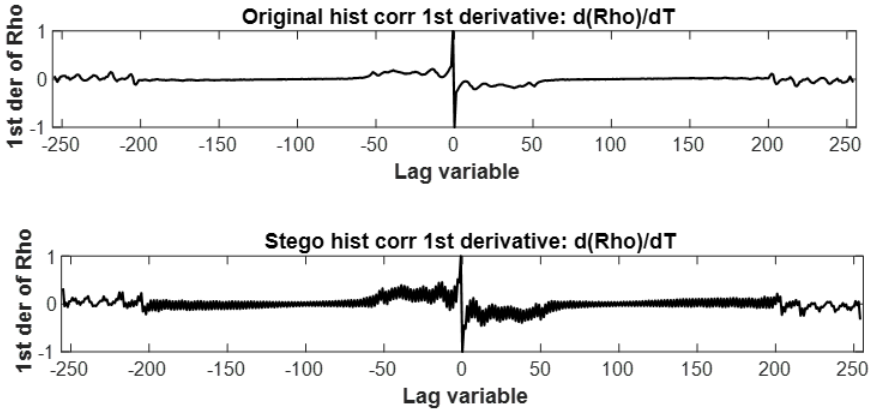

Figure (4.24): The 1st derivative of the histogram-correlative for stego and cover images.

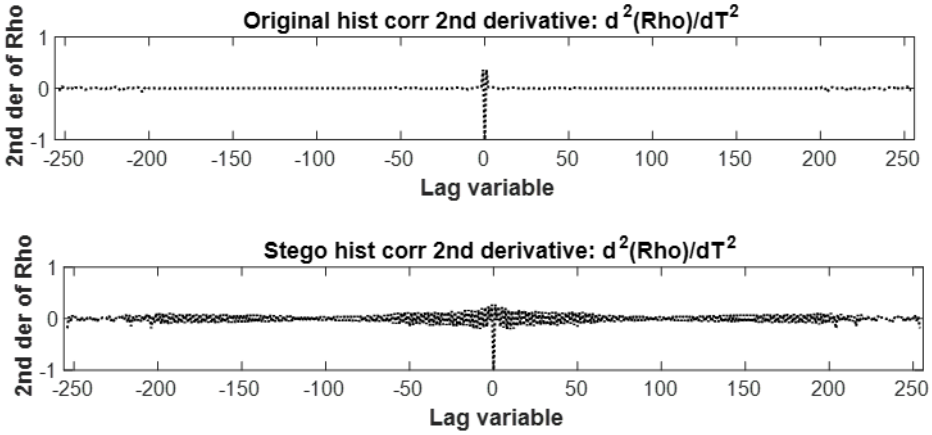

Figure (4.25): The 2nd derivative of the histogram-correlative for stego and cover images.

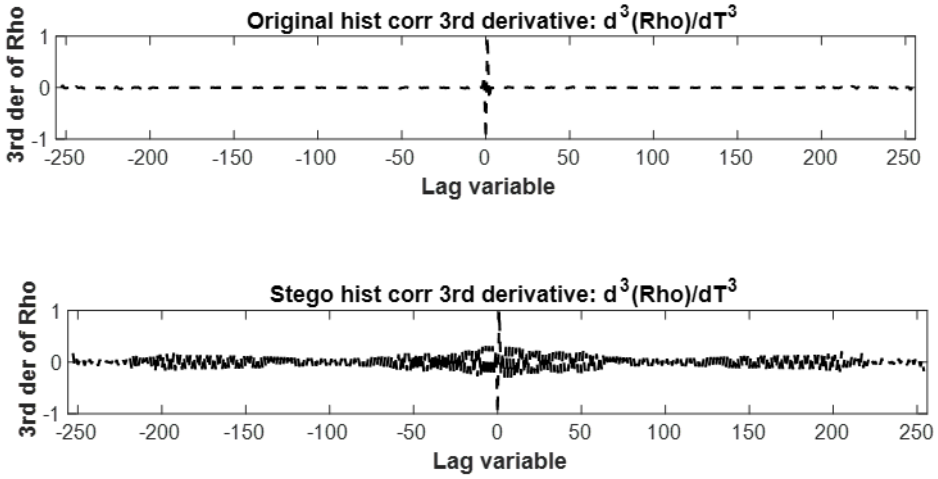

Figure (4.26): The 3rd derivative of the histogram-correlative for stego and cover images.

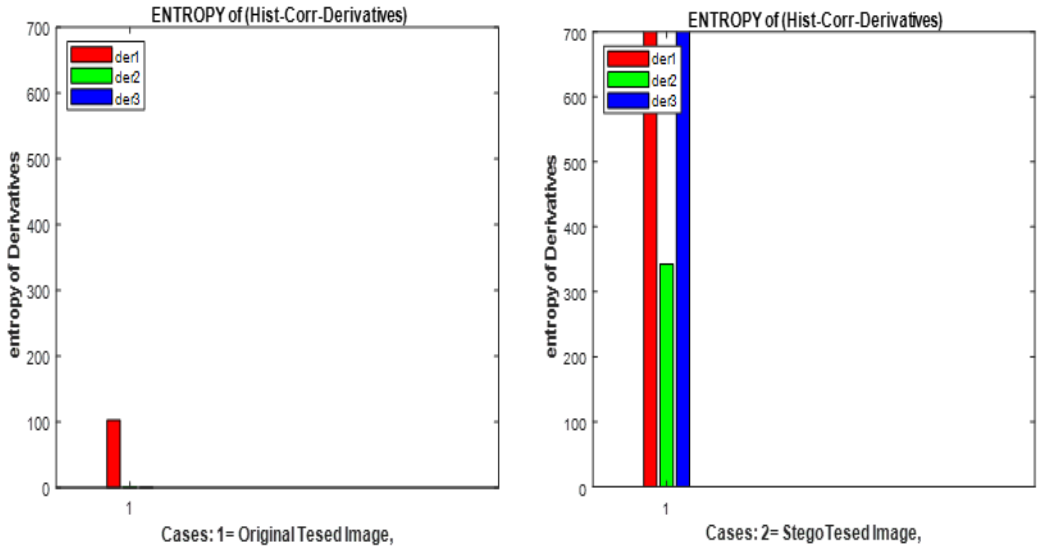

Figure (4.27): The value of derivative entropy of the histogram-correlative for stego and cover images.

In the following experiment, from Figure (4.28) to Figure (4.33) the suggested system evaluated for detecting the embedding image by the least significant bits method (LSB). The cover images, which were true color images (RGB), For the 3-channel embedding, the pixels of each image were embedded with secret data sequentially, where each channel in each pixel were embedded with 2 bits or 4 bits by replacing the least significant bits [82]. were taken from the Mendeley Data. The test was performed randomly on 100 images selected from 1,500 images that were 512 × 512 pixels in size. It is found that if the image is clear, then the value of the wavelet entropy of the first three derivatives of histogram correlative is tiny. When the value of the wavelet entropy of derivatives of histogram correlative is high, then the tested image is stego.

Figure (4.28): The true-color (RGB) cover and stego images.

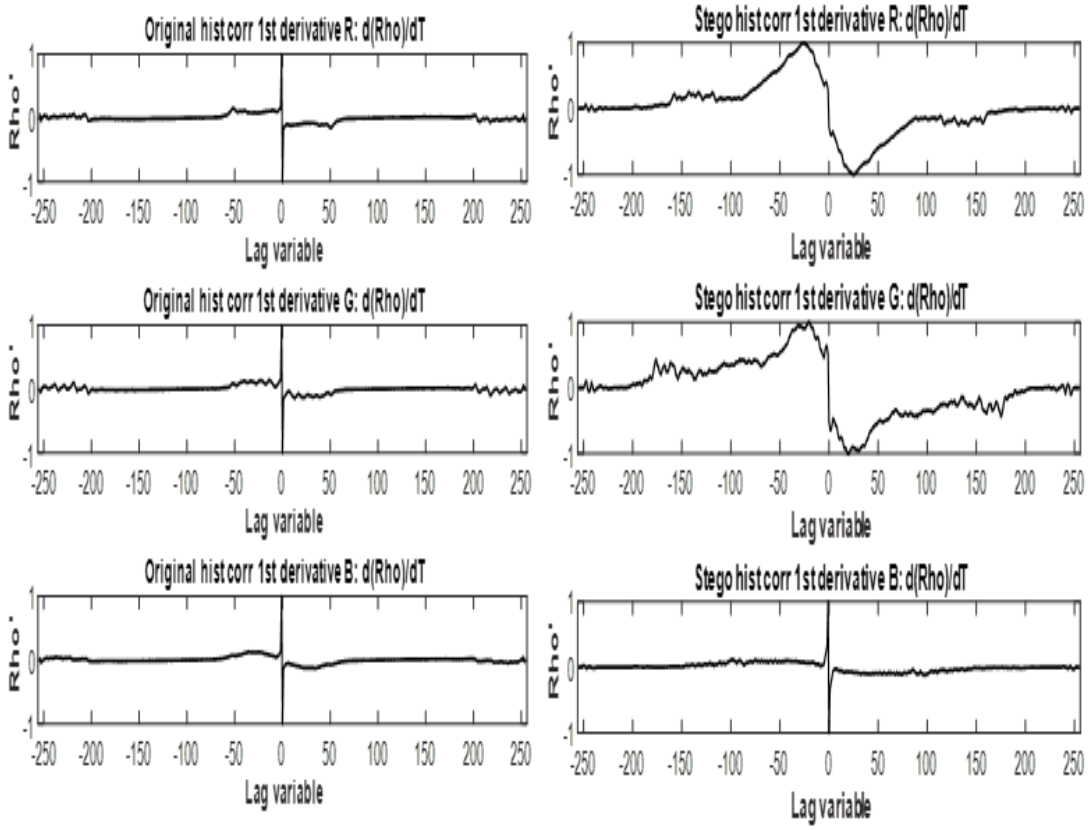

Figure (4.29): The 1st derivative of the histogram-correlative (RGB bands) for cover and stego images.

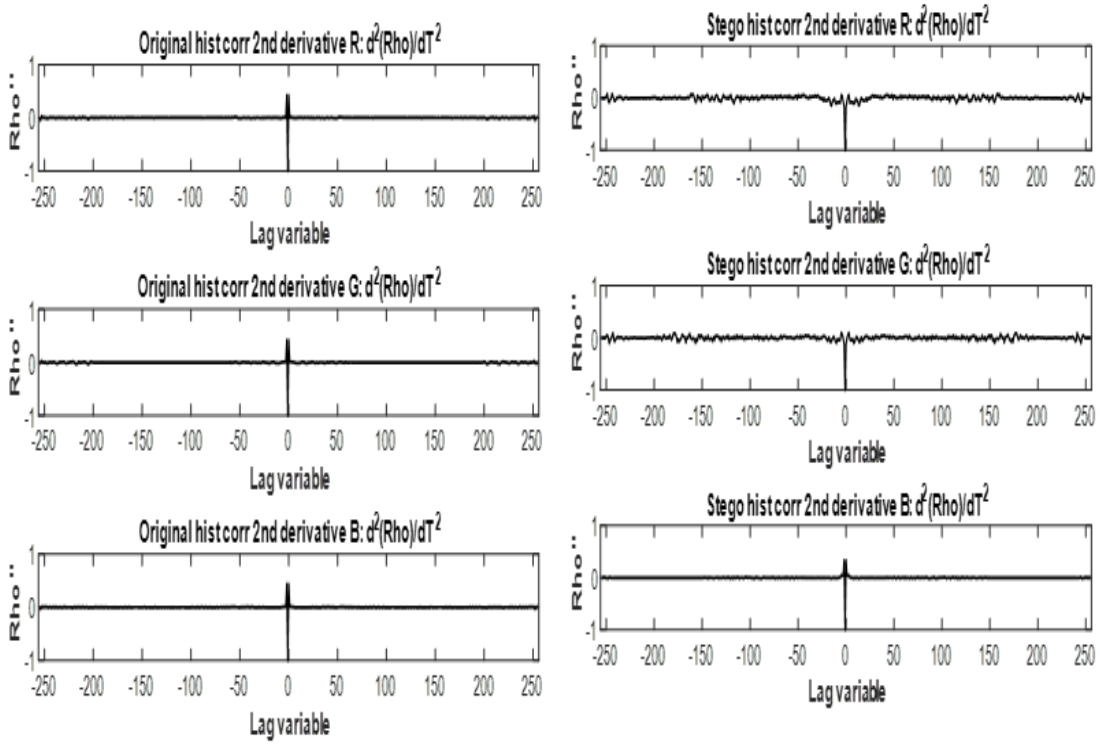

Figure (4.30): The 2nd derivative of the histogram-correlative (RGB bands) for cover and stego images.

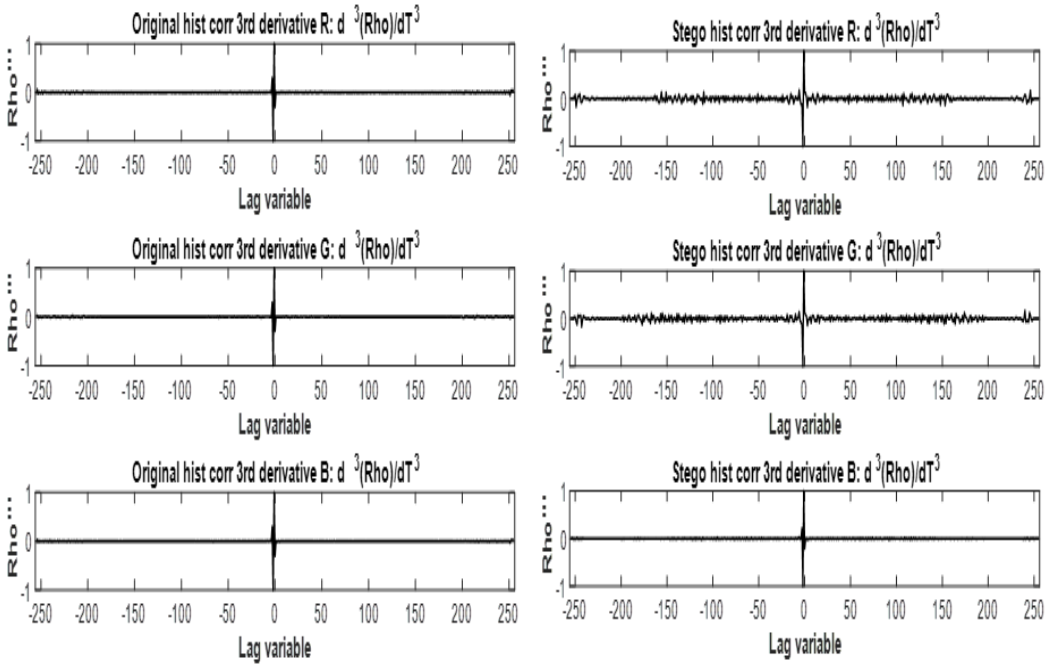

Figure (4.31): The 3rd derivative of the histogram-correlative (RGB bands) for cover and stego images.

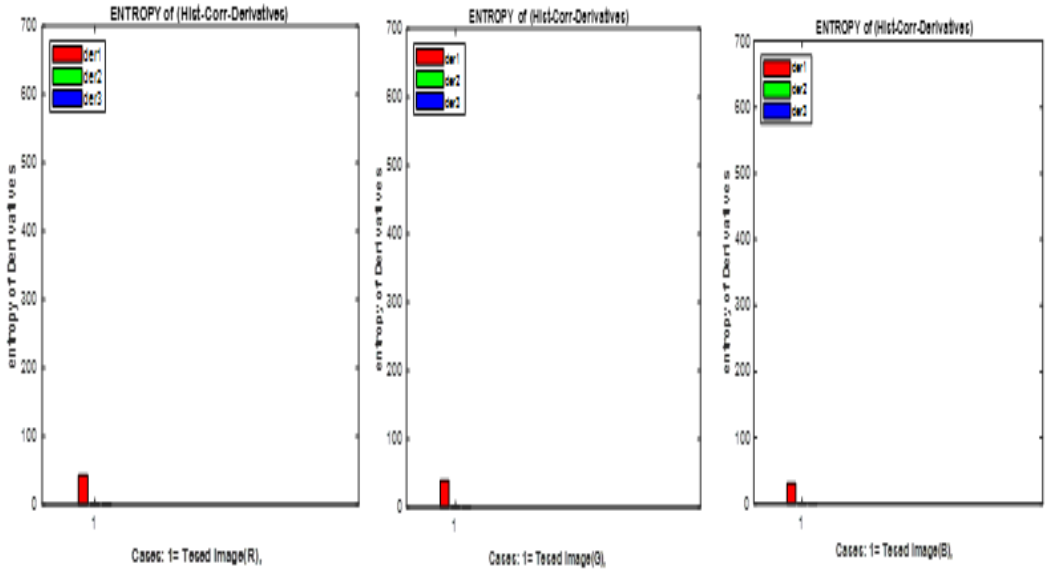

Figure (4.32): The value of derivative entropy of the histogram-correlative (RGB bands) for cover image.

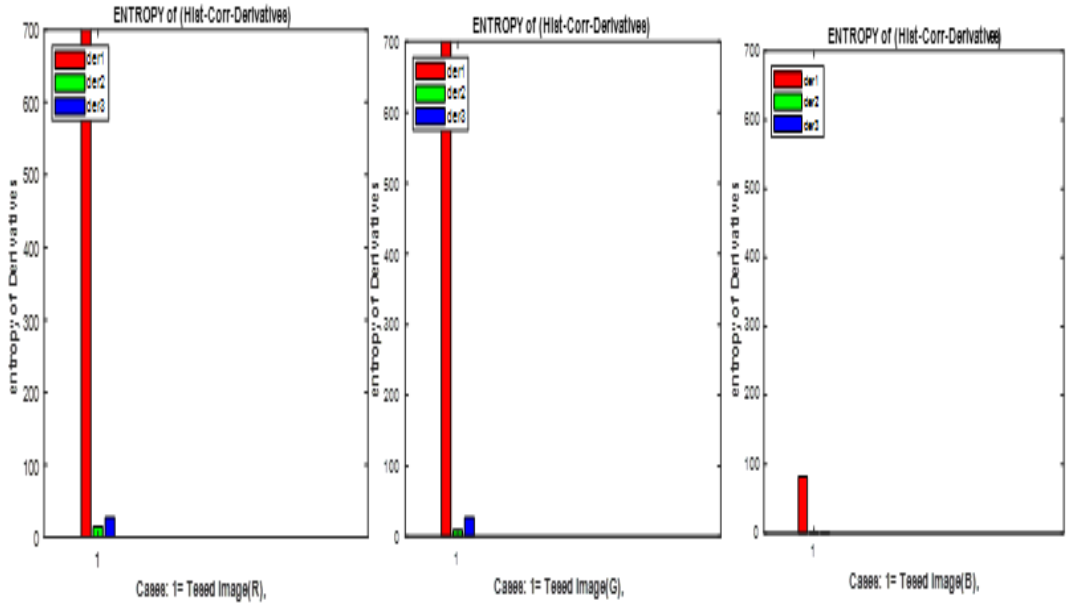

Figure (4.33): The value of derivative entropy of the histogram-correlative (RGB bands) for stego image.

In the following experiment, from Figure (4.34) to Figure (4.38) we evaluated the proposed system examined 1000 images using a different image steganography technique with a color image format (256). It is found that if the

image is clear, then the value of the wavelet entropy of the first three derivatives of histogram correlative is small. When the value of the wavelet entropy of derivatives of histogram correlative is high, then the input image is stego.

Figure (4.34): The color (256) cover and stego images.

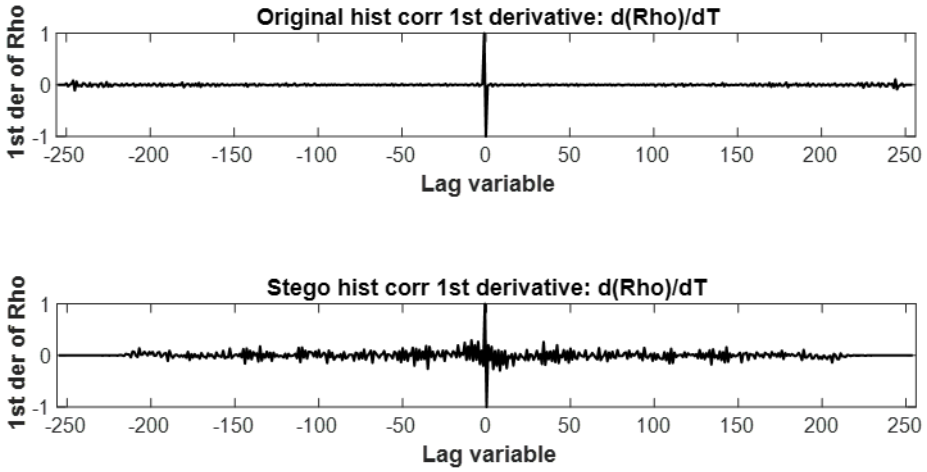

Figure (4.35): The 1st derivative of the histogram-correlative for cover and stego images.

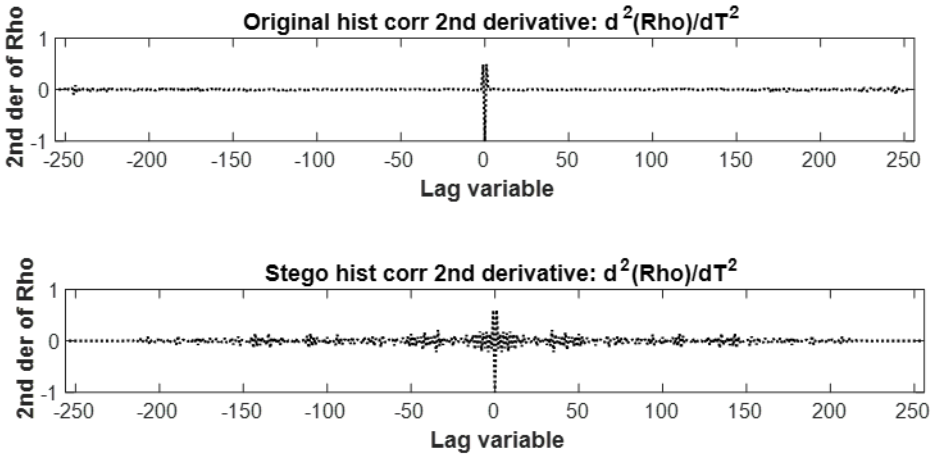

Figure (4.36): The 2nd derivative of the histogram-correlative for cover and stego images.

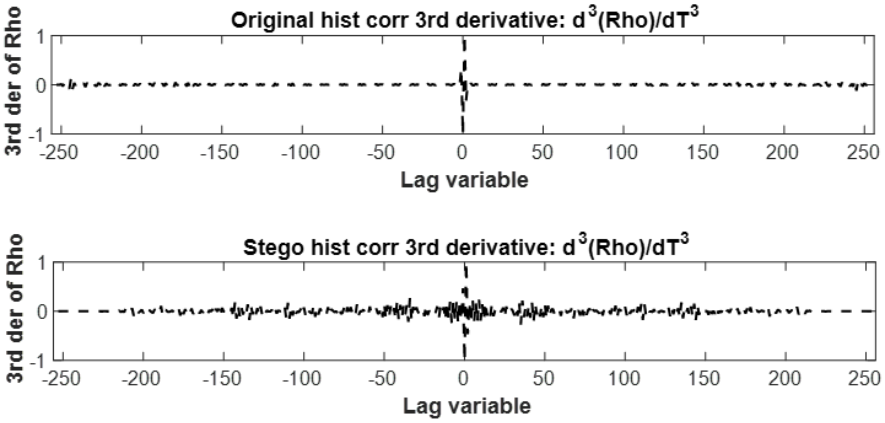

Figure (4.37): The 3rd derivative of the histogram-correlative for cover and stego images.

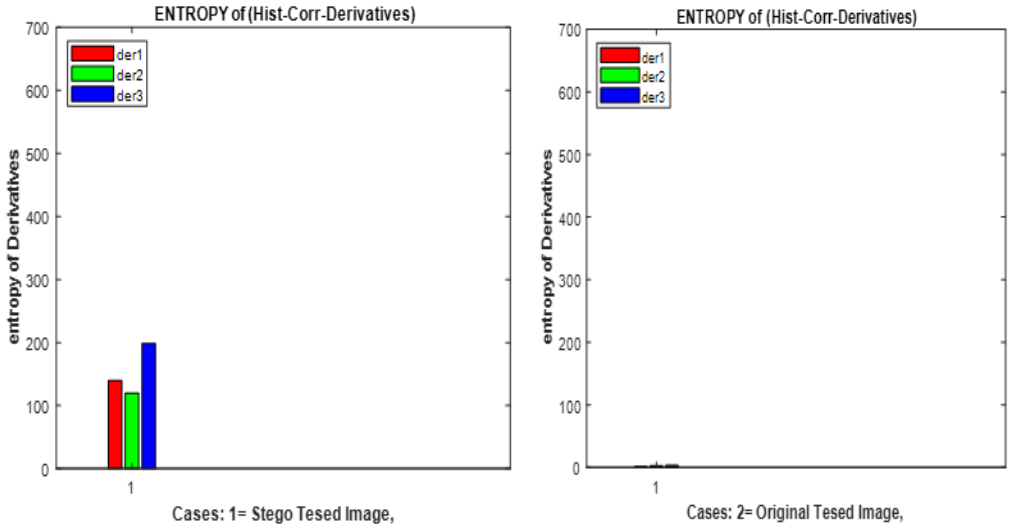

Figure (4.38): The value of derivative entropy of the histogram-correlative for stego and cover images.

4.8 Comparison with the Related Works

It is clearly shown that the best detection is the proposed system using the specific method on LSB steganography. Now compare the performance of our proposed method with those of three specific methods [4,5,6] in terms of recognizing embedded messages. In [4] uses The Chi-squared approach, the Chi-squared method consistently recognizes sequentially embedded messages, but it fails miserably when the embedding is random. In [5] uses the Raw Quick Pair (RQP) approach that a significant degree of detection reliability might be achieved even with secret message capacities of 0.1–0.3 bits per pixel. In [6] uses the RS method for detecting LSB embedding in color and grayscale images. The RS Technique is more effective for arbitrarily spreading data throughout the stego image. The results of the proposed technique show detect sequentially and randomly embedded message capacities of (R_m > 0.01). Table (4.5) show comparative between the proposed method and three specific techniques.

Table (4.5) comparison the specific methods with the proposed method.

Ref. No.	**Method**	**Image**	**Detect message**	**recognizing embedded messages**
[4]	Specific method A .Westfeld , A . Pfitzmann (2000)	JPEG	_____	Sequentially
[5]	Specific method J . Fridrich, M . Long (2000)	RGB true color	0.1–0.3 bits per pixel	Randomly
[6]	Specific method J .Fridrich, M. Goljan (2001)	Color and Grayscale	0.05 bits per pixel	Randomly
	Specific method The proposed method	Color and Grayscale	R_m > 0.01	Sequentially and Randomly

The proposed method is universal but doesn't use a dataset in the processing phase to classify images into cover or stego images. It is universal steganalysis, the first attempt to detect images without adopting the dataset.

The proposed evaluated technique for embedding ratios as low as 20%. The tests were performed at random on 100 or 1,000 images chosen at random from a database of 10,000 images that were 512 × 512 pixels in size. The images utilized in the testing were from the BOSS image database.

Table (4.6) summarizes the results of comparing the suggested method with three techniques for detecting the embedding image by the modified pixel-value differencing steganography method (MPVD). The table shows that the proposed technique performed better with the BOSS database than other methods.

Table (4.6) Comparison of accuracies for BOSS database images.

Embedding ratio (%)	**Proposed method**	**W. Lin et al.'s method [21]**	**Sabeti et al.'s method [84]**	**Bui et al.'s method [85]**
20	**0.975**	**0.959**	**0.902**	**0.898**

Chapter Five

Conclusions and Future Works

Chapter Five
Conclusions and Future Works

5.1 Conclusions

The following conclusions explain the most significant characteristics that have been obtained from the results of the proposed system:

1. The Pearson correlation idea is a suitable statistical indicator for testing and discovering the correlated value of the tested image histogram. It is clear that introducing a message in the cover image will reduce the natural correlation between its histogram and that this truth becomes clearer in histogram correlation than in pixel correlation.

2. Derivatives on the image histogram autocorrelation is an efficient method for revealing image tampering. Because the stego image will introduce significant ripples when the autocorrelation function of the image histogram is subjected to derivation. On the other hand, it is found that using higher derivatives may not be necessary, as they do not introduce a significant difference from the first derivative in revealing image tampering when the ripples are clear in the first derivative.

3. Experimental results have shown that the proposed system provides significant tools to discover these ripples, which represent the decision on the tested image whether it is stego or clear.

4. Experimental results have demonstrated that the suggested system fails if the message size is very small ($R_m < 0.01$). The use of higher derivatives is almost certainly required. Since the first derivative is vague if the message is limited in comparison to the cover (R_m), we can use the second and third derivatives.

5. Experimental results have shown that the proposed system is validated with various LSB steganography approaches and it is applied to other steganography approaches. The proposed method will determine whether an image was tampered with or not without relying on the original image.
6. The proposed system can detect different steganography methods. It is a universal steganalysis in the spatial domain.

5.2 Future Works

In this dissertation, the correlation statistics, derivatives of the image histogram autocorrelation, and decision threshold of the proposed system are implemented. The following are suggestions for future works:

1. Designing a model for stego image detection based on the Internet of Things (IoT) infrastructure.

2. Developing the proposed steganalysis system to be a hybrid system consisting of both Correlation image pixels and Correlation image histogram to benefit from the advantages of each type.

3. Developing the proposed system for detection of most limited secrete messages using higher derivatives.

4. Designing an efficient model for detection based on other digital media like movie, sound and text.
5. Diversify the decision of the proposed system to include the SVM classifier to detect the stego image.

REFERENCES

REFERENCES

[1] M. Raggo , C. Hosmer," Data hiding," *syngress*, pp. 350, Nov. 2012.

[2] W. You, H. Zhang, X. Zhao," A Siamese CNN for Image Steganalysis," IEEE Transactions on Information Forensics and Security, Vol. 16,2021.

[3] S. M. Badr, G. I. Salama, G. M. I. Selim, A. H. Khalil, "A Review on Steganalysis Techniques: From Image Format Point of View," International Journal of Computer Applications (0975 – 8887), Vo. 102– No.4, September 2014.

[4] A. Westfeld, A. Pfitzmann,"Attacks on steganographic systems," Berlin Heidel- berg: Springer; pp. 61–76. 2000.

[5] J. Fridrich, M. Long," Steganalysis of LSB encoding in color images," In: IEEE international conference on multimedia and Expo. ICME2000. Proceedings of the latest advances in the fast changing world of multimedia (Cat. No.00TH8532), Vol. 3. pp. 1279–82,2000.

[6] J. Fridrich, M. Goljan, R. Du," Reliable detection of LSB steganography in color and grayscale images," In: Proceedings of the 2001 workshop on multimedia and security new challenges - (MM&Sec '01), pp. 27, 2001.

[7] T. Zhang, X. Ping, "Reliable detection of LSB steganography based on the difference image histogram," In: Proceedings of the IEEE international conference on acoustics, speech, and signal processing, (ICASSP '03), Vol. 3III-545-8, 2003.

[8] M. Celik, G. Sharma, A. Tekalp, "Universal image steganalysis using rate-distortion curves," In: Proc. SPIE - International Society for Optics and Photonics, pp. 467–76, Jun. 2004.

[9] A. Ker, " A General Framework for Structural Steganalysis of LSB Replacement," International Workshop on Information Hiding, pp.296–311, 2005.

[10] H. Malekmohamadi, S. Ghaemmaghami," Steganalysis of LSB based image steganography using spatial and frequency domain features,". In: IEEE international conference on multimedia and expo, (ICME 2009); August.2009.

[11] T. Zhang, W. Li, Y. Zhang, E. Zheng, X. Ping," Steganalysis of LSB matching based on statistical modeling of pixel difference distributions," Information Sciences (NY),180(23):4685–94,2010.

[12] S. Cho, B-H. Cha, j. Wang, C-C. Jay Kuo," Block-based image steganalysis: algorithm and performance evaluation," Journal of Visual Communication and Image Representation, 24(7):846–56, 2013.

[13] Z. Xia, X. Wang, X. Sun, B. Wang," Steganalysis of least significant bit matching using multi-order differences," Security and Communication Networks 7(8):1283–91, Aug.2014.

[14] M. Goljan, J. Fridrich, R. Cogranne," Rich model for Steganalysis of color images,". In: 2014 IEEE international workshop on information forensics and security (WIFS 2014); p. 185–90, Apr.2015.

[15] W. Tang, H. Li, W. Luo, J. Huang," Adaptive steganalysis based on embedding probabilities of pixels,". IEEE Transactions on Information Forensics and Security 11(4): pp.734–745, 2016.

[16] R. Nouri, A. Mansouri," Digital image steganalysis based on the reciprocal singular value curve," Multimedia Tools and Applications 76(6): pp.8745–56, 2017.

[17] B. Li, W. Wei, A. Ferreira, and S. J. I. S. P. L. Tan, "ReST-Net: Diverse activation modules and parallel subnets-based CNN for spatial image steganalysis," vol. 25, no. 5, pp. 650-654, 2018.

[18] F. Liu, X. Yan, and Y. J. I. A. Lu, "Feature selection for image steganalysis using binary bat Algorithm," vol. 8, pp. 4244-4249, 2019.

[19] T. Pevny, P. Bas, J. J. I. T. o. i. F. Fridrich, and Security, "Steganalysis by subtractive pixel adjacency matrix," vol. 5, no. 2, pp. 215-224, 2010.

[20] Z. Xiang, J. Sang, Q. Zhang, B. Cai, X. Xia, and W. J. I. A. Wu, "A new convolutional neural network-based steganalysis method for content-adaptive image steganography in the spatial domain," vol. 8, pp. 47013-47020, 2020.

[21] W. Lin, T. Lai, C. Chou, " Chi-square-based steganalysis method against modified pixel-value differencing steganography," In: Arabian Journal for Science and Engineering: Springer; published online, April. 2021.

[22] B. Shavers, J. Bair," Hiding Behind the Keyboard: Uncovering Covert Communication Methods with Forensic Analysis," Syngress, pp.236, 2016.

[23] E. Sodipo, "The Art of Security and Information Hiding," Lulu, pp.192, Mar. 2008.

[24] F.A.P. Petitcolas, R.J. Anderson, M.G. Kuhn, "Information Hiding – A Survey," IEEE Transactions of Proceedings of Theory 87 (7), pp. 1062–1078, July 1999.

[25] P. Moulin, J.A. O'Sullivan," Information-theoretic analysis of information hiding," IEEE Transactions on Information Theory 49 (3), pp. 563–593, 2003.

[26] C. Cachin, "An Information-Theoretic Model for Steganography," 2nd Workshop on Information Hiding, LNCS 1525, Springer-Verlag, pp. 306–321, 1998.

[27] I. S. Moskowitz, G. E. London, L. Chang, “A new Paradigm Hidden in Steganography," Proceedings of the 2000 workshop on New security paradigms, pp.41-50, Feb.2001.

[28] E. Zielinska, W. Mazurczyk, K. Szczypiorski, "Trends in steganography," Communications of the ACM57 (3), pp. 86–95, 2014.

[29] M. Kasapbasi, "A new chaotic image steganography technique based on Huffman compression of Turkish texts and fractal encryption with post-quantum security," IEEE Access 7, pp. 148495–148510, 2019.

[30] K. Wu, C. Wang, "Steganography using reversible texture synthesis," IEEE Transactions on Image Processing 24 (1), pp. 130–139, 2014.

[31] E. Lin, E. Delp, " A Review of Data Hiding in Digital Images," PICS Conference, pp. 274-278, 1999.

[32] G. Kipper, “Investigator's Guide to Steganography,” CRC Press, 2003.

[33] F. Y. Shih, "Digital Watermarking and Steganography: Fundamentals and Techniques," CRC Press, 2017.

[34] S. Katzenbeisser, F. Petitcolas, "Information Hiding Techniques for Steganography and Digital Watermarking," Artech House, 2000.

[35] D. Bucerzan, C. Ratiu, "Testing methods for the efficiency of modern steganography solutions for mobile platforms," in: 6th International Conference on Computers Communications and Control, IEEE, pp. 30–36, 2016.

[36] D. Frith, "Steganography approaches options and implications," Network Security 8 (20), PP. 4–7, 2007.

[37] P. V. K. Borges, J. Mayer, E. Izquierdo, "Robust and transparent color modulation for text data hiding," IEEE Transactions on Multimedia 10 (8), PP.1479–1489, 2008.

[38] I. J. Kadhim, P. Premaratne, P. J. Vial, "High capacity adaptive image steganography with cover region selection using dual-tree complex wavelet transform," Cognitive Systems Research 60, PP.20–32, 2020.

[39] V. Korzhik, N. D. Cuong, G. M.Luna, "Cipher modification against steganalysis based on NIST tests," in: 24th Conference of Open Innovations Association (FRUCT), IEEE, pp. 179–186, 2019.

[40] M. Hussain, A.W. A. Wahab, Y. I. B. Idris, A. T.S.Ho, K.-H. Jung, "Image steganography in spatial domain: a survey," Signal Processing: Image Communication 65, PP.46–66, 2018.

[41] I. J. Kadhim, P. Premaratne, P. J. Vial, B. Halloran, "Comprehensive survey of image steganography: techniques, evaluations, and trends in future research," Neurocomputing 335, vol.335, pp. 299–326, Mar.2019.

[42] J. Wiley and Sons, "Applied Cryptography protocols, algorithms and source code in C," 20th Anniversary Edition, Wiley, Mar.2015.

[43] M. Hassaballah, M.A. Hameed, M.H. Alkinani, " Introduction to digital image steganography," Digital Media Steganography, pp.1-15, 2000.

[44] L. M. Marvel, C. G. Boncelet, C.T. Retter, "Spread spectrum image steganography," IEEE Transactions on Image Processing 8 (8), pp.1075–1083, 1999.

[45] S. Sajasi, A.-M. E.Moghadam," An adaptive image steganographic scheme based on noise visibility function and an optimal chaotic based encryption method," Applied Soft Computing30, pp.375–389, May 2015.

[46] G. Swain, "Adaptive pixel value differencing steganography using both vertical and horizontal edges," Multimedia Tools and Applications 75 (21), pp.13541–13556, 2016.

[47] W. Hong, T.-S. Chen, "A novel data embedding method using adaptive pixel pair matching," IEEE Transactions on Information Forensics and Security 7 (1), pp.176–184, 2012.

[48] D.-C. Wu, W.-H. Tsai, "A steganographic method for images by pixel-value differencing," Pattern Recognition Letters 24 (9), pp.1613–1626, 2003.

[49] C.-K. Chan, L.-M. Cheng, "Hiding data in images by simple LSB substitution," Pattern Recognition37 (3), pp.469–474, 2004.

[50] L. Zhang, Ha. Wang, R. Wu, "A high-capacity steganography scheme for JPEG2000 baseline system," IEEE Transactions on Image Processing 18 (8), pp.1797–1803, 2009.

[51] R. Yadav, R. Saini, G. Chawla, “A novel approach for image steganography in spatial domain using last two bits of pixel value," Int. J. Security, 5:51-61, 2011.

[52] O.N. Kadhim, Z. M. Hussain, " Information Hiding using Chaotic-Address Steganography," Journal of Computer Science, 14 (9): pp.1247.1266, 2018.

[53] S. Al-Janabi, I. Al-Shourbaji,"A Hybrid Image Steganography Method based on Genetic Algorithm, "7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), 2016.

[54] X. Zhang, S. Wang, "Efficient steganographic embedding by exploiting modification direction," IEEE Communications Letters 10 (11), pp.781–783, 2006.

[55] K. S. Shet, A.R. Aswath, M.C. Hanumantharaju, X.-Z. Gao, "Design and development of new reconfigurable architectures for LSB/multi-bit image steganography system," Multimedia Tools and Applications 76 (11), pp.13197–13219, 2017.

[56] T. D. Nguyen, S. A.-Int, N. A.-Int, "An adaptive multi bit-plane image steganography using block data-hiding," Multimedia Tools and Applications 75 (14), pp.8319–8345, 2016.

[57] A. Cheddad, J. Condell, K. Curran, P. M. Kevitt, "Digital image steganography: survey and analysis of current methods," Signal Processing 90 (3), pp.727–752, 2010.

[58] J. Fridrich, R. Du, "Secure steganographic methods for palette images," in: International Workshop on Information Hiding," Springer, pp. 47–60, 1999.

[59] N. Ghoshal, J. K. Mandal," A novel technique for Image Authentication in Frequency Domain using Discrete Fourier Transformation Technique (IAFDDFTT)," Malaysian Journal of Computer Science 21 (1), pp. 24-32, 2008.

[60] D. R. Wootten, " A graphic user interface for rapid integration of steganography software," Storming Media,p.142, Jan.1996.

[61] N. F. Johnson, Z. Duric, S. Jajodia, " Information Hiding: Steganography and Watermarking-Attacks and Countermeasures," Springer, Boston, MA, p.137, 2001.

[62] N.F. Johnson, S. Jajodia," Exploring steganography: Seeing the unseen," IEEE Computer 31 (2), pp. 26–34, Feb.1998.

[63] J. Fridrich, M. Goljan, R. Du, "Detecting LSB steganography in color and gray-scale images," IEEE Multimedia Magaz., Special Issue on Security, pp.22-28, October–November 2001.

[64] H. Wang, S. Wang," Cyber warfare: Steganography vs. steganalysis," Commun. ACM 47, pp.76-82, Oct. 2004.

[65] K. Karampidis, E. Kavallieratou, G. Papadourakis, "A review of image steganalysis techniques for digital forensics, "Journal of Information Security and Applications. Elsevier,40:217–235, 2018.

[66] T. D. Sairam, K. Boopathybagan, " Computational intelligence-based steganalysis comparison for RCM-DWT and PVA-MOD methods, " Special Issue: Computational Intelligence and Capsule Networks, vol. 60, no. 3, 285–293, 2019.

[67] P. Schober, C. Boer, L. A. Schwarte," Correlation Coefficients: Appropriate Use and Interpretation," Anesthesia and Analgesia 126(5), pp. 1763-1768, 2018.

[68] R. GUNESH," CORRELATION ANALYSIS," Lecturer in Mathematics and Statistics at Damelin College, Durban, p.7, 2016. http://www.rajgunesh.com/resources/notes.htm

[69] C. Chatfield," The Analysis of Time Series: An Introduction," Sixth Edition, Chapman & Hall, 1996.

[70] J. Rafiee, P.W. Tse," Use of autocorrelation of wavelet coefficients for fault diagnosis," Mechanical Systems and Signal Processing Vol.23, pp.1554–1572, 2009.

[71] United States Naval Academy," Auto Correlation," Ocean engineering Programs.

https://www.usna.edu/Users/oceano/pguth/md_help/html/time57ou.htm

[72] R. C. Gonzalez, R.E.Woods," Digital Image Processing," Second Edition, Prentice Hall, 2008.

[73] A. Alazzawi, H. Alsaadi, A. Shallal, S. Albwi, " EDGE DETECTION-APPLICATION OF (FIRST AND SECOND) ORDER DERIVATIVE

IN IMAGE PROCESSING," Diyala Journal of Engineering Sciences, pp. 430-440, 2015.

[74] Y . Wua, et al.," Local Shannon entropy measure with statistical tests for image randomness," In: Information Sciences, Vol. 222. pp. 323–342, 2013.

[75] T. Kvalseth, "On the Measurement of Randomness (Uncertainty):A More Informative Entropy," In: MDPI Proceedings Journals , pp.1-15, April. 2016.

[76] Z. M. Myo, Z. M. Aung, Z. M. Naing, " Design and Implementation of Active Band-Pass Filter for Low Frequency RFID (Radio Frequency Identification) System", Proceedings of the International MultiConfrence of Engineers and Computer Scientists, Hong Kong, Vol. 1, March 18-20,2009.

[77] A. S. Sedra, K. C. Smith, " Microelectronics Circuits", fifth edition, OXFORD UNIVERSITY PRESS, 2004.

[78] V. Srivastava, L. Maurya, E. R. Mehra, " Detection of Noise in High Pass Butterworth IIR Filter using MATLAB," International Journal of Engineering Research & Technology (IJERT), Vol. 3 Issue 2, pp.1232-1235, 2014.

[79] L. B. Jackson, "Roundoff-Noise Analysis for Fixed- Point Digital Filters Realized in Cascade or Parralel Form," IEEE Transactions on Audio and Electroacoustics, vol. AU- 18, no. 2, pp. 107-122, June 1970.

[80] L. S. Derong, " Application of Fourier Transform in Signal Processing," Signal and Information Processing, Vol.1, No. 1, pp.1-5, 2018.

[81] http:// agents. fel. cvut. cz/ stego data/. Accessed 15 July 2019.

[82] Z. I. Rasool, M. M. Al-Jarrah, S. Amin, " Steganalysis of RGB Images Using Merged Statistical Features of Color Channels," 2018 11th International Conference on Developments in eSystems Engineering (DeSE), 2-5 Sept, IEEE Xplore Press, Cambridge, UK, 2018.

[83] M. Al-Jarrah, “RGB-BMP Steganalysis Dataset”, Mendeley Data, v1, 2018

http://dx.doi.org/10.17632/sp4g8h7v8k.1.

[84] V. Sabeti, S. Samavi, M. Mahdavi, S. Shirani," Steganalysis and payload estimation of embedding in pixel differences using neural networks," Pattern Recognition, vol. 34, no. 1, pp. 405-415, 2010.

[85] C. Bui, H. Lee, J. Joo, H. K. Lee," Steganalysis method defeating the modified pixel-value differencing steganography," Int. J. Innov. Comput. Inf. Control, vol. 6, no. 7, pp. 3193–3203, 2010.

المستخلص

الاخفاء وتحليل الاحفاء وجهان لعملة واحدة. تم تطوير تقنيات إخفاء المعلومات والتحليل في نفس الوقت. الاخفاء هي تقنية لإخفاء البيانات السرية وجعلها غير مرئية من خلال تضمينها في نطاق وسائط متعددة. تحليل إخفاء الصور هو تقنية لتحديد ما إذا كانت الصور تحتوي على معلومات مخفية أم لا. نظرًا لأن أساليب إخفاء المعلومات المتقدمة يمكنها تحميل رسائل سرية صغيرة في ألاغلفة، فإن تقنيات تحليل إخفاء الصور الحديثة مع تحسين تقنية إخفاء المعلومات لا يمكنها التمييز بين صور الغلاف وصور **stego** أو يتم اكتشافها بأداء أقل دقة، مما يجعل الأمر صعبًا.

أثبتت الأبحاث الحديثة فعالية استخدام الشبكات العصبية لاكتشاف صور **stego**. ومع ذلك، نظرًا لأن الوصول إلى قاعدة البيانات أمر معقد، وهو أمر ضروري في عملية التصنيف لاكتشاف ما إذا كانت الصورة عبارة عن صورة غلاف أم صورة **stego**. تقدم الدراسة الحالية تقنية لمعالجة هذه المشاكل للكشف عن صورة **stego** في المجال المكاني.

يعتمد النظام المقترح على الخصائص الإحصائية للصورة المدخلة. يعتمد النظام على اشتقاق وظيفة الارتباط التلقائي للرسم البياني للصورة، ثم تطبيق مرشح تمرير عالي كعتبة لتقويم النظام المقترح. يمكن استخدام هذه التقنية لتحديد الصورة التي هي غلاف أو **stego** دون اعتماد الصورة الأصلية. على الرغم من أن هذه الدراسة قد ركزت على إخفاء المعلومات في البتات الأقل أهمية (**LSB**)، فقد وجد أنه يمكن تطبيق النهج المقترح بنجاح على إخفاء **LSB** بالتسلسل والعشوائي مع أوامر مختلفة من مشتقات ارتباط المدرج التكراري. أيضًا، يتم النظر في نسبة صورة **stego** إلى حدود الغلاف، حيث يمكن للنسب الصغيرة أن تتخطى طريقة الكشف هذه ما لم يتم تعديلها. لقد تم فحص هذه الإستراتيجية أيضًا بحثًا عن طرائق إخفاء الصور الأخرى.

كشفت النتائج في النهاية عن فعالية هذا النظام الشامل لتحليل الإخفاء. تم استخدم عتبات مختلفة لتقويم النظام المقترح. تعتمد هذه العتبات على تنسيق صورة الإدخال. توصل النظام المقترح من خلال النتائج أن مرشح التمرير العالي أكثر ملاءمة من العتبات الأخرى عند استخدام التدرج الرمادي كصورة إدخال وأن مشتقات ارتباط الانتروبيا في الرسم البياني للصورة مناسبة لصورة ملونة. أن النظام المقترح أفضل عند مقارنة أدائه مع طرائق أخرى محددة من حيث التعرف على الرسائل المضمنة بشكل تسلسلي أو عشوائي وكذلك سعة الرسائل المكتشفة. الطريقة المقترحة عامة ولكنها لا تستخدم مجموعة بيانات في مرحلة المعالجة لتصنيف الصور إلى صور غلاف أو صور مخفية.

يتم تقويم نتائج الطريقة المقترحة باستخدام خمسة مناهج إخفاء المعلومات. الطريقة المقترحة قادرة على اكتشاف الرسالة السرية إذا كان حجم الرسالة صغيرًا (**Rm < 0.01**). لغة البرمجة للنظام المقترح هي MATLAB R2021a (64-bit).

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة بابل- كلية تكنولوجيا المعلومات
قسم البرمجيات

كشف الاخفاء بالصورة بالاعتماد على التحليل المترابط

أطروحة مقدمة
الى مجلس كلية تكنولوجيا المعلومات في جامعة بابل وهي جزء من متطلبات نيل درجة
الدكتوراه فلسفة في تكنولوجيا المعلومات / برمجيات

من قبل
ناطق مطشر عبدعلي حسين

إشراف
أ. د. زاهر محسن حسين

2022 م

1443 هـ