



Faculty of Information and Communication Technology

**MULTILAYER REVERSIBLE DATA HIDING VIA HISTOGRAM
SHIFTING**

Hamida Mohamed Almangush

Doctor of Philosophy

2016

MULTILAYER REVERSIBLE DATA HIDING VIA HISTOGRAM SHIFTING

HAMIDA MOHAMED ALMANGUSH

**A thesis submitted
in fulfillment of the requirements for the degree of Doctor of Philosophy**


Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2016

DECLARATION

I declare that this thesis entitled Multilayer Reversible Data Hiding via Histogram Shifting is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : 

Name : Hamida Mohamed Al Mangush

Date : 11 November 2016

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in term of scope and quality for the award of Doctor of Philosophy.

Signature :

Name : Professor Dr. Mohd Khanapi Bin Abd Ghani

Date : 11 November 2016

DEDICATION

To my family, especially my husband and my children.

ABSTRACT

Concealing messages from unauthorised people has been desired since written communication first began. With advancements in digital communication technology and the growth of computer power and storage, the difficulty of ensuring the privacy of individuals and the protection of copyright has become increasingly challenging. Steganography finds a role in attempting to address these growing concerns. Problems arise in the steganography method because of the trade-off between capacity and imperceptibility whereby increasing the embedding capacity increases the distortion in the stego object and it thus becomes suspect. Another problem is concerned with non-retrieval of the original cover object whereby misplacing data could be crucial for example in the case of medical images. Reversible data hiding technique based on histogram shifting addresses the problem of retrieving the original cover. Embedding the secret message by shifting the histogram between the pair of the peak and minimum points wastes the embedding capacity and does not control the distortion in the stego image for various secret messages sizes. In this research, a technique for reversible data hiding is proposed which enables the retrieval of both the hidden secret message and the original image at the receiver's side. The proposed technique considers the size of the secret message and the distribution of the colour values within the cover image to determine the value of the optimal pair or set of container and carried colours within the best sub image instead of the pair of peak and minimum points. The experimental results show that the proposed technique increases the embedding capacity within the cover image and produces a stego image with a high peak signal-to-noise ratio value. In addition, the experimental results show that by using the proposed re-shifting and extraction formulas, the technique has the ability to extract the hidden data and retrieve the original images from the stego images. In comparison to the traditional histogram-shifting techniques, the proposed technique significantly improves the stego image quality and the embedding capacity. Thus, this research has contributed to two principles, namely improvements in capacity and quality.

ABSTRAK

Menyembunyikan mesej daripada orang yang tidak dibenarkan telah diinginkan sejak komunikasi bertulis pertama bermula. Dengan kemajuan teknologi komunikasi digital dan pertumbuhan kuasa perkomputeran dan penyimpanan, kesukaran dalam memastikan rahsia individu dan perlindungan hak cipta menjadi semakin mencabar. Stenografi mengambil peranan dalam usaha untuk mengatasi masalah yang semakin membimbangkan ini. Masalah timbul dalam kaedah stenografi kerana keseimbangan di antara kapasiti dan ketakbolehkeliwatan, iaitu peningkatan kapasiti membenam meningkatkan keherotan pada objek stego dan menjadikannya suspek. Satu lagi masalah yang membimbangkan ialah ketidakbolehan dalam mendapatkan semula objek penutup asal yang dengannya kehilangan data boleh menjadi penentu, contohnya dalam kes imej-imej perubatan. Teknik penyembunyian data berbalik yang berasaskan anjakan histogram dapat mengatasi masalah mendapatkan semula penutup asal. Membenamkan mesej rahsia tersebut dengan menganjak histogram di antara pasangan titik puncak dan titik minimum mengurangkan kapasiti membenam dan tidak mengawal keherotan pada imej stego untuk berbagai saiz imej rahsia. Dalam kajian ini, teknik penyembunyian data berbalik dicadangkan bagi membolehkan dapatan semula kedua-dua mesej rahsia yang tersembunyi dan imej asal di sebelah penerima. Cadangan teknik ini mempertimbangkan saiz mesej rahsia tersebut serta taburan nilai warna dalam imej penutup untuk menentukan nilai pasangan optimum atau set bekas dan warna yang dibawa dalam sub imej terbaik menggantikan pasangan titik puncak dan titik minimum. Keputusan eksperimen menunjukkan cadangan teknik ini meningkatkan kapasiti membenam di dalam imej penutup dan menghasilkan imej stego dengan satu nilai puncak nisbah isyarat-hingar yang tinggi. Tambahan lagi, keputusan eksperimen menunjukkan melalui cadangan formula anjakan semula dan pengestrakan, teknik ini berupaya untuk mengekstrak data tersembunyi dan mendapatkan semula imej-imej asal daripada imej stego. Berbanding dengan teknik tradisional anjakan histogram, cadangan teknik ini jelas sekali memperbaiki kualiti imej stego dan kapasiti membenam. Lantaran itu, kajian ini telah menyumbang kepada dua prinsip iaitu peningkatan kapasiti dan kualiti.

ACKNOWLEDGEMENT

In the Name of Allah, Most Gracious, Most Merciful,

I would like to express my sincere appreciation and heartfelt thanks to my supervisor, Prof Dr. Mohd Khanapi Bin Abd Ghani, for his wisdom, endurance, invaluable guidance, suggestions and full support during this research work. His philosophies and way of thinking to lead a meaningful life has taught me not only how to conduct my research but also how to face future challenges.

Great thanks from my heart my husband for his patience and support, my son and daughters for their prayers, and my brothers, sisters, friends, and colleagues for their tremendous encouragement.

Also, I would like to thank all the lecturers, administrators, and staff of Universiti Teknikal Malaysia Melaka and all academic and non-academic staffs of the Graduate school for their help and support.

Finally, I am deeply indebted to my family, my husband and my children for their patience and encouragement during the period of this research.

Hamida Mohamed Almangush, Melaka, October 2016

TABLE OF CONTENTS

	PAGE
DECLARATION	
APPROVAL	
DEDICATION	
ABSTRACT	i
ABSTRAK	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	viii
LIST OF FIGURES	x
LIST OF ALGORITHMS	xv
LIST OF APPENDICES	xvi
LIST OF ABBREVIATIONS	xviii
LIST OF PUBLICATIONS	xx
CHAPTER	
1. INTRODUCTION	1
1.1 Overview	1
1.2 Problem Statement	4
1.3 Research Questions	7
1.4 Research Objectives	7
1.5 Research Scope	8
1.6 Contributions of the Research	8
1.7 Thesis Organization	10
1.7.1 Chapter One	10
1.7.2 Chapter Two	10
1.7.3 Chapter Three	10
1.7.4 Chapter Four	10
1.7.5 Chapter Five	11
1.7.6 Chapter Six	11
1.7.7 Chapter Seven	11
2. LITERATURE REVIEW	12
2.1 Introduction	12
2.2 General Model of a Steganography System	14
2.3 Basic Features for Data Hiding Systems	15
2.3.1 Embedding Capacity	16
2.3.2 Robustness	16
2.3.3 Invisibility	16
2.3.4 Undetectability	17
2.3.5 Security	17
2.3.6 Complexity	17
2.4 Types of Digital Data Hiding Techniques	18
2.4.1 Detectable and Readable Data Hiding Techniques	18
2.4.2 Blind and Non-blind Data Hiding Techniques	19
2.4.3 Reversible and Non-reversible Data Hiding Techniques	20
2.4.4 Spatial, Transform and Quantisation Data Hiding Techniques	21
2.4.5 Robust, Semi-Fragile and Fragile Data Hiding Techniques	22
2.4.6 Public and Private Data Hiding Techniques	23

2.4.7	Symmetric and Asymmetric Data Hiding Techniques	23
2.4.8	Visible and Invisible Data Hiding Techniques	24
2.5	Steganography and Watermarking Applications	24
2.5.1	Copyright Protection	25
2.5.2	Fingerprint (Traitor-Tracie)	25
2.5.3	Identify Digital Media	25
2.5.4	Usage in Modern Printers	26
2.5.5	Commercial Use	26
2.5.6	Broadcast Monitor	27
2.5.7	Network Traffic Monitor	27
2.5.8	Use by Military Agencies	27
2.5.9	Use by Terrorists	27
2.5.10	Use by Healthcare Applications	28
2.6	Reversible Data Hiding	30
2.7	Features and Evaluation Measurements of Reversible Data Hiding	31
2.8	Applications of Reversible Data Hiding	32
2.9	Approaches for Reversible Data Hiding in Digital Images	36
2.9.1	Reversible Data Hiding Based on Modulo Arithmetic	37
2.9.2	Reversible Data Hiding Based on Difference Expansion	37
2.9.3	Reversible Data Hiding Based on Histogram Shifting	38
2.9.4	Reversible Data Hiding Based on Contrast Mapping	40
2.10	Related Work in Reversible Data Hiding Based on Histogram Shifting	41
2.10.1	Summary of Review for Related Work in Reversible Data Hiding Based on Histogram Shifting	59
2.11	Steganalysis	67
2.12	Classification of Steganography Attacks	67
2.13	Detection and Destruction Steganography Attacks	68
2.13.1	Detection Attack	68
2.13.2	Destruction Attack	69
2.14	Steganographic Systems Evaluation	70
2.14.1	Evaluation of Capacity (Payload)	71
2.14.2	Evaluation of Imperceptibility	71
2.14.3	Evaluation of Steganographic Robustness	76
2.15	Summary	77
3.	RESEARCH METHODOLOGY	78
3.1	Introduction	78
3.2	Research Framework	78
3.2.1	Review of Current Works through Literature	79
3.2.2	Data Collection	80
3.2.3	Design and Develop the Proposed Technique	81
3.2.4	Testing Performance and Validate the Proposed Technique	89
3.2.5	Data Analysis	91
3.3	Summary	91
4.	CONCEPTUAL DESIGN	93
4.1	Introduction	93
4.2	Analysis Phase	93
4.3	Design Phase	93
4.3.1	Convert Binary to Trinary Model	94
4.3.2	Selection Module	95

4.3.3	Shifting and Embedding Modules	100
4.3.4	Extraction and Recovery Modules	101
4.4	Mathematical Formulations	102
4.4.1	Mathematical Formula for Converting Binary to Trits and Vice Versa	102
4.4.2	Mathematical Formula for Shifting Pixel Values	106
4.4.3	Mathematical Formula for Embedding Data	109
4.4.4	Mathematical Formula for Extracting Data	110
4.4.5	Mathematical Formula for Recovery	110
4.5	Overhead Information Structure	111
4.5.1	Primary Overhead Information	111
4.5.2	Boundaries Map	113
4.5.3	Location Map	114
4.6	Summary	114
5.	PROPOSED TECHNIQUE	115
5.1	Introduction	115
5.2	Selection Phase	117
5.2.1	Determining the Colour Frequencies, Colour Positions and Bit per Container BPC Value	117
5.2.2	Initial Selection of the Container Colours	119
5.2.3	Determining the Sub Images of the Container Colours	123
5.2.4	Filtering the Container Colours	126
5.2.5	Selection of the Carried Colours	130
5.2.6	Creating the Pairs and Sets of Colours	135
5.2.7	Selecting the Best Pair and Set of Colours	139
5.3	Embedding Phase	139
5.3.1	Composing the Payload	140
5.3.2	Determining the Shifting Ranges and the Shifting Direction Values	140
5.3.3	Embedding the Payload	141
5.3.4	Embedding Overhead Information	143
5.4	Extraction and Recovery Phase	146
5.4.1	Extracting Overhead Information	146
5.4.2	Determining the RE-shifting Ranges and the RE-shifting Direction Values	147
5.4.3	Extracting the Payload and Recovery of the Shifting Pixels	148
5.4.4	Decomposing the Payload and Recovery of the Original Image	150
5.5	Summary	153
6.	IMPLEMENTATION AND EXPERIMENTAL RESULTS	154
6.1	Introduction	154
6.2	Cover Images Chosen	154
6.3	Secret Messages Chosen	156
6.4	Discussion of the Proposed Technique	157
6.5	Testing and Analysis of the Results	158
6.5.1	Stego Image Quality Test	158
6.5.2	Embedding Capacity Test	182
6.5.3	Extracting Test	190
6.5.4	Reversibility Test	191
6.5.5	Robustness Test	192
6.6	Comparative Study	198

6.6.1	Comparison of Stego Image Quality	200
6.6.2	Comparison of Embedding Capacity	202
6.7	Conclusion	204
6.8	Summary	205
7.	CONCLUSIONS	206
7.1	Introduction	206
7.2	Summary of Completed Work	206
7.3	Novelty and Contributions	207
7.4	Conclusions Related to the Research Objectives	209
7.5	Future Research	211
	REFERENCES	213
	APPENDICES	233

LIST OF TABLES

TABLE	TITLE	PAGE
2.1	Operational table for private and public watermarks	23
2.2	The Advantage and Disadvantage of Chosen Reversible Data Hiding Schemes based on Histogram Shifting	61
2.3	Quality, Impairment, Continuous Quality and Comparison Scales	75
3.1	Initial Operational Framework	80
3.2	Operational Framework	82
4.1	Reserved Number of Bits to Represent the Primary Overhead Information	112
6.1	Cover Images	155
6.2	Cover Images Capacities	156
6.3	Different Selection of Pairs	159
6.4	Number of the Shifted Pixels within the Cover Images	160
6.5	Sub Images of the Peak Point Colours	161
6.6	Container Colours for Different Embedding Rates (ER)	163
6.7	Sub Images for Different Embedding Rates (ER)	165
6.8	PSNR Obtained for Different Embedding Rates and Selection of Pairs of Colours	170
6.9	Shifted Pixels in the Marked Images for Different Embedding Rates	174
6.10	Different Selections of Sets of Colours	176
6.11	PSNR Obtained for Different Embedding Rates and Selection of Sets of Colours	177
6.12	Crests of the PSNR Values	182
6.13	Location Map Sizes for the Cover Images and the Sub Images	183
6.14	Maximum Embedding Rate of Different Selecting of Sets of Colours	185

6.15	Optimum Embedding Rate for Different Bits per Container Values	187
6.16	Improvement of the Embedding Capacity	189
6.17	Location Map Size for Image Number 4	190
6.18	PSNR and SSIM Values of the Recovered Images	192
6.19	Extracted Images after Applying Crop Attack	194
6.20	Extracted Images after Applying Salt and Pepper Attack on Image Number10	196
6.21	Comparison of Marked Image Quality (PSNR) between the Available Techniques and the Proposed Technique	202
6.22	Comparison of Maximum Embedding Rate (MER) between the Available Techniques and the Proposed Technique	203

LIST OF FIGURES

FIGURES	TITLE	PAGE
1.1	Conflicting Requirements for Data Hiding	5
2.1	Taxonomy of Information Hiding	13
2.2	General Steganographic Model	14
2.3	(a) Detectable and (b) Readable Data Hiding Techniques	19
2.4	(a) Non-blind and (b) Blind Data Hiding Techniques	20
2.5	(a) Non-Reversible and (b) Reversible Data Hiding Techniques	21
2.6	(a) Symmetric and (b) Asymmetric Data Hiding Techniques	24
2.7	Fujitsu exploitation of steganography: (a) a sketch representing the concept and (b) the idea deployed into a mobile phone shown at an exhibition	26
2.8	Taxonomy of Digital Data Hiding	31
2.9	Zidan and Abdulsattar's Data Embedding Process	35
2.10	An Example of an Image Histogram	42
2.11	Histogram Shifting using JinHa Hwang et al.'s Mechanism	43
2.12	(a) Original Pixels Values and (b) Pixel Values after Shifting	43
2.13	Difference Pair Pattern	44
2.14	(a) Original Difference Histogram; (b) Histogram Shifting; (c) Message Embedding (L=2); (d) Message Embedding (L=1); (e) Message Embedding (L=0)	46
2.15	Four Ways of Block Division	48
2.16	(a) Original Histogram ;(b) Generalised Histogram Shifting ; (c) Stego Histogram	50
2.17	Histogram Modification for EL = 2	52

2.18	(a) Original Histogram and (b) Histogram Packed Histogram	53
2.19	Rhombus Prediction	54
2.20	Histogram Shifting using Li et al.'s Mechanism	56
2.21	(a) Original Histogram, (b) Shifted Histogram, (c) Stego Histogram	58
3.1	Research Framework.	79
3.2	Selection and Embedding Stage	86
3.3	Extraction and Recovery Stages	88
4.1	Ternary Format	95
4.2	Items of the Pair and Set of Colours	96
4.3	Inputs and Outputs of the Selection Procedure	97
4.4	An example of Overlapping Sub-Images	97
4.5	Sub-image of a container colour	99
4.6	(a) Histogram of the Cover Image (b) Histograms of the MSP and LSP	99
4.7	Comparison between Shifting the Cover Image and the Sub-Image Histograms. (a) Shifting the Cover Image Histogram (b) Shifting Sub-Image Histogram	100
4.8	Inputs and Outputs of the Embedding Procedure	101
4.9	Inputs and Outputs of the Extraction Procedure	102
4.10	Conversion between Binary Data and Ternary Data	103
4.11	Shifting Directions within Category 1	107
4.12	Shifting Directions within Category 2	107
4.13	Shifting Direction within Category 3	107
4.14	Shifting Direction within Category 4	108
4.15	Shifting Direction within Category 5	108
4.16	The Embedding Location (a) Sub-Image Type 1 (b) Sub-Image Type 2	113
5.1	Selecting and Embedding Procedures	116
5.2	Extraction and Recovering Procedures	116
5.3	Scanning the Cover Image Pixels	117

5.4	Flowchart of Determining the Bit per Container and the Multi values	119
5.5	Container Colours	119
5.6	Primary Pure Capacities of a Colour (a) Primary Pure Capacity of the First Sub-Image (b) Primary Pure Capacity of the Second Sub-Image.	120
5.7	Pure Capacity of a cover image	122
5.8	Flowchart for Selecting the Initial Container Colours	123
5.9	Pure Capacity of a Container Colour (a) Pure Capacity of a Container Colour in Set W_1 (b) Pure Capacity of a Container Colour in Set W_2	127
5.10	Flowchart of Filtering the Container Colours from Sets W_1 and W_2	129
5.11	Carried Colours (a) Carried Colours in the First Sub-Image (b) Carried Colour in the Second Sub-Image	131
5.12	Selecting of the Carried Colours from the Left Side of the Container Colour	137
5.13	Selection of the Carried Colours from the Left and Right Sides of the Container Colour	138
5.14	Cover Image Pixels	144
5.15	Scanning Embedding Location	144
5.16	Shifted Pixels and Embedding Data in the Embedding Location	145
5.17	Stego Image Pixel Values	146
5.18	Stego Image Pixels	150
5.19	RE-shifted Pixels and Extracted Data in the Extracting Location	152
5.20	Recovered Original Image	152
6.1	Minimum Frequencies of the Container colours	164
6.2	Number of Container Colors for Different Embedding Rates	164
6.3	Number of the Pairs of Colours for Different Embedding Rates	167
6.4	Minimum Number of Shifted Pixels for Different Embedding Rates	167
6.5	Minimum Values of the Shifted Pixels for Different Selected Pairs	168
6.6	Size of the Overhead Information for Different Selection Pairs of Colour and	172

	Embedding Rates	
6.7	PSNR Value for Embedding Rate =0.25	172
6.8	PSNR Value for Embedding Rate = 0.50	173
6.9	PSNR Value for Embedding Rate = 0.75	173
6.10	Overhead Information Size for ER=1	178
6.11	Overhead Information Size for ER=5	179
6.12	Overhead Information Size for ER=10	179
6.13	PSNR Value for Embedding Rate =1	180
6.14	PSNR Value for Embedding Rate =5	180
6.15	PSNR Value for Embedding Rate =11	181
6.16	Maximum Embedding Rate of Different Selections of Sets	185
6.17	Maximum Capacity of Different Selecting of Sets	186
6.18	Maximum Embedding Rate for Different Bits per Container Values	187
6.19	Location Map Size for Different Embedding Locations within Image Number 4	190
6.20	Level of Salt and Pepper Attack and PSNR	196
6.21	Level of Salt and Pepper Attack and SSIM	197
6.22	Level of Salt and Pepper Attack and NCC	197
6.23	Level of Salt and Pepper Attack and BER	198
6.24	Level of Salt and Pepper Attack and BCR	198

LIST OF ALGORITHMS

ALGORITHM	TITLE	PAGE
5.1	Determining the Cover Image Histogram, the High Frequency, the First Positions of the Colours From 0 to 255.	118
5.2	Determining the Top and Bottom Boundaries of a Sub Image	125
5.3	Determining the Histogram of a Sub Image	126
5.4	Determining the Ranges of the Shifted Pixels and the Shifting Direction Values	141
5.5	Shifting Pixels Values and Embedding the Payload	142
5.6	Determining the Extracted Ranges, the Ranges of the Re-Shifted Pixels and the Re-Shifting Direction Values	147
5.7	Extracting the Hiding Secret Message and Recover the Shifted Pixels	149

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	Convert Binary Code to Trinary Code	234
B	Convert Trinary Code to Binary Code	237
C	Collection of Selected Additional Medical Images	240
D	Cover Images Capacities	243
E	Maximum Embedding Rate MER	244
F	Peak Signal to Noise Ratio PSNR Obtained for Different Embedding Rates	245

LIST OF ABBREVIATIONS

APD	Adjacent Pixel Difference
ASCII	American Standard Code for Information Interchange
BCH	Bose, Chaudhuri, and Hocquenghem
BCR	Bit Correct Ratio
BER	Bit Error Ratio
BM	Boundaries Map
BMP	Bitmaps
BPB	Bit PerBit
BPC	Bit Per Continuer
BPF	Bits Per Frame
BPP	Bits Per Pixel
BPS	Bits Per Second
DCT	Discrete Cosine Transform
CD	Compact disc
CDCS	Class Dependent Coding Scheme
CT	Computerized Tomography
DE	Difference Expansion
DWT	Discrete Wavelet Transform
ECI	Embedding Capacity Improvement
EL	Embedding Level
EPR	Electronic Patient Record
ER	Embedding Rate

FBI Federal Bureau of Investigation

FR Full Reference

H Height

HF High Frequency

LIST OF PUBLICATIONS

Hamida Mohammed Almangush, Mohd K. AbdGhani, Ahmed B. Abugharsa. (2012)"A Novel Reversible Data Hiding Technique with High Capacity and Less Overhead Information". International Journal of Computer Applications, 43(19), Pages 42-47.

Hamida Mohamed Almangush, MohdKhanapiAbdGhani, Ahmed Bashir Abugharsa. (2013)"A New Mechanism to Control Marked Image Quality of Reversible Data Hiding Based on Histogram Shifting". Australian Journal of Basic and Applied Sciences, 7(14), Pages: 195-203.

Hamida Mohammed Almangush, Mohd K. AbdGhani, Ahmed B. Abugharsa.,(2015) "MultiLayer Reversible Data Hiding Based on Histogram Shifting with High Quality and Capacity". International Review on Computers and Software (I.RE.CO.S.), 10(8), Pages 820-828.

CHAPTER 1

INTRODUCTION

1.1 Overview

As long as there has been written communication, humans have had the desire to conceal their secret messages from the curious eyes of others. Information hiding techniques have become the newest hot spot in security research (Rudramath and Madki, 2012). New applications and new technologies bring new threats, thus new protection mechanisms have to be invented. Moreover, the need for confidentiality of valuable information, private and sufficiently secure communications in several applications such as e-banking, e-trading, mobile telephony, medical data interchange, the military, intelligence agencies etc., is rapidly increasing (Souvik et al., 2011; Por et al., 2008). With these forces driving the need, research into information hiding has grown explosively.

Steganography is the art and science of hiding communication and it has been used throughout history for secret communications. The word steganography comes from the Greek “Steganos”, which means covered or secret and “-graphy” meaning writing or drawing. Therefore, steganography means, literally, covered writing (Jayaram et al., 2011; Cheddad et al., 2010). The emergence of the computer and the evolution of the sciences and techniques have breathed life again into this art, with the use of new ideas and techniques that draw on computer characteristics in the way that data is represented (Bachrach and Shih, 2011). The most well-known computer representation of all data including images files, audio files and video files is binary. These binary files include redundant bits that can be modified without causing any awareness by human senses by means of hearing or sight (Naji et al., 2009; Jayaramu and