جـــامـعـة الـــشـرق الأوسـط
MIDDLE EAST UNIVERSITY
Amman - Jordan

Steganalysis of Color Images for Low Payload Detection

تحليل الاخفاء للصور الملونة لكشف الخزن المنخفض

By:

Renad Mekhled Al-Manaseer

Supervisor:

Dr. Mudhafar Al-Jarrah

A Thesis Submitted in Partial Fulfillment of the Requirements for the Master Degree in Computer Science

Department of Computer Science

Faculty of Information Technology

Middle East University

Jan. 2019

Authorization

I, **Renad Mekhled Al-Manaseer**, hereby authorize Middle East University to supply copies of my thesis to libraries, organizations or individuals when required.

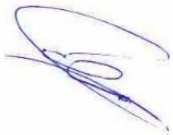
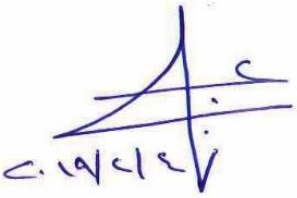
Name: Renad Mekhled Al-Manaseer.

Date: 26 / 01 / 2019

Signature: 

Thesis Committee Decision

This thesis titled "**Steganalysis of Color Images for Low Payload Detection**" was successfully defended and approved on 26/01/2019

Thesis Committee Members	**Signature**
(Supervisor) Dr. Mudhafar Al-Jarrah Middle East University	
(Head of the Committee and Internal Examiner) Dr. Bassam Al-shargabi Middle East University	
(External Examiner) Dr. Adnan Hadi Al-Helali Al-Israa University	

Acknowledgment

(اللهم لك الحمد والشكر كما ينبغي لجلال وجهك الكريم وعظيم سلطانك وعلو مكانك)

First, I would like to thank Allah for the strength and patience he had given me to finish this work. This work could not have been achieved without having faith that Allah is there to support and help me. May he bless everyone who was there for me during my studying period.

My words cannot describe how grateful I am to **Dr. Muthafar Al-Jarrah**, whose recommendations, devotion, advocacy, patience, encouragement, and support have led me to achieve this work. I cannot express how lucky I am having him to supervise my thesis. Also, I would like to express my deepest gratitude to all the respectable lecturers at the Faculty of Information Technology, Middle East University.

The Researcher
Renad Al-Manaseer

بسم الله الرحمن الرحيم

"وقل ربي زدني علما"

Dedication

This thesis is dedicated to my whole family;

Especial thanks to my precious Father and my one and only my Mother, who always proud of me and supported me in every step of my life, no words can describe what they have done for me, thank you for your endless love.

My brothers Ammar, Khaled, Mohammad and Ahmad, who are one part of my life.

My best friends, who always been there for me during difficult and stressful times, particularly, **Ghofran Mahmood**.

My cousins Maha and Dania, who support me with their precious words and love me, and I thank Allah for their presence in my life.

My friends, who supported me with their nice words, who were the cause of my happiness during my last days at university, and who loved them and will stay in my heart.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviations	Meaning
1-LSB	One Least Significant Bit
2-LSB	Two Least Significant Bit
3-LSB	Three Least Significant Bit
4-LSB	Four Least Significant Bit
ANN	Artificial Neural Network
BMP	Bitmap
BPC	Bit Per Channel
BPP	Bit Per Pixel
CGCM	Color Gradient Co-Occurrence Matrix
DT	Decision Tree
FN	False Negative
FP	False Positive
GIF	Graphics Interchange Format
GLCM	Gray Level Co-Occurrence Matrix
GLCM-d	Extended Gray level co-occurrence matrix
JPEG	Joint Photographic Exchange Group
LSB	Least Significant Bit
PNG	Portable Network Graphics
PPM	Portable Pixel Map
PSNR	Peek-Signal-to-Noise-Ratio
RGB	Red, Green, Blue
RHB	Right Half Byte
SVM	Support Vector Machine
TIFF	Tag Image File Format
TN	True Negative
TP	True Positive

Steganalysis of Color Images for Low Payload Detection

By: Renad Mekhled Al-Manaseer

Supervisor: Dr. Mudhafar Al-Jarrah

Abstract

Steganography, the science of embedding secret data in an appropriate cover object such as video, audio, network, and images. Steganalysis, aims to detect the presence of hidden data inside the cover object, is a countermeasure against information hiding techniques that can be used for illegitimate purposes. The work in this thesis introduces a steganalysis model that uses statistical texture features and the machine learning approach to detect the presence of hidden data in RGB color images. The work analyzes features of an RGB image in PPM format as a composite unit. The feature set used in this study consists of 120 features per color channel, which includes the basic and extended Gray Level Co-Occurrence Matrix (GLCM) features of correlation, contrast, homogeneity and energy, calculated for full bytes, half-bytes, 3-bit, 2-bit, and 1-bit fragments. The features are applied to single channels, and the single channel features are merged into three-channel image feature sets. The machine learning binary classifier that is selected for this work is the Support Vector Machine (SVM) algorithm. A public dataset of 10,000 uncompressed PPM clean images is used, and seven stego image datasets of 10,000 images each were created from the clean images, which were embedded with random secret data at payload ratios from 0.01 to 0.5 bit per channel, using 1LSB steganography technique. For the classification phase, a set of 5000 clean images were randomly selected and 5000 stego images were randomly selected for each payload ratio. The steganalysis results, using the Support Vector Machine, showed detection accuracy values ranging from 56.18% for 0.01 bit per channel to 91.00% for 0.5 bit per channel.

The results showed that the model achieved higher true positive detection than true negative in most of the payloads ratios, indicating that the model is more effective in detecting stego images, which is the purpose of steganalysis. MATLAB 2016a was used in the implementation of the image analysis and classification parts of the proposed model.

Keywords: Steganalysis, Steganography, Stego Image, SVM Classifier, GLCM, Feature Set, Detection Accuracy, PPM Image, Payload Ratio, Secret Data, Bit Per Channel.

تحليل الاخفاء للصور الملونة لكشف الخزن المنخفض

اعداد: رناد مخلد المناصير

اشراف: الدكتور مظفر الجراح

الملخص

إخفاء المعلومات هو علم وتكنولوجيا لإخفاء المعلومات السرية بداخل عدة اشكال من وسائل الاعلام الرقمية مثل الفيديو، الصوتيات والصور. اما تحليل غطاء علم الاخفاء هو علم و تكنولوجيا تهدف الى كشف اخفاء المعلومات داخل وسائل الاعلام الرقمية التي قد تستخدم لاغراض غير مشروعة. البحث المقدم في هذه الاطروحة يعرض نموذج تحليل غطاء الإخفاء الذي يستخدم الملامح الإحصائية ونهج التعلم الآلي للكشف عن وجود البيانات المخفية في الصور الملونة RGB. يحلل البحث خصائص صورة RGB من نوع PPM كوحدة مركبة. تتألف مجموعة الخصائص من 120 خاصية لكل قناة والتي تتضمن ميزات مصفوفة مستوى الرمادي الاساسية والممتدة (GLCM) للارتباط والتباين والتجانس والطاقة، محسوبة لحالات البايت الكاملة ونصف البايتات، 3. بت، 2 بت و1 بت. يتم تطبيق الميزات على القنوات الفردية، ويتم دمج ميزات القناة الواحدة في مجموعات صور ثلاثية القنوات. خوازمية التصنيف المستخدمة في هذا البحث التصنيف الثنائي هي خوارزمية ماكنة متجه الدعم(SVM). تم استخدام مجموعة البيانات عامة في هذا البحث والتي تتكون من 10,000 صورة غير مضغوطة من نوع PPM، كما تم إنشاء سبع مجموعات من صور الستيجو باستخدام 10,000 صورة من الصور النظيفة، والتي تم تضمينها مع بيانات سرية عشوائية في نسب الحمولة من 0.01 إلى 0.5 بت لكل قناة، باستخدام تقنية إخفاء المعلومات LSB−1. بالنسبة لمرحلة التصنيف، تم اختيار مجموعة من 5000 صورة نظيفة بشكل عشوائي وتم اختيار 5000 صورة بشكل عشوائي لكل نسبة حمولة.

وقد تم العمل التجريبي على 5000 صورة مختارة بشكل عشوائي من صور النظيفة والصور المتضمنة للإخفاء (ستيجو) لجميع الحمولات. وأظهرت نتائج تحليل الخوارزمية، باستخدام خوارزمية ماكنة متجه الدعم (SVM)، قيم دقة للكشف تتراوح من 56.18٪ لـ 0.01 بت لكل قناة إلى 91.00٪ لـ 0،5 بت لكل قناة. وأظهرت النتائج أن النموذج قد حقق اكتشافًا إيجابيًا (TP) أعلى من السلبية الحقيقية (TN)في معظم نسب الحمولات، مما يشير إلى أن النموذج أكثر فاعلية في الكشف عن الصور الثابتة(الستيجو)، وهو الغرض من Steganalysis. تم استخدام MATLAB 2016a في تنفيذ أجزاء تحليل وتصنيف الصور الخاصة بالنموذج المقترح.

الكلمات المفتاحية: كشف الاخفاء، اخفاء المعلومات، صورة مضمنة ، المصنفSVM ، مصفوفة مستوى الرمادي، مجموعة ميزات، دقة الكشف، نسب الحمولات، بيانات سرية، تضمين، استخراج.

Chapter One

Introduction

1.1 Research Context

The research in this thesis focuses on the detection of embedded random secret information in RGB (Red, Green, and Blue) cover images, using a statistical feature-based approach. The texture features of the cover images will be extracted and analyzed to detect the existence of hidden data, based on textures of the color image and components of the image.

1.2 Background

A problem of considerable interest in recent years has been the exchange of secret data embedded in innocuous-looking images, audio and video clips (Fridrich, 2009), with the aim of hiding the existence of the secret data from possible observers who might capture the transmitted cover image. Steganography and steganalysis, the hiding and detection of a covert payload within an innocent cover object, started to receive attention from the computer science, engineering, and mathematics communities (Bohme, 2010).

The common technique of hiding information in digital media such as video, audio, and images are steganography (Bohme, 2010), this technique can be used by individuals and businesses to send their private and confidential information in a secure way by embedding the information within a cover media (stego).

Steganography, the science of embedding secret data in an appropriate cover object, is an important research issue in the computer security field (Anderson & Petitcolas, 1998).

However, the problem with steganography is that it can be used for illegal purposes, such as the exchange of messages by criminals, and the attempts by business insiders to send private documents to competitors. The steganalysis technique (Bohme, 2010), is a countermeasure that aims to detect the existence of hidden data inside media files, and in some cases to extract the hidden data. The steganalysis technique is mainly used in monitoring malicious communications by terrorists and criminals, but it can help in detecting harmful malware inside documents exchanged over the internet.

1.3 Problem Statement

The problem tackled in this thesis is that low payload steganography can be difficult to detect. The distortion caused by embedding a small secret data might not be easily observed through the analysis of the carrier media (the cover).

1.4 Aim and Objectives

The aim of this thesis is to enhance the detection of embedded low payload secret information in color images based on extending GLCM features.

The following objectives are considered:

1. Select texture features to be measured.
2. Design and implement a detection model.
3. Select a classifier for the steganalysis model.
4. Build and/or acquire image datasets for the experiment.
5. Evaluate the detection accuracy of the proposed model.

1.5 Motivation

This thesis is motivated by the need to prevent the transmission of illegal information that has increased lately due to the large availability and ease of access to steganography tools over the internet, that attempt to avoid detection through reducing secret data size.

1.6 Significance of Research

This thesis enhances the detection of steganography at low payload size in uncompressed color images based on basic GLCM and extending GLCM features, which will help cybersecurity administrators to prevent the sending and receiving of images loaded with hidden data.

1.7 Research Questions

The problem in this thesis can be filtered in the following questions:

1. What are the image features that will be used in the steganalysis model to detect reduced size secret data?
2. How will the features of the color channels and other segments of the image be combined to enhance the detection process?
3. What is the classification method that will be used?
4. What is the dataset that will be used in evaluating the proposed model?
5. What is the evaluation method and what are the accuracy results?

1.8 Scope of Research

The scope of this thesis covers the following points:

1. Using the LSB steganography method to embed secret data in LSB of color channels of some pixel in RGB color images.
2. Extracting the statistical image texture features.

3. Selecting a color image dataset.
4. Embedding random secret data of varying sizes in images of the selected dataset.
5. Analyzing the clean and stego images.
6. Selecting a binary classifier for the detection process.
7. Evaluating the proposed model through experimental work.

1.9 Limitations of the Research

This research is limited to steganalysis of uncompressed RGB color images with a low size of the payload. Further work will be needed to adapt the proposed model for other media files such as audio, video and compressed color images.

1.10 Thesis Organization

This thesis consists of five chapters that covered the following topics, as below:

Chapter one is the introduction, which introduced the research context of the thesis, background of the study, problem statement, aim and objectives, motivation, the significance of work, research questions, the scope of work and limitations of the research.

Chapter two presents a literature review, overview about the concept and main topics of steganography, embedding algorithms, steganalysis, and its attacks, methods, and approaches, also define some types of feature set, some types of classifiers and the types of images and the related work.

Chapter three presents the methodology and the proposed model which introduced the methodology approach outlines of the proposed model, statistical features selection, the classifier, and the proposed model and the evaluation metrics.

Chapter four presents experimental results and discussion which introduced the introduction, objectives of the experimental work, the image dataset, experimental work and results and discussion of the reference work and the proposed work.

Chapter five presents the conclusion and future work which introduced the conclusion and suggestions for future work.

Chapter Two

Literature Review and Related Work

2.1 Introduction

This chapter presents an overview of the concept and main topics of steganography, embedding algorithms, also presents the definition of steganalysis including its attacks, methods and the approaches. This chapter will also define some types of feature set, some types of classifiers and the types of images.

2.2 Steganography

Steganography is defined for arbitrary communication channels, only those where the cover media consist of multimedia objects, such as image, video or audio files (Bohme, 2010). Steganography and cryptography are two types in hiding data techniques. Cryptography is the art of protecting information by transforming it into an unreadable format, called ciphertext. To decipher this unreadable format, a secret key is required. On the other hand, Steganography can be defined as the hiding of information by embedding messages within other, seemingly harmless messages, graphics or sounds (Siper, Farley & Lombardo, 2005). According to Johnson, Jajodia (1998), Steganography in security is to supplement cryptography, not replace it. If a hidden message is encrypted, it must also be decrypted if discovered, which provides another layer of protection. Batch steganography (Ker & Pevný 2012) has been used to lower secret data size by splitting a secret document into a batch of small pieces that are less likely to be detected.

Steganography is an important research issue in the computer security field (Anderson & Petitcolas 1998).

Provos and Honeyman (2003) also defined steganography as the art and science of hiding communication; a steganographic system thus embeds hidden content in unremarkable cover media so as not to arouse an eavesdropper suspicion.

Figure 2.1 below shows a simple explanation of the steganography approach:

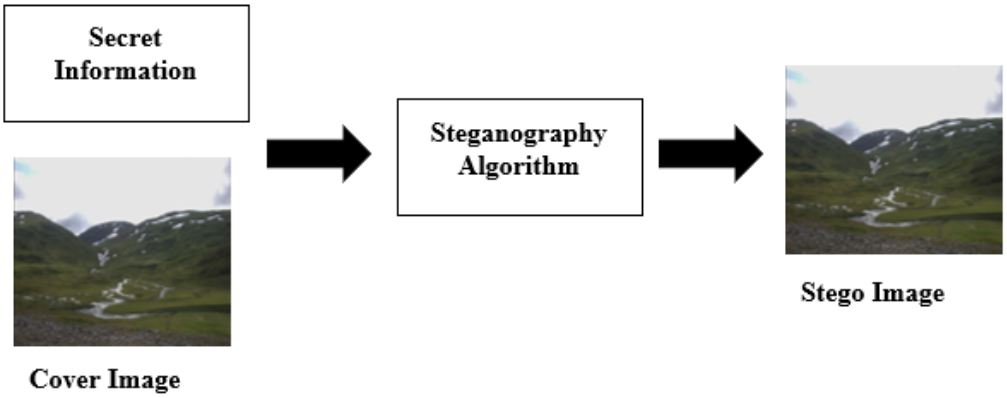


Figure 2.1 Steganography Approach

2.2.1 History of Steganography

Steganography comes from the Greek, the first steganographic technique was developed in ancient Greece. Steganography continued development in the early 1600s as Sir Francis Bacon used a variation in typeface to carry each bit of the encoding. Steganography continued over time to develop into new levels. During times of war, steganography is used extensively.

During the American Revolutionary War, both the British and American forces used various forms of Invisible Inks. Invisible Ink involved common sources, this included milk, vinegar, fruit juice, and urine, for the hidden text (Siper, Farley & Lombardo, 2005). Additionally, in the past, Egyptians used a technique involving illustrations to conceal secret messages. And not long ago in Saudi Arabia, a project was initiated at the King Abdul-Aziz City of Science and Technology to translate from secret writing into English a large number of ancient Arabic manuscripts which were believed to have been written over 1200 years ago (Anderson & Petitcolas 1998; Johnson & Jajodia 1998).

2.2.2 Steganography Types

- **Text Steganography:** can be achieved by altering the text formatting, or by altering certain characteristics of textual elements (e.g., characters). The goal in the design of coding methods is to develop alterations that are reliably decodable (even in the presence of noise) yet largely indiscernible to the reader (Nosrati, Karimi & Hariri, 2011).
- **Image Steganography:** hiding information inside images is a popular technique nowadays. To hide a message inside an image without changing its visible properties, the cover source can be altered in noisy areas with many color variations, so less attention will be drawn to the modifications. The most common methods to make these alterations involve the usage of the least significant bit or LSB, masking, filtering and transformations on the cover image. These techniques can be used with varying degrees of success in different types of image files (Nosrati, Karimi & Hariri, 2011).

- **Audio, Video Steganography:** in audio steganography, the secret message is embedded into a digitized audio signal which results in the slight altering of the

binary sequence of the corresponding audio file. There are several methods are available for audio steganography such as LSB Coding, Phase Coding, Spread Spectrum and Echo Hiding (Nosrati, Karimi & Hariri, 2011).

- **Network Steganography:** which employs hiding data in the network datagram level in a TCP/IP based network like Internet. Network Covert Channel is the synonym of network steganography. The overall goal of this approach to make the stego datagram is undetectable by Network watchers like a sniffer, Intrusion Detection System (IDS) etc. In this type of information to hide is placed in the IP header of a TCP/IP datagram. Some of the fields of IP header and TCP header in an IPv4 network are chosen for data hiding (Das, Das, Bandyopadhyay & Sanyal, 2008).

2.2.3 Steganography Methods

Steganography methods categories into two domains:

1. **Spatial Domain** defined as the secret data and the cover medium modified in the spatial domain, which involves encoding at the level of the LSB. This method although simpler has a larger impact compared to the other two types of methods. (Cheddad, Condell, Curran & Kevitt, 2010)
2. **Frequency Domain** defined as the transform is applied on the cover image and the secret message bits are hidden inside the coefficients of the transformed cover image. This method has been proved to be more robust than spatial domain techniques but is complex as compared to LSB techniques. Most commonly used transforms are Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT). (Joshi, Bokil, Jain & Koshti, 2012).

2.2.4 Embedding Algorithms

The embedding algorithm is used to hide a secret data inside a cover media, which results in a stego media. An example of these algorithms is:

- The Least Significant Bit (LSB) substitution algorithm, is the most commonly used spatial domain technique. In LSB substitution technique the least significant bit of each pixel of the cover is replaced by the secret message bits. (Joshi, Bokil, Jain & Koshti, 2012)
- There are other algorithms that used in the research such as The Spatial-Universal Wavelet Relative Distortion (S-UNIWARD) steganography algorithm, the Wavelet Obtained Weights WOW steganography algorithm, and the Synchronizing the Selection Channel Synch-HILL steganography algorithm. (Abdulrahman, Chaumont, Montesinos & Magnier, 2016)

2.3 Steganalysis

The aim of steganalysis is to detect the existence of hidden information in digital media. Steganalysis is the art and science of detecting secret messages hidden using steganography (Fridrich, Goljan, & Du, 2001; Johnson & Jajodia, 1998). The goal of steganalysis is to collect enough evidence about the presence of an embedded message and to break the security of its carrier.

2.3.1 Steganalysis Approaches

Steganalysis approaches can be classified as:

- **Specific/ Target Steganalysis:** A specific steganalytic method fully utilizes the knowledge of a targeted steganographic technique and may only be applicable to such a kind of steganography and often takes advantage of the insecure aspect of a steganographic algorithm (Li, He, Huang & Shi, 2011).
- **Generic / Blind / Universal Steganalysis:** A universal steganalytic method can be used to detect several kinds of steganography. Usually, universal methods do not require the knowledge of the details of the embedding operations. Therefore, it is also called a blind method (Li, He, Huang & Shi, 2011).

2.3.2 Steganalysis Attacks

Johnson and Jajodia (1998) classified attacks in six main categories as in the following:

1. **Stego Only:** The evaluation must be conducted using only the stego item.
2. **Known Cover:** The original cover object and the stego object are available for analysis.
3. **Known Message:** The secret content is accessible when contrasted with the stego-item.
4. **Chosen Stego:** The stego-algorithm and stego-item are both accessible for the evaluation.
5. **Chosen Message:** Uses selected content to create a stego-item for additional evaluations.
6. **Known Stego:** The stego-algorithm, concealing content, and stego-item are known and can be employed for evaluation.

2.3.3 Steganalysis Methods

- **Visual Method:** By analyzing the images visually, like considering the bit images and try to find the difference visually in these single bit images (Al-Taie, 2017).
- **Structural Method:** The format of data file often changes as the data to be hidden is embedded, identifying these characteristic structural changes can detect the existence of image, for example in palette based steganography the palette of the image is changed before embedding data to reduce the number of colors so that the adjacent pixel color difference should be very less. This shows that groups of pixels in a palette have the same color which is not the case in normal images (Al-Taie, 2017).
- **Statistical Method:** The statistical analyses of the images by some mathematical formulas is done and the detection of hidden data is done based on these statistical results. Generally, the hidden message is more random than the original data of the image thus finding the formulae to know the randomness reveals the existence of data (Al-Taie, 2017).

2.4 Feature Set

In steganography research, a lot of algorithms have proposed a collection of statistical features for the cover media to enhance detectability process.

2.4.1 Co-Occurrence Matrix

A co-occurrence matrix is a matrix that is defined over an image to be the distribution of co-occurring pixel values (grayscale values, or colors) at a given offset.

2.4.1.1 GLCM

Gray level Co-occurrence Matrix also rarely called the Grey Tone Spatial Dependency Matrix (GLCM) was proposed in 1973 by Haralick.

Gray level co-occurrence matrix (GLCM) known as the gray-level spatial dependence matrix is a statistical method of examining texture that considers the spatial relationship of adjacent pixels, it is classified as second order statistics. The GLCM functions characterize the texture of an image by storing the number of pixel neighborhoods in an image that has a grayscale combination. GLCM function result can be either logical or numeric, and it must contain real, non-negative, finite integers. The texture filter functions provide a statistical view of texture based on the image histogram. These functions can provide useful information about the texture of an image. GLCM texture considers the relation between two pixels at a time, called the reference and the neighbor pixel (Hall-Beyer, 2017).

GLCM has three orders based on the relationship between pixels, first order texture measures are statistics calculated from the original image values, like variance, and do not consider pixel relationships. Second order which considers the relationship between groups of two pixels in the original image and third and higher order textures considering the relationships among three or more pixels (Hall-Beyer, 2017).

Kekre, Athawale, and Patki (2011) extract 31 features from several combinations of GLCM.

The figure below shows how the GLCM matrix process:

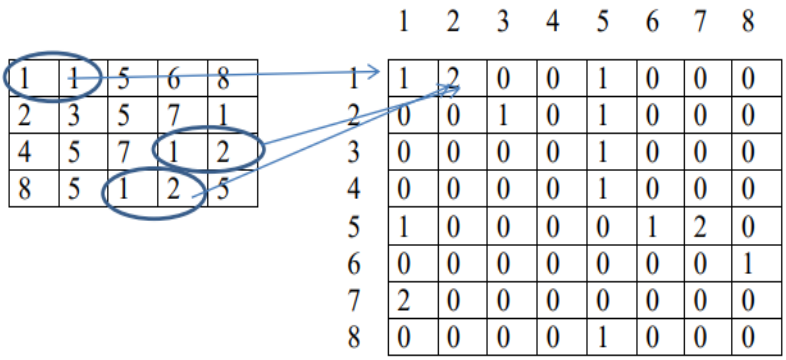


Figure2.2: GLCM Matrix Process (Al-Taie, 2017)

The properties of GLCM are:

- **Contrast:** In short form, it is called a CON. 'Sum of Square Variance' is another name of contrast. It defers the calculation of the intensity contrast linking pixel and its neighbor over the whole image (Sharma, Priyanka, Kalsh & Saini, 2015).
- **Correlation:** It passes the calculation of the correlation of a pixel and its neighbor over the whole image means it figures out the linear dependency of gray levels on those of neighboring pixels (Sharma, Priyanka, Kalsh & Saini, 2015).
- **Energy:** Since energy is used for doing work, Thus orderliness. It makes use for the texture that calculates orders in an image. It gives the sum of square elements in GLCM (Sharma, Priyanka, Kalsh & Saini, 2015).
- **Homogeneity:** In the short term it is going by the name of HOM. It passes the value that calculates the tightness of distribution of the elements in the GLCM to the GLCM diagonal (Sharma, Priyanka, Kalsh & Saini, 2015).

2.5 Classifiers

The classifier determines if the image is clean or stego, where a set of training dataset with selected input features are used to train the classifier to extract the information using labeled datasets in training phase, then in testing phase the classifier used unlabeled datasets to classify it and give the final result. There are a lot of classification methods such that Decision Tree, Artificial Neural Network (ANN) and Support Vector Machine (SVM).

- **Decision Tree (DT)** used for supervised learning algorithm and used for classification. A decision tree is applicable in both cases that are continuous and categorical output and input variables. In the DT algorithm, the significant differentiator in input variables is used to split the samples of the dataset into homogeneous sets. (Gaur & Chouhan, 2017)
- **Artificial Neural Network (ANN)** A neural network model is like a human nervous system. The artificial neural network is taught through a dataset. This dataset may be known to us then ANN is trained in a supervised manner, and it learns precisely and quickly about the pattern buried in the dataset. And trained ANN is used to identify the patterns for which it is trained. But if the dataset is not known to us in advance then the unsupervised training is used. The neural network consists of neurons that are correlated together to convert inputs into useful output. (Gaur & Chouhan, 2017)
- **Support Vector Machine (SVM)** comes in the category of supervised learning. The SVM used for regression and classification. But it is popularly known for classification. It is a very efficient classifier. Every object or item is represented by a point in the n-dimensional space.

The value of each feature is represented by the coordinate. Then the items divided into classes by finding hyper-plane (Gaur & Chouhan, 2017).

The diagram shows support Vectors that represent the coordinate of each item. The SVM algorithm is a good choice to segregates the two classes (Gaur & Chouhan, 2017).

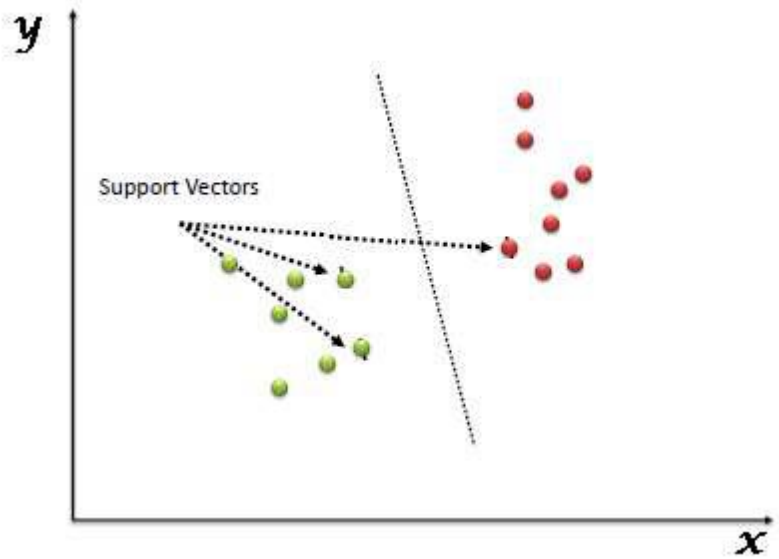


Figure 2.3: SVM Process (Gaur & Chouhan, 2017)

2.6 Types of Images

Images are classifying as uncompressed, lossless compressed and compressed, images have many formats such as BMP, JPEG, TIFF, GIF, PNG, and PPM.

- **BMP (Windows Bitmap Format):** BMP files are device-independent files most frequently used in Windows systems, Based on the RGB color model. Header region contains info about size and color depth. Data region contains the values of each pixel in a line. (Badr, Salama, Selim, & Khalil, 2014)

- **JPEG (Joint Photographic Exchange Group):** JPEG image is a popular cover image format used in steganography. The compression method is usually loosely compression, meaning that some visual quality is lost in the compression process and cannot be restored. (Badr, Salama, Selim, & Khalil, 2014)
- **TIFF (Tag Image File Format):** TIFF is a common format for exchanging raster graphics (bitmap) images between applications programs, A TIFF file can be identified as a file with a ".tiff" or ".tif" file name suffix. TIFF format supports RGB, indexed color, and grayscale images with alpha channels and Bitmap mode images without alpha channels. TIFF is a flexible bitmap image format supported by all paint, page layout, and image editing. TIFF documents have a maximum file size of 4 GB. TIFF image format allows for lossless compression (Rasool, 2017).
- **GIF (Graphics Interchange Format):** The GIF format Introduced in 1987 by CompuServe, it supports up to 8 bits per pixel and the color of the pixel is referenced from a palette table of up to 256 distinct colors mapped to the 24-bit RGB color space. GIF is still a popular image format on the internet because the image size is relatively small compared to other image compression types. (Badr, Salama, Selim, & Khalil, 2014)
- **PNG (Portable Network Graphics):** PNG Introduced in 1999. PNG supports three main image types: true color, grayscale, and palette-based "8-bit" (Badr, Salama, Selim, & Khalil, 2014). It created to replace GIF because it has smaller file sizes than GIF.

- **PPM (Portable Pixel Map):** The PPM format is one of the simplest formats and devised to be an intermediate format for use in developing file format conversion systems. The PPM format is a lowest common denominator color image file, which allows very little information about the image beside basic color PPM is used as an intermediary format. The file extension for ppm images are. ppm, PGM, PBM, PNM (Pawar, Halgaonkar, Bakal, & Wadhai, 2011).

2.7 Reasons to Choose the Detection of Secret Data in Images

There are many reasons behind the attention given to steganalysis of images compared to other media. First, the images are the most available type of cover to hide secret data over the internet and the human eye cannot easily detect if the image is clean or has been embedded with data. The second reason is to prevent illegal information transmission by criminals who use steganography for their illegal purposes.

2.8 Previous Works

Steganalysis of RGB images introduced in many previous work such as (Fridrich, Goljan & Du,2001; Gong, Wang, 2012; Aljarf, Amin, Filippas & Shuttelworth, 2013; Goljan, Fridrich & Cogranne, 2014; Abdulrahman, Chaumont, Montesinos & Magnier, 2015; Abdulrahman, Chaumont, Montesinos & Magnier, 2016; Abdulrahman, Chaumont, Montesinos & Magnier, 2016; Al-jarf, 2016; Rasool, 2017; AlTaie, 2017; Al-Jarrah, Al-Taei & Aboarqoub, 2017).

1. Fridrich et al. (2001) proposed a reliable method for detecting least significant bit (LSB) non-sequential embedding in digital images. The Kodak DC260 digital camera used to convert a color 1536×1024 image ‘kyoto.bmp’ to grayscale with 384×256 pixels. A series of stego-images were created from the original image by randomizing the LSB of 0−100% pixels in 5% increments. The experimental results obtained by RS steganalysis.
2. Gong, et al. (2012) introduced a steganalysis algorithm based on colors gradient co-occurrence matrix (CGCM) for GIF images. CGC Misco structure with colors matrix and gradient matrix of the GIF image, and 27-dimensional statistical features of CGCM, which are sensitive to the color-correlation between adjacent pixels.
3. Aljarf et al. (2013) have proposed a system to detect steganography. At this paper three steganography tools to create stego images which are OpenStego, StegHide and F5 algorithm. The model used for both gray and color images and it is based on four features which are contrast, energy, homogeneity, and correlation. The initial test in this paper done using the co-occurrence function in MATLAB. It supports the gray images only, so it must be used with each single-color channel in the color image.
4. Goljan et al. (2014) proposed an extension of the spatial rich model (SRM), which was designed for color images, to allow for more accurate detection of steganography in color images. At this work, SRMQ1 features are augmented by a collection of symmetrized 3D co-occurrences of residuals between color channels, the Color Rich Model (CRMQ1) of dimension 5404. While these additional features help detection only marginally in color images. Their experiment is done on three versions of BossBase 1.01: BossBaseRes,

BossBasePPG and BossBaseAHD and the embedding algorithms they used are the non-adaptive LSB matching (LSBM) and the content-adaptive WOW with different sizes of payloads.

5. Abdulrahman et al. (2015) proposed steganalysis method based on color feature correlation. The feature set consist of two subsets, the first set consisting of 18157 features. The second set consists of 3000 features obtained from the correlation of different R, G, B channel gradients. Color filter array, Demosaicking algorithm and Bayer pattern were used to crop original raw images into five crops. The classifier that used is the ensemble classifier to train and test the dataset which is built from two databases. The first database is BossBase 1.0 and the second database is Dresden image dataset and the final dataset consist of 10,000 PPM color images of size 512x512.

   The steganography algorithms that used to obtained stego images are S-UNIWARD and WOW steganography algorithm. The comparison based on embedding in one channel (Green) and embedding in three channels with different payloads and the results show that the detection if the embedding is in one channel better than three channels. The detection rates achieved higher performance by registering 87.54% and 86.63% for S-UNIWARD and WOW respectively with the payload 0.5 bit per channel.

6. Abdulrahman et al. (2016) have introduced steganalysis method extension to their previous work at color images: Color image steganalysis using correlations between RGB channels at 2015. At this work, new features have obtained from color rich model and features from the first method in addition to two new features. The first type of feature reflects local Euclidean transformations, and the second one reflects mirror transformations. The total number of features in this

paper is 24157 features and the dataset was same as the first method. The steganography algorithms that used were: S-UMIWARD, WOW, and Synch-HIL, also ensemble classifier used in this paper. The detection rates achieved higher performance by registering 88.76%, 87.93%, and 88.07% for S-UNIWARD, WOW, and synch-HILL, respectively (with the payload 0.5 bit per channel) which it is better than their first method.

7. Abdulrahman et al. (2016) have proposed extension to their previous works by developing another steganalysis system to enriches the features from the Color Rich Model by adding new features obtained by applying steerable Gaussian filters and then computing the co-occurrence of pixel pairs to increase the detection of hidden messages, the total number of features in this paper is 22,563 features. The dataset that used was a public dataset that consists of 10,000 PPM color images, where the experimental work was on a randomly chosen dataset that consist of 5000 PPM clean and stego images, the steganographic algorithms that used to hide messages in the image are S-UNIWARD, WOW, and Synch-HILL. The ensemble classifier used to analyse the features and get the accuracy results for all payloads.
8. Aljarf (2016) has presented a detection system that combines three different steganalysis techniques. All technique addresses blind image steganalysis rely on the extraction CGCM and histogram features. The proposed detection system was trained to classify grey or color clean images and grey or color stego images, which were created using LSB and the F5 steganography algorithm. The database included color and grey images in various formats using both lossless and lossy compression.

The proposed detection system was trained and tested to distinguish stego images from clean ones using the Discriminant Analysis (DA) classification method and Multilayer Perceptron neural network (MLP).

9. Rasool (2017) has proposed a new method in detection the existence of hidden messages that embedding in RGB images. The statistical texture features set that used consist of 26 features for each channel. Two datasets were used, one dataset that consists of 1500 BMP uncompressed color images for training the classifier and one dataset that consists 1000 PNG uncompressed color images for testing, the classifiers that used SVM and Discriminant Analysis (DA) compare with the first classifier. The experiment approach done on MATLAB, the embedding algorithms that used were 2LSB and 4LSB. The results expressed on the individual channel (blue), dual channel (RG, RB, BG) and on the combination of three channels (RGB). The 3-fold cross validation used to calculate the accuracy, where the detection accuracy obtained a very high accuracy over 99% for the combined RGB channels features as well as for dual channel combinations and single channels, and the SVM classifier achieves better performance than DA.
10. Al-Jarrah et al. (2017) have proposed a steganalysis model for grayscale images with a new texture feature set, that is based on statistical texture features of images including gray level co-occurrence matrix (GLCM), Entropy, and additional statistical image features. The steganography methods that used to embedded secret data were 2-LSB, 3-LSB and 4-LSB bit. The dataset that used was the public BossBase1.01 which consists of 10,000 PGM images. The experimental work was on MATLAB using SVM classifier. the detection accuracy results of the validation phase were 99.41% for the combined clean and 4LSB images and 99.02% for the clean and 2LSB stego images.

11. Al-Taie (2017) has proposed a statistical model for steganalysis to enhance the detection of the existence of hidden data inside 8-bit depth gray-scale BMP images that are based on an enhanced GLCM feature set, in the analysis of gray-scale one channel images. The research included experimental results of analyzing many gray-scale images from public datasets that contain 5000 images for testing and 1500 images for training. The average of the detection accuracy ranged from 97.50% to 98.73% in the validation test and 97.82% to 98.28% in the field test.

Table 2.1 summarizes the brief description of the related work.

Table 2.1 Summary Description of the Related Work

Papers	Year	Description
Fridrich, Goljan, and Du	2001	Proposed a reliable method for detecting least significant bit (LSB) non-sequential embedding in digital images.
Gong, Wang	2012	Proposed a steganalysis algorithm based on colors gradient co-occurrence matrix (CGCM) for GIF images CGC Misco structure with colors matrix and gradient matrix of the GIF image, and 27-dimensional statistical features of CGCM
Aljarf	2013	Proposed a steganalysis system for both gray and color images based on four features which are contrast, energy, homogeneity, and correlation, using grey images for steganography has many limitations
Goljan, Fridrich, and Cogranne	2014	Proposed an extension of the spatial rich model (SRM), which was designed for color images, to allow for more accurate detection of steganography in color images.
Abdulrahman, Chaumont,	2015	Proposed steganalysis method for color images based on color feature correlation, they use two set of features. The first set consists of 18157 features.

Montesinos, and Magnier		The second set consisted of 3000 features obtained from the correlation of different R, G, B channel gradients
Abdulrahman, Chaumont, Montesinos, and Magnier	2016	Introduced steganalysis method extension to his previous work at color images. At this work, they use new features from color rich model and features from the first method in addition to two new features.
Abdulrahman, Chaumont, Montesinos, and Magnier	2016	Proposed extension to their previous works by developing another steganalysis system to enriches the features from the Color Rich Model by adding new features obtained by applying steerable Gaussian filters and then computing the co-occurrence of pixel pairs to increase the detection of hidden messages, the total number of features in this paper is 22,563 features
Al-jarf	2016	presented a detection system for color and gray images that combines three different steganalysis techniques. All technique address blind image steganalysis are relied on the extraction CGCM and histogram features
Rasool	2017	Proposed a new method in detection the existence of hidden messages that embedding in RGB

		images. He used statistical texture features set that consist of 26 features for each channel
AlTaie	2017	Proposed a statistical model for steganalysis to enhance the detection of the existence of hidden data inside 8-bit depth gray-scale BMP images that is based on an enhanced GLCM feature set
Al-Jarrah, Al-Taei, and Aboarqoub	2017	Proposed a steganalysis model for grayscale images with a new texture feature set, that is based on statistical texture features of images including gray level co-occurrence matrix (GLCM), Entropy, and additional statistical image features.

Chapter Three

Methodology and the Proposed Model

3.1 Methodology Approach

The methodology approach in this thesis is experimental. The proposed research develops a feature-based model to enhance the detection process of embedded data in uncompressed RGB images in the spatial domain, using various low payloads. The experimental work will use a public dataset of 10,000 uncompressed color images. The results are evaluated using detection accuracy metrics.

3.2 Outline of the Proposed Model

The proposed model is developed to enhance the detection accuracy of data hiding in uncompressed RGB cover images that have low embedding payload. We use a dataset of stego and clean images to train a binary classifier based on a selected set of statistical texture features that are extracted from the training and testing images.

3.3 Statistical Features Selection

The proposed model contains a hybrid feature set combining the basic GLCM model features, and an extension of the GLCM features, calculated for the individual color channels and combined into an image feature set, that is used in the evaluation of an RGB image to determine if it is a stego or a clean image. The combined feature set includes:

- **Basic GLCM Feature Set**

  The Basic GLCM features will be calculated for the three channels of an image separately, using a one-pixel horizontal distance, as each channel will be stored in a one-dimensional array.

  The basic GLCM feature set will be calculated for the full channels as well as parts of channels (Byte, 4-bit LSB, 3-bit LSB, 2-bit LSB, 1-bit LSB).

  Sample of feature vectors tables for basic GLCM (d=1) show in Appendix B.

  This feature includes four properties:

  1. Contrast
  2. Correlation
  3. Energy
  4. Homogeneity

The feature set for a single channel is shown in table 3.1

Table 3.1 List of the Selected Single Channel Features for Basic GLCM, Distance = 1

Feature Name	Feature Description
GLCM-Byte	Contrast, Correlation, Homogeneity Energy, of full bytes
GLCM-4bit LSB	Contrast, Correlation, Homogeneity Energy, of 4bit-LSB
GLCM-3bit LSB	Contrast, Correlation, Homogeneity Energy, of 3bit-LSB
GLCM-2bit LSB	Contrast, Correlation, Homogeneity Energy, of 2bit-LSB
GLCM-1bit LSB	Contrast, Correlation, Homogeneity Energy, of 1bit LSB
Total number of basic features per channel = 20	

- **Extended GLCM Feature Set**

  This feature set extends the distance between the reference pixel and the neighbor pixel to more than one. This work is proposing to use several distances and experimentally determines the extended set that provides better detection accuracy.

  Sample of feature vectors tables for extended GLCM (d=2, 3, 4, 5, 6) show in Appendix B.

  Table 3.2 shows the extended GLCM features for distances of 2, 3, 4, 5 and 6 pixels:

Table 3.2 List of the Selected Single Channel Features for Extended GLCM (Distance = 2, 3, 4, 5, 6)

Feature Name	Feature Description
GLCM-d2-Byte	Contrast, Correlation, Homogeneity Energy, of full bytes
GLCM-d2-4bit-LSB	Contrast, Correlation, Homogeneity Energy, of 4bit-LSB
GLCM-d2-3bit-LSB	Contrast, Correlation, Homogeneity Energy, of 3bit-LSB
GLCM-d2-2bit-LSB	Contrast, Correlation, Homogeneity Energy, of 2bit-LSB
GLCM-d2-1bit-LSB	Contrast, Correlation, Homogeneity Energy, of 1bit-LSB
GLCM-d3-Byte	Contrast, Correlation, Homogeneity Energy, of full bytes
GLCM-d3-4bit-LSB	Contrast, Correlation, Homogeneity Energy, of 4bit-LSB
GLCM-d3-3bit-LSB	Contrast, Correlation, Homogeneity Energy, of 3bit-LSB
GLCM-d3-2bit-LSB	Contrast, Correlation, Homogeneity Energy, of 2bit-LSB
GLCM-d3-1bit-LSB	Contrast, Correlation, Homogeneity Energy, of 1bit-LSB
GLCM-d4-Byte	Contrast, Correlation, Homogeneity Energy, of full bytes
GLCM-d4-4bit-LSB	Contrast, Correlation, Homogeneity Energy, of 4bit-LSB
GLCM-d4-3bit-LSB	Contrast, Correlation, Homogeneity Energy, of 3bit-LSB
GLCM-d4-2bit-LSB	Contrast, Correlation, Homogeneity Energy, of 2bit-LSB

GLCM-d4-1bit-LSB	Contrast, Correlation, Homogeneity Energy, of 1bit-LSB
GLCM-d5-Byte	Contrast, Correlation, Homogeneity Energy, of full bytes
GLCM-d5-4bit-LSB	Contrast, Correlation, Homogeneity Energy, of 4bit-LSB
GLCM-d5-3bit-LSB	Contrast, Correlation, Homogeneity Energy, of 3bit-LSB
GLCM-d5-2bit-LSB	Contrast, Correlation, Homogeneity Energy, of 2bit-LSB
GLCM-d5-1bit-LSB	Contrast, Correlation, Homogeneity Energy, of 1-bit-LSB
GLCM-d6-Byte	Contrast, Correlation, Homogeneity Energy, of full bytes
GLCM-d6-4bit-LSB	Contrast, Correlation, Homogeneity Energy, of 4bit-LSB
GLCM-d6-3bit-LSB	Contrast, Correlation, Homogeneity Energy, of 3bit-LSB
GLCM-d6-2bit-LSB	Contrast, Correlation, Homogeneity Energy, of 2bit-LSB
GLCM-d6-1bit-LSB	Contrast, Correlation, Homogeneity Energy, of 1bit-LSB
Total number of extended features per channel = 100	

3.4 The Embedding Method

In this thesis the secret data is generated randomly, using a common seed for all payloads, to be embedded in each color channel, by leaving gaps between the pixels of a channel to achieve various embedding payloads. The payload ratios that are selected in this work are {0.01 bpc, 0.05 bpc, 0.1 bpc, 0.2 bpc, 0.3 bpc, 0.4 bpc, 0.5 bpc}, where the differences between payload ratios are due to the gaps between locations of embedding pixels.

- For 0.01 payload the gap between pixels is 100 pixels.
- For 0.05 payload the gap between pixels is 20 pixels.
- For 0. 1 payload the gap between pixels is 10 pixels.
- For 0.2 payload the gap between pixels is 5 pixels.

- For 0.3 payload the gap between pixels is 3 for odd pixels and 4 for even pixels.
- For 0.4 payload the gap between pixels is 2 for odd pixels and 3 for even pixels.
- For 0.5 payload the gap between pixels is 2 pixels.

3.5 The SVM Classifier

The Support Vector Machine (SVM) method will be used in our model, where the training phase uses a labeled dataset of clean and stego images of equal numbers, and the testing phase uses unlabeled clean and stego images that are not part of the training set. Implementation of the proposed model will use the SVM classifier that is available in MATLAB. The SVM classifier has been chosen due to its excellent performance in similar binary classification applications (Prakash, 2006).

3.6 Experimental Work

The research is based on an experimental evaluation of the proposed model using a dataset of color cover images. The cover images are uncompressed in the format of PPM with 512x512 dimensions and 24-bit depth.

The selected cover images will be embedded with randomly generated secret messages. The clean images set and the stego images set will be used data to train the SVM classifier during the training phase. An additional test set of images (clean and stego), similar in format to the cover images, will be used in evaluating the detection accuracy of the proposed model.

3.7 Training and Testing Steps

- Dataset creation: a dataset of clean images will be embedded with the same sequence of random numbers for all embedding payloads to create the stego dataset. The payload ratio ranges from 0.01 bpc to 0.5 bpc.
- Feature extraction of training dataset: using the selected feature set, the clean and stego images will be processed and the features extracted, which will result in two feature datasets: clean images feature dataset and stego images feature dataset.
- Feature extraction of the testing dataset: A set of unlabeled clean and stego images will be used, whose features will be extracted.
- Training and classification: The classifier (SVM) will be given the training feature set, labels of the training dataset, and the testing feature set. Based on the training data the classifier will evaluate each testing image as clean or stego.
- The accuracy of the proposed model will be evaluated based on the testing set evaluation metrics.

3.8 The Proposed Model

- Embedding flowchart for one image: to embed random secret data in a clean image to get a stego image for each payload ratios, as shown in figure 3.1.
- This module is executed for all clean images to get the stego images dataset.

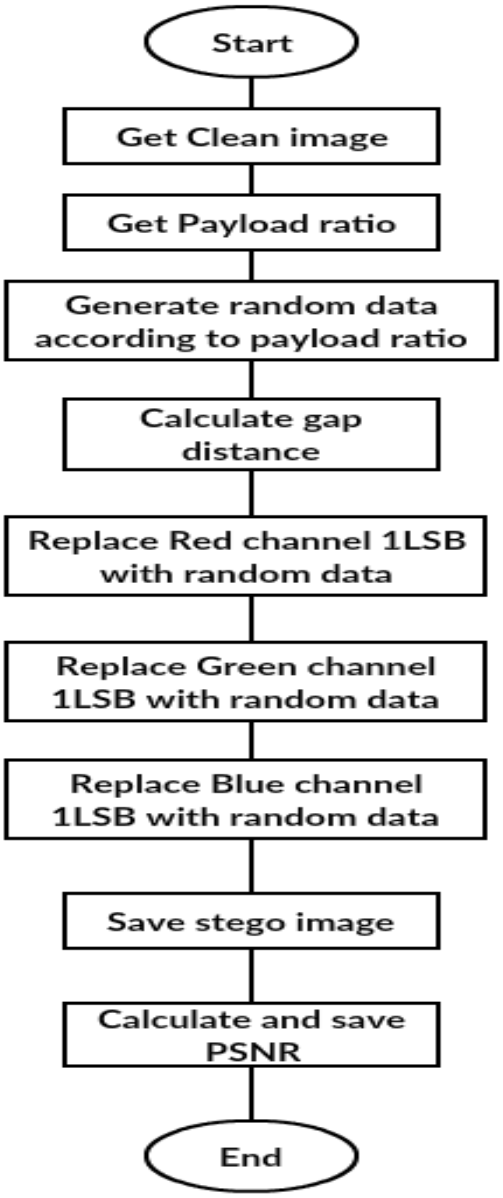


Figure 3.1 Flowchart of Batch Embedding

- Feature extraction flowchart: to extract the features from clean and stego images and store feature vectors in a CSV file, as shown in figure 3.2.

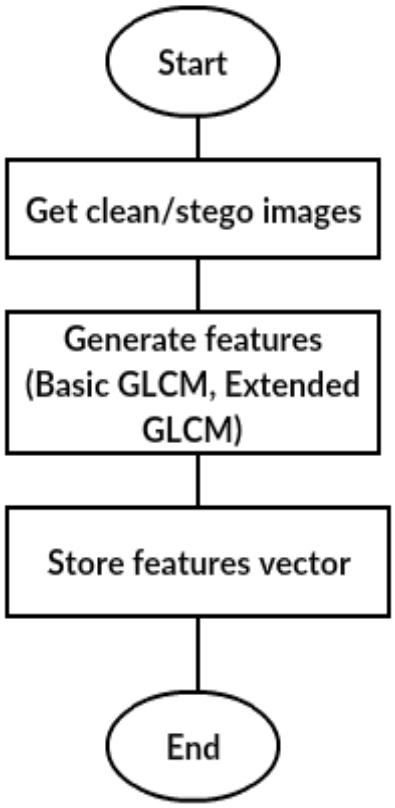


Figure 3.2 Flowchart of Feature Extraction

- Classifier flowchart: to train the SVM classifier using training and label datasets, then using testing dataset to get the results of classification, as shown in figure 3.3.

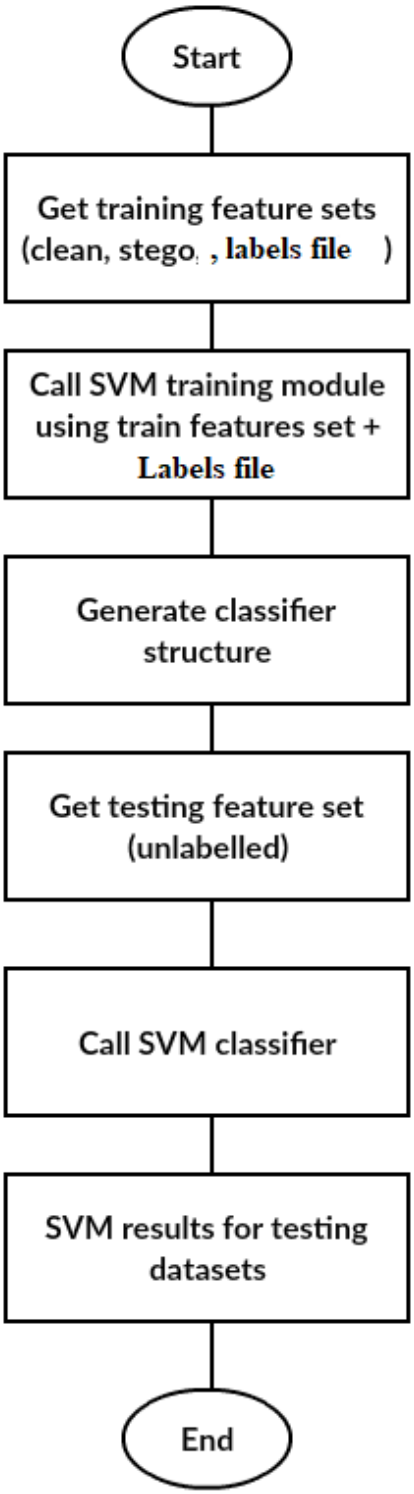


Figure 3.3 Flowchart of Classifier

3.9 Evaluation Metrics

The following metrics will be used in evaluating the detection performance of the proposed model:

- True Negative (TN): The ratio of true negative detections to the number of clean images.
- True Positive (TP): The ratio of true positive detections to the number of stego images.
- False Negative (FN): The ratio of false negative detection to the number of stego images.
- False Positive (FP): The ratio of false positive detection to the number of clean images.
- Detection Accuracy: The ratio of correctly detected clean and stego images to the total number of clean and stego images (James, Witten, Hastie, & Tibshirani, 2013):

  Accuracy = (TN + TP) / (TN + TP + FP + FN) * 100%.

Chapter Four

Experimental Results and Discussion

4.1 Introduction

In this work, the proposed system used a one-bit spatial steganography model to create stego images, the 1-LSB model which embeds 1 bit per channel (1 bpc). The proposed model was implemented in MATLAB as a working system with a title "PPM Steganalysis Work". The experimental work included the embedding of randomly generated secret data with seed, in three channels (Red, Green, Blue) of the clean dataset to generate the stego dataset, then extract features of the clean and stego images and use the extracted feature sets for training and testing to determine the detection accuracy of the proposed model.

4.2 Objectives of the Experimental Work

The experimental work is aimed to evaluate the detection accuracy using the proposed model at various payloads ratios. The following objectives are considered:

1. Selection of the clean images dataset.
2. Creation of the stego images datasets at payload ratios from 0.01 to 0.5 bit per channel.
3. Extraction of the feature vectors for the clean images dataset and the seven stego datasets.
4. Classification and detection accuracy calculation for each payloads ratio.

4.3 The Image Dataset

The selected clean cover image type is the uncompressed PPM-RGB in three channels. The selected public dataset contains 10,000 PPM color images that were downloaded from LIRMMBase website: (http://www.lirmm.fr/~chaumont/PPG-LIRMM-COLOR.html).

Abdulrahman, Chaumont, Montesinos and Magnier (2016) collected row images from two subsets, the Dresden Image Databases, and the Break Our Steganographic System database (BOSSbase), Afterwards, the RGB color images were obtained by using the Patterned Pixel Grouping (PPM) demosaicking algorithm named "dcraw", then from each color RGB image, they randomly extracted five cropped images measuring 512x512, the final number of RGB cropped images are 10,000.

Samples of the LIRMM PPM color database images show in Appendix C.

Image number 10000.ppm

Figure 4.1 Sample of Cover Images

4.4 Experimental Work

The experimental system consists of three main modules:

1. Batch Embedding
2. Feature Extraction
3. Classification

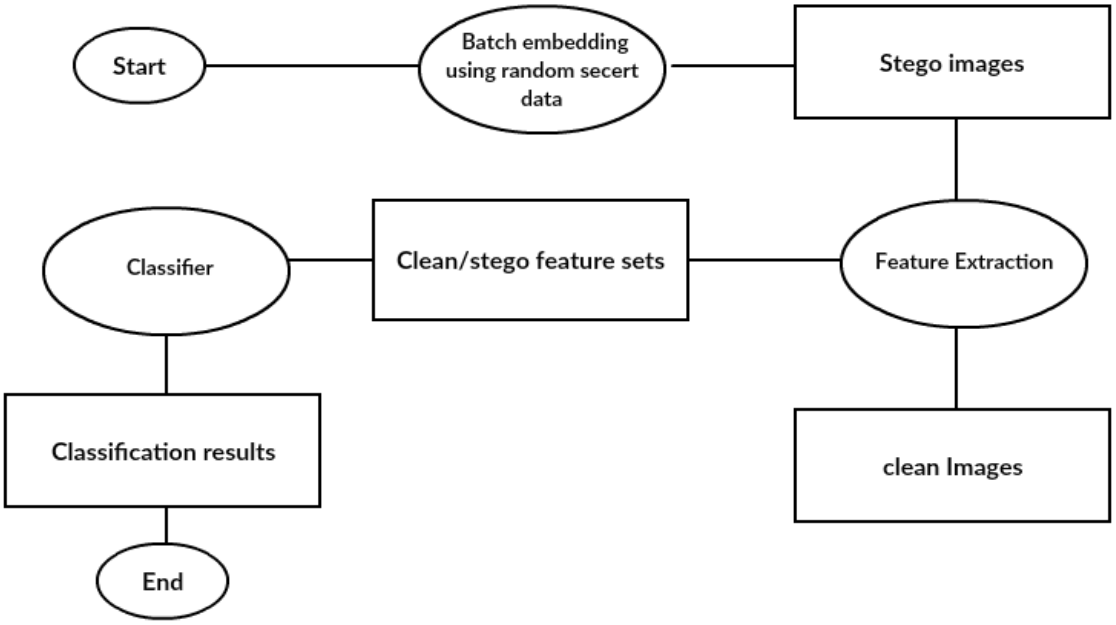


Figure 4.2 Modules of the Experimental Work

4.4.1 Batch Embedding

In this module the clean images are embedded with a randomly generated data, using a fixed seed for all images and payloads, with various gaps between pixels according to the size of payloads. The embedding steganography model is the spatial domain 1-LSB method for each channel.

The Peak-Signal-to-Noise Ratio (PSNR) also calculated which is a standard metric to compare imperceptibility because of embedding. The signal, in this case, is the original data, and the noise is the error introduced by embedding.

4.4.2 Feature Extraction

In this module, the selected features are extracted from each channel, and the features of the three channels of an image are combined into one feature vector. The feature vector for the image dataset is saved in CSV format for processing by the classifier. The extracted features are based on the basic GLCM and the extended GLCM, which consist of four properties (Contrast, Correlation, Energy, Homogeneity). The features are extracted from five fragments the byte of each channel as well sub-bytes as well as the right half byte (RHB), 3LSB, 2LSB, and 1LSB. For each object, the basic GLCM and extended GLCM features are calculated using MATLAB functions, resulting in 120 feature set elements in the feature vector.

4.4.3 Classification

In this module, the feature vectors of the training and testing images are analyzed, to determine for each image whether it is a stego or a clean image. The clean feature dataset and the stego feature dataset were divided into two sets, the first one for training which contained 60%, and the second one for testing which contained 40%.

The SVM classifier has two stages; the first stage takes the training dataset and its labels (clean or stego) to train it to recognize the clean images from the stego images, and the second stage in which the unlabeled testing feature dataset is analyzed, based on the previous training, to determine the category of a test image (clean or stego). The detection accuracy is calculated for each payload based on the outcome of classification.

For each payload ratio, the training feature dataset consists of features of 6000 images (3000 clean and 3000 stego), while the testing feature dataset consists of features of 4000 images (2000 clean and 2000 stego).

4.5 Results and Discussion

This section presents the experimental results of the Steganalysis of the PPM images that were embedded with random data using the proposed feature sets. The results are compared with the work of Abdulrahman, Chaumont, Montesinos and Magnier (2016).

4.5.1 Summary of the LIRMM Work

Abdulrahman, et al. (2016) developed a steganalysis system to enrich the features of the Color Rich Model by adding new features obtained by applying Steerable Gaussian Filters and then computing the co-occurrence of pixel pairs to increase the detection of hidden messages. The total number of features are 22,563 features. They used a public dataset that consists of 5000 PPM color images, which are randomly chosen from the 10,000. The steganographic algorithms that were used to hide messages in cover image are S-UNIWARD, WOW, and Synch-HILL.

The payloads ratios ranged from 0.01 to 0.5. The Synch-HILL steganography model achieved the highest detection accuracy for the payloads 0.01, 0.05, 0.1, 0.2, 0.3. The S-UNIWARD steganography model achieved the highest detection accuracy for the payloads 0.4 and 0.5.

Table 4.1 shows the results using three different embedding models, where the combined detection accuracy is shown without true positive and true negative values.

Table4.1: Accuracy Results of LIRMM Work

Steganography Model	Payload (bpc)	Accuracy %
S-UNIWARD	0.01	53.63
	0.05	61.65
	0.1	70.61
	0.2	78.36
	0.3	84.41
	0.4	87.98
	0.5	88.83
WOW	0.01	53.13
	0.05	61.46
	0.1	69.09
	0.2	77.31
	0.3	83.15
	0.4	86.23
	0.5	87.94
Synch-HILL	0.01	53.49
	0.05	63.53
	0.1	70.54
	0.2	78.87
	0.3	84.64
	0.4	87.06
	0.5	88.75

4.5.2 Analysis of Results of the Proposed Model

Steganalysis of the selected dataset using the proposed model, has resulted in detection accuracy according to each payload ratio (0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5). The embedding was on 10,000 clean images with various embedding payloads of random data according to the 1LSB steganography model, with various gaps. The experimental work was performed on 5000 clean images and 5000 stego images of each payload, that were randomly selected from the 10,000 clean datasets and the generated 10,000 stego dataset for each payload.

Table 4.4 shows the results of the proposed model, where each payload achieves an approximately similar or higher detection accuracy than the LIRMM work. The proposed model calculates the values true negative (TN) and true positive (TP) error metrics, then the average of the two values gave the final detection accuracy.

Table 4.2 shows the accuracy results of our work for each payload ratios compared with LIRMM work.

The accuracy results for each payload ratio (TN, TP, FN, FP, Detection Accuracy) using the SVM classifier shown in Appendix A.

Table 4.2: Accuracy Results of our Work

Payload	TN%	TP%	Accuracy%	LIRMM work Accuracy%
0.01	55.70%	56.65%	56.18%	53.49%
0.05	62.80%	64.40%	63.60%	63.53%
0.1	64.35%	69.90%	67.13%	70.54%
0.2	75.65%	80.20%	77.93%	78.87%
0.3	79.20%	83.15%	81.18%	84.64%
0.4	85.85%	90.35%	88.10%	87.98%
0.5	88.20%	93.80%	91.00%	88.83%

4.5.3 PSNR Results

During the batch embedding process, the PSNR value is calculated for the image, and the average PSNR for each payload ratio is calculated. The average PSNR value shows a consistent decrease as the payload increase. Table 4.3 shows the average PSNR for each stego dataset, according to the payload ratio.

The table below shows the average effect of payload on PSNR for all payloads ratios.

Table 4.3: The Effect of Payload on PSNR

Payload	Average PSNR
0.01 bpc	71.14154
0.05 bpc	64.1515
0.1 bpc	61.14127
0.2 bpc	58.1305
0.3 bpc	56.58175
0.4 bpc	55.12045
0.5 bpc	54.15135

Chapter Five

Conclusion and Future Work

5.1 Conclusion

This thesis presented a steganalysis model to detect the existence of low payload hidden data inside RGB color images, using statistical texture features of the clean and stego images. The selected feature sets were extracted from datasets of clean and stego images and classified using the Support Vector Machine algorithm. The focus of this work was on enhancing the effectiveness of the GLCM function by using the basic GLCM and the extended GLCM as a feature set to enrich the detection process in three channels. A public dataset of 5000 clean images and 5000 stego images for seven payload ratios (0.01 to 0.5 bpc) were used, where 6000 clean and stego images were used for training, while testing was performed on 4000 images clean and stego images, for each payload ratio. The steganography algorithm that was used is the spatial domain 1LSB method, where random data embedded in 1LSB of each channel, and the embedding was spaced by gaps according to payload ratios. The steganalysis results showed that the detection accuracy values ranged from 56.18% for 0.01 bpc to 91.00% for 0.5 bpc, where the detection accuracy increased with the payload ratio. Comparison with the related work which analyzed the same dataset showed higher detection accuracy in our work although smaller feature set was used in our work. Also, the results showed that the model achieved higher true positive detection than true negative, which indicates that the model is more effective in detecting stego images, which is the purpose of steganalysis model.

5.2 Suggestion for Future Work

In the research fields, there is not complete research, but each research work can provide new ideas for another work. Based on the outcome of the present research, the following ideas are suggested for future work:

- The images can be split horizontally into partitions or segments where the partitions are analyzed using the selected features. An image is considered as stego if any of its partitions is detected as such.
- Enhancing the feature set by adding inter-channel correlation features.
- Using other machine learning models to enhance the detection accuracy.

References

Abdulrahman, H., Chaumont, M., Montesinos, P., & Magnier, B. (2015). Color image steganalysis using correlations between RGB channels. ***Availability Reliability and Security (ARES), 10th International Conference*** on (pp. 448-454). IEEE.

Abdulrahman, H., Chaumont, M., Montesinos, P., & Magnier, B. (2016). Color images steganalysis using rgb channel geometric transformation measures ***Wiley Journal on Security and Communication Networks (SCN).***

Abdulrahman, H., Chaumont, M., Montesinos, P., & Magnier, B. (2016). Color image steganalysis based on steerable gaussian filters bank. ***Proceedings of the 4th ACM workshop on Information Hiding and Multimedia Security.***

Al-Jarrah, M., Al-Taie, Z.H., Abuarqoub, A. (2017). Steganalysis using LSB-focused statistical features. **In Proceedings of ICFNDS**, Cambridge, United Kingdom.

Aljarf, A., Amin, S., Filippas, J., & Shuttelworth, J. (2013). Develop a detection system for grey and colour stego images. ***International Journal of Modeling and Optimization***, 3(5), 458.

Al-jarf, A (2016). **Development of a detection system for colour steganographic images based on extraction of colour gradient co-occurrence matrix features and histogram of different image** (Doctoral dissertation), Coventry University.

Andersen, R.J. & Petitcolas, F.A.P. (1998). "On the limits of steganography, ***IEEE Journal of Selected Areas in Communications***, No.4.

Al-Taie, Z.H. (2017). **Statistical steganalysis detector model for 8-bit depth images** (Unpublished master thesis), Middle East University, Amman, Jordan.

Badr, S., Salama, G., Selim, G., & Khalil, A. (2014). A Review on Steganalysis Techniques: From Image Format Point of View. ***International Journal of Computer Applications***, No. 4.

Bohme, R. (2010). **Advanced statistical steganalysis.** Berlin Heidelberg: Springer-Verlag.

Cheddad, A., Condell, J., Curran, K., & Kevitt, P.M. (2010). Digital Image Steganography: Survey and Analysis of Current Methods, ***Signal Processing***, Volume 90, Issue 3.

Das, S., Das, S., Bandyopadhyay, B., & Sanyal, S. (2008). Steganography and Steganalysis: Different Approaches, ***International Journal of***

Computers Information Technology and Engineering (IJCITAE), Vol. 2, No 1.

Fridrich, J., Goljan, M., & Du, R. (2001). Reliable Detection of LSB Steganography in Color and Grayscale Images**.** ***Proceedings*** **of the *workshop* on *Multimedia* and *security***, Ottawa, Ontario, Canada.

Fridrich, J. (2009). **Steganography in digital media: principles, algorithms, and applications**, (1st edition). Cambridge University Press.

Gaur, R., & Chouhan, V.S. (2017). Classifiers in Image processing, **International *Journal on Future Revolution in Computer Science & Communication Engineering*,** Vol.3 (6).

Goljan, M., Fridrich, J., & Cogranne, R. (2014). Rich Model for Steganalysis of Color Images. ***IEEE International Workshop on Information Forensics and Security (WIFS)***, Atlanta, GA, USA.

Gong, R., & Wang, H. (2012) Steganalysis for GIF images based on colors-gradient cooccurrence matrix. ***Optics Communications***, (vol. 285), no. 24, pp. 4961-4965.

Hall-Beyer, M. (2017). GLCM texture: tutorial. ***Arts Research and Publications***, V (3.0).

Haralick, R.M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification, ***IEEE Trans. on Systems, Man and Cybernetics***.

Joshi, S.V., Bokil, A.A., Jain, N.A., & Koshti, D. (2012). Image steganography combination of spatial and frequency domain, **International *Journal of Computer Applications,*** (0975 – 8887) Volume 53– No.5.

Johnson, N. F., & Jajodia, S. (1998). Steganography: Seeing the Unseen, ***IEEE Computer***, pp. 26-34.

Kaur, M., & Kaur, G. (2014). Review of Various Steganalysis Techniques, ***International Journal of Computer Science and Information Technologies***, Vol.5 (2).

Kekre, H.B., Athawale, A.A., & Patki, S.A. (2011). Steganalysis of LSB Embedded Images Using Gray Level Co-Occurrence Matrix. ***International Journal of Image Processing (IJIP)***, Volume (5), Issue (1).

Ker, A.D., Pevny. T. (2012). Batch Steganography in the real world. ***Proceedings of the on Multimedia and Security***, PP. 1-10.

Li, B., He, J., Huang, J., Shi, Y.Q. (2011). A Survey on Image Steganography and Steganalysis, ***Journal of Information Hiding and Multimedia Signal Processing***, Volume 2, Number 2.

Nosrati, M., Karimi, R., Hariri, M. (2011). An introduction to steganography methods, ***World Applied Programming***, Vol (1), No (3).

Prakash, G.S. (2006). **Measures for classification and detection in steganalysis** (Doctoral dissertation*)*, Indian Institute of Science, Bangalore.

PPM LIRMM Color Database Images, http://www.lirmm.fr/~chaumont/PPG-LIRMM-COLOR.html.

Provos, N. and Honeyman, P., (2003) Hide and seek: An introduction to steganography. ***IEEE Security & Privacy,*** 1 (3), 32-44.

Rasool, Z.I. (2017). **The detection of data hiding in RGB images using statistical steganalysis** (Unpublished master thesis), Middle East University, Amman, Jordan.

Sharma, K., Priyanka., Kalsh, A., Saini, K. (2015). GLCM and its Features, ***International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)***, Volume 4, Issue 8.

Siper, A., Farley, R., & Lombardo, C. (2005). The Rise of Steganography, ***Proceedings of Student/Faculty Research Day***, CSIS, Pace University.

Appendix A

Detection Accuracy Results for Each Payload Ratio Using the SVM Classifier

Table A1.1 Detection Accuracy Results for 0.01 BPC

Metric	Result
FN	43.35%
FP	44.30%
TN	55.70%
TP	56.65%
Detection Accuracy	56.18%

Table A1.2 Detection Accuracy Results for 0.05 BPC

Metric	Result
FN	35.60%
FP	37.20%
TN	62.80%
TP	64.40%
Detection Accuracy	63.60%

Table A1.3 Detection Accuracy Results for 0. 1 BPC

Metric	Result
FN	30.10%
FP	35.65%
TN	64.35%
TP	69.90%
Detection Accuracy	67.13%

Table A1.4 Detection Accuracy Results for 0.2 BPC

Metric	Result
FN	19.80%
FP	24.35%
TN	75.65%
TP	80.20%
Detection Accuracy	77.93%

Table A1.5 Detection Accuracy Results for 0.3 BPC

Metric	Result
FN	16.85%
FP	20.80%
TN	79.20%
TP	83.15%
Detection Accuracy	81.18%

Table A1.6 Detection Accuracy Results for 0.4 BPC

Metric	Result
FN	9.65%
FP	14.15%
TN	85.85%
TP	90.35%
Detection Accuracy	88.10%

Table A1.7 Detection Accuracy Results for 0.5 BPC

Metric	Result
FN	6.20%
FP	11.80%
TN	88.20%
TP	93.80%
Detection Accuracy	91.00%

Appendix B

Samples of Basic GLCM and Extended GLCM Features

- B1: Basic GLCM features for the first 10 images (Distance= 1)

Table B1.1 Basic GLCM Features (d=1)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Basic GLCM-Byte				Basic GLCM-4bit-LSB				Basic GLCM-3bit-LSB				Basic GLCM-2bit-LSB				Basic GLCM-1bit-LSB			
2	0.660403	0.869876	0.096221	0.793789	5.518091	0.009071	0.786423	0.901463	10.46925	0.003243	0.617638	0.813049	18.13488	0.003094	0.395626	0.676163	24.32177	0.006974	0.250164	0.565683
3	0.239911	0.935544	0.147887	0.901748	5.48538	0.043418	0.783558	0.902047	10.45748	0.017481	0.614914	0.813259	18.11413	0.010454	0.393403	0.676533	24.01653	0.019731	0.250099	0.571133
4	0.422483	0.947523	0.084929	0.835162	5.678282	0.009666	0.780531	0.898602	10.70907	0.000331	0.610588	0.808767	18.33937	0.000267	0.391435	0.672511	24.37523	0.005092	0.250007	0.564728
5	0.123635	0.988582	0.341271	0.978339	4.013191	0.24898	0.815752	0.928336	8.196854	0.193294	0.653335	0.853628	16.58867	0.096459	0.401382	0.703774	24.01447	0.019602	0.250208	0.57117
6	0.144433	0.970481	0.489472	0.965014	3.224759	0.113401	0.864291	0.942415	7.342252	0.090185	0.707916	0.868888	17.22495	0.025249	0.411408	0.692412	24.28457	0.00874	0.250046	0.566347
7	0.153412	0.93997	0.284049	0.939652	4.917514	0.149767	0.791679	0.912187	9.577639	0.10307	0.62482	0.828971	17.31168	0.054228	0.397964	0.690863	23.06734	0.058471	0.250857	0.588083
8	0.279393	0.93538	0.141732	0.907554	5.216588	0.085805	0.78842	0.906847	10.22121	0.047467	0.615925	0.817478	18.2474	0.008664	0.390632	0.674154	24.43561	0.002621	0.250005	0.56365
9	0.074997	0.99211	0.174969	0.964991	4.718257	0.117745	0.803839	0.915745	9.586798	0.067708	0.632771	0.828807	17.3556	0.031808	0.405427	0.690079	23.03015	0.058832	0.251517	0.588747
10	0.162602	0.924562	0.509457	0.934076	5.677161	0.063192	0.773887	0.898622	10.16588	0.042499	0.6189	0.818466	17.98983	0.021795	0.392333	0.678753	23.90699	0.024194	0.250152	0.573089

- B2: Extended GLCM features for the first 10 images (Distance= 2)

Table B2.1 Extended GLCM Features (d=2)

	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO
1	GLCM d2-Byte				GLCM d2-4bit-LSB				GLCM d2-3bit-LSB				GLCM d2-2bit-LSB				GLCM d2-1bit-LSB			
2	-0.00157	0.250152	0.561946	24.45187	0.001663	0.250152	0.56336	24.51365	-0.00086	0.250151	0.562256	24.52365	-0.00127	0.250151	0.562078	24.44822	0.001812	0.250152	0.563425	18.12747
3	-0.00072	0.250002	0.562186	24.48402	0.000649	0.250002	0.562785	24.53346	-0.00137	0.250002	0.561903	24.52832	-0.00116	0.250002	0.561994	24.50168	-7.17E-05	0.250002	0.56247	18.34374
4	-0.00059	0.25	0.562243	24.47748	0.000919	0.25	0.562902	24.55458	-0.00223	0.250001	0.561525	24.46944	0.001247	0.250001	0.563046	24.50729	-0.0003	0.25	0.56237	18.38953
5	0.013746	0.250158	0.568608	24.08251	0.016825	0.250182	0.569955	24.16204	0.013578	0.250157	0.568535	24.15111	0.014024	0.25016	0.56873	24.13325	0.014753	0.250166	0.569049	17.30595
6	0.004876	0.250033	0.564656	24.35074	0.006039	0.250036	0.565165	24.37532	0.005036	0.250033	0.564726	24.40869	0.003674	0.25003	0.564131	24.38485	0.004647	0.250032	0.564556	17.47511
7	-0.00016	0.250002	0.562433	24.47579	0.000983	0.250003	0.562932	24.42822	0.002925	0.250005	0.563782	24.5014	-6.19E-05	0.250002	0.562475	24.4772	0.000926	0.250003	0.562907	18.22467
8	0.00058	0.250004	0.562757	24.48084	0.000775	0.250004	0.562842	24.54561	-0.00187	0.250004	0.561686	24.44084	0.002408	0.250005	0.563556	24.4542	0.001862	0.250004	0.563318	18.29962
9	0.030581	0.250869	0.576402	23.84624	0.025482	0.250795	0.574174	23.79605	0.027533	0.250823	0.575071	23.91895	0.02251	0.250758	0.572876	23.75941	0.029031	0.250845	0.575725	17.70933
10	0.000386	0.250006	0.562674	24.48757	0.000496	0.250006	0.562722	24.45626	0.001774	0.250006	0.563281	24.53785	-0.00156	0.250006	0.561824	24.53944	-0.00162	0.250006	0.561796	18.39514

- B3: Extended GLCM features for the first 10 images (Distance= 3)

Table B3.1 Extended GLCM Features (d=3)

	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH
1	GLCM d3-Byte				GLCM d3-4bit-LSB				GLCM d3-3bit-LSB				GLCM d3-2bit-LSB				GLCM d3-1bit-LSB			
2	18.12747	0.003504	0.395664	0.676295	18.20006	-0.00048	0.39528	0.674999	18.16256	0.001575	0.395479	0.675669	18.20357	-0.00068	0.395264	0.674936	18.19541	-0.00023	0.395306	0.675082
3	18.34374	-0.00209	0.392202	0.672433	18.24511	0.003304	0.392711	0.674194	18.32033	-0.0008	0.392321	0.672851	18.33703	-0.00172	0.392236	0.672553	18.33972	-0.00186	0.392221	0.672505
4	18.38953	-0.00247	0.391177	0.671615	18.34699	-0.00014	0.391394	0.672375	18.4164	-0.00393	0.391041	0.671136	18.34582	-7.98E-05	0.3914	0.672396	18.32196	0.001223	0.391521	0.672822
5	17.30595	0.057388	0.39687	0.690965	17.53817	0.044742	0.3955	0.686818	17.85246	0.027621	0.39372	0.681206	18.05889	0.01638	0.392593	0.67752	18.29673	0.003428	0.39134	0.673273
6	17.47511	0.011095	0.409917	0.687944	17.49854	0.009766	0.409782	0.687526	17.58273	0.005005	0.409293	0.686023	17.57644	0.005352	0.409332	0.686135	17.59931	0.00406	0.409199	0.685727
7	18.22467	0.004352	0.392844	0.674559	18.23558	0.003758	0.392786	0.674365	18.25004	0.002966	0.392711	0.674106	18.2632	0.00225	0.392642	0.673871	18.27822	0.001431	0.392563	0.673603
8	18.29962	0.005824	0.390362	0.673221	18.3782	0.001558	0.389958	0.671818	18.50837	-0.00551	0.3893	0.669493	18.40974	-0.00015	0.389796	0.671255	18.42664	-0.00107	0.389709	0.670953
9	17.70933	0.012078	0.403373	0.683762	17.75911	0.009303	0.403091	0.682873	17.73096	0.010871	0.403251	0.683376	17.72729	0.011078	0.40327	0.683441	17.74006	0.010368	0.403197	0.683213
10	18.39514	-0.00024	0.390202	0.671515	18.32343	0.00366	0.390568	0.672796	18.37846	0.00067	0.390285	0.671813	18.41498	-0.00132	0.390101	0.671161	18.36794	0.001242	0.390339	0.672001

- B4: Extended GLCM features for the first 10 images (Distance= 4)

Table B4.1 Extended GLCM Features (d=4)

	BI	BJ	BK	BL	BM	BN	BO	BP	BQ	BR	BS	BT	BU	BV	BW	BX	BY	BZ	CA	CB
1	GLCM d4-Byte				GLCM d4-4bit-LSB				GLCM d4-3bit-LSB				GLCM d4-2bit-LSB				GLCM d4-1bit-LSB			
2	18.12747	0.003504	0.395664	0.676295	18.20006	-0.00048	0.39528	0.674999	18.16256	0.001575	0.395479	0.675669	18.20357	-0.00068	0.395264	0.674936	18.19541	-0.00023	0.395306	0.675082
3	18.34374	-0.00209	0.392202	0.672433	18.24511	0.003304	0.392711	0.674194	18.32033	-0.0008	0.392321	0.672851	18.33703	-0.00172	0.392236	0.672553	18.33972	-0.00186	0.392221	0.672505
4	18.38953	-0.00247	0.391177	0.671615	18.34699	-0.00014	0.391394	0.672375	18.4164	-0.00393	0.391041	0.671136	18.34582	-7.98E-05	0.3914	0.672396	18.32196	0.001223	0.391521	0.672822
5	17.30595	0.057388	0.39687	0.690965	17.53817	0.044742	0.3955	0.686818	17.85246	0.027621	0.39372	0.681206	18.05889	0.01638	0.392593	0.67752	18.29673	0.003428	0.39134	0.673273
6	17.47511	0.011095	0.409917	0.687944	17.49854	0.009766	0.409782	0.687526	17.58273	0.005005	0.409293	0.686023	17.57644	0.005352	0.409332	0.686135	17.59931	0.00406	0.409199	0.685727
7	18.22467	0.004352	0.392844	0.674559	18.23558	0.003758	0.392786	0.674365	18.25004	0.002966	0.392711	0.674106	18.2632	0.00225	0.392642	0.673871	18.27822	0.001431	0.392563	0.673603
8	18.29962	0.005824	0.390362	0.673221	18.3782	0.001558	0.389958	0.671818	18.50837	-0.00551	0.3893	0.669493	18.40974	-0.00015	0.389796	0.671255	18.42664	-0.00107	0.389709	0.670953
9	17.70933	0.012078	0.403373	0.683762	17.75911	0.009303	0.403091	0.682873	17.73096	0.010871	0.403251	0.683376	17.72729	0.011078	0.40327	0.683441	17.74006	0.010368	0.403197	0.683213
10	18.39514	-0.00024	0.390202	0.671515	18.32343	0.00366	0.390568	0.672796	18.37846	0.00067	0.390285	0.671813	18.41498	-0.00132	0.390101	0.671161	18.36794	0.001242	0.390339	0.672001

- B5: Extended GLCM features for the first 10 images (Distance= 5)

Table B5.1 Extended GLCM Features (d=5)

	CC	CD	CE	CF	CG	CH	CI	CJ	CK	CL	CM	CN	CO	CP	CQ	CR	CS	CT	CU	CV
1	GLCM d5-Byte				GLCM d5-4bit-LSB				GLCM d5-3bit-LSB				GLCM d5-2bit-LSB				GLCM d5-1bit-LSB			
2	5.556618	0.002156	0.785814	0.900775	5.56449	0.000746	0.785689	0.900634	5.553296	0.002759	0.785866	0.900834	5.55799	0.001891	0.785794	0.90075	5.557638	0.001957	0.7858	0.900756
3	5.686529	0.008344	0.780388	0.898455	5.703	0.005475	0.78013	0.898161	5.734051	6.40E-05	0.779644	0.897606	5.742484	-0.00143	0.779515	0.897456	5.723066	0.001958	0.779818	0.897802
4	5.721296	0.002167	0.779857	0.897834	5.733655	1.54E-05	0.779664	0.897613	5.739284	-0.00096	0.779575	0.897513	5.747905	-0.00246	0.77944	0.897359	5.731477	0.000406	0.779696	0.897652
5	4.253027	0.204076	0.811686	0.924053	4.345757	0.186725	0.810125	0.922397	4.443534	0.168404	0.808491	0.920651	4.523928	0.153362	0.80715	0.919216	4.617408	0.13587	0.805598	0.917546
6	GLCM d5-4	0.074652	0.861801	0.939898	3.375817	0.071877	0.861623	0.939718	3.419944	0.059749	0.860847	0.93893	3.443696	0.053177	0.860433	0.938505	3.45156	0.051018	0.860295	0.938365
7	5.465587	0.05501	0.782863	0.9024	5.560939	0.038527	0.781355	0.900698	5.615167	0.029154	0.7805	0.899729	5.630517	0.026504	0.780259	0.899455	5.61035	0.029994	0.780575	0.899815
8	5.355677	0.061433	0.786193	0.904363	5.436074	0.047347	0.784913	0.902927	5.498901	0.03634	0.783917	0.901805	5.537429	0.029592	0.783307	0.901117	5.601004	0.018454	0.782304	0.899982
9	5.136979	0.039453	0.797012	0.908268	5.238311	0.020508	0.795382	0.906459	5.289548	0.010931	0.79456	0.905544	5.277605	0.013168	0.794751	0.905757	5.293326	0.010232	0.794499	0.905476
10	5.956443	0.01711	0.769541	0.893635	5.99572	0.010632	0.768934	0.892934	6.010323	0.008226	0.768709	0.892673	6.036141	0.003969	0.768311	0.892212	6.046258	0.002303	0.768155	0.892031

- B6: Extended GLCM features for the first 10 images (Distance= 6)

Table B6.1 Extended GLCM Features (d=6)

	CW	CX	CY	CZ	DA	DB	DC	DD	DE	DF	DG	DH	DI	DJ	DK	DL	DM	DN	DO	DP
1	GLCM d6-Byte				GLCM d6-4bit-LSB				GLCM d6-3bit-LSB				GLCM d6-2bit-LSB				GLCM d6-1bit-LSB			
2	1.701482	0.664746	0.069455	0.692938	2.466817	0.513946	0.06123	0.648568	2.793069	0.449664	0.057244	0.625858	2.96693	0.415404	0.054765	0.611441	3.164154	0.376543	0.052671	0.597506
3	0.555714	0.8507	0.114071	0.833112	0.843916	0.773271	0.09608	0.787145	1.059243	0.715421	0.085011	0.753664	1.238709	0.667207	0.077303	0.727157	1.405969	0.622271	0.071521	0.704892
4	0.983486	0.877839	0.062748	0.738815	1.359989	0.831072	0.05535	0.694685	1.487434	0.81524	0.052011	0.673046	1.547954	0.807721	0.050098	0.661158	1.633456	0.7971	0.048955	0.653516
5	0.263804	0.975637	0.333487	0.969157	0.387753	0.96419	0.328994	0.962924	0.511849	0.95273	0.32551	0.957465	0.63432	0.94142	0.321936	0.951926	0.755106	0.930265	0.317946	0.946141
6	0.322177	0.934153	0.472939	0.946917	0.48274	0.901337	0.459941	0.932281	0.640925	0.869008	0.448199	0.918695	0.79229	0.838072	0.438136	0.907164	0.933424	0.809227	0.428977	0.896475
7	0.338763	0.86744	0.249187	0.900346	0.468679	0.8166	0.235365	0.882193	0.536637	0.790004	0.230799	0.873891	0.593155	0.767888	0.228084	0.868111	0.663544	0.740344	0.223139	0.85953
8	0.488796	0.886948	0.12898	0.876832	0.618419	0.856967	0.123876	0.862023	0.760304	0.82415	0.11903	0.847487	0.893629	0.793313	0.114681	0.834613	1.014721	0.765304	0.111121	0.823607
9	0.136491	0.985641	0.157163	0.937291	0.17808	0.981266	0.147443	0.920219	0.205921	0.978337	0.142083	0.909759	0.230298	0.975772	0.137888	0.901039	0.254446	0.973232	0.134018	0.892662
10	0.375369	0.825848	0.464661	0.887964	0.565978	0.737415	0.438104	0.860595	0.69456	0.677759	0.423006	0.844249	0.779262	0.638463	0.413214	0.833277	0.840405	0.610097	0.405482	0.824653

Appendix C

LIRMM PPM Color Database Images