COVRE OPTIMIZATION FOR IMAGE STEGANOGRAPHY BY USING IMAGE FEATURES

ZAID NIDHAL KHUDHAIR

A dissertation submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Science (Computer Science)

Faculty of Computing
Universiti Teknologi Malaysia

JULY 2013

This dissertation is dedicated to my beloved parents, wife, daughter, brother, sisters and my friends.

ACKNOWLEDGEMEN

All praise is to Allah and my peace and blessing of Allah be upon our prophet, Muhammad and upon all his family and companions. In particular, I wish to express my sincere appreciation to my thesis supervisor, **PROF. DR. DZULKIFLI MOHAMAD**, and to all my friends for encouragement, guidance, critics, advices and supports to complete this research. I really appreciate his ethics and great deal of respect with his students, which is similar to brothers in the same family.

In addition, I am extremely grateful to my father for unlimited support and encouragement during this research. I would like to thanks my beloved family: mother, wife, daughter, brother, and sisters who I am always beholden to them for their everlasting patience, support, encouragement, sacrifice, and love which have been devoted to me sincerely, so I could endure for being away and eventually my master program came to fruition. For that, I ask Allah to bless all of them.

Besides, I would like to thank the authority of Universiti Teknologi Malaysia (UTM) and Faculty of Computing (FK) for providing me with a good environment and facilities such as computer laboratory to complete this project with software which I need during process.

ABSTRACT

Steganography is the art of science to hide digital media in any other digital media using different techniques. Steganography uses cover to embedded secret data, currently the cover chooses randomly, and for the same secret data every one can choose different cover without a prior knowledge which one is better, because there are no rules or measurements used for choosing suitable cover. Chooses the suitable cover is one of the major problems of steganography. This thesis develops techniques for discriminating between images which used as steganography cover in image in image steganography. Proposed algorithm is based on the hypothesis that a particular message embedding scheme leaves statistical evidence or structure that can be exploited for detection with the aid of proper selection of image features analysis. We pointed out the features of image that should be taken more seriously into account in the design of more successful steganography. This thesis suggests rules or measurements to select the proper cover for specific embedded message. It relay on the image features and also some technique to determine the expected secrecy and robustness. For each embedded image, we suggest many images to use as covers, the image features are determined for both covers and embed image. The embedded features are compared with all the covers features, and according to suggest relations in this thesis we will determine specific weight for each cover, best cover is with highest weight for that secret data. The proposed algorithm tested by using LSB image steganography, stego-image compared with the origin one which gives the promised results. Also proposed algorithm compared with other similar works and give better results.

ABSTRAK

Steganografi adalah seni sains untuk menyembunyikan media digital dalam mana-mana media digital yang lain menggunakan teknik-teknik yang berbeza. Steganografi menggunakan penutup untuk membenam data rahsia. Kini, penutup memilih secara rawak dan bagi data rahsia yang sama, setiap satunya boleh memilih penutup yang berbeza tanpa pengetahuan awal yang mana satu yang lebih baik kerana tidak ada peraturan atau ukuran yang digunakan untuk memilih penutup yang sesuai. Pemilihan penutup yang sesuai adalah salah satu masalah utama steganografi. Tesis ini membangunkan teknik untuk membezakan imej yang digunakan sebagai penutup steganografi dalam imej dalam steganografi imej. Algoritma yang dicadangkan adalah berdasarkan kepada hipotesis bahawa skema pembenaman mesej tertentu meninggalkan bukti statistik atau struktur yang boleh dieksploitasi untuk mengesan dengan bantuan pemilihan yang betul analisis ciri-ciri imej. Kami mengetengahkan ciri-ciri imej yang perlu diambil kira dengan lebih serius dalam reka bentuk steganografi yang lebih berjaya. Tesis ini mengemukakan peraturan-peraturan atau ukuran untuk memilih jenis penutup yang sesuai bagi mesej benaman tertentu. Ia membabitkan ciri-ciri imej dan juga beberapa teknik untuk menentukan kerahsiaan dan kekukuhan yang dijangkakan. Bagi setiap imej benaman, kami mencadangkan banyak imej digunakan sebagai penutup, ciri-ciri imej ditentukan bagi kedua-dua penutup imej dan imej benaman. Ciri-ciri benaman tersebut dibandingkan dengan semua ciri-ciri penutup dan mengikut cadangan hubungan di dalam tesis ini, kami akan menentukan bebanan khusus bagi setiap penutup, penutup terbaik adalah penutup yang mempunyai bebanan tertinggi bagi data rahsia tersebut. Perbandingan antara algoritma yang dicadangkan yang diuji menggunakan steganografi imej LSB, imej-stego dengan imej asli memberikan hasil yang memberangsangkan. Hasil yang lebih baik juga diperolehi daripada perbandingan algoritma yang dicadangkan dengan hasil kajian lain.

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	DECLARATION	ii
	DEDICATION	iii
	ACKNOWLEDGEMENT	iv
	ABSTRACT	v
	ABSTRAK	vi
	TABLE OF CONTENTS	vii
	LIST OF TABLES	x
	LIST OF FIGURES	xi
	LIST OF ABBREVIATIONS	xii
1	**INTRODUCTION**	**1**
	1.1 Introduction	1
	1.2 Background of the Study	2
	1.3 Related Work	5
	1.4 Problem Statment	6
	1.5 Aim of Thesis	7
	1.6 Objectives of Research	7
	1.7 Scope of study	8
	1.8 Thesis Outlines	8
2	**LITERATURE REVIEW**	**10**
	2.1 Introduction	10

2.2 History of Steganography 11

2.3 Steganography Applications 12

2.4 Cover Media Selection 13

2.4.1 Steganography In Text 13

2.4.2 Steganography in Images 15

2.4.3 Steganography in Audio 16

2.4.4 Steganography In Video 16

2.4.5 Steganography in Protocol 16

2.5 Steganography Strength 17

2.6 Least Significant Bit 18

2.7 Steganalysis 20

2.8 Data Hiding 21

2.8.1 Steganography 21

2.8.2 Watermark 22

2.8.3 Encryption 23

2.9 Image Steganography 24

2.9.1 Image Definition 24

2.9.2 Image Layout 24

2.10 Image Data Types 25

2.11 Optimization 27

2.12 Classification of Optimization Problems 28

2.12.1 Continuous Problems 29

2.12.2 Discrete Problems 29

2.12.3 Combinatorial Problems 29

2.13 Image optimization 30

2.14 Image Optimization needed 30

2.15 Most raleated work 31

2.16 Some Comments about the Related Works 33

3 METHODOLOGY 36

3.1 Introduction 36

	3.2	Measuring the features for each image	40
		3.2.1 Entropy	40
		3.2.2 Capacity	42
		3.2.3 Mean	43
		3.2.4 Variance	44
		3.2.5 Histogram	45
		3.2.6 Energy	46
		3.2.7 Robustness	47
		3.2.8 Expected secrecy	48
	3.3	Results evaluate	52
4	**RESULET AND DISCUSSION**		**55**
	4.1	Introduction	55
	4.2	Thesis goal and contributions	55
	4.3	The contribution of this thesis	56
	4.4	Statistics features	56
	4.5	Decision making	57
	4.6	Implementation	67
	4.7	Comparison of proposed technique with previous techniques for PSNR	65
5	**CONCLUSION AND FUTURE WORKS**		**68**
	5.1	Introduction	68
	5.2	Conclusions	68
	5.3	Future Works	70
	REFERENCES		**7**

LIST OF TABLES

No.	TITLE	PAGE
2.1	Table describe the related works and comments about each one	34
3.1	The best and worst values for the features used in this research	54
4.1	Comparing results when hiding embedded 1 in covers	63
4.2	Comparing results when hiding embedded 2 in covers	63
4.3	Comparing results when hiding embedded 1 in covers	64
4.4	Comparing results when hiding embedded 2 in covers	65
4.5	A Comparison of the different techniques	66

LIST OF FIGURE

No.	TITLE	PAGE
1.1	Classification of a security system	3
1.2	Goals of data security	4
2.1	Steganography types	13
2.2	Kinds of Text Steganography	14
2.3	Left figure shows the main properties for good steganography, while the right one is expanded to all properties.	18
2.4	Optimization types	27
3.1	Framework of the study	38
3.2	Block diagram of the proposed algorithm	39
4.1	The covers and secret images that were used in experiment	58
4.2	Images specifications (cover and embedded)	58
4.3	Final weight when calculating features with both embedded 1, and embedded 2	60
4.4	Comparing cover image before and after hiding embedded 1	61
4.5	Histogram for origin and stego-image when hiding embedded 1 in cover image	62
4.6	Shows the data sate image (Lena, Baboon, Peppers)	66

LIST OF ABBREVIATIONS

DWT	Discrete Wavelet Transform
DCT	Discrete Cosine Transform
LSB	Last Significant Bit
RGB	Red Green Blue
PSNR	Peak Signal Noise Ratio
BMP	Bitmap
TCP/IP	Transmission Control Protocol Internet Protocol
DNA	Database Network Associates
PCFG	Probabilistic Context-free Grammar

CHAPTER 1

INTRODUCTION

1.1 Introduction

Sensitive information is constantly exposed to threats by malicious peoples. The important factors in this case are the cost and the very big losses of data, when the information gets in the wrong hands. Hackers work hard to steal and destroy the confidential data and information. Visible confidential data is exposed to stealing and destroying by hackers. What if the confidential information is hidden? There is an illusion that compliance equals security, which has led both individuals and organizations to excessively spend on compliance at the detriment of the security that steganography attempts to solve in this dilemma. The main goal of steganography is to hide digital messages in other digital media by using robust algorithm which provides a stego-object look innocent file, this prevents the enemy to detect that a secret message exists.

Steganography is the art and science of hiding information by embedding data into another digital data. Widespread use of computers has led to the development of new ways for implementation steganography, this is what it is called Electronic steganography techniques which use digital ways of hiding and detecting processes. Normally, the detection process works inversely of the hiding process.

Many steganography techniques are suggested, some of them simple and most people are familiar with, and others are more complex. Steganography becomes more

complex by using the computer facilities. It transfers from simple tools like, invisible inks and microdots to electronic techniques. With computers and networks, there are many other ways of hiding information, such as:

- Covert channels by using header of packet to hide data without effect on the header's information.
- Hidden text within Web pages, images, audio and video.
- Hiding files in "plain sight" (e.g., what better place to "hide" a file than with an important sounding name in the c:\winnt\system32 directory?)
- Different ways of ciphers.

Nowadays, Steganography has become more developed than the above suggested examples, allowing a user to increase the capacity of hiding data within digital media such as; image and audio files. Steganography is shared with cryptography to protect the information by encrypting the secret message and then hiding it in the cover, this adds more complexity to recover data hiding by an unauthorized person.

There are a number of uses for steganography besides the mere novelty. One of the most widely used applications is for so-called digital watermarking. A watermark is like a stamp used to protect the originality of digital media. The nears application to watermark is the steganography. Steganography is widely used in the medical field to hide important information in medical images such as; x-ray, it is also used for copyright and communications. Many digital media are used for steganography such as; image, text, video and audio protocols.

1.2 Background of the Study

Two main techniques are considered for the security cryptography and information hiding, as shown in figure 1.1.

Nowadays, information is rapidly available through the Internet. Companies have the ability to communicate with a worldwide audience through the World Wide Web. So Information hiding techniques are receiving increasing attention due to:

a. The availability of low price computers, multimedia and digital form objects.
b. The availability of fast transmission media such as internet.
c. The need to present solutions to stealing information and copyright problems.

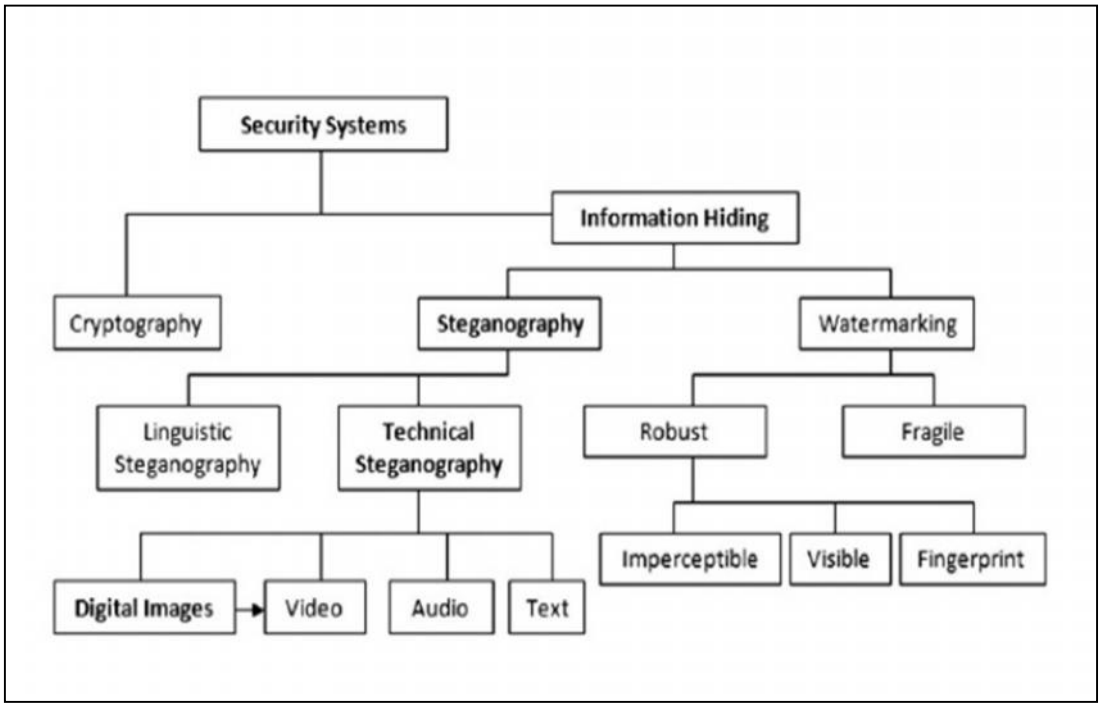

Figure 1.1: Classification of a security system

From figure 1.1 and figure 1.2 (Cheddad, 2010), it is clear there are many form of mediums used for steganography to hide and transmit information. Any digital media used for the transmission of messages can be used for steganograohy. The digital media used to carry the secret message is called ‘cover’, while the secret message which is embedded in the cover is called ‘embedded data’. Many different techniques are used to embed secret message depending on the type of cover. There are three aspects for data security, confidentiality, integrity, and availability.

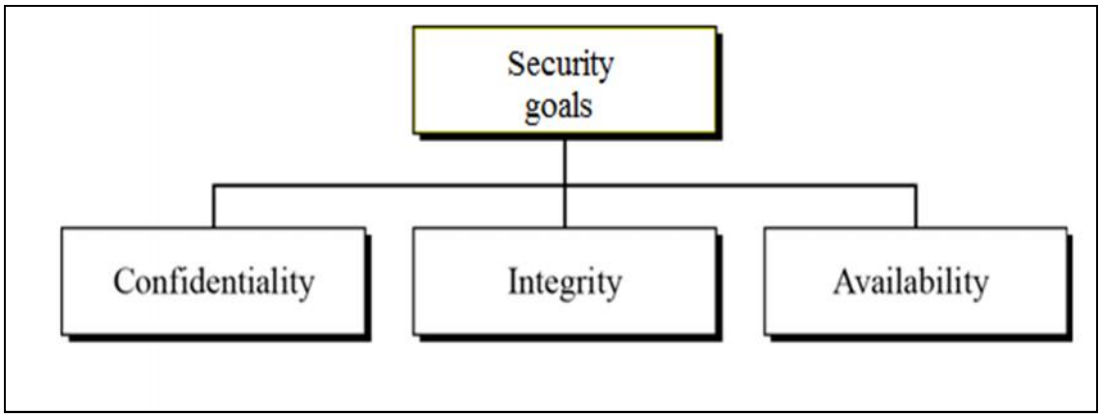

Figure 1.2: Goals of data security

a. Confidentiality

Confidentiality is the process of keeping data secure and not to permit unauthorized people to know the existence of important data. An organization needs to use some techniques to guard sensitive data against those malicious actions that endanger the confidentiality.

b. Integrity

Integrity means that the data don't change or aren't destroyed in the transmission, and any alert to the data should be done through authorized channel and by an authorized user.

c. Availability

The information created and stored by an organization needs to be available to authorized users and for applications. Information is useless if it is not available. Information needs to be changed constantly, which means that it must be accessible

to those authorized to access it. Unavailability of information is just as harmful to an organization as a lack of confidentiality or integrity.

1.3 Related Work

Study of steganography cover is new, so, there are few researches dealing with this problem:

Kharrazi,(2006) studied the performance of steganalysis techniques, and suggested some measurements to measure the image properties before and after hiding the data, which affected on the steganalysis performance. He studied three scenarios effected on the cover selection, in which the secret data contain either no knowledge, partial knowledge, or full knowledge of the steganalysis technique.

These measures are divided into two categories. First, cover image properties (Changeable Coefficients, JPEG Quality Factor) and second, cover-stego based distortion measures (Number of Modifications to the cover image, Mean Square Error, Prediction Error, Watson's metric, Structural Similarity Measure).

Zhang,(2007) used two least significant bits of cover to hide data. A double-layered embedding scheme that introduces the wet paper coding mechanism to determine the selection of addition or subtraction is proposed. The aim of this way is to improve the embedding efficiency and embedding rate by exploiting the second bit in addition to the first bit, this will help to carry additional bit.

Sajedi,(2009) presented a new adaptive steganography method aimed to increase the amount of embedding message based on contourlet transform. The proposed method is called ContSteg. This method tried to represent edge in an image by a large magnitude coefficient. In ContSteg, these coefficients are considered for data hiding because human eyes are less sensitive in edgy and non-smooth regions of

images due to highly distributed colors. For embedding the secret data, contourlet sub bands are divided into 4×4 blocks. Each bit of secret data is hidden by exchanging the value of two coefficients in a block of contourlet coefficients. This method embed high amount of secret data without causing noticeable change of stego-object. It is considered better than steganography based on wavelet transformation.

Chen-Feng,(2008) introduced a new data hiding method based on hiding two secret digits in seven most significant bit, LSB used as indicator. This method provided high capacity due to repeating embed secret data, it preserved the image quality and ensured the secrecy. Therefore, two secret digits can be embedded into one pixel-group without using the original cover image and extra data.

Kaur,(2012) presented 2k correction method & edge detection method to be used in steganography. This technique has good ability to carry more payloads without leaving visible changes to the stego-object. Hence, this method is classified as a better method than the previous methods. Due to the nature of the human eyes which cannot easily detect the distortion at edge, so it used the edge areas to embed more data.

Now the question is "**How to choose cover to embed secret data**", there are no rules to choose cover. For the same secret data, there are unlimited covers that can be chosen randomly and it's all hide the secret data and result stego-image.

1.4 Problem Statement

The issue of this research is to choose the best cover for the secret image because steganography is dependent on using cover to embedded secret message on it, the cover absolutely chooses randomly. For one secret message, it could choose different covers by different persons.

If there are more than one cover to embed secret data in, then "**which one is the best in regards to capacity, security, and robustness**", it is clear not all the covers chosen have the same behavior for capacity, security, and robustness. The issue is the absence of rules or measurements to choose the best cover to embed specific secret date.

1.5 Aim of Thesis

Many easy tools to use steganography are available to hide secret messages on one side of communication and detect hidden information on the other side. Steganography uses cover to embed secret data, this cover is randomly chosen and for the same secret data, everyone can choose different covers without a prior knowledge of which one is better, because there are no rules or measurements used for choosing suitable cover.

The aim of this thesis is to propose many features that can be used as measurements to choose the best cover among many suggested covers for an embedded secret data (using image in image steganography); independent of the embedding operations.

1.6 Objectives of Research

In order to accomplish the aim of this study, the following objectives have been identified:

1. To propose a new measurements scheme for choosing the best cover for embedding secret data based on the proposed selected features.
2. To evaluate the performance of the proposed measurement scheme based on capacity, robustness, and imperceptibility.

1.7 Scope of study

The aim of this algorithm is to find best cover for an embedded secret data, it concentrates on image in image steganography, for that:

1. Color image of 24 bits is used as cover.
2. Any secret image can be embedded.
3. The evaluation will be based the following parameters capacity, robustness, PSNR, entropy.

1.8 Thesis Outlines

The contents of the remaining chapters of this thesis are as follows:

Chapter 1: In this chapter we will have, introduction for data hiding and steganogragh, background of study, problem statement, aim of thesis objectives and the scope of study.

Chapter 2: Chapter two introduces the basic concept of steganography and the most related works. The history of steganography, the LSB algorithm, the RGB color image are explained and a comparison is made between steganography and cryptography.

Chapter 3: The design and implementations of proposed methods introduced in this chapter and we define all the features that are used in the methodology and how to choose the best cover for the secret image.

Chapter 4: This chapter presents the results and analysis of the proposed methods and how the decision is made, also the implementation of the methodology and lastly comparing with the previous techniques.

Chapter 5: This chapter introduces the objectives of this thesis.A summary of the work and suggestions for future works and how to develop this work, is discussed in the conclusion.

REFERENCES

Al-Shatnawi, A. M. (2012). A New Method in Image Steganography with Improved Image Quality. *Applied Mathematical Sciences*. 6(79), 3907-3915.

Burger, W. and M. Burge (2009). Principal of Digital Image Processing. *Springer-Verlag* London Limited.

Cachin, C. (1998). An information theoretic model for steganography. *LNCS: 2nd Int'l Workshop on Inf*ormation *Hiding*. 306–318.

Chan, C. K. and Cheng L. M. (2004). Hiding data in images by simple LSB substitution. *The journal of the Pattern Recognition society*. 469 – 474.

Chandramouli, R. and Memon N. (2001). Analysis of LSB based image steganography techniques. *International Conference on*.

Cheddad, A., J. C. (2010). Digital image steganography: Survey and analysis of current methods. *Signal Processing* 90(3). 727-736.

Chin-Feng, L., C. Lin-Yi. (2008). Hiding Information Employing Reduplicating Embedding. Asia-Pacific Services Computing Conference. *APSCC '08. IEEE*.

Fridrich, J. and Goljan M. (2004). On Estimation of Secret Message Length in LSB Steganography in Spatial Domain.

Fridrich J., M. G. D. and Soukal D. (2003). Quantitive steganalysis of digital images: Estimating the secret message length. *ACM Multimedia Systems Journal, Special issue on Multimedia Security*. 9(3). 288–302.

Jacob and Jackson. (2003). Blind Steganography Detection Using a Computational Immune System. *International Journal of Digital Evidence*. 4 (1).

Kaur, S. P. and Singh S. (2012). A New Image Steganography Based on 2k Correction Method and Canny Edge Detection. *Proceedings of 'I-Society 2012' at GKU*. Talwandi Sabo Bathinda .

Kharrazi, M. H. T. S. (2006). Cover Selection for Steganographic Embedding. *IEEE International Conference on*. Image Processing.

Lin and Delp. (1999). A review of data hiding in digital images. Proceedings of the Image Processing. *Image Capture Systems Conference, PICS'99, the Society for Imaging Science and Technology*. 274–278.

Lin, C. C. and Tsai W. H. (2004). Secret image sharing with steganography and authentication. *Journal of Systems and Software*. 73(3). 405-414.

Lisa, M. (1998). Hiding Information in Images. Image Processing. *ICIP 98. Proceedings*. 396-398.

Morkel, J. H. P. (2005). An Overview of Image Steganography. *Proceedings of the Fifth Annual Information Security South Africa Conference (ISSA2005)*. Sandton, South Africa.

Niimi (2003). A Framework of Text-based Steganography Using SD-Form Semantics Model. *Pacific Rim Workshop on Digital Steganography*. Kyushu Institute of Technology, Kitakyushu, Japan.

Reddy, L. (2011). Implementation of LSB Steganography and its Evaluation for Various File Formats. International Journal of Advanced Networking and Applications. 2(5). 868-872.

Russ, J. (2011). The Image Processing Handbook. by Taylor and Francis Group, LLC.

Sajedi, H. and Jamzad (2010). Contourlet-Based Steganography Using Cover Selection. *International Journal of Information Security*. 9(5). 337-345.

Sharma, V. K. and Shrivastava, V. (2012). A Steganography Agorithm for Hiding Image in Image by Improved LSB Substitution by Minimize Detection. *Journal of Theoretical and Applied Information Technology*. 36(1).

Shirali-Shahreza, M. (2008). Text Steganography by Changing Words Spelling. *ICACT 10th International Conference on*. *Advanced Communication Technology*.

Weiming, Z., W. S. (2007). Improving Embedding Efficiency of Covering Codes for Applications in Steganography. *Communications Letters, IEEE*. 11(8). 680-682.

Xiangwei, K., W. Z. (2005). Steganalysis of Palette Images: Attack Optimal Parity Assignment Algorithm. *Information, Communications and Signal Processing, 2005 Fifth International Conference on*.

Xinpeng, Z. and Shuozhong W. (2006). Efficient Steganographic Embedding by Exploiting Modification Direction. *Communications Letters, IEEE*. 10(11). 781-783.

Xu, Y. H. J. (2009). Viewpoint invariant texture description using fractal Analysis. *International Journal of Computer Vision*. 83(1). 85-100.

Yadav, R. (2011). Analysis of Various Image Steganography Techniques Based Upon PSNR Metric. *International Journal of P2P Network Trends and Technology*. 1(2).

Zhang, Z. (2007). Efficient double-layered steganographic embedding. *Electronics Letters*. 43(8).