Information hiding for intellectual property protection of digital media

Yu, Dan

2005

Yu, D. (2005). Information hiding for intellectual property protection of digital media.
Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/3919>

<https://doi.org/10.32657/10356/3919>

Nanyang Technological University

Downloaded on 26 Apr 2023 21:22:07 SGT

INFORMATION HIDING FOR INTELLECTUAL PROPERTY PROTECTION OF DIGITAL MEDIA

YU DAN

SCHOOL OF ELECTRICAL & ELECTRONIC ENGINEERING
NANYANG TECHNOLOGICAL UNIVERSITY

2005

Information Hiding for Intellectual Property Protection of Digital Media

YU DAN

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

2005

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me and has not been submitted for a higher degree to any other University or Institute.

.....

Date

.....

YU DAN

*To Mom and Dad,
for their encouragements and love.*

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Farook Sattar, for his guidance, encouragement and support in every stage of my Ph.D study. I could count on him to provide guidance and insight when I was stuck, but he also allowed me considerable freedom to explore various topics to develop my abilities in research and problem solving. His knowledge, kindness, and vision as a professional researcher have provided me lifetime benefits.

I am grateful to Prof. Kai-Kuang Ma for contributing useful discussion regarding the use of ICA in image watermarking, and his thoughtful comments and encouragement on my research are highly appreciated. I would like to thank Prof. N. G. Kingsbury for his valuable suggestions and comments that help to improve the proposed watermarking scheme for transaction tracking in multimedia distribution applications. I would also like to thank Profs. Braham Barkat and Sirajudeen Gulam Razul for providing the enlightening discussions and comments with their abundant knowledge to improve my works. Special thanks to Prof. Alex Kot for his valuable comments and suggestions in broadening my vision and understanding in data hiding field. I want to take this opportunity to thank Profs. Lap Pui Chau, Anthony Ho, Eric Sung and Krishnan Shankar M. for their supports and encouragements on my research.

Special thanks to Zhu Yan, Liu Bibo and Dr. Cai Zhi for their friendships and countless helps during my whole graduate studies. I would also like to express my appreciation to Dr. Liu Mingchun, Dr. Wang Qing, Dr. Lu Hong, Liang Yongqing, Moe Pwint and Gao Xinting for their kind helps and enjoyable discussions in a wide range. I would like to thank all my fellow graduate students and technicians in the Information System Research lab and Media Technology lab. My memorable days of

life in Nanyang Technological University can never be complete without them.

I am so grateful to my sister, Yu Jing, and my brother, Yu Shouzhi for their encouragements and supports all the time in my study. Lastly I must mention two persons – my parents without whose love and nurturing I could never accomplish this thesis. Thanks Mom for her care and love of her daughter. Thanks Dad for his love, nurturing, encouragements and supports all along my 20-year school education. I dedicate this thesis to my parents.

Contents

Acknowledgements	i
List of Figures	viii
List of Tables	xii
List of Abbreviations	xiv
Abstract	xv
1 Introduction	1
1.1 Motivation	3
1.2 Objectives	4
1.3 Thesis Contributions	5
1.4 Thesis Organization	6
2 Fundamentals of Information Hiding	8
2.1 Introduction on Information Hiding	8
2.2 Information Hiding Framework	11
2.3 Challenging Issues and Benchmarking	14
2.4 Communication Model of Information Hiding	17
Part I: Information Hiding for Transaction Tracking	19
3 Introduction	20
3.1 Independent Component Analysis	20
3.1.1 Statistical Independence	21

3.1.2	Definition of ICA	22
3.1.3	Fundamental Restrictions	23
3.1.4	Summary of ICA Algorithms	23
3.2	Motivation	24
3.3	Objective	26
4	Watermark Detection and Extraction Using Independent Component Analysis Method	27
4.1	Watermark Design	28
4.2	Proposed Watermarking System	29
4.2.1	Watermark Embedding Scheme	29
4.2.2	Proposed Watermark Detection and Extraction Scheme	30
4.3	Robust Batch Algorithm for Watermark Detection and Blind Extraction	31
4.3.1	AR modelling Process for Image Sources	32
4.3.2	Dewatermarking Process using Robust Batch Algorithm	32
4.4	Evaluation	37
4.4.1	Performance Measures	37
4.4.2	Simulation Results	38
4.4.3	Performance Analysis with ICA	42
4.4.4	Performance against Embedding Weighting Coefficients	49
4.4.5	Performance Comparisons with Various ICA Methods	50
4.5	Robustness Test	54
4.5.1	Geometric Transformations	55
4.5.2	JPEG Compression	56
4.5.3	Low Pass Filtering	56
4.5.4	Noise Addition	58
4.5.5	Multiple Marks	58
4.5.6	Collusion	58
4.6	Chapter Summary	61
5	A Wavelet-Domain Blind Watermarking Scheme Based on ICA	63
5.1	Introduction	63

5.1.1	Wavelet-domain Embedding	64
5.1.2	Features of Proposed Scheme	65
5.1.3	Perceptual Mask	66
5.2	Proposed Watermark Embedding Scheme	68
5.3	Proposed Watermark Extraction Scheme Using ICA Method	72
5.4	Robustness Test	75
5.5	Evaluation of the Extraction Performance with ICA	79
5.6	Chapter Summary	82
6	Transaction Tracking for Multimedia Distribution Through ICA-based Watermarking Technique	84
6.1	Multimedia Distribution Framework	85
6.2	Limitations of Existing Watermarking Schemes and The Solution . . .	88
6.3	ICA-based Watermarking Scheme for Transaction Tracking	91
6.3.1	Embedding Channel Selection of Color Images	91
6.3.2	Watermark Embedding Scheme	93
6.3.3	Watermark Extraction Scheme Using ICA Method	99
6.4	Performance Evaluation	101
6.4.1	Robustness Against Cropping and RST Attacks	102
6.4.2	Discussions	105
6.5	Concluding Remarks and Future Trends	108
	Part II: Information Hiding for Ownership Verification	111
7	Transparent Robust Information Hiding for Ownership Verification	112
7.1	Motivation and Objective	112
7.2	Proposed Adaptive Watermark Embedding Scheme	115
7.2.1	Adaptive Watermark Embedding process	116
7.2.2	Automatic Embedding Range Selection	120
7.2.3	Threshold Setting	124
7.2.4	Dual Key Generation	125
7.2.5	Watermark Embedding Results	126
7.3	Proposed Blind Watermark Extraction Scheme	129

7.4	Performance Evaluation	130
7.4.1	Imperceptibility	131
7.4.2	Robustness Test	132
7.5	Discussions	137
7.5.1	Binary versus Gray-level Watermark	137
7.5.2	Additive versus Replacement Embedding	138
7.5.3	Whole DCT Domain versus Block DCT Domain Embedding . .	138
7.5.4	Security Issues	139
7.5.5	Applications	140
7.6	Chapter Summary and Conclusion	141
Part III: Information Hiding for Content Authentication		143
8	Introduction	144
8.1	Authentication Watermarking	144
8.1.1	Requirements of Authentication	145
8.1.2	Review of Authentication Techniques	147
8.2	Complex Chirp Signals	150
8.3	Objective	153
9	Multiresolution Fragile Watermarking Using Complex Chirp Signals for Content Authentication	155
9.1	Multiresolution Representation	156
9.1.1	Wavelet Decomposition of Host Image	156
9.1.2	Generation of Multiresolution Watermark	156
9.2	Proposed Multiresolution Watermark Embedding Scheme	160
9.2.1	Watermark Embedding Locations	160
9.2.2	Quantization Process	163
9.2.3	Watermark Embedding Algorithm	166
9.2.4	Watermark Embedding Results	168
9.3	Watermark Extraction and Performance Evaluation	171
9.3.1	Watermark Extraction and Verification Scheme	171
9.3.2	Performance Evaluation: Robustness versus Tamper-Sensitivity	173

9.4 Applications	179
9.4.1 Content Integrity Verification Without Resorting the Original Watermark	180
9.4.2 Blind Quality Assessment Using Multiresolution Fragile Water- marking Technique	186
9.5 Chapter Summary and Conclusion	190
10 Final Remarks	192
10.1 Future Works	192
10.2 Summary and Concluding Remarks	196
Appendices	198
Appendix A: JADE Algorithm	199
Appendix B: Infomax Algorithm	203
Appendix C: FastICA Algorithm	206
Author's Publications	210
Bibliography	212

List of Figures

2.1	A proposed classification of information hiding techniques.	9
2.2	A framework of information hiding system.	12
2.3	Trade-offs between three requirements of an information hiding system	15
2.4	A communication model of information hiding system.	17
4.1	Special images used as keys or watermarks	28
4.2	Watermark embedding process.	30
4.3	Watermark extraction process.	31
4.4	Column- and row-wise scanning orders for image source	33
4.5	Various zig-zag scanning orders for the spatial source image	33
4.6	Proposed watermark detection and extraction scheme.	34
4.7	Block diagram of the robust batch ICA	35
4.8	Watermark embedding example	39
4.9	Watermark extraction result for watermarked case	39
4.10	Watermark extraction result for unwatermarked case	39
4.11	Convolutional watermark embedding example	41
4.12	Watermark extraction results for convolutional embedding example . . .	41
4.13	Watermark extraction results using text image as watermark	43
4.14	Improved watermark embedding scheme	45
4.15	Another watermark extraction example	46
4.16	Running window to compute average variance	47
4.17	Performance against embedding weighting coefficients	49
4.18	Watermark embedding examples using Boat image	50
4.19	Robustness against RST attacks	57
4.20	Extraction result under attack of JPEG compression	59

4.21	Extraction result under attack of low pass filter	59
4.22	Extraction result under attack of median filter	59
4.23	Extraction result under attack of additive Gaussian noise	60
4.24	Extraction result under attack of multiple marks	60
4.25	Extraction result under attack of collusion	60
5.1	Noise Visibility function (NVF)	67
5.2	Second level wavelet decomposition of Lena image.	68
5.3	A watermark embedding example	69
5.4	Watermark embedding scheme	70
5.5	Owner's copy of Lena image	72
5.6	Proposed blind watermark extraction scheme	73
5.7	Extracted watermark using JADE ICA method	74
5.8	Results against JPEG compression attacks	77
5.9	Results against JPEG2000 compression attacks	77
5.10	Results against color quantization	77
5.11	Watermarking example 1 using a text image as watermark	79
5.12	Watermarking example 2 using a text image as watermark	80
5.13	The watermarked Lena images	80
6.1	A multimedia distribution system	85
6.2	Proposed multimedia distribution framework	87
6.3	Flowchart of watermark embedding scheme for color image	94
6.4	Wavelet decomposition of Y component	94
6.5	A color image watermark embedding example	95
6.6	Proposed watermark embedding algorithm	96
6.7	Owner's copy of color Lena image	98
6.8	Proposed blind watermark extraction scheme	100
6.9	Extracted user's watermark using JADE ICA method	101
6.10	Extraction result under cropping attack	104
6.11	Extraction results under RST attacks	105
7.1	A watermark embedding example	116

7.2	Watermark embedding process.	117
7.3	The search engine for watermark embedding.	117
7.4	Adaptive watermark embedding algorithms	118
7.5	Selection of <i>End</i> Index for embedding range	121
7.6	Selection of <i>Initial</i> Index for embedding range	123
7.7	Watermarked image quality wrt <i>Initial</i> Index	123
7.8	A watermark embedding example	126
7.9	More watermark embedding examples	127
7.10	Watermark extraction process.	129
7.11	Extracted watermark	130
7.12	Extraction result under JPEG compression	133
7.13	Extraction result under quantization attack	135
7.14	Extraction result under additive Gaussian noise	135
7.15	Extraction result under geometric distortions	135
7.16	Extraction result under random geometric distortions	136
7.17	Extraction result under blurring attack	136
7.18	Extraction result under sharpening attack	136
8.1	Time-frequency representations of FM chirp signals	154
9.1	Three level wavelet decomposition of Lena image	156
9.2	Quadratic FM complex chirp watermark signal	158
9.3	Multiresolution chirp watermark generation	160
9.4	Watermark embedding locations.	161
9.5	Clusters of wavelet coefficients for embedding	162
9.6	The quantization function for the wavelet coefficients.	164
9.7	Rules for embedding of watermark bits	165
9.8	The watermark embedding process.	167
9.9	Watermark embedding example 1	169
9.10	Watermark embedding example 2	170
9.11	The watermark extraction process.	171
9.12	Quality of watermarked image wrt width of secure embedding zone . .	176

9.13	BERs of extracted watermark wrt width of secure embedding zone . . .	177
9.14	Watermarked image using Kundur's scheme	178
9.15	Column-wise location indexes	181
9.16	The detector response when there is no attacks.	181
9.17	A tampered Lena image by one pixel's value	182
9.18	Tamper detection and localization when one pixel is altered	183
9.19	A tampered image by replacement	185
9.20	Tamper detection and localization when multiple blocks are altered . .	185
9.21	Curves for magnitude of extracted complex chirp signal	187
A.1	Inverting $\mathbf{A}$ by chaining a whitener and a unitary matrix.	200
B.1	Illustration for the problem of blind source separation.	204

List of Tables

4.1	Performance evaluation using Gold-like sequence as watermark	40
4.2	Normalized correlation coefficient wrt convolutive filter size	41
4.3	Normalized correlation coefficients using text image as watermark . . .	42
4.4	Normalized correlation coefficients between sources.	43
4.5	Extraction performance comparison for Spatial and DCT domain embedding	45
4.6	Extraction performance comparison for column and row-zz wise scanning	47
4.7	Relative comparison of average variances	48
4.8	Extraction performance for watermarked Boat image	51
4.9	Extraction performance for another watermarked Boat image	51
4.10	Computational complexities for various ICA methods	54
5.1	PSNR of watermarked image	71
5.2	Extraction results using JADE and RICA algorithms	81
5.3	Normalized correlation coefficients between sources.	81
6.1	PSNR of watermarked image	97
7.1	More watermark embedding results	127
7.2	Characteristics of various DCT-based robust watermarking schemes . .	131
7.3	PSNR of watermarked image	132
7.4	Robustness under JPEG lossy compression attacks	133
9.1	BERs of extracted watermarks under JPEG compressions	174
9.2	BERs of extracted watermarks under various signal processing attacks .	175
9.3	PSNR of watermarked image under various signal processing attacks . .	175

9.4	BERs of extracted watermarks under JPEG compression for Kundur's scheme	179
9.5	Quality assessment under JPEG compressions	189
9.6	Quality assessment under various signal processing	189

List of Abbreviations

ABBREVIATIONS	FULL EXPRESSIONS
A/D	Analog-to-Digital
AM	Amplitude Modulation
AR	Autoregressive
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BSS	Blind Source Separation
D/A	Digital-to-Analog
DCT	Discrete Cosine Transform
DRM	Digital Rights Management
DWT	Discrete Wavelet Transform
EZW	Embedded Zerotree Wavelet
FastICA	Fast fixed-point ICA algorithm
FM	Frequency Modulation
FT	Fourier Transform
HVS	Human Visual System
ICA	Independent Component Analysis
IDCT	Inverse Discrete Cosine Transform
IDWT	Inverse Discrete Wavelet Transform
IE	Instantaneous Energy
IES	Instantaneous Energy Spread
IMF	Instantaneous Mean Frequency
IMFS	Instantaneous Mean Frequency Spread
Infomax	Information Maximization
IP	Intellectual Property
IPP	Intellectual Property Protection
IPR	Intellectual Property Rights
IWT	Inverse Wavelet Transform

JADE	Joint Approximate Diagonalization of Eigen-matrices
JND	Just Noticeable Distortion
JPEG	Joint Photographic Experts Group
LAN	Local Area Network
LSB	Least Significant Bit
LUT	Look-Up Table
MAN	Metropolitan Area Network
MPC	Multimedia Personal Computer
MSB	Most Significant Bit
MSE	Mean Square Error
NC	Normalized Coefficient
NVF	Noise Visibility Function
PC	Personal Computer
PCA	Principle Component Analysis
PCM	Pulse Code Modulation
PSNR	Peak-Signal-to-Noise Ratio
RST	Rotation, Scaling, and Translation
QF	Quality Factor
QIM	Quantization-Index-Modulation
RMSE	Root-Mean-Squared Error
RICA	Robust batch ICA algorithm
RLC	Run-Length Coding
SNR	Signal-to-Noise Ratio
SPIHT	Set Partitioning in Hierarchical Trees
TFR	Time-Frequency Representation
VW2D	Variable-Watermark Two-Dimensional
WAN	Wide Area Network
WD	Wigner Distribution
wPSNR	weighted Peak-Signal-to-Noise Ration
wrt	with respect to
WT	Wavelet Transform
WVD	Wigner-Ville Distribution
WWW	World Wide Web
zz	zig-zag

Abstract

The large-scale communication of multimedia data motivates the intellectual property (IP) protection of digital media against illegal duplication and manipulation. This thesis concentrates on designing and developing information hiding and retrieval techniques, as well as their applications to enforce IP rights of the digital media. Information hiding for protection of the cover media is also defined in term of *digital watermarking*. The following three aspects are investigated: information hiding for transaction tracking (Part I); information hiding for ownership verification (Part II); and information hiding for content authentication (Part III).

In Part I, we propose a new watermark detection and extraction method by using Independent Component Analysis (ICA) technique. The feasibility of using ICA for *dewatermarking* is evaluated, and the factors that influence the extraction performance are analyzed. A blind wavelet-domain watermarking system with ICA method for extraction is also proposed. A multimedia distribution framework for transaction tracking through the proposed ICA-based watermarking scheme is then presented. It has shown that our proposed scheme can overcome the assumptions/restrictions of other watermarking schemes in watermark generation, embedding and extraction, hence, it is more efficient for transaction tracking in multimedia distribution applications.

In Part II, a novel transparent robust watermarking technique has been proposed for ownership verification. To enforce copyright protection through ownership verification, the essential requirement of a watermarking system would be its robustness against attacks. The adaptive replacement embedding method in the DCT-domain can hide the gray level watermark transparently into the host media. The use of adaptive replacement technique and the automatic selection of embedding range has successfully

compromised the trade-off between two contradictory requirements of an information hiding system – the imperceptibility and the robustness.

In Part III, a wavelet-domain multiresolution fragile watermarking scheme is presented for content authentication. The FM complex chirp signal is employed as watermark to embed in the randomly selected wavelet coefficients by an improved quantization-index-modulation (QIM) technique. The use of a secure embedding zone successfully improves the Kundur's QIM embedding scheme in terms of the quality of the watermarked image and the bit error rates (BERs) of the extracted watermark. The promising feature by using complex chirp watermark signal is that no original watermark is required during authentication, thus a truly 'blind' authentication process is achievable. Two potential applications of the proposed scheme are demonstrated for content integrity verification and blind quality assessment.

Chapter 1

Introduction

We are now in a digital information age. The digital information technology has changed our society as well as our lives. The information revolution may be understood to take place in two forms:

- data/information retrieval/representation;
- data/information dissemination/communication.

Digital presentation of data allows the information recorded in a digital format; and thus it brings easy access to generate and replicate the information. It is such an easy access that provides the novelty in the current phase of the information revolution. Digital technology allows primarily use with the new physical communications media, such as satellite and fiber optic transmission. Therefore, PCs, email, MPCs, LANs, WANs, MANs, Intranets and the Internet have been evolved rapidly since the 1980's. The *Internet* has a world-wide broadcasting capability, a mechanism for information distribution, and a medium for collaboration and interaction between individuals and their computers without regard to geographic location. This allows researches and professionals to share the relevant data and information with each other.

As image, audio, video and other works become available in digital form, it may be ease for which perfect copies can be made. The widespread use of computer networks and the global reach of the World Wide Web have added substantially an astonishing abundance of information in digital form, as well as offering unprecedented ease of access to it. Creating, publishing, distributing, using, and reusing information have become many times easier and faster in the past decade. The good news is the enrichment that this explosive growth in information brings to society as a whole. The bad news is the enrichment that it can also bring to those who take advantage of the properties of digital information and the Web to copy, distribute, and use information illegally. The Web is an information resource of extraordinary size and depth, yet it is also an information reproduction and dissemination facility of great demand and capability.

Therefore, there has been currently significant amount of research in Intellectual Property (IP) protection issues involving the multimedia content distribution through Internet. The multimedia security covers such topics as *access control*, *content authentication*, and *copyright protection* [1, 2]. The secure networking environment requires the tools that provide

- secure and fast encryption and decryption to achieve the secure data transmission channel,
- content authentication functionality of the received data by the recipient, and
- robust and trustworthy marks indicating copyright and the legal ownership.

Encryption and decryption techniques [3] are developed to enhance the secure transmission, which is only one requirement of a secure networking. This thesis addresses information hiding techniques [1, 2, 4, 5, 6] offering an efficient tool for content authentication and the copyright protection, which is unsolvable by cryptographic terminology.

Information hiding techniques received much more attention due to the high demands of IP rights protection for digital media. Particular information is hidden into the digital media to be protected, which is retrieved later on to enforce the ownership identification, copyright protection and content authentication of the cover media.

1.1 Motivation

The recent growth of the networked multimedia systems has increased the need for the protection of digital media. This is particularly important for the protection and enforcement of intellectual property rights. The *intellectual property protection* (IPP) is becoming an increasing importance nowadays. Copyright protection involves the ownership authentication of the data, and the identification of illegal copies of a possibly forged data. Techniques are needed to prevent the copying, forgery and unauthorized distribution of image, audio and video. Without such methods, placing images or video sequences on a public network puts them at risk of theft and alteration.

The basic scenario is as follows: a user has created an electronic multimedia data at some effort and expense, and wants to make it available on a communications network. When unauthorized copies or forgeries of the image appear elsewhere on the network, the user needs to prove that the data belongs to them. One also needs to determine whether and by how much the data has been changed from the original.

Copyright laws have been raised to protect the intellectual property of the individuals and groups. These laws protect the creators of the contents by restricting the distribution and modification of the contents without permission. The effectiveness of the copyright laws, however, depends on some important requirements:

- The owner of the original content must be able to provide the trustful proof that

he/she is the rightful owner of the content.

- The owner must be able to track the distributions of the content, so that he/she is able to find the person who should be responsible for the illegal distribution.
- The owner should be able to detect any illegal attempts to alter the content.

Without mechanisms to support above three basic requirements, the copyright owners can not provide proof to claim that he/she is the rightful owner and somebody else violate the owner's intellectual property.

1.2 Objectives

Motivated by the importance of IP protection in our digital world, the works dedicated in my thesis focus on information hiding for protection of digital media towards the following three objectives:-

- **Transaction Tracking:** to propose an efficient multimedia distribution framework, which can trace the transactions of the digital media in the public networks. To overcome the restrictions of the existing watermarking schemes for tracking application, more general watermarking and dewatermarking techniques would be investigated.
- **Ownership Identification:** to propose transparent robust information hiding technique for providing trustful proof of ownership even when the content suffers heavy degradation. The robust watermarking scheme would achieve excellent robustness while preserving high quality of the watermarked content.
- **Content Authentication:** to propose a fragile watermarking scheme, which would be able to detect any possible tampers in the watermarked content with

localization capability. Blind authentication process is desirable.

In this thesis, the information hiding and retrieval techniques are described in the context of watermarking still images, but they are applicable to the general multimedia content, for example, audio and video.

1.3 Thesis Contributions

The original contributions of this thesis include the following:

- The proposal of a new watermark detection and extraction method by using Independent Component Analysis (ICA) technique. (Chapter 4)
- The evaluation of the feasibility of ICA technique for watermark detection and extraction, using a simple non-blind spatial-domain watermarking scheme. The main factors that affect the watermark extraction performance of ICA methods are also investigated. Moreover, the possible solutions to improve the watermark extraction performance by ICA methods are proposed. (Chapter 4)
- The proposal of a blind wavelet-domain robust watermarking scheme with ICA method for watermark extraction, which improves the robustness performance of previous non-blind spatial domain watermarking scheme. (Chapter 5)
- The introduction of a multimedia distribution framework through the proposed ICA-based watermarking scheme to enable transaction-tracking functionality. It has shown that our proposed ICA-based watermarking scheme can overcome the assumptions/restrictions of the other watermarking schemes in terms of watermark generation, embedding and extraction, hence, it is more efficient for transaction tracking of multimedia distribution applications. (Chapter 6)

- The proposal of a novel transparent robust information hiding technique for ownership verification. The proposed robust DCT-based information-hiding scheme achieves excellent robustness performance against attacks, while preserving high quality of the watermarked content. (Chapter 7)
- The proposal of a multiresolution fragile watermarking scheme for content authentication applications. The proposed scheme further improves the performance of Kundur's Quantization-Index-Modulation (QIM) scheme for both the quality of the watermarked content and the bits error rate (BER) of the extracted watermark. (Chapter 9)
- The use of the complex chirp signals as watermarks in our proposed multiresolution fragile watermarking scheme for content integrity verification and blind quality assessment. Without resorting the original watermark and the original content, the received content can be authenticated blindly. (Chapter 9)

1.4 Thesis Organization

Chapter 1 introduces the motivation and objectives of this thesis. The rest of the thesis is organized as follows. Chapter 2 presents the fundamentals of information hiding. The main works of the thesis are organized into three parts: *Information Hiding for Transaction Tracking* (Part I), *Information Hiding for Ownership Verification* (Part II), *Information Hiding for Content Authentication* (Part III). The summary of the thesis, future works as well as the concluding remarks are presented in Chapter 10.

Part I consists of four chapters – Chapter 3, 4, 5 and 6. Chapter 3 is an introduction of Part I. Chapter 4 proposes a new watermark detection and extraction method by using Independent Component Analysis (ICA) technique. Then the feasibility of ICA

technique for watermark detection and extraction is evaluated, using a simple non-blind spatial-domain watermarking scheme. In Chapter 5, a new blind wavelet-domain robust watermarking scheme based on ICA is proposed. Chapter 6 further applies the proposed ICA-based blind watermarking scheme for transaction tracking to enforce the intellectual property protection of digital media. Part II has only one chapter – Chapter 7, which proposes a novel transparent robust information hiding technique for ownership verification. Part III consists of Chapter 8 and 9. Chapter 8 is an introduction of Part III. Chapter 9 proposes a multiresolution fragile watermarking scheme including its applications for content authentication.

Chapter 2

Fundamentals of Information Hiding

2.1 Introduction on Information Hiding

Over thousand years, our ancestor has thought about multitude of methods and variation to hide information with the objective of hiding of the secret messages within another seemingly innocuous carrier to conceal the very existence of the hidden messages. Recently, information hiding [1, 4, 5, 7], also called as data hiding techniques, received rapidly increasing attention with wider applications. Information hiding or data hiding is a more general term covering a wide range of problems including steganography and watermarking.

In Figure 2.1, we propose a classification scheme of current information hiding techniques. First of all, based on the goals, it is classified into two categories:

1. Information hiding for covert communication – to transfer the secret information by hiding it within the cover media;

Figure 2.1: A proposed classification of information hiding techniques.

2. Information hiding for protection of the cover media – to identify, trace, locate, or authenticate the digital media (cover media) by hiding message (watermark) into it.

The first category contains the concept of keeping the existence of the information secret, such as the topics on maintaining anonymity while using a network, keeping part of the database secret from unauthorized users. Steganography [1] is another subdiscipline of information hiding under the first category. It is usually interpreted as hiding supplemental information into multimedia data. Examples include sending a message to a spy by marking secret information in a newspaper using invisible ink, and adding sub-perceptible echo at certain places in an audio recording. The information hiding under this category falls out of the scope of this thesis.

The works of this thesis dedicate on information hiding techniques in the second category, i.e., to enforce the intellectual property protection of the cover media by the means of information hiding. This category of information hiding techniques can be further classified into two different types according to their protection applications:

1. *Robust copyright marking* [1, 2] – the hidden message (watermark) in the cover

media is robust against various attacks. Therefore, it can efficiently identify the copyright ownership of the protected digital media by extraction of the embedded watermark.

2. *Authenticate marking* [1, 2] – the hidden message (watermark) is fragile/semi-fragile, which is distorted or removed totally under attacks. It is used to authenticate the possible alteration in the protected digital media based on the extracted watermark.

The robustness of the hidden information against possible attacks is one most important requirement for a copyright marking system. The robust marks include two types, namely, *fingerprints* and *robust watermarks*. Fingerprints [7, 8] (also called *labels* or *signatures*) are the hidden serial numbers to identify the user or customer of a particular copy of the content. The fingerprints are to identify the customers or the users, whereas, robust watermarks are aimed to identify the owner of the content. The watermarks could be either *visible* or *invisible*. Visible watermarks could be some visible patterns, such as a company logo or a copyright sign, overlaid on the digital data. Most of the current researches focus on the invisible (or *transparent*) watermarking that is a more difficult task than visible one. The scope of this thesis is also on invisible watermarking.

Authenticate marking, as opposed to copyright marking, requires the hidden information has good sensitivity against forgery. They are used to authenticate whether an object has been ‘touched’ or manipulated, and might be useful as evidence in court. Authenticate marks can be further classified into *fragile watermarks* [9] and *semi-fragile watermarks* [10] based on the degree of the attacks they can resist. Fragile watermarking algorithms function as a strict tamper detection tool, in that they are destroyed even when there are very slight alterations. On the other hand, semi-fragile water-

marks are designed to trace changes in the protected content. In other words, they would sustain a moderate degree of the desirable (or innocent) attacks.

The term digital watermarking was used for the first time probably in the early 1990s. The copy prevention and copyright protection have been the major driving forces in the research of digital watermarking originally, which is identical with the concept of the robust copyright marking as discussed above. The concept and applications of watermarking is rather limited at the beginning. However, a number of profound applications such as broadcasting monitoring, transaction tracking, authentication, copy control and device control, are suggested in the recent research of digital watermarking. These challenging applications make digital watermarking rapidly becoming so popular in the research community and in the industry. The concepts of watermarking now comprise not only the robust watermarking for copyright protection, but also fragile/semi-fragile watermarking for authentication. It has been extended to a more broad definition of digital watermarking – the *imperceptible, robust, secure communication* of information by embedding it in and retrieving it from the digital cover media to enforce copyrights or ensure the integrity of the cover media [11].

The information hiding discussed in this thesis is equivalent to the term of *watermarking* with the definition in a broad sense. Watermarking shares many same technical approaches as information hiding. This thesis uses these two terms *information hiding* and *watermarking* of the same concept, referring to embedding data into the cover media for the protection of the cover media.

2.2 Information Hiding Framework

Figure 2.2 shows a typical information hiding framework for the protection of digital media. The original digital media (**I**) such as image, audio or video are to be protected.

Figure 2.2: A framework of information hiding system.

Information or a watermark ($\mathbf{w}$) is embedded for uniquely identifying the ownership or authenticating the digital content. The information to be hidden can be either pseudorandom sequences, binary/grayscale images (for example, company logo or owner's signature image¹) or signals with special characteristics (for example, complex chirp signals²).

An information hiding (or watermarking) system comprises two generic building blocks: an information hiding or watermark embedding scheme and an information recovery scheme, also known as watermark extraction or watermark decoder, or *dewatermarking* process defined in this thesis [1, 14, 15].

The information is embedded in either spatial or transform domain of the original

¹In part I and part II of this thesis, grayscale images are used as the watermarks.

²In part III of this thesis, complex FM chirp signals [12] having special patterns in time-frequency representation [13] are used as fragile watermarks.

data. A transformation process of the original data is necessary if the information hiding is conducted in a transformed domain. The pre-processing of the watermark to-be-embedded may also be required. For example, a visual perceptual model would be necessarily applied to modulate the watermark signal in order to ensure its invisibility after embedding; the transformation of the watermark could be needed, if the transformed coefficients of the watermark rather than its spatial amplitudes are to be embedded. The inputs to the embedding block include the original digital media, the watermark and an optional public or secret key. The key is used to enforce security, that is, to prevent unauthorized party from recovering or manipulating the hidden watermark. The key could be generated based on random generators, or may be generated from the embedding process, such as, embedding locations and strength, working as side information for watermark detection and extraction. The output of the information hiding scheme is the watermarked data or the stego-media, $\mathbf{X}$.

The marked data may further go through compression process before distribution over the public networks. During the channel transmissions, the digital media could be distorted due to the imperfect channel or the intentionally attacks by the illegal attackers. The received data thus could be the same as the original data $\mathbf{X}$ or a distorted version of the original data, $\mathbf{X}'$.

Inputs to the detection/extraction block include the received data, the embedding key. Some watermarking algorithms also require the original watermark and/or the original unwatermarked data for the detection/extraction. The output of the detection/extraction block is either a confidence measure indicating how likely the given watermark is present in the received data (for detection) or the recovered watermark $\mathbf{w}'$ (for extraction). For watermark extraction, the corresponding reverse processing is then applied to the extracted watermark, $\tilde{\mathbf{w}}'$, to obtain the final extracted watermark, $\mathbf{w}'$.

The following types of watermarking systems can be identified, according to the difference of the nature and combination of inputs and outputs [1, 14, 15].

- Private watermarking (also called non-blind watermarking) systems require both the key and the original data for watermark detection or extraction.
- Semiprivate watermarking (or semiblind watermarking) does not use the original data for detection and extraction.
- Public watermarking (also referred to as blind or oblivious watermarking) requires neither the original data nor the embedded watermark.

Although blind watermarking is a more difficult task, it has drawn more attentions of researchers because many practical applications require blind schemes.

2.3 Challenging Issues and Benchmarking

It is now well accepted that an effective data hiding system must balance the requirements of three issues: *imperceptibility* – *robustness* – *capacity* [1, 2, 6, 7, 15]. *Imperceptibility* requires that the information hidden in the original cover data be such that the marked data and the original data should be perceptually indistinguishable. *Robustness* requires that the embedded information should be reliably detected or retrieved when the marked data is altered due to either channel transmission or intentionally attacks. *Capacity* refers to the amount of the information that is being embedded into the host cover data. The relationship between these three requirements involves relatively complex trade-offs, as illustrated in Figure 2.3. If one of the requirements is determined, the other two are inversely proportional. For instance, if the amount of information to be embedded is decided, there always exists a trade-off between the visual quality of the marked data and robustness of the watermark. In general, the higher the embed-

Figure 2.3: Trade-offs between the amount of information hidden, its invisibility and its robustness in an information hiding system.

ding strength, the better robustness a watermarking system can achieve. However, at the same time it may result in a poorer visual quality of the marked image.

It is not easy to satisfy all the above contradictory requirements simultaneously for an information hiding system. Therefore, some of the requirements of a practical information hiding system can be reduced or neglected, depending on the application. For instance, in a robust copyright watermarking system the robustness against attacks is of prime importance. The robustness of a watermarking system for ownership verification needs to be very high, so that the owner can always prove the ownership of the digital media even after watermarked image has been degraded significantly. On the contrary, an authentication mark must be very sensitive to the alterations in the marked data.

Many robust watermarking schemes are proposed for copyright protection. A systematic benchmarking [16, 17, 18, 19] provides a common criteria for robustness performance evaluation against attacks. However, the evaluation should also include measurement of distortions in the watermarked data introduced by watermark embedding.

It is well known that there exists a trade-off between the watermark imperceptibility (measured by the quality of the watermarked data) and its robustness. Thus for a fair performance evaluation, the watermarked image quality degradation needs to be taken into account. After designing a new digital watermark method, proper performance evaluation is conducted according to the same benchmark, so that we can ensure that the methods under investigation are examined under comparable conditions. Existing benchmarking tool, such as StirMark [20], integrates a number of attacks including geometric distortion (e.g., affine transforms, cropping, row/column removal and random local distortions) and common signal processing (e.g., lossy compression, quantization, dithering, filtering, denoising, noise addition, multiple watermarks and collusion). More possible attacks are suggested in [19], i.e., denoising attacks and wavelet compression, watermark copy attack, synchronization removal, denoising/compression followed by perceptual remodulation, denoising and random deformation.

As for fragile/semi-fragile watermarks in the authentication applications, the requirement of robustness is reduced or is not required at all. Fragile watermark should be removed or distorted whenever the marked data is altered. Semi-fragile watermarks should be able to survive a moderate degree of desirable attacks such as lossy compressions or content-preserving manipulations but would be removed easily under other types of attacks. The sensitivity of the authentication watermarks under various attacks, in particular, the crop-and-replacement and etc., would be evaluated. The fragility tests rather than the robustness tests are required for authentication marking system. However, the line of demarcation between the malicious (undesirable) and non-malicious attacks is very application dependent.

2.4 Communication Model of Information Hiding

Information hiding process can be modelled as a communication system [21, 22, 23, 24, 25, 26]. We wish to communicate a message from the watermark embedder end to the watermark receiver end through communication channels. Figure 2.4 shows a way to map watermarking system into a framework of communication system where the watermark message, $\mathbf{s}$, is communicated. Let $\mathbf{I}$ be the original cover data, to which a message $\mathbf{s}$ is added. The modified data $\hat{\mathbf{I}} = \mathbf{I} + \mathbf{s}$, which is perceptually indistinguishable from its original version, $\mathbf{I}$. In this model of information hiding, there are two sources of noise $\mathbf{n}$ – the cover image $\mathbf{I}$ and the noise due to the processing $\mathbf{P}$ (such as compression and decompression, and the transmission channel noise). At the embedder side, the message is hidden into the cover data based on a watermark key. The detector can detect and/or extract the message hidden based on the key with or without the help of the original cover work. And $\hat{\mathbf{s}}$ denotes the recovered message at the receiver end.

The concept of communication theory is used to characterize information hiding system: one tries to hide the maximum amount of information into the cover work invisibly to maximize the robustness of the hidden message against attacks of removal. The amount of information allowed to hide in the channel is defined as channel capacity

Figure 2.4: A communication model of information hiding system.

of embedding. For simplicity, we assume the information hiding channel is Additive White Gaussian Noise (AWGN) channel, which is valid as long as the statistics of noises are Gaussian and the noises are simply added to the signal. The theoretical embedding capacity, C , under overall AWGN noise is achieved by applying Shannon's well-know channel capacity bound [27],

$$C = \frac{1}{2} \log_2(1 + \frac{S}{N}), \quad (2.1)$$

where $\frac{S}{N}$ is the signal to noise ratio, and for information hiding channel it is the ratio of the message variance to the noise variance. Note that the sources of noise include both the cover data and the processing. Let σ_s^2 , σ_i^2 and σ_p^2 denote the variance of the message to be transmitted, the cover data and the processing noise, respectively. The capacity of the data-hiding channel can be expressed as [21, 22, 23, 24]

$$\begin{aligned} C &= \frac{1}{2} \log_2(1 + \frac{\sigma_s^2}{\sigma_i^2 + \sigma_p^2}) \text{ bits/pixel}, & \text{or} \\ C &= \frac{m \times n}{2} \log_2(1 + \frac{\sigma_s^2}{\sigma_i^2 + \sigma_p^2}) \text{ bits}, \end{aligned} \quad (2.2)$$

where m , n are the row number and column number of the cover data, respectively. In general, the channel noise is mainly due to the host cover data, because the power of the original cover data is much stronger than that of the processing noise, i.e., $\sigma_i^2 \gg \sigma_p^2$. Thus the noise due to processing σ_p^2 can be neglected in the analysis of channel capacity for simplicity.

For the purpose of covert communication, the most important task is to maximize the hiding capacity of a channel. However, for the purpose of IP rights protection of the cover media (as in this thesis), the issue of balancing the following two requirements – the imperceptibility of the embedded watermark and its robustness against attacks – is of the prime importance.

Part I

Information Hiding for Transaction Tracking

Chapter 3

Introduction

To enforce Intellectual Property protection, the tracking of the distributed multimedia data is one of the most important tasks. When the Work is subsequently redistributed illegally, the owner is therefore able to find who should be responsible for the illegal distribution before taking legal action to protect his/her intellectual property rights (IPRs). One popular application of digital watermarking is for data monitoring or tracking, in which the user is interested in monitoring data transmission in order to control royalty payments, or simply track the distribution to localize the data for market and sales purposes. The works of Part I in this thesis concentrate on information hiding and retrieval techniques for transaction tracking applications.

3.1 Independent Component Analysis

The basic goal of Independent Component Analysis (ICA) is to find a transformation in which the components $\{\mathbf{s}_k, k = 1, \dots, n\}$ are statistically as independent from each other as possible [28, 29, 30]. ICA is one of the most commonly used method for performing blind source separation, in which it is used to separate statistically indepen-

dent unknown source signals $\mathbf{s}_k$ from their linear mixtures $\{\mathbf{x}_j, j = 1, \dots, m\}$ without knowing the mixing process [28, 29, 30]. ICA has widely applied in many areas such as audio processing, biomedical signal processing, telecommunications [28, 29, 30].

3.1.1 Statistical Independence

The fundamental assumption for ICA is that the source components are statistically independent. The blind separation performance can thus be determined by the verification of the statistical independence of the recovered sources.

Statistical independence can be defined by the probability densities. The joint probability density function, $f(x, y)$, for statistically independent random variables can always be factored into the product of two marginal density functions as [31]

$$f(x, y) = f_X(x)f_Y(y), \quad (3.1)$$

where f_X and f_Y are the two marginal densities for the random variables X and Y . Hence, one of the important consequences to be obtained is

$$E[XY] = E[X]E[Y]. \quad (3.2)$$

That is, the expected value of the product of two statistically independent random variables is the product of their mean values.

According to (3.2), if the variables are independent, they are uncorrelated. Equivalently, two random variables X and Y are said to be uncorrelated, if their covariance is zero [31]; i.e.,

$$E[XY] - E[X]E[Y] = 0. \quad (3.3)$$

A weaker form of independence is uncorrelatedness. On the other hand, uncorrelatedness does not imply independence [31].

3.1.2 Definition of ICA

ICA model consists of two parts: the mixing process and unmixing process.

In the mixing process [28, 29, 30], the observed linear mixtures $x_1, \dots, x_m$ of n number of independent components are defined as

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \dots + a_{jn}s_n; 1 \leq j \leq m, \quad (3.4)$$

where $\{s_k, k = 1, \dots, n\}$ denote the source variables, i.e., the independent components, and $\{a_{jk}, j = 1, \dots, m; k = 1, \dots, n\}$ are the mixing coefficients. In vector-matrix form, the above mixing model can be expressed as

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \quad (3.5)$$

where $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$ is the mixing matrix [28, 29, 30], $\mathbf{x} = [x_1 x_2 \cdots x_m]^T$, $\mathbf{s} = [s_1 s_2 \cdots s_n]^T$, and T is the transpose operator.

For unmixing process [28, 29, 30], after estimating the matrix $\mathbf{A}$, one can compute its inverse – the unmixing matrix, $\mathbb{W}^1$, and the independent components are obtained as:

$$\mathbf{s} = \mathbb{W}\mathbf{x}. \quad (3.6)$$

The basic simplest BSS model assumes that there are the same number of linear and instantaneous mixtures as the independent components or sources. The basic model has been extended in different directions, for instance, more sensors than sources, noisy observations, complex signals and mixtures, convolutive mixtures and so on [32].

¹The symbol $\mathbb{W}$ is used to differentiate the symbol $\mathbf{W}$ for the watermark signal.

3.1.3 Fundamental Restrictions

To assure the identifiability of the ICA model, following fundamental restrictions are imposed [28, 29]:

- The source signals in the mixing process should be principally statistically independent.
- All the independent components $\mathbf{s}_k$, with the possible exception of one component, must be non-Gaussian.
- The number of observed linear mixtures m must be at least as large as the number of independent components n , i.e., $m \geq n$.
- The matrix $\mathbf{A}$ must be of full column rank.

3.1.4 Summary of ICA Algorithms

There are many various ICA algorithms which have been proposed recently [28, 29, 30, 33, 34, 35, 36, 37, 38, 39, 40, 41]. Based on theories in probabilistic models, information theory and artificial neural networks, several unsupervised learning algorithms are developed that can perform ICA. The seemingly different theories such as infomax, maximum likelihood estimation, negentropy maximization, nonlinear PCA, Bussgang algorithm and cumulant-based methods are used to derive different objective functions whose optimization enables the estimation of the ICA model.

To summarize, the ICA algorithms can be classified basically from two aspects. Firstly, the process is to estimate all the components at the same time, or only a subset of them, possibly one-by-one. Secondly, the algorithm could be either adaptive or batch-mode (or block). In practice, on-by-one component estimating and the batch algorithm is more interesting and preferable, due to the computational and other

considerations [28].

Specially, four popular ICA algorithms are employed for performance evaluation of watermark extraction in Part I of the thesis. They are Cardoso's JADE [42, 43] (see Appendix A), Bell and Sejnowski's Infomax [44] (see Appendix B), Hyvärinen and Oja's FastICA [45] (see Appendix C) and Cichocki and Barro's RICA [35, 36] (see Section 4.3.2).

3.2 Motivation

The basic principles of most watermarking methods are applying small, pseudo random changes to the selected coefficients in the spatial or transform domain. Most of the watermark detection schemes use some kinds of correlating detector to verify the presence of the watermark with some confidence measure. The watermark can be recovered with information of the key and the original (i.e., un-watermarked) image.

One of the first and most cited watermarking schemes that make use of the original image for watermark detection and extraction is proposed by Cox *et al.* in [46]. In [46], the watermark coefficients are zero-mean, unit variance i.i.d samples having a Gaussian distribution. Watermark detection and extraction have been done by correlating the watermark image with the estimate of the embedded watermark that is obtained by subtracting the original image from the watermarked image. It is observed that the above technique as well as many other popular techniques is based on several assumptions/restrictions in watermark detection and extraction. For example, the watermark embedding locations, the original watermark and/or the original image are required for watermark detection and extraction.

Thus, the aim of Part I is to propose a new more general watermarking scheme.

Especially, it shall have a more general dewatermarking process, and is robust to a wide range of image manipulations.

ICA technique is one of the most widely used methods for performing Blind Source Separation (BSS). It is a very general-purpose statistical technique to recover the independent sources given only sensor observations that are linear mixtures of independent source signals. With the recent increase of interest in ICA, it has lots of interesting application in many areas. Applications of ICA can be found in areas such as audio processing, biomedical signal processing, image processing and telecommunications. We are motivated to further apply ICA technique in applications of digital watermarking as a tool for watermark detection and extraction.

The basic idea behind our work is to use some specific images, such as some special patterns or signature images as the key and the watermark for watermark embedding in order to create a watermarked image, which can be considered as an observed mixture image with the original image, the key and the watermark; hence, watermark recovery can be viewed as a blind source separation problem. The motivations and advantages of using ICA technique in watermark detection and extraction could be the following:

- Any of the three source images, i.e., the watermark, the key and the original image, can be recovered by the separation process, which is just like a reverse process of watermarking, hence, this process can be defined as “*dewatermarking*”.
- The watermark can be embedded either instantaneously or convolutively, that is, the observed watermarked image can be either instantaneous or convolutive mixture.
- The watermarked image can be extracted with or without the original image. When using both key and original image for extraction, there will be three mixtures can be generated (see Chapter 4 for detailed explanation), which is the case where the number of observed mixtures are as many as the independent source signals. Consid-

ering watermark extraction using only the key or the original image, there will be less observed mixtures than sources which is a special case defined as *overcomplete* ICA [33].

The simple subtraction method for watermark detection and extraction in [46] makes several assumptions regarding values of the mixing coefficients, distributions of the watermark as well as the mixing process. However, in our proposed BSS based method, we do not have restrictions on the mixing process as well as the mixing coefficients. The proposed method is more flexible (less restrictions) in terms of watermark detection and extraction process. It can also work for both the instantaneous and convolutive mixtures. For convolutive mixing case, the coefficients are not only multiplied, but also shifted/delayed. In such cases, it is quite difficult to extract unknown watermark using the existing simple methods, e.g. using [46].

3.3 Objective

The objective of Part I is thus to introduce ICA technique for watermark detection and extraction (defined as *dewatermarking*) scheme in digital watermarking. The feasibility of ICA method for watermark detection and extraction is investigated (in Chapter 4). The main factors for watermark extraction results using ICA methods are analyzed, and the possible solutions to further improve the extraction results are suggested (in Chapter 4). To improve the robustness performance of the watermarking system, a blind, wavelet-domain robust watermarking scheme is proposed based on ICA method for watermark extraction (in Chapter 5). A multimedia distribution framework by using the proposed ICA-based blind watermarking scheme for multimedia tracking, is further presented (in Chapter 6).

Chapter 4

Watermark Detection and Extraction Using Independent Component Analysis Method

Applications of Independent Component Analysis (ICA) can be found in many different areas. We further introduce ICA technique in digital watermarking for blind watermark detection and extraction. To present the basic principle of this new watermark extraction technique, our watermarking and dewatermarking experiments are firstly conducted with restrictions to the simplest ICA model: real source signals, linear and instantaneous/convolutive mixtures and as many mixtures as sources (i.e., the system requires both the key and original images for watermark detection and extraction).

To examine the feasibility of using ICA technique for watermark detection and extraction in a watermarking system, we firstly proposed a simple spatial-domain non-blind watermarking scheme [47, 48]. Specially, a robust, batch ICA algorithm [35, 36] is employed for the dewatermarking processing to illustrate the watermark detection and extraction process.

4.1 Watermark Design

Some early watermarking systems used to embed the watermark by *least significant bit* (LSB) insertion or manipulation [1]. It is a quick and easy way to hide information but is vulnerable to small distortions from image processing. In most of the previous works [1, 2, 7], binary patterns or random sequences are used as watermarks, which can only be detected by employing detection theory. During the verification process, the watermark is recovered by subtraction method [46], and compared with the original watermark. The presence of a watermark is determined by comparing the similarity measure between the extracted watermark and the original watermark with a pre-defined threshold.

Figure 4.1: (a) A special image pattern of size 256×256 , (b) a text image – ‘Nanyang Technological University’ logo (abbreviate as ‘NTUlogo’), (c) a Gold-like sequence image [47].

In our proposed watermarking system, some specific *grayscale* images, such as some special patterns (for example, Figure 4.1(a), (c)) or signature images (for example, Figure 4.1(b)), are used as a key or a watermark for embedding. In daily life, people claim the personal or organization properties by signing signature or stamping seal or company logo. This type of recognizable image pattern is more intuitive and unique than the random sequence to identify the ownership of the Work. By using grayscale

watermark, it is found more robust against attacks, because the grayscale images can always preserve a certain degree of structural information, which are meaningful and recognizable and also can be easier to be verified by human eyes rather than some objective similarity measurements.

4.2 Proposed Watermarking System

4.2.1 Watermark Embedding Scheme

In the generic watermark embedding scheme, the inputs to the system include the original data, the watermark and an optional public or secret key. The key is used to enforce the security, that is, to prevent unauthorized party from recovering and manipulating the watermark. Our proposed image watermarking system exploits a watermark, $\mathbf{W}$, and a secret key, $\mathbf{K}$, for the purpose of conducting two levels of security, by using the special images as the watermark and the key, with the same the size as the original image, $\mathbf{I}$, to be embedded. Spatial domain is used to perform the hiding operation.

Figure 4.2 illustrates the watermark embedding process. Both the watermark $\mathbf{W}$ and the key $\mathbf{K}$ are inserted in the spatial domain of the original image $\mathbf{I}$. The watermarked image $\mathbf{X}$ is a linear mixture of the original image $\mathbf{I}$, key $\mathbf{K}$ and watermark $\mathbf{W}$ with both $\mathbf{K}$ and $\mathbf{W}$ having signal energy sufficiently less (at least 10-100 times less) than the energy of the original image $\mathbf{I}$ in order to make them invisible. That is,

$$\mathbf{X} = \mathbf{I} + a\mathbf{K} + b\mathbf{W}, \quad (4.1)$$

where a and b are small weighting coefficients, or

$$\mathbf{X} = \mathbf{I} + a\mathbf{K} + b \otimes \mathbf{W}, \quad (4.2)$$

Figure 4.2: Watermark embedding process.

where $\mathbf{b}$ is the small filter coefficients and ' $\otimes$ ' denotes *convolution*.

4.2.2 Proposed Watermark Detection and Extraction Scheme

Private (or non-blind) watermarking system is adopted in the watermark recovery process, that is, the system requires both the key and the original data for watermark detection and extraction. The authorized party can detect and/or verify the watermark using both the secret key and the original data.

The watermarked image can be viewed as a mixture observation. There are totally three sources or independent components in this observation – the original image, the key and the watermark. To exploit ICA methods for image separation, it is required that the number of observed linear mixture inputs is at least equal to or larger than the number of independent sources, in order to assure the identifiability of ICA model [29]. Therefore, in this watermark extraction scheme, another two linear mixtures of the three independent sources are needed. By using both the key image $\mathbf{K}$ and the original image $\mathbf{I}$, two more mixed images are generated by adding them into the watermarked

Figure 4.3: Watermark extraction process.

image $\mathbf{X}$, respectively,

$$\begin{aligned}\mathbf{X}_1 &= \mathbf{X}, \\ \mathbf{X}_2 &= e\mathbf{X} + c\mathbf{K}, \\ \mathbf{X}_3 &= e\mathbf{X} + d\mathbf{I},\end{aligned}\tag{4.3}$$

where c and d are arbitrary non-zero real numbers, and e is an arbitrary real number. The proposed watermark extraction process is illustrated in Figure 4.3.

These three images are to be re-arranged into three row vectors $\mathbf{X}_j(k)$, with length of p each, (where $j = 1, 2, 3$ and $k = 1, 2, \dots, p$), to satisfy the input data requirements of a nonlinear blind extraction algorithm – robust batch ICA algorithm, which is used for dewatermarking process. The proposed watermark detection and extraction scheme is explained in the next section.

4.3 Robust Batch Algorithm for Watermark Detection and Blind Extraction

With the recent increase of interest in ICA, various algorithms have been proposed, based on probabilistic models, information theory, artificial neural networks and so on

[28, 29, 30, 34]. In the following experiments, a robust batch algorithm (RICA) [35, 36] which is based on the second-order statistics is used for watermark detection and blind extraction.

4.3.1 AR modelling Process for Image Sources

It is well known that images are spatially correlated. The pixel in the image is always highly correlated with its neighboring pixels. Images can be characterized as 1-D or 2-D autoregressive (AR) models in the time series analysis [49, 50, 51, 52]. In our algorithm, 1-D AR model for images is adopted to satisfy the input data requirements of robust batch ICA algorithm. The pixel processing is correlated with its past pixels, in another word, the image sequence is temporally correlated sequence. Image can be scanned in either column- or row-wise (shown in Figure 4.4) and re-arranged into 1-D sequence. To obtain a more temporal correlated sequence, it can be scanned in zig-zag column-wise or row-wise manner, as shown in Figures 4.5(a) and 4.5(b) respectively. Figure 4.5(c) shows another way for performing zig-zag ordering for the 2-D DCT transformed coefficients of a image. It is identical to the zig-zag process performed in the JPEG standard, since it can consistently yield a more temporally correlated data sequence for DCT-transformed image [53]. Different images have different scanning order to obtain their most temporally correlated sequences. The dewatermarking performance using different scanning sequences is illustrated later in Section 4.4.3.

4.3.2 Dewatermarking Process using Robust Batch Algorithm

A robust batch algorithm is presented in this section, where the source signals are modelled as an autoregressive process (AR) [35, 36]. Hence, the algorithm is an effective blind separation approach for temporally correlated sources. Images can be well

Figure 4.4: (a) Column-wise and (b) row-wise scanning for image source.

Figure 4.5: (a) Column-wise zig-zag ordering for the spatial source images, (b) row-wise zig-zag ordering for the spatial source images, (c) frequency zig-zag ordering for the DCT transformed source images.

represented as temporally correlated sequences. That is why, in the following batch algorithm, the use of AR model for image source is appropriate.

The authorized party can detect or extract the watermark only with the secret key. For the proposed watermark detection and extraction scheme [47], three mixtures can be generated as explained in Section 4.2.2. These three mixture images are then re-arranged into three row vectors $\mathbf{X}_j(k)$, with length of p each, (where $j=1,2,3$ and $k=1,2,\dots,p$), as the input of the nonlinear blind extraction algorithm, which depends on the second-order statistics of the source signals. This dewatermarking process consists of following two stages, i.e., PCA whitening process for watermark detection followed by the robust batch ICA algorithm for watermark extraction (shown in Figure 4.6).

Figure 4.6: Proposed watermark detection and extraction scheme.

PCA Whitening - Watermark Detection

Standard Principle Component Analysis (PCA) is often used for whitening, because it can compress information optimally in the mean-squared error sense, while filtering possible noise simultaneously [34]. The PCA whitening matrix is given by,

$$\mathbf{V} = \mathbf{D}^{-1/2} \mathbf{U}^T, \quad (4.4)$$

where, $\mathbf{D}$ is the diagonal matrix of the data covariance matrix $E[\mathbf{X}_j \mathbf{X}_j^T]$, and $\mathbf{U}$ is its eigenmatrix, and $E[\cdot]$ denotes the expectation operator. After whitening or sphering process, the data have been centered, i.e., with zero-mean. With the whitened data, the algorithm always converges better.

PCA whitening also provides a convenient means for estimating the number of sources or independent components, n , from the rank of the diagonal matrix $\mathbf{D}$ [34]. For watermarked case, the rank of $\mathbf{D}$ is equal to three, since there are totally three image sources – the original image, the key image and the watermark. On the other hand, if the image is unwatermarked, the three mixtures are actually the combinations of the original and key images only; hence, the rank of $\mathbf{D}$ will be reduced to two. Thus PCA whitening process provides an efficient means to distinguish the watermarked objects from the non-watermarked ones.

Figure 4.7: Block diagram of the robust batch ICA algorithm(m : number of mixtures, n : number of sources).

Robust Batch ICA Algorithm - Watermark Extraction

After pre-whitening process, the sources are recovered by iteratively estimating the unmixing matrix $\mathbb{W}^1$ through a simple batch learning algorithm. The convergence of the learning process is controlled by $\varepsilon_i(k)$ via a FIR filter whose transfer function is $\mathbf{B}_i(z)$ [35, 36]. Figure 4.7 shows the block diagram of the robust batch algorithm.

The output vector $\mathbf{y}(k) = \mathbb{W}\mathbf{x}(k)$. The error from the i -th iteration is defined as (sample index k is dropped for simplicity)

$$\varepsilon_i = y_i - \underbrace{\mathbf{b}_i^T}_{\tilde{y}_i} \tilde{\mathbf{y}}_i, \quad (4.5)$$

where $\mathbf{b}_i = [b_{i1}, b_{i2}, \dots, b_{iL}]^T$, $y_i = \mathbf{w}_i^T \mathbf{x}_i$, and $\tilde{\mathbf{y}}_i = [y_{i1}, y_{i2}, \dots, y_{iL}]^T$.

The cross-correlation matrix $\mathbf{R}_{\varepsilon_i \mathbf{x}_i}$ is computed in [35, 36],

$$\mathbf{R}_{\varepsilon_i \mathbf{x}_i} = E[\varepsilon_i \mathbf{x}_i^T] = \mathbf{w}_i^T \mathbf{R}_{\mathbf{x}_i \mathbf{x}_i} - \mathbf{b}_i^T \mathbf{R}_{\tilde{\mathbf{y}}_i \mathbf{x}_i}, \quad (4.6)$$

¹The symbol $\mathbb{W}$ is used to denote the unmixing matrix, which is different from the symbol $\mathbf{W}$ for watermark.

where $\mathbf{R}_{\mathbf{x}_i\mathbf{x}_i} = E[\mathbf{x}_i\mathbf{x}_i^T]$ and $\mathbf{R}_{\tilde{\mathbf{y}}_i\mathbf{x}_i} = E[\tilde{\mathbf{y}}_i\mathbf{x}_i^T] = \mathbf{R}_{\mathbf{x}_i\tilde{\mathbf{y}}_i}^T$.

Based on the principle of decorrelation, the optimum $\mathbf{w}_{i,opt}$ is obtained when the cross-correlation matrix $\mathbf{R}_{\varepsilon_i\mathbf{x}_i} = 0$,

$$\mathbf{w}_{i,opt} = \hat{\mathbf{R}}_{\mathbf{x}_i\mathbf{x}_i}^{-1} \hat{\mathbf{R}}_{\mathbf{x}_i\tilde{\mathbf{y}}_i} \mathbf{b}_i, \quad (4.7)$$

where $\hat{\mathbf{R}}$ denotes the estimated matrix of $\mathbf{R}$.

By applying standard PCA and normalization of signals to unit variance after each deflation procedure [35, 36], the autocorrelation matrix $\mathbf{R}_{\mathbf{x}_i\mathbf{x}_i}$ will be the identity matrix. Hence, the optimum $\mathbf{w}_{i,opt}$ can be further simplified as

$$\mathbf{w}_{i,opt} = \hat{\mathbf{R}}_{\mathbf{x}_i\tilde{\mathbf{y}}_i} \mathbf{b}_i. \quad (4.8)$$

For updating the vectors $\mathbf{b}_i$, the optimal minimum mean-squared error $E[\varepsilon_i^2]$ can be written as

$$E[\varepsilon_i^2] = E[y_i^2] + \mathbf{b}_i^T \mathbf{R}_{\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i} \mathbf{b}_i - 2\mathbf{R}_{\tilde{\mathbf{y}}_i y_i} \mathbf{b}_i^T. \quad (4.9)$$

The gradient of this function related to $\mathbf{b}_i$ will be given by,

$$\frac{\partial E[\varepsilon_i^2]}{\partial \mathbf{b}_i} = 2\mathbf{R}_{\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i} \mathbf{b}_i - 2\mathbf{R}_{\tilde{\mathbf{y}}_i y_i}. \quad (4.10)$$

By using the Wiener filtering, we have the optimum vector $\mathbf{b}_{i,opt}$ [35, 36],

$$\mathbf{b}_{i,opt} = \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i}^{-1} \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i y_i}. \quad (4.11)$$

Thus, in [35, 36], a heuristical algorithm is obtained that the updating vector $\mathbf{b}_i$ can be further simplified by removing the matrix inverse:

$$\mathbf{b}_{i,opt} = \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i y_i}. \quad (4.12)$$

Hence, the optimum updating vectors for $\mathbf{w}_i$ and $\mathbf{b}_i$ adopted in this algorithm are (4.8) and (4.12).

4.4 Evaluation

In this section, the performance measures for evaluation of the proposed watermarking system are presented (Section 4.4.1). Illustrative simulation results of dewatermarking are shown in Section 4.4.2 for watermark embedding both instantaneously and convolutively. These results are obtained when the watermark and the key are inserted in the spatial domain of the original image and the mixture images are all scanned into 1-D sequences in column wise (see Figure 4.4(a)). The watermark extraction performance is evaluated in term of *normalized correlation coefficient*. In Section 4.4.3, the main factors affecting the performance of watermark extraction by using ICA method are analyzed. The extraction results with embedding in the different domain and using different sequencing orders are compared and discussed. Watermark extraction performance against watermark embedding weighting coefficients and the use of various ICA methods are also analyzed and compared in Sections 4.4.4 and 4.4.5, respectively.

4.4.1 Performance Measures

After watermark embedding, the quality of the watermarked image is measured to evaluate the imperceptibility of the embedded watermark. Generally, a high quality measure of the watermarked image is desirable to ensure the transparent watermark embedding. The root-mean-squared error (RMSE) and peak signal-to-noise (PSNR) are used as error metrics to measure the watermarked image quality. The RMSE between the original image $\mathbf{I}$ (of $N_1 \times N_2$ pixels) and the watermarked image $\mathbf{X}$ is given by [2, 14]

$$\text{RMSE} = \sqrt{\frac{1}{N_1 \times N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} [\mathbf{I}(i, j) - \mathbf{X}(i, j)]^2}, \quad (4.13)$$

which represents the standard deviation of the error image. For an 8-bit (0-255) image, the PSNR (in dB) of the watermarked image is then computed as [2, 14]

$$\text{PSNR} = 20 \log_{10} \left(\frac{255}{\text{RMSE}} \right). \quad (4.14)$$

Embedding a watermark into a digital image, video, or audio signal should not reduce its perceived quality. Normally, the PSNR of the watermarked image should be above 40dB [21].

The robustness performance of watermark extraction is evaluated by *normalized correlation coefficient*, r , of the extracted watermark $\mathbf{W}'$ and the original watermark $\mathbf{W}$ (of $M1 \times M2$ pixels) [31, 54]

$$r = \frac{\sum_{i=1}^{M1} \sum_{j=1}^{M2} \widetilde{\mathbf{W}}(i, j) \times \widetilde{\mathbf{W}}'(i, j)}{\sqrt{\sum_{i=1}^{M1} \sum_{j=1}^{M2} \widetilde{\mathbf{W}}^2(i, j) \times \widetilde{\mathbf{W}}'^2(i, j)}}, \quad (4.15)$$

where $\widetilde{\mathbf{W}}$ and $\widetilde{\mathbf{W}}'$ are respectively the normalized original and extracted watermark by subtracting their corresponding mean values. The magnitude range of r is $[0, 1]$, and the unity holds if the extracted watermark perfectly matches the original one.

4.4.2 Simulation Results

Simulation experiments are conducted to demonstrate the feasibility of using ICA method for watermark detection and extraction. Some illustrative results for watermark detection as well as extraction are illustrated. Figure 4.8 illustrates an example about how a Cameraman image (of size 64×64 pixels) is watermarked. The watermarked image is quite identical to the original Cameraman image because of the small weighting coefficients a and b . Figure 4.9(f) shows extracted watermark from a watermarked Cameraman image of size (64×64) pixels. It has been found that the key and the extracted watermark are the reverse of their original images. For unwatermarked

Figure 4.8: (a) Cameraman - original image, (b) secret key image, (c) watermark image (Gold-like sequence), (d) watermarked image (PSNR=40.31dB).

Figure 4.9: Watermarked case: (a) watermarked image, (b)-(c) generated mixture images, (d) extracted key, (e) extracted Cameraman image, (f) extracted watermark.

Figure 4.10: Unwatermarked case: (a) unwatermarked image (original image), (b)-(c) generated mixture images, (d) extracted key, (e) extracted Cameraman image.

Table 4.1: Performance evaluation using Gold-like sequence as the watermark.

Test Conditions	Correlation Coefficient, r		
	Cameraman	Key	Watermark (Gold-like sequence)
Watermarked case	0.9978	-0.9945	-1.0000
Unwatermarked case	0.9978	-0.9946	N/A

case (Figure 4.10), only two images are extracted after separation – the original image and the key.

Table 4.1 shows the normalized correlation coefficients between the original and the extracted images for the examples described in Figures 4.9 and 4.10. The correlations for extracted Cameraman image and extracted key image are all very close to 1 or -1. For watermarked case, the watermark is perfectly extracted from the watermarked image except that a reverse version of the original watermark image was produced; thus, using minus sign to indicate such case. The robust batch ICA algorithm separates the images from the mixtures successfully. The extraction performance is quite satisfactory, which indicates the efficiency of the use of ICA method.

Figure 4.11 shows a watermarked image with convolutive embedding method using (4.2). Figure 4.11(a) is the frequency response of a 2-D filter, $\mathbf{b}$, of size $N \times N$ ($N=4$), which is used to convolve the watermark image shown in Figure 4.8(c). The original and watermarked Cameraman images are displayed in Figures 4.11(b) and 4.11(c), respectively. Figure 4.12 shows extraction results for convolutive mixing using the filter $\mathbf{b}$ shown in Figure 4.11(a). Figure 4.12(f) is the extracted watermark.

The correlation coefficient, r , between the extracted watermark in Figure 4.12(f) and its original, is found as 0.7063. Table 4.2 shows the performance index, r , for watermark extraction with respect to the size $N \times N$ of $\mathbf{b}$ filter. As N increases, the

Figure 4.11: (a) Frequency response of a filter, $\mathbf{b}$, in (4.2), (b) original Cameraman image, (c) watermarked image (PSNR=39.12dB) using the 2-D filter in (a).

Figure 4.12: Results of watermark extraction for convolutive watermark embedding model in (4.2): (a) watermarked image, (b)-(c) generated mixture images, (d) extracted Cameraman image, (e) extracted key, (f) extracted watermark.

filter bandwidth reduces due to sharper cut-off. Therefore, according to Table 4.2, the values of r become smaller with larger N . It is also noticed that r values are getting closer as N values increase from 6 to 48.

The batch algorithm that is based on second-order statistics performs well, when the source signals are temporally correlated [35]. This algorithm can be efficiently applied for natural images since they are not purely random but contain structures, so the pixels are highly correlated with their neighbors. It has been found that the results using Gold-like sequence as watermark gives an especially good extraction results. The

Table 4.2: Normalized correlation coefficient, r , with respect to size $N \times N$ of $\mathbf{b}$ filter.

$N \times N$	4×4	6×6	8×8	12×12	16×16	24×24	32×32	48×48
r	0.7063	-0.5775	0.5686	-0.5307	-0.5175	0.5079	-0.5054	-0.4999

Gold-like sequence [54] is generated by modulo-two addition of a pair of maximal linear sequences, with the output being the same length of the two maximal codes, which gives highly correlated sequence.

4.4.3 Performance Analysis with ICA

Extensive experiments have been carried out for different types of images including the original image, key and watermark. However, not all types of images can be well separated. This can be explained through following two main factors: the statistical dependence between image sources and the temporal correlation of each source.

1) Performance wrt Source Dependence

- Effects of Source Dependence

The basic starting point of ICA model is that the source components are statistically independent. The blind separation performance is thus to be determined by the verification of the statistical independence of the sources going to be recovered.

Another extraction example shown in Figure 4.13 demonstrates the effects of the source dependence. In this example, a text image instead of Gold-like sequence is used as the watermark. The normalized correlation coefficients for extracted images are measured against their original images, and shown in Table 4.3.

Comparing the two examples in Figures 4.9 and 4.13, it can be seen that the re-

Table 4.3: Normalized correlation coefficients, r , using text image as the watermark.

Camerman	Key	Watermark(text)
0.9859	-0.9945	-0.9447

Figure 4.13: Watermark extraction results using text image as watermark: (a) original Cameraman image, (b) key image, (c) text watermark, (d) watermarked image (PSNR=36.68dB), (e)-(f) generated mixture images, (g) extracted key, (h) extracted Cameraman image, (i) extracted watermark.

Table 4.4: Normalized correlation coefficients between sources.

Sources	Watermark	
	Gold-like sequence	Text
Cameraman & Key	-0.0820	-0.0820
Cameraman & Watermark	0.0089	0.1505
Key & Watermark	-0.0041	0.0052

sults using Gold-like sequence as the watermark gives a better extraction result than using text image. This can be explained from the viewpoint of statistical independence of the sources. The degree of independence can be again measured in terms of normalized correlation coefficients in (4.15). As mentioned before, independence implies un-correlatedness. If the image sources are independent, they are uncorrelated, so the correlation coefficients between three image sources should be zero. Table 4.4 shows the correlations between the sources in above two examples (Figures 4.9 and 4.13). The correlation between Cameraman and text image is 0.1505, which is quite high comparing with the correlation between Cameraman and Gold-like sequence, therefore, it

leads a worse extraction result.

- **Spatial Domain versus DCT Domain Watermark Embedding**

To reduce the dependence between the image sources, the key and the watermark to be embedded should be designed specially to keep each source highly temporally correlated, while maintain very low correlation between them – that is, high *intra*-correlation within each source and low *inter*-correlation between different sources. However, the correlations of the original image with the key or the watermark image cannot be guaranteed always very small. This can be improved by inserting the watermark and the key on the transform (for example, DCT) domain of the original image instead of its spatial domain.

Figure 4.14 illustrates an example about how an image $\mathbf{I}$ (Cameraman with size of 64×64 pixels) is watermarked in the DCT domain, using the same key and watermark images as Figure 4.13. The key image $\mathbf{K}$ and the watermark image $\mathbf{W}$ with very low energy level each are added into the DCT-transformed Cameraman image $\tilde{\mathbf{I}}$ to obtain a watermarked Cameraman image in the DCT domain, $\tilde{\mathbf{X}}$. Expectedly the recovered watermarked Cameraman image $\mathbf{X}$, after performing the inverse DCT transform of $\tilde{\mathbf{X}}$, would be identical to the original image $\mathbf{I}$. Thus the DCT domain image of the watermarked Cameraman image is

$$\tilde{\mathbf{X}} = \tilde{\mathbf{I}} + a\mathbf{K} + b\mathbf{W}, \quad (4.16)$$

where a and b are small weighting coefficients for reducing the energy of $\mathbf{K}$ and $\mathbf{W}$.

Figure 4.14(b) is a typical output image of the DCT transform process. Dark regions indicate less energy. Most of the energies are highly concentrated on the top-left corner. This kind of special image pattern is very different from the normal images such as the key image and the watermark image, therefore the correlations between the sources can be substantially reduced: the correlation between transformed Cameraman and the key

Figure 4.14: Improved watermark embedding scheme: (a) original Cameraman image, (b) DCT transformed Cameraman image, (c) secret key image, (d) text watermark, (e) mixture image (using (4.16)) for the images in (b)-(d), (f) watermarked Cameraman image (i.e. the inverse DCT transform of the image in (e)) with PSNR=39.81dB by using $a=0.005$ and $b=0.008$.

Table 4.5: Comparison of watermark extraction performance between spatial and DCT domain watermark embedding schemes.

Test Conditions	Normalized Correlation Coefficients, r		
	Cameraman	Key	Watermark (text)
Spatial domain embedding	0.9859	-0.9945	-0.9557
DCT domain embedding	1.0000 (DCT) 0.9991 (Spatial)	0.9996	0.9997

is now -5.4884×10^{-6} ; the correlation between transformed Cameraman and the text watermark is 0.0050. Using the images of Figures 4.14(b)-(d) for experiments, the normalized correlation coefficients between the original and the extracted images for both the spatial domain and DCT domain watermark embedding cases are presented in Table 4.5, where all the results are obtained with the same column-wise image sequencing order for comparison.

From Table 4.5, it shows that for the case where the dependence between the

sources is high, the dewatermarking performance can be improved by using DCT-domain watermark embedding, which is because the statistical dependence has been successfully reduced by adopting the DCT transform of the Cameraman image as compared with the spatial-domain of the Cameraman image.

2) Performance wrt Source Temporal Correlation

To demonstrate the importance of source temporal correlation, another watermark extraction example is illustrated in Figure 4.15, using the Boat image as the original image, the handwriting “Baltimore” as the key, and the Gold-like sequence as the watermark. The performance for column-wise and row-wise zig-zag sequencing methods with both spatial domain embedding has been examined. The performance measured in *normalized correlation coefficient* is shown in Table 4.6. For this example, the row-wise zig-zag sequencing makes significant improvements for extracted Boat and extracted key images. The extraction results have improved remarkably, which is a good example to verify the importance of temporally correlated sources in the robust batch ICA algorithm on blind separation.

The robust batch ICA algorithm models the source signals by unknown but stable AR process [37]. In the application of the algorithm on images, an image is characterized by a 1-D signal that appears at the output of a raster scanner, that is, a sequence of rows or columns. One property of AR model is that the linear prediction of the

Figure 4.15: Another watermark extraction example: (a) original image – unwatermarked Boat image, (b) key image (a handwriting image of “Baltimore”), (c) watermark (Gold-like sequence).

Table 4.6: Comparison of watermark extraction performance for column wise and row-zz wise image scanning using the watermarked Boat image measured with respect to the corresponding original images (zz denotes zig-zag).

Test Conditions	Normalized Correlation Coefficients, r		
	Boat	Key (Baltimore)	Watermark (Gold-like sequence)
Spatial - Column	-0.8936	0.9295	0.9996
Spatial - Row-zz	-1.0000	0.9962	-0.9996

Figure 4.16: The running window $\mathcal{W}$ of size (1×3) used to compute the average variance of the source.

current sample depends on the previous samples. An estimation of the sources is obtained based on the correlation of this sample with its previous samples. If the sources are more temporally correlated, the adjustment of the updating vectors is easier and faster to converge, hence, to complete the source separation.

The temporal correlation of a source can be measured from the average variance of the source using a running window $\mathcal{W}$ of size (1×3) as illustrated in Figure 4.16, where a source sequence is indexed as 1, 2, ..., 4096. Firstly the variance of the first three sample values as indexed 1, 2, 3 is calculated. Then for each time the window $\mathcal{W}$ is shifted right by one sample and the variance of the three windowed samples is calculated and recorded until the end of the sequence. The average variance is then obtained by taking the sum of all the variances and dividing the result by the total number of variances, which is the total number of samples minus the window size plus one. The higher the temporal correlation of the sources is, the lower the average variance will be.

Table 4.7: Relative comparison of average variances for column-wise sequence, column-wise zig-zag sequence, row-wise sequence, and row-wise zig-zag sequence (using a window of size (1×3)).

Image sources	column : column-zz : row : row-zz
Boat	1 : 0.9962 : 0.5619 : 0.4655
Baltimore	1 : 1 : 1.2638 : 1.2612
Gold-like sequence	1 : 1.0030 : 1.0062 : 1.0072

The average variances of the example in Figure 4.15 (in spatial domain) for column-wise sequence, column-wise zig-zag sequence, row-wise sequence, and row-wise zig-zag sequence are compared, and shown in Table 4.7. By using zig-zag sequence, the average variance is usually lower than without using zig-zag sequence, which indicates the temporal correlation can be increased by zig-zag sequencing for most of the images. For the Boat image the variance of the zig-zag sequence in row-wise order is about half the variance of sequence in column-wise order, because the pixels in each row are much more correlated than the pixels in the column. For this case, the *inter*-row dependence is much more significant than *inter*-column dependence. That is why higher temporal correlation is obtained for Boat image by row-wise zig-zag sequence by ignoring the *inter*-column dependence. From Table 4.7, column-wise zig-zag sequence of “Baltimore” handwriting image is more temporally correlated than its row-wise zig-zag sequence. The scanning order for the Gold-like sequence watermark does not affect its source temporal correlation, since the Gold-like sequence watermark is generated with highly correlated sequence. Depending on its content, each image has different sequencing order to obtain the most temporally correlated sequence. The image sequencing order should be appropriately chosen to achieve the optimum dewatermarking performance.

4.4.4 Performance against Embedding Weighting Coefficients

Figure 4.17 illustrates the extraction results of the watermarked image against the mixing coefficients a and b (in (4.1)), by using the embedding sources in watermarking example given in Figure 4.9. It shows clearly the performance depends mainly on the weighting coefficient for the watermark, b . The watermark extraction performance is quite consistent before the coefficient b decreases to 0.003. If the energy level of the embedded watermark does not go too low (i.e., $b > 0.003$), the performance of watermark extraction using ICA method is independent of the watermark embedding strength. As discussed earlier, the performance of watermark extraction is affected by statistical dependence of the sources and the source temporal correlation, rather than the watermark embedding strength.

Figure 4.17: Performance against embedding weighting coefficients: a , b are weighting coefficients for key $\mathbf{K}$ and watermark $\mathbf{W}$, respectively, as in (4.1).

4.4.5 Performance Comparisons with Various ICA Methods

1) Extraction Results Comparisons

Four ICA methods, namely, Cardoso's JADE [42, 43], Bell and Sejnowski's Infomax [44], Hyvärinen and Oja's FastICA [45] and Cichocki and Barro's RICA [35], are employed and compared in terms of watermark extraction performance in this section. Tables 4.8 and 4.9 show the extraction performance for watermarked Boat images in Figures 4.18 (b) and (c), respectively.

Figure 4.18: (a) Original Boat image of 256×256 pixels, (b) watermarked Boat image (PSNR=39.41dB) with Figure 4.1(a) as key and Figure 4.1(b) as watermark, (c) another watermarked Boat image (PSNR=41.70dB) with Figure 4.1(a) as key and Figure 4.1(c) as watermark.

The FastICA algorithm is a computationally highly efficient method for performing the estimation of ICA. It uses a fixed-point iteration scheme that has been found to be 10-100 times faster than conventional gradient descent methods for ICA [29]. JADE is an off-the-shelf (no parameter tuning) ICA algorithm based on the (joint) diagonalization of cumulant matrices. 'Good' statistical performance is achieved by involving all the cumulants of order 2 and 4, while a fast optimization is obtained by the process of joint diagonalization [42]. Both FastICA and JADE methods can obtain satisfactory watermark extraction performance.

Table 4.8: Watermark extraction performance for watermarked Boat image in Figure 4.18(b) (measured in term of normalized correlation coefficient, r).

ICA method	Performance index, r		
	Boat	Key	Watermark (NTUlogo)
FastICA	0.9272	-0.9433	-0.9579
Infomax	-0.9878	-0.9959	-0.9578
JADE	0.9766	0.9950	0.9580
RICA	-0.9846	0.9880	-0.8817

Table 4.9: Watermark extraction performance for watermarked Boat image in Figure 4.18(c).

ICA method	Performance index, r		
	Boat	Key	Watermark (Gold-like sequence)
FastICA	-0.8893	0.8952	0.9909
Infomax	-0.9556	-0.7250	-0.6469
JADE	0.9738	-0.9767	0.9909
RICA	-0.9994	0.9989	-0.9907

Infomax is a self-organizing learning algorithm, which maximizes the information transferred in a network of non-linear units. Classical Infomax algorithm is well suited to separate mixtures of super-Gaussian sources [44]. The sources are defined as super-gaussian, when their kurtosis, or 4th order standardized cumulant, is greater than zero. In these two examples, the kurtosis of Gold-like sequence is relatively low (1.0000), comparing with the kurtosis of ‘NTUlogo’ (9.9823), hence, it can be found that the watermark extraction performance for watermarked Boat image using Gold-like sequence image in Figure 4.18(c) (see Table 4.9) is not very satisfactory.

RICA is a simple second order statistic batch algorithm, which is able to efficiently extract various temporally correlated sources even the sources have small or even zero kurtosis [35]. Comparing the performance of the above four ICA methods, the results by using RICA algorithm for watermarked Boat image in Figure 4.18(c) (see Table

4.9), provide the best overall performance, although it is rather sensitive to source dependence (worse extraction results as in Table 4.8) [47]. It is noted that the source correlation between the original Boat image and Gold-like sequence is 5.061×10^{-4} , whereas the correlation between the original Boat image and ‘NTUlogo’ is quite large as 0.2119.

2) Stability Analysis

The experiments show that the JADE and RICA algorithms are always computationally stable in watermark extraction, despite of the small values of the mixing weighting coefficients. On the contrary, the stability (or the convergence) of Infomax and FastICA algorithms are much dependent on the values of the mixing coefficients.

For Infomax algorithm, providing the mixing matrix, $\mathbf{A}$, is non-singular, the original sources can be recovered if the weight matrix, $\mathbb{W}$ ², is such that $\mathbb{W}\mathbf{A}$ is a ‘permutation’ matrix containing only one high value in each row and column. Referring to the derivations in paper [44], the learning rule for an individual weight, w_{ij} , is

$$\Delta w_{ij} \propto \frac{\text{cof } w_{ij}}{\det \mathbb{W}} + x_j(1 - 2y_i), \quad (4.17)$$

where x denotes the mixture signal and y is the separation output; $\text{cof } w_{ij}$, the cofactor of w_{ij} , is $(-1)^{i+j}$ times the determinant of the matrix obtained by removing the i -th row and the j -th column from $\mathbb{W}$. When the weight vectors for two output units become too similar, $\det \mathbb{W}$ becomes small and the natural dynamic of learning causes these weight vectors to diverge. Therefore, the mixing coefficients for the key $\mathbf{K}$ and the watermark $\mathbf{W}$ would not be very close in order to perform the stable extraction by using Infomax ICA algorithm.

²The symbol $\mathbb{W}$ is used to denote the weight matrix in order to differentiate the symbol $\mathbf{W}$ for watermark.

The FastICA algorithm finds the independent components one at a time, which estimates only certain desired independent components, provided there are sufficient prior information of the weight matrices corresponding to those components [29]. However, in the watermarking systems, the watermark as well as the key is embedded by multiplying with small weighting coefficients to make them invisible. Therefore, the extraction of such weakly weighted watermark signal may fail. Moreover, the extraction by using the FastICA algorithm are much depended on the initial guess of the weight vector [29]. Extensive experiments also show that the extraction process using FastICA algorithm is not always converged to the same value, and it sometimes may even fail to retrieve the watermark due to the small weighting coefficients for embedding.

In conclusion, from the viewpoint of stability, it is more appropriate to choose JADE or RICA algorithms, rather than Infomax and FastICA algorithms, for the watermark extraction in digital watermarking.

3) Computational Complexity Comparisons

The computational complexities by using four above-mentioned ICA methods, e.g., FastICA, Infomax, JADE and RICA, are compared by varying image sizes of the original, the key and the watermark together. The complexity is measured by computation time in second as shown in Table 4.10. A Cameraman image (Figure 4.8 (a)), a pattern image (Figure 4.8(b)) and a Gold-like sequence image (Figure 4.8(c)) are used as the original, the key and the watermark image, respectively. Note that all results are obtained using PC with Pentium 4 2.6G processor. From the viewpoint of computational complexity, Infomax algorithm is not efficient for watermark extraction as image size increases.

Table 4.10: Computational complexity comparisons for various ICA methods by varying the image sizes (measured by time in second).

ICA method	Image Size (pixels)			
	64×64	128×128	256×256	512×512
FastICA	0.4680	0.6400	1.5160	4.5940
Infomax	1.6560	17.6720	245.5940	3.4907×10^3
JADE	0.0630	0.0780	0.2500	0.9370
RICA	0.0780	0.3130	1.9380	33.8290

4.5 Robustness Test

The watermarking system should be robust against data distortions introduced by the standard data processing and various attacks. It should be virtually impossible for the unauthorized users to remove it; and practically the image quality must be degraded explicitly before the watermark is lost. A list of attacks has been proposed, against which image watermarking system could be evaluated. These attacks include geometric transformations, JPEG compression, low pass filtering, additive noise, multiple marks, collusion and so on [16, 17, 18]. Figures 4.18(a) and 4.18(b) show an original Boat image and a watermarked Boat images using mixing model in (4.1) with $a=0.005$ and $b=0.008$. Various attacks are applied to this watermarked image to evaluate if the above-proposed watermarking system can recover the embedded watermark, thus measuring the robustness of the watermarking system against these types of attacks. The watermark extraction results for various attacks, which are obtained by using RICA algorithm as well as instantaneous watermark embedding, are illustrated as follows.

4.5.1 Geometric Transformations

The robustness testing using the watermarking example given in Figure 4.18(b) under attacks of some geometric transformations – Scaling, Cropping and Rotation, are presented in Figures 4.19(a), (b) and (c), respectively (using the absolute values of the normalized correlation coefficients). The system requires employing a synchronization template to detect the specific geometrical transformations, because it is needed to apply the same transformations to the key and the original images to generate another two observed mixtures. This is a restriction in handling the geometrical transformations for the proposed watermarking system. Nevertheless, experiments have been carried out with such synchronization template, proving that the information contained in the images after geometric transformations are still sufficient to retrieve the watermark by using ICA method.

- Scaling

Figure 4.19(a) shows an excellent behavior of the watermark extraction performance against scaling with factors from 0.5 to 2. Only when the scaling factor becomes very small and consequently the total number of pixels is very few, the extraction performance will be degraded. Due to the intrinsic robustness of the proposed watermarking scheme against this particular type of geometric distortion, the watermark extraction turns out to be quite resistant against all types of practical resizing algorithms.

- Cropping

The watermark and the key are permanently embedded into the original image by modifying intensity of all the pixels, therefore, the subparts of the watermark and the key are still present in the cropped watermarked image. Experiments have been carried out by supposing the position of the subimage can be determined properly. The performance is measured by comparing the correlations between the extracted

subimages and their corresponding subpart of the original image, respectively. Figure 4.19(b) shows that the information contained in a subimage is still sufficient to extract the embedded subparts of the key and the watermark by using ICA method.

- Rotation

The watermark extraction performance against rotation angle from 0 to 90 degrees is shown in Figure 4.19(c). The quality of the extracted Cameraman image is almost immune to the rotation angle. The worst case is when the rotation angle is 45 degrees, the normalized correlation coefficients for extracted key and watermark are 0.6794 and 0.6327, respectively, which can still be easily verified visually.

4.5.2 JPEG Compression

The watermarked Boat image is compressed by a quality factor of 80%, shown in Figure 4.20(a). The ‘NTUlogo’ is still visually perceptible in the extracted watermark image (Figure 4.20(b)). The proposed spatial domain embedding method can survive JPEG compression attacks up to quality factor 70% only.

4.5.3 Low Pass Filtering

Figures 4.21(a) and 4.22(a) are two watermarked Boat images filtered with a 2-D low-pass Gaussian and a 2-D median filter of size 5×5 , respectively. The tests (see Figures 4.21(b) and 4.22(b)) demonstrate that the watermarking system can survive against these types of low pass filtering attacks.

Figure 4.19: Robustness tests of the proposed dewatermarking scheme against (a) scaling, (b) cropping and (c) rotation attacks. (Note: The central portion of the image is kept after cropping, for example, 20 border lines are removed, meaning 10 lines are removed from top, bottom, left and right sides, respectively.)

4.5.4 Noise Addition

The watermarked Boat image is corrupted by the additive Gaussian noise with zero-mean and variance $\sigma^2 = 1000$, shown in Figure 4.23(a). Although the image seems to be highly degraded (comparing to its original image in Figure 4.18(b)), the watermark can still be recovered, which is presented in Figure 4.23(b). The maximum acceptable noise level is limited by the total energy of the embedded watermark. The tests show that the extracted watermark will become imperceptible, when the additive noise energy level goes up to 40–50 times higher than the energy level of the watermark.

4.5.5 Multiple Marks

Figure 4.24(a) shows a Boat image after inserting five different watermarks, which is another form of attack aiming to make original mark unreadable. Figure 4.24(b) shows the extracted watermark, which is clearly recognizable as the logo image ‘NTUlogo’ (see Figure 4.1(b)).

4.5.6 Collusion

In this experiment in order to simulate collusion attack, five separately watermarked Boat images are generated and averaged them to obtain another Boat image shown in Figure 4.25(a). The watermark ‘NTUlogo’ is still well exist in the extracted watermark image as shown in Figure 4.25(b).

Figure 4.20: (a) Watermarked Boat image after JPEG compression with quality factor 80% (PSNR=39.05dB), (b) extracted watermark from (a).

Figure 4.21: (a) Low pass Gaussian filtered (5×5) watermarked Boat image (PSNR=23.59dB), (b) extracted watermark from (a).

Figure 4.22: (a) Median filtered (5×5) watermarked Boat image (PSNR=25.07dB), (b) extracted watermark from (a).

(a)

(b)

Figure 4.23: (a) Watermarked Boat image (PSNR=20.85dB) with additive Gaussian noise (variance $\sigma^2=1000$), (b) extracted watermark from (a).

(a)

(b)

Figure 4.24: (a) Watermarked Boat image with five different watermarks (PSNR=28.96dB), (b) extracted watermark from (a).

(a)

(b)

Figure 4.25: (a) Watermarked Boat image after averaging together five independently watermarked versions of the Boat image (PSNR=42.93dB), (b) extracted watermark from (a).

4.6 Chapter Summary

In this Chapter, Independent Component Analysis (ICA) technique has been proposed for watermark detection and extraction in digital watermarking. Extensive experiments have shown the efficacy and efficiency of ICA method for performing watermark detection and extraction. In the presented simple non-blind spatial-domain watermarking scheme, the watermark is readily detected by Principle Component Analysis (PCA) whitening process. The watermark can be further separated from the mixed source using a robust batch ICA algorithm. The robust batch algorithm is an effective blind source separation approach for temporally correlated sources, hence, the image sources are characterized as 1-D temporal correlated sequences. The RICA algorithm is shown to be able to achieve overall best watermark extraction results for the proposed simple spatial-domain watermarking system. The performance of the proposed method can be evaluated in terms of normalized correlation coefficient. Simulation results show satisfactory dewatermarking performance with ICA method. It has also been analyzed how the performance of watermark extraction is affected by the statistical dependence of the sources and the source temporal correlation. Experimental results demonstrate the moderate robustness performance of the non-blind spatial-domain watermarking scheme to geometrical transformations (with a proper synchronization template), low pass filtering, multiple marks, collusion, and additive noise attacks.

However, the limitations of the simple non-blind spatial-domain watermarking system include:

- 1) Non-blind watermark detection and extraction could be a disadvantage in the practical applications where the original data is not available during verification stage;
- 2) The spatial-domain watermarking is comparatively easy to implement but less robust against compression. For copyright protection application, the robustness re-

quirement of a watermarking system would be very high. The spatial domain embedding is not able to achieve sufficient robustness performance against attacks.

3) The simple additive embedding the watermark and the key images is rather impractical. Without employing suitable masking function based on human visual system (HVS), the embedding of watermark could result in noticeable image distortions.

Blind watermarking scheme with frequency domain (in particular, wavelet-domain) embedding, is proposed in the next chapter to further improve the robustness performance, where the ICA method is employed for watermark extraction.

Chapter 5

A Wavelet-Domain Blind Watermarking Scheme Based on ICA

5.1 Introduction

In this Chapter, we propose a new blind watermarking scheme by embedding a digital image signature/watermark in the wavelet domain and using Independent Component Analysis (ICA) technique for blind watermark extraction. The watermark is inserted in the middle frequency sub-bands. A visual masking function is used to modify the digital signature to assure the perceptual quality of the watermarked image. Independent Component Analysis (ICA) method is employed for blind watermark extraction without requiring the original image. The presented technique has been successfully evaluated and compared with other wavelet-based blind watermarking techniques against the most prominent attacks, such as JPEG and JPEG2000 compression and quantization.

5.1.1 Wavelet-domain Embedding

In general, digital watermarking can be classified according to embedding domain, i.e. spatial-domain and frequency-domain watermarking. The spatial-domain watermarking is comparatively easy to implement, but less robust against compression and filtering. Frequency domain watermarking shows better robustness under most of the signal processing attacks. Moreover, the wavelet-based watermarking technique satisfies the requirement for Internet distribution – the progressive transmission property. Some existing blind wavelet-domain watermarking techniques are summarized in the following.

Wang *et al.* [55] proposed an adaptive watermarking method to embed watermarks into selected significant sub-band wavelet coefficients. A blind watermark retrieval technique was proposed by truncating selected significant coefficients to some specific value. Inoue *et al.* [56] classified wavelet coefficients as insignificant and significant by zerotree which is defined in the embedded zerotree wavelet (EZW) algorithm. Information data embedding algorithms in the locations of significant and insignificant coefficients have been developed. Information data are detected using the position of the zerotree's root and the threshold value after decomposition of the watermarked image. Dugad *et al.* [57] added the watermark in selected coefficients with significant image energy in the transform domain in order to ensure non-erasability of the watermark. During watermark detection, all the high pass coefficients above the threshold are chosen and correlated with the original copy of the watermark. Kundur *et al.* [58] presented a novel technique for the digital watermarking of still images based on the concept of multiresolution wavelet fusion, which is robust to a variety of signal distortions. In [59], Xie *et al.* developed a blind watermarking digital signature for the purpose of authentication. The signature algorithm is first implemented in wavelet transform domain and is coupled within the SPIHT compression algorithm.

5.1.2 Features of Proposed Scheme

Most of the existing watermarking schemes are always based on some assumptions for watermark detection and extraction. Some schemes require the previous knowledge of watermark location, strength or some threshold. In some algorithms, the watermark is estimated with the help of the original watermark information. Firstly, to ensure the robustness and invisibility of a watermark, the optimum embedding location are generally different for different images. For a large image database, it could be a disadvantage if requiring watermark location and strength information for detection and extraction, since a large amount of information are needed to be stored. In the Internet distribution application, the owner can always distribute the multimedia data by assigning different watermarks to different users in order to prevent illegal redistribution of data by a legal user. In such scenario, the watermark detection/extraction algorithms requiring the information of watermark location and strength, or the original watermark should fail, since the users do not know exactly which watermark is embedded in this copy of the watermarked image.

Thus a new blind watermarking scheme, which overcomes the problems discussed above, is proposed in this chapter. Some features of the proposed watermarking scheme with ICA-based watermark extraction are as follows:

- (a) The proposed method has a general framework for copyright protection and ownership identification, since extraction algorithm does not require any information on watermark locations and strengths. Only requiring an owner's copy of the original image and the key, the presented method can extract the watermark embedded in different copies of the watermarked image for ownership identification.
- (b) Instead of random sequences, some readable signature or pattern image is used as watermark. By using such a watermark, it is more robust against attacks,

because the signature or image pattern can always preserve a certain degree of structural information, which are meaningful and recognizable, can be more easy to be verified by human eyes rather than some objective similarity measurements.

- (c) A visual masking – Noise Visibility Function (NVF) [60] is used during watermark insertion to ensure the perceptual quality of the watermarked image.

5.1.3 Perceptual Mask

The goal of image watermarking techniques is to embed information that is transparent to human eyes. According to the perceptive model of human vision [61, 62], human eyes are more sensitive to the contrast between signals of one frequency and one orientation. Two most commonly used perceptual masks are:

- Noise Visibility Function (NVF): paper [60] presents the computation of NVF that characterizes the local image properties, identifying textured and edge regions where the mark should be more strongly embedded.
- Just Noticeable Distortion (JND): in the paper [63], JND mask for the images to be embedded is implemented by analyzing the noise sensitivity of every pixel based on the local region image content, such as texture, edge and luminance information. Each bit of the watermark is spread spatially and shaped by a pseudo-noise sequence such that its amplitude is kept below the noise sensitivity of the pixel into which it is embedded.

Implementation of NVF Mask

The most known form of the empirical NVF is widely used in image restoration applications [64]:

$$\text{NVF}(i, j) = \frac{1}{1 + \theta \sigma_x^2(i, j)}, \quad (5.1)$$

Figure 5.1: (a) The original Lena image (of 256×256 pixels), (b) the Noise Visibility Function (NVF) mask of Lena image.

where i and j are the index for the row and the column of the image; θ is a tuning parameter which must be chosen for every particular image. The version of NVF was the basic prototype for a lot of adaptive regularization algorithms. In paper [60], based on the non-stationary Gaussian model, θ is image-dependent and was proposed as

$$\theta = \frac{D}{\sigma_{x_{max}}^2}, \quad (5.2)$$

where $\sigma_{x_{max}}^2$ is the maximum local variance for a given image and $D \in [50, 100]$ is an experimentally determined parameter. The level of NVF varies from 0 to 1. Figure 5.1(b) gives the obtained NVF mask of the original Lena image. In very flat region NVF approaches 1 (white color), and NVF approaches 0 for highly textured region.

The final watermark $\mathbf{W}$ to be embedded is formulated according to the content adaptive strategy as

$$\mathbf{W} = (1 - \text{NVF})\mathbf{P}, \quad (5.3)$$

where $\mathbf{P}$ denotes the assigned signature or image pattern for generation of the watermark to be embedded. The watermark should embed in highly textured areas containing edges than in flat regions. By using (5.3), the strength of the embedded watermark approaches to zero in very flat region as its NVF measure approaches to one.

5.2 Proposed Watermark Embedding Scheme

This section presents the watermark embedding scheme in the wavelet domain. Watermark embedding locations are selected. The watermark generation and embedding procedure are presented in detail.

Figure 5.2 shows a second level wavelet decomposition of Lena image into four bands – low frequency band (LL), high frequency band (HH), low-high frequency band (LH), and high-low frequency band (HL). Sub-bands LL and HH are not suitable for watermark embedding among these four sub-band components. The image quality can be degraded if watermark is embedding in LL sub-band, since it contains most important information of an image. Sub-band HH is insignificant compared to LH and HL sub-bands, and watermark embedding in such sub-band is difficult to survive against lossy JPEG compression attacks. Watermark embedding in the two sub-bands (e.g. LH2, HL2 of the second level wavelet decomposition) consisting the middle frequency pair are demonstrated in this chapter.

Some digital signature/pattern or company logo (**S**), for example, a text image in Figure 5.3(a), can be used to generate the watermark (**W**) to be embedded. As pointed out earlier, such watermarks are more robust against attacks. They are very unique

Figure 5.2: Second level wavelet decomposition of Lena image.

Figure 5.3: (a) A text signature (64×64), (b) a NVF masking function, (c) modified text watermark based on the visual mask in (b), (d) original Lena image (256×256), and (e) a watermarked Lena image (PSNR=45.5dB).

to identify the ownership, and can be much easier to be verified by human eyes rather than some objective similarity measurements. A masking function – Noise Visibility Function [60] is obtained to characterize the local image properties, identifying the textured and edge regions where the information can be more strongly embedded. Such high activity regions are generally highly insensitive to distortion. With the visual mask, the watermark strength can be reasonably increased without visually degrading the image quality.

Figure 5.4 illustrates the watermark embedding procedure using second level decomposed middle frequency pair (LH2 and HL2):

[Step 1] Perform the second level discrete wavelet decomposition of the original image, **I**. Sub-bands LH2 and HL2 are selected for the watermark insertion.

[Step 2] The NVF masking function [60], **M**, of the original image is generated. Figure 5.3(b) shows a NVF mask of the Lena image. For the homogeneous region, NVF approaches 1 (white color), and the strength of the embedded watermark

Figure 5.4: Watermark embedding scheme (for the second level wavelet decomposition).

approaches to zero. The watermark should embed in highly textured regions containing edges instead of homogeneous regions. The original signature image S , is modified according to the masking function to assure the imperceptibility of the watermark embedded. The final watermark is quantized into $[0-7]$ gray levels. The expression for watermark generation is given as

$$\mathbf{W} = \mathbf{Q}_8[(1 - \mathbf{M}) \cdot \mathbf{S}], \quad (5.4)$$

where $\mathbf{Q}_8$ denotes the quantization operator with 8 gray levels. Figure 5.3(c) shows a text watermark generated using the NVF masking function shown in Figure 5.3(b).

[Step 3] The watermark $\mathbf{W}$ is embedded in the following way:

$$\begin{aligned} \mathbf{X}'_{LH2} &= \mathbf{X}_{LH2} + \alpha_{LH2} \cdot \mathbf{W} = \mathbf{X}_{LH2} + \alpha_{\mathbf{X}} \cdot \mu(|\mathbf{X}_{LH2}|) \cdot \mathbf{W}; \\ \mathbf{X}'_{HL2} &= \mathbf{X}_{HL2} + \alpha_{HL2} \cdot \mathbf{W} = \mathbf{X}_{HL2} + \alpha_{\mathbf{X}} \cdot \mu(|\mathbf{X}_{HL2}|) \cdot \mathbf{W}, \end{aligned} \quad (5.5)$$

where $\mathbf{X}$ and $\mathbf{X}'$ are respectively the wavelet transform coefficients of the original image and the watermarked image, α_{LH2} and α_{HL2} denote respectively the weighting coefficients of the watermark embedding in sub-bands LH2 and HL2, $\mu(|\cdot|)$ denotes the operation of mean of the absolute value, and a common control parameter α_X is used to adjust the watermark embedding strength to maintain the quality of the watermarked image within 40-50dB in terms of PSNR [65].

Table 5.1: PSNR (in dB) of the watermarked image with respect to α_X .

α_X	0.01	0.05	0.10	0.15	0.20	0.25	0.30
PSNR(dB)	67.50	53.52	47.50	43.98	41.48	39.54	37.96

[**Step 4**] The watermarked image, $\mathbf{I}'$, is obtained by the inverse discrete wavelet transform. Figure 5.3(e) gives a watermarked Lena image with PSNR 45.5dB. Table 5.1 lists the quality of watermarked image (PSNR in dB) with respect to the control parameter α_X .

For the security purpose in multimedia, the original data are always kept in secret and should not be known to the users during their watermark detection/extraction. In this chapter, the technique used to solve the problem is that the user can keep a copy of the dataset embedding with an owner's watermark following the same procedure shown in Figure 5.4. The owner's watermark is also known as the *key* of the watermarking system, which is used for watermark extraction. Using only the owner's copy, $\mathbf{I}'_0$, and the key, $\mathbf{K}$, the owner is able to identify the ownership of any copy of a watermarked image. This is elaborated in Section 5.3 in detail.

Figure 5.5(a) is an owner's copy of the Lena image embedded with an owner's watermark as shown in Figure 5.5(b). The owner's watermark is generated from the Gold-like sequence image [48] in Figure 5.5(c) by multiplying with a NVF masking function in Figure 5.3(b). The owner's copy is obtained similarly by embedding $\mathbf{K}$ in the wavelet domain using the following equations,

$$\begin{aligned}\mathbf{X}'_{0_{LH2}} &= \mathbf{X}_{LH2} + \alpha_{LH2} \cdot \mathbf{K} = \mathbf{X}_{LH2} + \alpha_{\mathbf{X}_0} \cdot \mu(|\mathbf{X}_{0_{LH2}}|) \cdot \mathbf{K}; \\ \mathbf{X}'_{0_{HL2}} &= \mathbf{X}_{HL2} + \alpha_{HL2} \cdot \mathbf{K} = \mathbf{X}_{HL2} + \alpha_{\mathbf{X}_0} \cdot \mu(|\mathbf{X}_{0_{HL2}}|) \cdot \mathbf{K},\end{aligned}\quad (5.6)$$

where $\mathbf{X}_0$ and $\mathbf{X}'_0$ are respectively the wavelet transform coefficients of the original image and the owner's watermarked image, $\alpha_{\mathbf{X}_0}$ is a control parameter for the visual

Figure 5.5: (a) The owner's copy of the original Lena image (PSNR=47.6dB and 256×256), (b) the owner's watermark or the key (64×64), and (c) the original Gold-like sequence pattern.

quality of the watermarked image $\mathbf{I}'_0$.

5.3 Proposed Watermark Extraction Scheme Using ICA Method

This section proposes an ICA-based blind watermark extraction scheme. Instead of using the original image, only an owner's copy of the original image is required for watermark extraction. The new useful features of the proposed scheme are that the present method does not require the previous knowledge of watermark locations and watermark strength information for watermark extraction. The main idea is to consider two sub-bands ($\mathbf{X}'_{\mathbf{R}}$) of the watermarked image to have a mixture image of the wavelet transformed image ($\mathbf{X}_{\mathbf{R}}$) of the original image $\mathbf{I}$ and the watermark image ($\mathbf{W}$). Figure 5.6 shows the blind watermark extraction scheme.

[Step1] Perform the second level discrete wavelet decomposition of the watermarked image $\mathbf{I}'$ in order to obtain the wavelet coefficients, $\mathbf{X}'_{LH2}$ and $\mathbf{X}'_{HL2}$ for the two selected sub-bands of LH2 and HL2.

Figure 5.6: The proposed blind watermark extraction scheme (using the second level decomposed images).

[Step2] The first mixture signal, $\mathbf{Y}_1$, is obtained by

$$\mathbf{Y}_1 = \mathbf{X}'_{LH2} + \mathbf{X}'_{HL2}. \quad (5.7)$$

From (5.5), $\mathbf{X}'_{\mathbf{R}}$ ($\mathbf{R} \in [LH2, HL2]$) are the mixture observations of the wavelet transform of the original image ($\mathbf{X}_{\mathbf{R}}$) and the watermark ($\mathbf{W}$), therefore, (5.7) can be rewritten as

$$\mathbf{Y}_1 = \mathbf{X} + \alpha_1 \mathbf{W}, \quad (5.8)$$

where $\mathbf{X} = \mathbf{X}_{LH2} + \mathbf{X}_{HL2}$ and $\alpha_1 = \alpha_X \cdot [\mu(|\mathbf{X}_{LH2}|) + \mu(|\mathbf{X}_{HL2}|)]$. It is found that the first mixture image is a linear mixture of the two independent source images, i.e., $\mathbf{X}$ and $\mathbf{W}$.

[Step 3] Repeat the same procedure in Steps 1 and 2 for the owner's image $\mathbf{I}_0'$. The second mixture, $\mathbf{Y}_2$, is obtained by

$$\mathbf{Y}_2 = \mathbf{X}'_{0LH2} + \mathbf{X}'_{0HL2}. \quad (5.9)$$

Similarly $\mathbf{Y}_2$ is also a linear mixture of the wavelet transform of the original image ($\mathbf{X}_{\mathbf{R}}, \mathbf{R} \in [LH2, HL2]$) and the key/owner's watermark ($\mathbf{K}$). It can be written as

$$\mathbf{Y}_2 = \mathbf{X} + \alpha_2 \mathbf{K}, \quad (5.10)$$

where $\alpha_2 = \alpha_{\mathbf{X}_0} \cdot [\mu(|\mathbf{X}_{0LH2}|) + \mu(|\mathbf{X}_{0HL2}|)]$.

[Step 4] From (5.7) and (5.9), two mixture images can be obtained containing three source images or independent components in the observations – $\mathbf{X}$, key $\mathbf{K}$ and the watermark $\mathbf{W}$. As pointed out earlier, to exploit ICA methods for watermark extraction, it is required that the number of observed linear mixture inputs is at least equal to or larger than the number of independent sources, in order to assure the identifiability of ICA model [28, 29, 30]. Therefore, another linear mixture of the three independent sources is needed. The third mixture $\mathbf{Y}_3$ can be generated by linear superposition of $\mathbf{Y}_1$, $\mathbf{Y}_2$ and $\mathbf{K}$:

$$\mathbf{Y}_3 = \beta_1 \mathbf{Y}_1 + \beta_2 \mathbf{Y}_2 + \beta_3 \mathbf{K}, \quad (5.11)$$

where β_1 , β_2 and β_3 are arbitrary real numbers. The coefficient β_3 should be non-zero. Either β_1 or β_2 can be set to zero to efficiently reduce the computational load of ICA.

[Step 5] The three mixtures input into the JADE algorithm [42, 43], and the watermark image, $\mathbf{W}'$, is extracted. The ownership is able to be verified directly from the extracted signature. Figure 5.7 shows the extracted watermark from the watermarked image shown in Figure 5.3(e).

Noted that JADE algorithm [42, 43], instead of RICA algorithm [35], is employed for watermark extraction. The discussions on the issue of using JADE algorithm is

Figure 5.7: The extracted watermark image using JADE ICA method (normalized correlation coefficient, $r \approx 1.0000$).

presented in Section 5.5.

5.4 Robustness Test

The results for robustness of the proposed watermarking scheme are shown in this section by using the original Lena image of size 256×256 pixels where the simulations are performed in the MATLAB 6.5 software environment. A watermarked image (PSNR=45.5dB) in Figure 5.3(e) is generated by inserting 389 bits information (with $\alpha_{\mathbf{x}}=0.125$). In the experiments of watermark extraction, the parameters β_1 , β_2 and β_3 are set as 0, 1, 1, respectively, to simplify the computational load of ICA; Daubechies-1 (Haar) orthogonal filters are employed for wavelet decomposition. In order to investigate the robustness of the watermark, the watermarked image is attacked by various signal processing techniques, such as JPEG and JPEG2000 compression, color quantization. The watermark extraction is performed against those attacks to the watermarked image and compared the extracted watermark with the original watermark. For a robust watermark scheme, it should be able to survive against various attacks before the quality of the watermarked image would be degraded drastically.

In the following, the robustness of the proposed watermarking scheme is compared with some other blind wavelet-domain watermarking techniques [65] in terms of *normalized correlation coefficient* r given in Section 4.4.1. These techniques include Wang's algorithm [55], Inoue's blind algorithm (based on manipulating insignificant coefficients) [56], Dugad's algorithm [57], Kundur's algorithm [58] and Xie's algorithm [59].

Figure 5.8 shows the comparison results in terms of performance index against JPEG compression. For the proposed scheme, the extracted watermark's correlation decreases gradually as with the compression quality factor. The image quality has degraded significantly to 27dB PSNR when the compression quality becomes quite low

to 10%. In this difficult case, the watermark can be still extracted with the value of the performance index of 0.2553 for watermark embedding in second level wavelet decomposition. According to Figure 5.8, the present scheme performs better than Wang's and Kundur's methods, and much better than Inoue's method in terms of robustness against JPEG compression attack when the compression quality factor is very low.

Figure 5.9 is the extraction comparison against the JPEG2000 compression attacks. The robustness of the proposed scheme is demonstrated up to the compression factor 0.05 or compression rate 0.4 bpp (bit per pixel). The proposed scheme has better result than Kundur's method, and comparable result to Wang's method. The extraction performance of Inoue's method drops sharply when the JPEG2000 compression factor goes to 0.125. Embedding in the sub-bands of higher level wavelet decomposition (see curves for extraction performance of third level decomposition in Figures 5.8 and 5.9), can improve the robustness of the proposed scheme against compression attacks. The possible reason for higher robustness against compression attack is that the energy spectrum becomes more flat (less compact) as the decomposition level increases.

Figure 5.10 shows extraction results against color quantization from gray level 256 to 4 per pixel. The proposed scheme has very good robustness result against color quantization. The performance of the proposed scheme can be comparable to that of the Xie's method, and much better than the other methods.

From Figures 5.8 and 5.9, it is found that Xie's and Dugad's methods have excellent robustness against JPEG and JPEG2000 compression. In Xie's algorithm, the watermark is embedded solely in the approximation image (LL sub-band) of the host image [59]. Although LL sub-band embedding is robust against compression attacks, the image quality could be degraded visually because the coefficients of this portion always contain the most important information of an image [65]. It is claimed that the

Figure 5.8: Comparisons of results against JPEG compression attacks.

Figure 5.9: Comparisons of results against JPEG2000 compression attacks.

Figure 5.10: Comparisons of results against color quantization.

robustness of Xie's method is determined by the number of decomposition steps. Very good robustness result can be obtained by employing a five-level wavelet decomposition using Daubechies-7/9 bi-orthogonal filters [59, 65]. Dugad's method embeds the watermark in the significant coefficients of all the detail sub-bands [57], therefore it is more resistant to compression. During the watermark detection in Dugad's method, the original watermark is required to compute the correlation with the high pass coefficients above a threshold [57]. The presence of a watermark is determined by comparing this correlation with a threshold setting. It is not as general as our proposed scheme in the case when the watermark embedded in a watermarked image is previously unknown.

The experimental results show that the proposed scheme has good robustness against the most prominent attacks – JPEG and JPEG2000 compression, color quantization, and can be comparable to the existing blind wavelet-domain watermarking schemes. The new watermarking scheme shows its main advantage in terms of generality. No previous knowledge of watermark strength and location is required for watermark extraction. Experiments also show that the type of wavelet transform used is not critical for the robustness of the proposed watermarking scheme (the corresponding results are not included), which is crucial for Xie's method from robustness point of view [59, 65]. Moreover, most of the existing algorithms need to set a threshold for watermark detection while using the correlation method. The another advantage of the proposed ICA-based method is that without using a threshold, the extracted watermark could simply be verified from visually rather than by using some objective measures or some thresholds, since the embedded watermark is a readable digital signature image or logo image.

5.5 Evaluation of the Extraction Performance with ICA

To evaluate the performance of ICA algorithms in watermark extraction for the proposed wavelet-domain watermarking, two illustrative watermarking examples are shown and analyzed. The Lena image of 256×256 (in Figure 5.1(a)) is used as the original image. The final key image to be embedded (in Figures 5.11(b) and 5.12(b)) is generated from an image pattern (in Figure 4.8(b)) by multiplication with a NVF masking function (in Figure 5.3(b)). Two watermarking examples are demonstrated by using two different watermarks. In Example 1, a text image in Figure 5.3(a) is used to obtain the watermark, while a Gold-like sequence in Figure 5.5(b) is used in Example 2. Note that the mixing sources in the wavelet-domain embedding are of size 64×64 pixels, because the subbands of the second level wavelet decomposition of Lena image (256×256) are used to hide the watermark. Also note that in the wavelet-domain embedding model both the key and the watermark are modified by the NVF masking function to ensure the imperceptibility of the embedded information. Figures 5.11 and 5.12 show the three mixing sources in these two examples, respectively. The final watermarked Lena images for the two examples are shown in Figures 5.13 (a) and (b), respectively.

The watermark extraction results by using Cardoso's JADE [43] and Cichocki and

Figure 5.11: The three mixing sources used in watermarking Example 1: (a) the original wavelet transformed image $\mathbf{X}$, (b) the key image $\mathbf{K}$, and (c) the modified text watermark image $\mathbf{W}$.

Figure 5.12: The three mixing sources used in watermarking Example 2: (a) the original wavelet transformed image $\mathbf{X}$, (b) the key image $\mathbf{K}$, and (c) the modified Gold-like sequence watermark image $\mathbf{W}$.

Figure 5.13: The watermarked Lena images: (a) Example 1: a text image is used to obtain the watermark (PSNR=45.56dB); (b) Example 2: a Gold-like sequence is used to obtain the watermark (PSNR=41.93dB).

Barro's RICA [35] are shown in Table 5.2. Recall the stability analysis of various ICA algorithms in Section 4.4.5, it has been investigated that both the Infomax and the FastICA algorithms are not always guaranteed to converge for watermark extraction. Therefore, only the extraction performance using JADE and RICA algorithms are further studied here.

Firstly comparing the extraction results of wavelet-domain embedding (in Table 5.2) and spatial-domain embedding (in Table 4.9), it is found that the spatial-domain watermarking technique can achieve overall better extraction results using ICA algorithms compared with the wavelet-domain watermarking technique. The fundamental assumption of ICA for performing blind separation is that the mixing source components are statistically independent with each other [30]. The source dependence for the examples illustrated in Figures 5.11 and 5.12, measured in terms of *normalized*

Table 5.2: The watermark extraction results using JADE and RICA algorithms measured in terms of *normalized correlation coefficient*, r ($\mathbf{X}'$ denotes the extracted image of $\mathbf{X}$).

Example 1				Example 2			
Extracted Image	$\mathbf{X}'$	Key	Text	Extracted Image	$\mathbf{X}'$	Key	Gold-like Sequence
JADE	0.9967	-0.9333	0.9924	JADE	-0.9879	-0.6501	0.9302
RICA	0.9977	-0.8455	0.9557	RICA	0.9979	0.9876	0.8111

Table 5.3: Normalized correlation coefficients between sources.

Sources	Example 1	Example 2
$\mathbf{X}$ and $\mathbf{K}$	0.0245	0.0245
$\mathbf{X}$ and $\mathbf{W}$	0.0409	-0.0079
$\mathbf{K}$ and $\mathbf{W}$	0.2596	0.4526

correlation coefficient r , is shown in Table 5.3. The degree of dependence between the key and the watermark are significantly increased in the wavelet-domain embedding than that in the spatial-domain embedding, because both of them are multiplied by the same NVF visual masking function. Hence, the extraction with ICA method is a more difficult task for the wavelet-domain embedding model than the spatial-domain embedding model.

Secondly it is observed that the JADE algorithm obtains better extraction results than RICA algorithm. Cichocki and Barro's robust batch ICA algorithm [35] is an effective blind separation approach particularly for the temporally correlated sources, since it models the signal as an autoregressive (AR) process [35]. This RICA algorithm thus can always achieve good extraction results if the embedding and extraction are conducted in the spatial domain. This is because generally speaking the natural images are spatially correlated, and can be effectively modelled as temporally correlated sequences. However, in the proposed blind watermarking scheme of this chapter, the

watermark is embedded in the wavelet domain instead of the spatial domain. The AR modelling of the mixing sources, including the wavelet coefficients, the key and the watermark that are modified by the NVF visual masking function, could be not very appropriate any more. The experiments show that the JADE algorithm [42] outperforms the other ICA algorithms in such scenario. The good performance of JADE algorithm might be due to its use of higher order statistical parameters, such as fourth order cumulant, which can model the statistical behavior of wavelet coefficients as well as the key and the watermark signals more effectively [42].

Thirdly, comparing the extraction results of Examples 1 and 2 in Table 5.2, it is found that the watermarking example by using text image the watermark has better extraction performance than using Gold-like sequence. Again it is because the source dependence of the key and the watermark in Example 1 is lower than that in Example 2. Therefore, to ensure the good extraction by using ICA algorithms, the key and the watermark would be carefully designed or chosen such that the statistical dependence between them is maintained as low as possible. The other solution to improve the extraction results would be employing two different visual masking functions for the watermark and the key images, which could consequently reduce the large correlation value between them due to multiplication of the same masking function.

5.6 Chapter Summary

In this Chapter, a new blind image watermarking technique based on Independent Component Analysis (ICA) is presented. Watermark is obtained by modifying the signature image with a visual mask to prevent perceptual quality degradation of the watermarked image. The watermark is inserted in the two middle frequency sub-band pair at a higher decomposition level (say second/third decomposition level) of the

DWT of the original image. Without requiring any information on original watermark, watermark location and strength, the proposed scheme can extract the watermark by using an owner's copy of the image and the owner's watermark/key. Experimental results show that the proposed watermarking scheme can provide good resistance under attacks of image compression and color quantization.

The advantage of using ICA algorithm for this blind watermark scheme is that, unlike other methods, no *a priori* information about watermark location and strength as well as threshold, is needed for our blind watermark extraction scheme. Therefore, it is possible to extract the watermark from any copy of the watermarked image, where the embedded watermark is previously unknown. The generality of the proposed scheme implicates the method to be a quite useful tool for the ownership identification and copyright protection in the application of Internet distribution. The only disadvantage to achieve the generality using ICA-based technique could be the complexity of the ICA itself. However, it has been compromised by using only three mixing sources in our watermarking models.

Next chapter further presents the transaction tracking issue by means of digital watermarking terminology to enforce IP rights for multimedia distribution. One feasible solution using an ICA-based multiple watermarking technique is proposed for transaction tracking. More experimental results are also included to evaluate the resistance of the proposed ICA-based watermarking scheme against other types of attacks.

Chapter 6

Transaction Tracking for Multimedia Distribution Through ICA-based Watermarking Technique

This chapter presents transaction tracking issue in multimedia distribution applications through digital watermarking. The IP protection issue has evolved increasing importance in the digital world we live in today. Digital watermarking technique is addressed, which provides effective means to protect and enforce copyrights. The existing watermarking schemes are summarized and their assumptions as well as the limitations in multimedia distribution applications are analyzed. These restrictions involving in most of the current watermarking schemes limit their uses in the tasks of efficient transaction tracking in the public networking environment. In particular, an Independent Component Analysis (ICA) based multiple watermarking scheme is proposed, which can overcome the problems of the existing watermarking schemes.

6.1 Multimedia Distribution Framework

Let consider the scenario when an owner would like to sell or distribute the Work for the registered users only. To enforce the IPR, two primary problems are to be solved. First of all, the owner shall be able to prove that he/she is the legal owner of the distributed content. Secondly, if the data have been subsequently copied and redistributed illegally, how it is possible for the owner to find the person who is responsible for the illegal copying and redistribution of the data (see Figure 6.1).

Figure 6.1: A multimedia distribution system, where the digital content could be illegally re-distributed to an illegal user.

The first technology adopted to enforce protection of IP rights is cryptography. The cryptographic technology [3] provides an effective tool to secure the distribution process and control the legal uses of the contents that have been received by a user. The content to be delivered in the Internet is encrypted, and only the legal users who hold a decryption key is able to obtain the encrypted data, otherwise the data stream would be useless to a pirate without an appropriate decryption key. However, for the error-free transmission through the network, the contents, after the decryption in cryptography, will be exactly the same as the original data. The data contents can be replicated perfectly as many times and the user can also manipulate the contents.

Researchers and scientists then turn to search other technologies to counter copy-

right piracy on the global networks that are not solvable by cryptography. In this context, recently *digital watermarking* technology [2] has been drawn a lots of attentions. In the digital watermarking the information are transparently embedded into the Work, rather than a specific media format, such as the header of a file that could be lost during transmission or file format transformation. Digital watermarking technique thus provides an efficient means for transaction tracking of illegal copying and distribution of the multimedia information. In a typical transaction tracking application, the watermark identifies the first legal recipient of the Work. If it is subsequently discovered that the Work has been illegally redistributed, the watermark can help to identify the person who is responsible for it.

Figure 6.2 presents a multimedia distribution framework to enforce the intellectual property rights through digital watermarking technique. In the *Internet* distribution application, the users firstly send requests to the content provider if they are interested in the multimedia contents. Then the owner can always distribute the Work by signing a watermark to a registered user to uniquely identify the recipient of the Work, as shown in Figure 6.2. All the data sent to the registered user are embedded with this assigned watermark as well as an owner's watermark while maintaining a good perceptual visual quality of the marked content. By employing such multiple watermarking technique [66, 67], IP rights are enforced from two aspects:-

- Every copy of the Work contains the owner's watermark to identify the rightful ownership of the Work.
- The owner or authorized representative can uniquely identify the recipient or the legal user of a particular copy of the multimedia content according to the embedded unique user's watermark.

Consider the case when the owner needs to prove the rightful ownership of a copy

Figure 6.2: The multimedia distribution framework by inserting an owner's watermark to identify the ownership of the Work and a unique user's watermark to identify each unique legal user.

of the content. The owner can present its original data (without any marks) of the Work as well as the owner's watermark as evidences. The two embedded watermarks including one owner's watermark and one user's watermark are therefore able to be extracted by a simple subtraction method [2, 46]. One extracted watermark, that is the owner's watermark, is matched with the previously presented owner's watermark. The rightful ownership of the content is thus verified. It is an essentially important pre-requisite for IP protection to embed the owner's watermark into every copy of the Work distributed in the Internet.

The more difficult and challenging task discussed in this chapter is to identify the legal users efficiently in the absence of the original data, and hence to trace transactions of a particular copy of the multimedia content. For the security purpose in multimedia, the original data are always kept in secret and should not be known to the public during watermark detection and extraction. In some real applications, the owner may

need authorize representatives or service providers to perform the transaction tracking tasks. For security reason, the owner also shall not give the original data to those representatives. Therefore, there arises a challenging issue – how to extract the user’s watermark in the absence of the original data. This is the main problem in transaction tracking through digital watermarking, which is discussed in this chapter.

6.2 Limitations of Existing Watermarking Schemes and The Solution

A wide range of watermarking algorithms has been proposed. This chapter focuses on the robust watermarking. Robust watermarking, as opposed to authentication marking, requires the embedded watermark robust against all types of attacks or should be able to survive against attacks before the quality of the watermarked image would be degraded drastically. The major researches of current robust watermarking techniques include the following key factors [1, 14, 15, 46, 62, 68, 69]:

- The choice of a workspace to perform the hiding operation, mainly a spatial domain [68], or a transform domain such as full-frame Discrete Cosine Transform (full DCT) [46, 70, 71], block DCT [72, 73, 74, 75, 76, 77], Fourier Transform (FT) [78], Mellin-Fourier [79, 80], or Wavelet [55, 56, 57, 58, 59].
- The study of optimal watermark embedding locations based on the human visual system [60, 62, 63, 81].
- The signal embedding techniques by addition, signal-adaptive addition or modulation methods [46, 71].

- The watermark detection and extraction algorithms either in blind [71] or non-blind manner [46].

Watermark recovery is usually more robust if its original, unwatermarked data is available. For example, a simple subtraction method [46] is used for watermark extraction at the exact locations where watermark is embedded. The presence of a watermark is determined by cross-correlation of the original and the recovered watermark. In [71], the selected DCT coefficients for watermark embedding, are directly matched with all the possible watermarks stored in the database. Therefore, without resorting the original image, the presence of a watermark is determined by comparing with a pre-defined threshold.

From the view point of presence of a given watermark at the extraction or verification stage, there are two different types of watermarking systems in the literature [1]. The first type is to embed a specific watermark or information. Most watermarking techniques for copyright protection belong to this watermarking category, where it is assumed that the embedded watermark is previously known. The objective of this type of watermarking scheme is to verify the existence of the previously known watermark with or without the help of this watermark. The second type, embedding arbitrary information, is, for example, useful for tracking unique receipt of a copy of distributed data. In such scenarios, the watermark embedded in an image copy is previously unknown, therefore, no prior information regarding the embedded watermark is available for watermark extraction. It makes transaction tracking a more difficult task.

Assumptions as well as limitations of most existing watermarking schemes are summarized in the following, which cause the difficulties and ineffectiveness in applying these watermarking schemes for IP protection of multimedia distribution:-

(a) In some watermarking algorithms, the watermark detection and extraction requires the presence of the original content. This is not a desirable solution, since the original data should be always kept in secret and should not be shown to the public, or sometimes the original data are even not available immediately. Therefore, *blind* watermarking techniques are of great interests and concerns nowadays.

(b) Most of the existing watermarking schemes [1, 2, 46] are based on some assumptions for watermark detection and extraction, such as the previous knowledge of watermark location, strength or some threshold. However, in order to ensure the robustness and invisibility of the watermark, the optimum embedding locations as well as the embedding strength are generally different for different images. For a large image database, it could be a disadvantage if it requires watermark location and strength information for the detection and extraction of the watermark. As a result, a large amount of side information is needed to be stored.

(c) As explained above, Figure 6.2 shows a framework to prevent illegal redistribution of the data by a legal user. In such scenario, the current watermark detection and extraction algorithms requiring the information of watermark location and strength, or the original watermark could fail, because no one knows which watermark exists in a received copy of the Work.

(d) Furthermore, the general watermark detection algorithm is based on a match filter finding the maximum correlation of the recovered watermark with the stored watermarks in the database, which contains watermarks used to identify all the possible users. It is rather a time consuming and inefficient process, especially when a large database is needed for distribution among a large number of users.

In this chapter, an Independent Component Analysis (ICA) based blind multiple watermarking scheme [48, 82], which can overcome the existing problems as mentioned

above. No *a priori* knowledge such as watermark locations, strengths or a threshold setting is required for watermark extraction. This presented watermarking algorithm is found very effective in the application of legal data tracking against other watermarking algorithms. Therefore, by adopting this ICA-based watermarking approach, an efficient multimedia distribution framework for copyright protection is accomplished.

6.3 ICA-based Watermarking Scheme for Transaction Tracking

This section presents watermarking scheme for color images using the blind ICA-based watermarking scheme proposed in Chapter 5. Both the watermark embedding and extraction are conducted in the selected channel, of a chosen color space. Specially, multiple watermarking technique [66, 67] is employed. Rather than one owner's watermark, two watermarks are to be embedded into two wavelet transform sub-bands of the selected channel, respectively. One is the owner's watermark (or the key of the watermarking system), and the other is a unique watermark assigned for a unique legal user. ICA technique is applied for extraction of user's watermark with the help of side information.

6.3.1 Embedding Channel Selection of Color Images

Looking into the field of image watermarking, most researches are done for grayscale images. The extension for color images can be simply accomplished by marking the image luminance such that it is color space independent [69, 83, 84]. A digital watermark can be placed in a color image with different color models, for instance, RGB, YUV, YIQ, YCrCb, HSI, $L^*a^*b^*$, and will survive when an image is converted from

one color space to another.

Watermark embedding and detection can be achieved in either one color channel or all of them [85]. In [86], Chae *et al.* proposed an approach for fusing color signature images in larger images using wavelet transforms and lattice structures in the Y component of YUV color space. Kutter [87] proposed an amplitude modulation scheme where signature bits are repeatedly embedded by modifying pixel values in the blue channel of RGB color space, because human eye is less sensitive to the changes in the blue color band. Fleet *et al.* [88] proposed an embedding scheme based on S-CIELAB, which is an extension of CIELAB standard for measuring color reproduction errors. Yellow-Blue (YB) color band of the opponent-color representation is used to embed amplitude modulated sinusoidal signals.

Embedding and decoding algorithms involving the inter-correlation between color channels were also proposed in [89, 90]. Piva *et al.* [89] exploited the characteristics of HVS and the correlation between the RGB image channels, to hide the watermark into the image transparently and to retrieve it reliably. The final results demonstrate the superiority of system based on the inter-correlation between color channels with respective to the luminance-based ones [89]. Zheng *et al.* [90] made use of the statistics characteristics of the correlation value in between the DCT coefficients of color channels Cr and Cb. The proposed schemes are shown to provide very good results in terms of both image transparency and robustness against attacks.

There is no color difference formula yet which can consistently model the human perception for color images. There still exists many unsolved problems in color image watermarking, such as, the most suitable color model, color channel and embedding algorithms, to be further investigated. In this chapter, as an illustration, a simple luminance channel embedding is used to show the efficiency of the proposed digital watermarking technique in IPR protection.

6.3.2 Watermark Embedding Scheme

Firstly the overview of proposed multiple watermark embedding scheme for color image is introduced. Then the watermark generation and the detailed embedding algorithm are demonstrated, followed by the side information generation for identification of the recipient.

1) Overview of Color Image Watermarking Scheme

The color image watermark embedding flowchart is shown in Figure 6.3. Assume the original image, $\mathbf{I}$, is in RGB color space. Firstly transform the color image into selected color space. The watermark embedding can use one channel of the selected color space or, all of them. There exist tradeoffs by embedding in one or all of the channels. By embedding all the three channels, the robustness of a watermark can be improved in terms of detection and extraction, while the visual quality of the watermarked image may be easily degraded without the proper use of human visual masking function. In this chapter, we illustrate the proposed ICA-based watermarking scheme by embedding only one channel, $\mathbf{C}$. The watermark is inserted into the selected channel adaptively according to the embedding algorithm shown in Figure 6.3. This is achieved by maintaining the watermarked image quality, for instance, its PSNR measure is within a pre-defined range, while tuning the parameter of watermark embedding strength. The final watermarked color image denoted as, $\mathbf{I}'$, is obtained by inverting the marked channel, $\mathbf{C}'$, together with the other two unselected channels back to the RGB color space.

The Y channel (the luminance component) of color Lena image in YUV color space is selected for watermark embedding. Figure 6.4 shows a second level wavelet decomposition of Y component of color Lena image into four bands – low frequency band

Figure 6.3: The flowchart of watermark embedding scheme for color image.

Figure 6.4: Second level wavelet decomposition of Y component of color Lena image.

(LL), high frequency band (HH), low-high frequency band (LH), and high-low frequency band (HL). Similarly, watermark embedding in the two sub-bands (e.g. LH2, HL2 of the second level wavelet decomposition) consisting the middle frequency pair is to be demonstrated.

Again some digital signature/pattern or company logo (**S**), for example, a text image in Figure 6.5(b), is used to generate the watermark (**W**) to be embedded. The NVF visual mask [60] is generated to identify the suitable embedding locations, that is, the highly textured and edge regions.

Figure 6.5: (a) A NVF masking function, (b) a text signature (64×64), (c) modified text watermark based on the visual mask in (a), (d) an owner's watermark or key of watermarking system, (e) modified key based on (a), (f) original color Lena image (256×256), and (g) a watermarked color Lena image (PSNR=44.80dB).

2) Watermark Embedding Algorithm

Figure 6.6 illustrates the multiple watermark embedding procedure using second level decomposed middle frequency pair (LH2 and HL2):

[Step 1] Perform the second level discrete wavelet decomposition of Y-component of YUV color space. Sub-bands LH2 and HL2 are selected for the watermark insertion.

[Step 2] The NVF masking function [60], $\mathbf{M}$, of Y-component of the original color image is generated. Figure 6.5(a) shows the NVF mask of Y channel of the Lena image. Similarly, the original signature image $\mathbf{S}$, is modified according to the masking function, and is quantized into [0-7] gray levels. The final watermark to

Figure 6.6: The proposed watermark embedding algorithm in the selected channel C (for the second level wavelet decomposition).

be embedded is given as

$$W = Q_8[(1 - M) \cdot S], \quad (6.1)$$

where Q_8 denotes the quantization operator with 8 gray levels. Figure 6.5(c) shows a text watermark generated using the NVF masking function shown in Figure 6.5(a).

[Step 3] The key K that is also the owner's watermark, is pre-processed by multiplying the same visual mask M as

$$K' = Q_8[(1 - M) \cdot K], \quad (6.2)$$

where Q_8 denotes the quantization operator with 8 gray levels. Figure 6.5(d) gives a key image and Figure 6.5(e) shows the modified key after pre-processing by using the NVF masking function in Figure 6.5(a).

[Step 4] The watermark W and the modified key K' are inserted into the LH2 and HL2 sub-band, respectively, in the following way:

$$\begin{aligned} X'_{LH2} &= X_{LH2} + \alpha_{LH2} \cdot W = X_{LH2} + \alpha_X \cdot \mu(|X_{LH2}|) \cdot W; \\ X'_{HL2} &= X_{HL2} + \alpha_{HL2} \cdot K' = X_{HL2} + \alpha_X \cdot \mu(|X_{HL2}|) \cdot K', \end{aligned} \quad (6.3)$$

where $\mathbf{X}$ and $\mathbf{X}'$ are the wavelet transform coefficients of the original and the watermarked selected channel, respectively. In (6.3), α_{LH2} and α_{HL2} denote the weighting coefficients for the watermark and the key embedding in sub-bands LH2 and HL2, respectively, while $\mu(|\cdot|)$ denotes the operation of mean of the absolute value. A common control parameter α_X in (6.3) is used to adjust the watermark embedding strength to preserve a satisfactory quality of the final watermarked color image [65].

[Step 5] The watermarked channel, $\mathbf{C}'$, is obtained by the inverse discrete wavelet transform. The watermarked color Lena image, $\mathbf{I}'$, is generated from the inverse color transform of the marked selected channel together with the other two original un-selected channels.

[Step 6] The Steps 4 and 5 are repeated until the quality of the final watermarked color image are satisfied, for instance, its PSNR measure is within the range of 40-50dB. Specially the parameter α_X is tuned to adjust the watermark strength to obtain the desired embedding result. Table 6.1 lists the quality of watermarked color image (in dB) with respect to the control parameter α_X . Decreasing the magnitude of α_X results a better quality of final marked color image and vice versa. Figure 6.5(e) gives a watermarked Lena image with PSNR 44.80dB.

3) Side Information for Watermark Extraction

As discussed earlier, the original data may be unavailable in many real applications for the security purpose. In order to identify the legal users, some side information

Table 6.1: PSNR (in dB) of the watermarked image with respect to α_X .

α_X	0.01	0.05	0.10	0.15	0.20	0.25	0.30
PSNR(dB)	67.75	54.32	48.32	44.80	42.30	40.36	38.78

Figure 6.7: The owner's copy of color Lena image (256×256 and PSNR=43.26dB).

is required for user's watermark extraction in the absence of the original data. The technique used in the proposed scheme to solve the problem is that the user can keep a copy of the dataset embedding with only owner's watermark following the same procedure shown in Figure 6.6. The owner's watermark is also known as the *key* of the watermarking system, which is used for watermark extraction. Using only the owner's copy, $\mathbf{I}'_0$, and the key, $\mathbf{K}$, the owner is able to identify the user of any copy of a watermarked image by ICA methods. This is elaborated in Section 6.3.3 in detail.

Figure 6.7 is an owner's copy of the Lena image, embedded with an owner's watermark shown in Figure 6.5(e). The owner's copy is obtained similarly by embedding $\mathbf{K}'$ in the wavelet domain using the following equations,

$$\begin{aligned}\mathbf{X}'_{0_{LH2}} &= \mathbf{X}_{LH2} + \alpha_{LH2} \cdot \mathbf{K}' = \mathbf{X}_{LH2} + \alpha_{\mathbf{X}_0} \cdot \mu(|\mathbf{X}_{0_{LH2}}|) \cdot \mathbf{K}'; \\ \mathbf{X}'_{0_{HL2}} &= \mathbf{X}_{HL2} + \alpha_{HL2} \cdot \mathbf{K}' = \mathbf{X}_{HL2} + \alpha_{\mathbf{X}_0} \cdot \mu(|\mathbf{X}_{0_{HL2}}|) \cdot \mathbf{K}',\end{aligned}\quad (6.4)$$

where $\mathbf{X}_0$ and $\mathbf{X}'_0$ are respectively the wavelet transform coefficients of the original selected channel and the watermarked channel, $\alpha_{\mathbf{X}_0}$ is a control parameter for the visual quality of the watermarked image $\mathbf{I}'_0$.

Suppose an owner wants to authorize a third party named as *appointed representative* to do the tracing task. In such case, the appointed representative should also be assigned a unique watermark by the owner. This watermark would replace the existing owner's watermark in the HL2 wavelet sub-band. In such scenario, the representative's

watermark would be used as the key during watermark extraction. However, for ownership verification the owner's watermark has to embed as well which could be in any other wavelet sub-bands.

6.3.3 Watermark Extraction Scheme Using ICA Method

This section presents blind watermark extraction process using ICA technique for multimedia transaction applications. Instead of using the original image, only an owner's copy of the original image is required for watermark extraction. The useful features of the proposed ICA-based scheme for multimedia tracking are that the presented method does not require the previous knowledge of watermark locations and watermark strength information for watermark extraction. The main idea is to consider two sub-bands ($\mathbf{X}'_{\mathbf{R}}$, $\mathbf{R} \in [LH2, HL2]$) of the marked channel ($\mathbf{C}$) in the watermarked image to have a mixture image of the wavelet transformed channel ($\mathbf{X}_{\mathbf{R}}$) of the original channel $\mathbf{C}$, the watermark image ($\mathbf{W}$), and the modified key image ($\mathbf{K}'$). Again the received watermarked color image $\tilde{\mathbf{I}}'$, undergoes the color conversion into YUV color space. The symbol ($\sim$) is to indicate the received data may or may not be the same as its original watermarked data due to transmission errors or possibly pirate attacks. This symbol ($\sim$) is removed for simplicity in the following. The watermark extraction is conducted in the Y component that is selected for the watermark embedding previously. Figure 6.8 shows the proposed blind watermark extraction scheme.

[Step1] Perform the second level discrete wavelet decomposition of the watermarked channel $\mathbf{C}'$ in order to obtain the wavelet coefficients, $\mathbf{X}'_{LH2}$ and $\mathbf{X}'_{HL2}$ for the two selected sub-bands of LH2 and HL2.

[Step2] The first mixture signal, $\mathbf{Y}_1$, is obtained by

$$\mathbf{Y}_1 = \mathbf{X}'_{LH2} + \mathbf{X}'_{HL2}. \quad (6.5)$$

Figure 6.8: The proposed blind watermark extraction scheme (using the second level decomposed images).

From (6.3), $\mathbf{X}'_{\mathbf{R}}$ ($\mathbf{R} \in [LH2, HL2]$) are the mixture observations of the wavelet transform of the selected channel ($\mathbf{X}_{\mathbf{R}}$), the watermark ($\mathbf{W}$) and the modified key ($\mathbf{K}'$), therefore, (6.5) can be rewritten as

$$\mathbf{Y}_1 = \mathbf{X} + \alpha_1 \mathbf{W} + \alpha_2 \mathbf{K}', \quad (6.6)$$

where $\mathbf{X} = \mathbf{X}_{LH2} + \mathbf{X}_{HL2}$, $\alpha_1 = \alpha_{\mathbf{X}} \cdot \mu(|\mathbf{X}_{LH2}|)$ and $\alpha_2 = \alpha_{\mathbf{X}} \cdot \mu(|\mathbf{X}_{HL2}|)$. The first mixture signal is a linear mixture of the three independent sources, i.e., $\mathbf{X}$, $\mathbf{W}$ and the modified key $\mathbf{K}'$.

[Step 3] Repeat the same procedure in Steps 1 and 2 for the owner's image $\mathbf{I}'_0$. The second mixture, $\mathbf{Y}_2$, is obtained by

$$\mathbf{Y}_2 = \mathbf{X}'_{0LH2} + \mathbf{X}'_{0HL2}. \quad (6.7)$$

Similarly $\mathbf{Y}_2$ is also a linear mixture of the wavelet transform of the selected channel of the original image ($\mathbf{X}_{\mathbf{R}}$, $\mathbf{R} \in [LH2, HL2]$) and the key/owner's watermark ($\mathbf{K}$). It can be written as

$$\mathbf{Y}_2 = \mathbf{X} + \alpha_3 \mathbf{K}', \quad (6.8)$$

where $\alpha_3 = \alpha_{\mathbf{X}_0} \cdot [\mu(|\mathbf{X}_{0LH2}|) + \mu(|\mathbf{X}_{0HL2}|)]$.

[**Step 4**] Using (6.5) and (6.7), two mixture images can be obtained containing three sources or independent components in the observations – $\mathbf{X}$, the modified key $\mathbf{K}'$ and the watermark $\mathbf{W}$. Again, to exploit ICA methods for watermark extraction, another linear mixture of the three independent sources is needed, in order to fulfill the requirement of identifiability in ICA model [28, 29, 30]. The third mixture $\mathbf{Y}_3$ is then generated by linear superposition of $\mathbf{Y}_1$, $\mathbf{Y}_2$ and $\mathbf{K}'$:

$$\mathbf{Y}_3 = \beta_1 \mathbf{Y}_1 + \beta_2 \mathbf{Y}_2 + \beta_3 \mathbf{K}', \quad (6.9)$$

where β_1 and β_2 are arbitrary real numbers, and β_3 is a non-zero arbitrary real number. Either β_1 or β_2 can be set to zero to efficiently reduce the computational load of ICA. Note that the modified key $\mathbf{K}'$ can be easily obtained by re-generating the NVF visual mask and multiplying to the original owner's watermark $\mathbf{K}$.

[**Step 5**] The three mixtures input into the JADE ICA algorithm [42], and the watermark image, $\mathbf{W}'$, is extracted. The user of the particular image copy is able to be verified directly from the extracted signature. Figure 6.9 shows the extracted watermark from the watermarked image shown in Figure 6.5(g).

Figure 6.9: The extracted user's watermark image using JADE ICA method (normalized correlation coefficient, $r=0.9927$).

6.4 Performance Evaluation

The results for robustness of the proposed watermarking scheme are shown in this section by using color Lena image of size (256×256) . A watermarked color image

(PSNR=44.80dB) in Figure 6.5(g) is generated by setting the watermark strength control parameter $\alpha_{\mathbf{x}}$ as 0.15. In the experiments of watermark extraction, the parameters β_1 , β_2 and β_3 are set as 0, 1, 1, respectively, to simplify the computational load of ICA; Daubechies-1 (Haar) orthogonal filters are employed for wavelet decomposition. The watermark extraction has been performed against various attacks to the watermarked image and compared the extracted watermark with the original watermark. The good robustness of the proposed scheme against the most prominent attacks – JPEG and JPEG2000 compression, color quantization – have been shown in Section 5.4. More robustness performance of the proposed watermarking system, for instance, against cropping and geometric distortions, is included in the following. The discussions on the use of the proposed ICA-based multiple watermarking scheme for multimedia tracking are then presented.

6.4.1 Robustness Against Cropping and RST Attacks

Many watermarking techniques cannot survive geometric transformations such as rotation, scaling and translation (RST) and sometimes cropping attack as well due to the loss of synchronization of the watermark detector. A solution to such geometric attacks is a re-synchronization process in either a blind or non-blind way before watermark extraction. Non-blind solution requires the presence of the original data or at least some knowledge of the image features [91]. Davoine *et al.* [92] proposed a non-blind solution by splitting the original picture into a set of triangular patches. This mesh serves as reference mesh and is kept in memory for synchronization pre-processing. This proposed method, however, is only efficient in the case of involving minor deformations. Johnson *et al.* [93] proposed a method to invert affine transformations by estimating the difference in the least square sense between the salient image feature points in the original and transformed images. Kutter [94] proposed alternative methods to retrieve

the watermark from geometric distorted image without using the original data. The first method is to pre-set part of the watermark to the known values and use them for spatial resynchronization. This approach decreases the hiding capacity for the useful information, and is also computationally very expensive. The second method [94] proposed by Kutter is to use self-reference systems that embed the watermark several times at shifted locations.

Generally speaking, the tuning process is easier, more accurate and requires less computational load when the original data or reference feature points is available, though it may need extra memory to store those reference data. In our proposed watermarking scheme, original data is not available during extraction, however, an owner's or a representative's copy of the data is available. This image copy would be very close to the original data, thus it is convenient to use it directly as a reference for synchronization of geometric distorted or cropped data. By simple comparisons, the tampered data can be adjusted back to original size and position rapidly and accurately. In the following, the watermark extraction results against attacks of cropping and RST are shown. The effectiveness of employing synchronization pre-processing is demonstrated by showing the significant improvements of extraction results with and without synchronization.

As shown in Figure 6.10(a), the face portion of a marked color Lena image is cropped. By comparison with the owner's Lena image copy, it can be easily detected that the pixels within a square area, with row index from 121 to 180 and column index from 121 to 180, is corrupted. The absence of the watermark information in this corrupted region (by considering both rows and columns from 31 ($\lceil 121/4 \rceil$) to 45 ($\lceil 180/4 \rceil$)) results in an undesired over brightness effect for the extracted watermark due to the high values in the corrupted region. This makes both the subjective and the objective verification measurements quite poor (see Figure 6.10(b)). One simple

Figure 6.10: (a) A cropped color Lena image, (b) the extracted watermark ($r=0.2706$) and (c) the extracted watermark after improving the contrast of (b) ($r=0.9036$).

solution is to discard the corresponding undesired high valued pixels from the extracted watermark and replace them with zero valued pixels. In this way, according to Figure 6.10(c), the extracted watermark can be recovered mostly with some information loss in the corrupted region. Therefore, the normalized correlation coefficient r is found to increase from 0.2706 to 0.9036 showing the recovery of the low contrast watermark (compare Figures 6.10(b) and 6.10(c)).

Our proposed watermark extraction of the received data under geometric attacks may fail due to the loss of synchronization. A re-synchronization pre-processing of the received data is necessary to tune it back to the right positions or sizes before further input to watermark decoder. The side information in the proposed watermarking scheme – the owner's or the representative's copy of the original image – provides a good reference to assist the synchronization process. The watermark extraction performance is consequently improved. Figures 6.11 (b), (d) and (f) show the extraction results under attacks of rotation, scaling and translation respectively, with the use of synchronization operations. The extracted watermarks are of satisfactory qualities in both subjective visualization and objective similarity measurement.

Figure 6.11: (a) A color Lena image rotated by 10° , (b) its corresponding extracted watermark ($r=0.5847$); (c) a color Lena image down-sized by reducing the number of rows and columns by $\frac{1}{4}$, and (d) its corresponding extracted watermark ($r=0.4970$); (e) a color Lena image translated to the left and downwards by 10 and 36 lines, respectively, and (f) its corresponding extracted watermark ($r=0.5356$).

6.4.2 Discussions

1) Watermarking versus Fingerprinting

To enforce the IPR for multimedia distribution in the Internet, it requires not only to verify about the legal recipient but also to prove the rightful owner of the original Work. The term *fingerprinting* [1, 8, 6] is closely related to watermarking in the context of the traitor tracking problem. Fingerprinting technique involves the embedding of a different label into each distributed copy of the Work. The application of fingerprinting concerns identifying the legal recipient rather than the source of digital data. Thus, using fingerprinting technique solely is not sufficient to enforce IPR protection in multimedia distribution systems, as the owner could not provide trustful evidence for proof of ownership. This chapter presents the multiple digital watermarking framework that can achieve the above two requirements – identifying the owner as well as the recipient.

Furthermore, fingerprinting has another distinct interpretation, which does not involve the concept of digital watermarking at all. It refers the extraction of unique features, for example, semantically relevant or characteristic features from multimedia signals, in order to distinguish itself from other similar objects [1, 6]. The extracted features are normally stored separately as signatures for authentication of the content rather than inserting into the content as watermarks. This concept falls out of the scope of this chapter.

2) Security Issues

Here we discuss about the security related issues of the proposed scheme. Firstly, the owner's watermark and the watermarks assigned to each user should be kept in secret for security reason. Secondly, we would like to point out that the use of a key and an owner's copy for watermark extraction actually provide two levels of security for the system. Even if an attacker knows the key, it is still unable to retrieve the embedded watermark without knowing the owner's copy of the data.

Let consider the scenario that the representative distributes the data on the owner's behalf. The owner will give the representative the data with owner's watermark embedded. The representative will then take the given data, i.e., the original image embedded with the owner's watermark, as the new "original" data. Thus the owner's watermark is hidden in all the data copies distributed by the representative. When the owner wants to prove the ownership, he/she can produce the unmarked original image and retrieve the owner's watermark by simple subtraction method.

The objective of collusion and averaging attacks are to average the difference between every copy of the same content, and hence to remove the embedded watermarks. For these types of attacks, the owner's watermark plays its importance for the security

of the transaction tracking system in case the representative distributes the data. No matter how attackers manipulate the watermarked image and try to remove the watermark, the “original” image they try to recover still contains the owner’s watermark.

Holliman and Memon *et al.* proposed a collage attack on certain block-based oblivious watermarking schemes [95]. They show that oblivious watermarking techniques that embed information into a host image in a block-wise independent fashion are vulnerable to a counterfeiting attack. However, our proposed scheme is to embed the watermark in two wavelet sub-bands (that is, it is not a block-based scheme), and watermark embedded is a pattern of 8 gray levels modified by human visual masking (that is, the watermark is not binary). Therefore, the proposed transaction system has better resistance against collage attack than those block-wise independent watermarking techniques.

In the future, we need further study the security issue of proposed system against copy attack proposed by Kutter [96]. We also need study the importance issue of image dependent key [97], and possibly design the key from the image itself to further improve the security of the proposed system.

3) Summary and Future Work of the Proposed Watermarking Scheme

The proposed watermarking scheme shows its main advantage in terms of generality. Unlike other methods, no *a priori* information about watermark location and strength as well as threshold, is needed for our blind watermark extraction scheme. Therefore, it is possible to extract the watermark from any copy of the watermarked image, where the embedded watermark is previously unknown. Moreover, most of the existing algorithms need to set a threshold for watermark detection while using the correlation method. The other advantage of the proposed ICA-based method is that without using

a threshold, the extracted watermark could simply be verified from visually rather than by using some objective measures or some thresholds, since the embedded watermark is a readable digital signature image or logo image. The generality of the proposed scheme implicates the method to be a quite useful tool for the ownership identification and copyright protection in the application of Internet distribution.

In the future work, more experiments will be carried out in order to evaluate the resistance of the scheme against other types of attacks. The study on generation of the perceptual mask which can best simulate the human visual model to further improve the visual quality of the marked data. The attacks such as collusion, copy attacks and the possible counterattack solutions in multimedia distribution applications are to be investigated to improve the robustness of the system.

6.5 Concluding Remarks and Future Trends

This chapter has presented ICA-based multiple watermarking scheme for transaction tracking to enforce IP rights of distributed digital media. Two watermarks consisting of an owner's watermark for ownership identification and a user's watermark for unique recipient identification are embedded. Watermark is obtained by modifying the signature image with a visual mask to prevent perceptual quality degradation of the watermarked image. The watermarks are inserted in the two middle frequency sub-band pair at a higher decomposition level (say second/third decomposition level) of the DWT of the original image. Without requiring any information on watermark location and strength, the proposed scheme can extract the user's watermark by using an owner's copy of the image and the owner's watermark/key. Previous experimental results show that the proposed watermarking scheme can provide good resistance under attacks of image compression, color quantization. This chapter also demonstrates

the good extraction performance against cropping and geometric transformations, with the help of the available side information.

It has been elaborated in this chapter that digital watermarking is employed as an efficient tool for multimedia tracking to enforce the IPR protection. From a wider point of view, the challenging issues of watermarking in applications of transaction tracking in the Internet include

- The original data is not available during extraction of the recipient's watermark. Thus the watermarking technique shall be *blind*.
- No prior information about the embedded watermark and the corresponding locations shall be required for watermark extraction.
- In order to present as trustful evidence in court to litigate the pirate, a watermarking scheme of excellent robustness against common signal processing attacks as well as collusion attacks is needed.
- For some applications, for example, Web crawlers searching for pirated watermarked image, it is required that the watermarking scheme is able to extract the watermark easily and with a low complexity.

No doubt that transaction tracking is a more difficult task than copyright protection by means of digital watermarking. More general watermarking techniques are urged such that no original data and prior watermark information are required for embedding, providing the methods are reliable, robust and computationally efficient.

Another requirement to enforce the IPR could be that the owner should be able to detect any illegal attempts to alter the distributed content. Fragile watermark would be suggested to be inserted as well to protect the integrity of the content. The authentication mark is very sensitive to attacks and can locate the positions of any

possible modifications. In such scenario, three watermarks would be hidden in the Work in order to verify the owner, to identify the user and to authenticate the content, respectively. Since the information hiding capacity of the cover media is limited, further challenge issue is to investigate how to compromise the requirements of three parties involving capacity for the information hiding, the imperceptibility of the embedded watermark and its robustness.

There has been no rush yet to embrace any of the current watermarking schemes for IPR protection of multimedia distribution. A period of inspection and appraisal to explore better solutions of digital watermarking schemes is still required, but first of all the scientists and researchers have to fully understand the practical requirements associated with the problem. In the meantime, the challenge is for the researchers to develop ever more invisible, decodable, robust or fragile watermarking methods, and perhaps to meet even more requirements as practical multimedia distribution systems are demanded.

Part II

Information Hiding for Ownership Verification

Chapter 7

Transparent Robust Information Hiding for Ownership Verification

7.1 Motivation and Objective

Digital watermarking technology has evolved very rapidly in recent years. The most common application of watermarking is to solve the problem of ownership of digital multimedia documents (e.g., audio, image, and video). For a robust copyright watermarking system, the robustness against attacks is of prime importance. The robustness of a watermarking system for ownership verification needs to be very high, so that the owner can still extract the hidden watermark to prove the ownership of the digital media even after watermarked image has been degraded significantly.

Generally, digital watermarking can be classified into two types depending on the embedding domain, i.e., the spatial- and frequency-domain watermarking. The spatial-domain watermarking is comparatively easy to implement but less robust against compression and filtering. Frequency domain (commonly used domains include DCT and

wavelet) watermarking shows better robustness under most of the signal processing attacks. Also the blind watermarking is of our interests because of its prominent applications in cases where the original data is not available.

For the robust watermarking system, when the amount of information to be embedded is fixed, the trade-off between the visual quality of the marked data and robustness of the watermark needs to be investigated. In general, the higher the embedding strength, the better robustness a watermarking system can achieve. However, at the same time it may result in a poorer visual quality of the watermarked image. The systematic benchmarking, StirMark [16, 17, 20], provides a common criteria for robustness performance evaluation against attacks. However, for a fair performance evaluation it should also include measurement of the distortions of a watermarked data introduced by watermark embedding.

The scope of Part II of the thesis is on a *blind*, *robust* and *transparent* DCT-domain watermarking system for copyright protection. As DCT is a commonly used coding technique in JPEG standard [52, 53], many robust watermarking schemes are proposed based on the DCT technique [46, 71, 98, 99, 100, 101]. In the following, some existing DCT-based watermarking schemes are summarized for the comparisons with our proposed watermarking scheme.

Cox *et al.* [46] proposed a well-known DCT-domain watermarking method, where the digital watermark is embedded into the most perceptually significant DCT coefficients by a non-linear fashion, however, the original image must be needed during the watermark extraction. In Piva's method [71], the selected DCT coefficients where the watermark is inserted, are directly correlated with all the possible watermarks. Therefore, without resorting to the original image, the presence of a watermark is determined by comparing it with a pre-defined threshold. Hsu and Wu [99] presented a binary watermarking technique by means of 8×8 discrete cosine transformation. The

binary watermark is permuted by a pseudo random number generator. The middle-frequency coefficients of each block are selected for embedding by XOR logic operation on the watermark block and the polarity block of the 8×8 transformed block. However, the watermark extraction requires the original image, which could be a disadvantage in some practical applications. Niu *et al.* [100] proposed a DCT-based gray-level digital watermarking technique. By stack filter's threshold decomposition technique, a gray-level digital watermark is decomposed into a series of binary digital images for implementing multiple watermarking [100]. Both the original image and watermark are required for watermark extraction. It is reported that the quality of the watermarked image (PSNR - Peak Signal to Noise Ratio) is not very high due to the multiple watermarks. Tsai *et al.* [101] also proposed a robust 8×8 block DCT based watermarking scheme, by adopting gray-level watermarks. Both the original image and digital watermark are processed by DCT transformation. An adjustable quantization table is built to reduce the information size of the watermark. It is noted that the middle frequency coefficients of the original image are randomly selected to be replaced by the quantized watermark. The experimental results show that the embedded watermark can resist various attacks and have superior robustness performance than the existing DCT-based watermarking scheme, however, the quality measurement of the watermarked image (in PSNR) is not very high due to the random replacement of DCT coefficients.

Thus the objective of this chapter is to propose a new blind robust DCT-based information hiding scheme to further improve the performance in terms of robustness and imperceptibility, so that it is more efficient for copyright protection. Better robustness of the proposed watermarking scheme against attacks is achieved than the existing DCT-based schemes, while a better perceptual quality of the watermarked image is always preserved. Both watermark embedding and extraction processes of our proposed scheme are performed in the whole DCT domain. The proposed watermarking

system is blind since no original image nor watermark is required for watermark extraction. The replacement operation rather than the conventional addition technique is employed for watermark embedding. The high robustness is obtained by selecting the less sensitive DCT coefficients against compressions for embedding. Nearest matching and adaptive embedding by a threshold setting can always ensure a good visual quality of the watermarked image while preserving good robustness performance.

The rest of this chapter is organized as follows. The proposed transparent robust information hiding scheme is introduced of the embedding scheme in Section 7.2 and the extraction scheme in Section 7.3. In Section 7.4, performance of the proposed watermarking scheme including both the robustness and the imperceptibility, is compared with other DCT-based robust watermarking schemes as described above. Discussions on the features of our proposed watermarking scheme as well as the potential applications are also presented in Section 7.5. Finally, the conclusion and the future works are given in Section 7.6.

7.2 Proposed Adaptive Watermark Embedding Scheme

The embedding technique of our proposed method is by *replacement* operation rather than *addition*. That is, to replace the DCT coefficients of the original image adaptively by the DCT coefficients of a watermark. The DCT domain is chosen because it has good energy compaction capability. The coefficients are normally classified into significant and insignificant portions. The proposed scheme embeds the watermark into relatively more significant portion of the coefficients to achieve higher capability of resisting various attacks. The grayscale images such as some pattern, logo or signature images

are adopted as watermarks. This type of watermarks can also provide unique ownership verification and can be easily recognized visually rather than by using an objective similarity measure. Note that we use a grayscale watermark image rather than a binary watermark. Since two grayscale images have similar energy distribution of the DCT coefficients, the mismatch between the DCT coefficient of the original image and that of the watermark image can be very low for the nearest match. Thus a grayscale watermark image can be embedded more transparently.

Suppose $\mathbf{I}$ be the original image of size $N1 \times N2$, and $\mathbf{W}$ be the grayscale watermark image of size $M1 \times M2$. Figure 7.1 shows an original Lena image, $\mathbf{I}$, and a grayscale text watermark image, $\mathbf{W}$, as used to illustrate our watermark embedding and extraction processes.

Figure 7.1: (a) The original Lena image (256×256 pixels) and (b) the grayscale watermark (38×111 pixels, ‘NTU’ is the abbreviation form of ‘Nanyang Technological University’) to be embedded.

7.2.1 Adaptive Watermark Embedding process

Figure 7.2 illustrates the proposed watermark embedding process, which includes the following steps:

[Step 1] Firstly, perform the 2-D discrete cosine transform (DCT) for both the origi-

Figure 7.2: Watermark embedding process.

Figure 7.3: The search engine for watermark embedding.

nal image $\mathbf{I}$ and the watermark $\mathbf{W}$. The DCT coefficients of $\mathbf{I}$ and $\mathbf{W}$ are denoted as $\mathbf{C_I}$ and $\mathbf{C_W}$, respectively.

[Step 2] A 1-D sequence, S_I , is then generated from the 2-D matrix of $\mathbf{C_I}$ by a zig-zag scanning rearranging the coefficients keeping the *most* significant coefficient first while the *least* significant coefficient to be the last [52, 53]. Thus the first coefficient of S_I sequence would be the DC component followed by the AC components with frequencies varying from low to high. According to Figure 7.3, the coefficients within the range of $[Initial, End]$, where *Initial* denotes the starting index and *End* denotes the ending index, are chosen for watermark embedding. Similarly, another sequence of S_W is generated by a column-wise scanning of the watermark DCT coefficient matrix $\mathbf{C_W}$.

Figure 7.4: The adaptive watermark embedding algorithms for i -th component of watermark DCT coefficient.

[Step 3] The embedding algorithm searches for the nearest match of each watermark coefficient $S_{\mathbf{W}}(i)$ in the selected range of the original image's DCT coefficients $S_{\mathbf{I}}$, in terms of their absolute amplitudes. To achieve good robustness of the watermarking system, low to middle frequency components (i.e., more significant coefficients) are chosen for embedding. The detailed adaptive watermark embedding algorithm is shown in Figure 7.4.

- As shown in Figure 7.4, the closest match for the i -th component of the watermark's coefficient $S_{\mathbf{W}}(i)$ in the selected range of the original image's coefficients is found, and the location of this nearest match of $S_{\mathbf{I}}$ is recorded in a vector – $location(i)$. For example, the location of nearest match for the

DC component $S_{\mathbf{W}}(1)$ is recorded in $location(1)$.

- Let define the difference, d , between the two nearest matches as

$$d(i) = | |S_{\mathbf{I}}(location(i))| - |S_{\mathbf{W}}(i)| |, \quad i = 1, 2, \dots, M1 \times M2, \quad (7.1)$$

where $S_{\mathbf{I}}(location(i))$ is the original image's DCT coefficient which is the nearest match for the i -th watermark coefficient.

- A threshold T_0 is set to control the quality of the marked image. If the difference d is smaller than T_0 , $S_{\mathbf{I}}(location(i))$ is replaced by the watermark coefficient $S_{\mathbf{W}}(i)$ directly. However, if the difference d is not less than T_0 , the coefficient $S_{\mathbf{I}}(i)$ is kept unchanged, instead a scaling factor, $scale(i)$, between the $S_{\mathbf{I}}(location(i))$ and $S_{\mathbf{W}}(i)$ is generated for a perfect watermark recovery in the watermark extraction. To represent mathematically, the watermarked image's DCT coefficients' sequence, $S_{\mathbf{X}}(j)$ ($\mathbf{X}$ denotes the final watermarked image, $j = 1, 2, \dots, N1 \times N2$), is

$$S_{\mathbf{X}}(j) = \begin{cases} \text{sign}\{S_{\mathbf{I}}(j)\} \cdot |S_{\mathbf{W}}(j)| & \text{if } j \in location(i), \quad i = 1, 2, \dots, M1 \times M2 \\ & \text{and } d(i) < T_0 \\ S_{\mathbf{I}}(j) & \text{else} \end{cases}, \quad (7.2)$$

and the scaling factors – $scale$ – are defined as follows:

$$scale(i) = \begin{cases} 1 & \text{if } d(i) < T_0 \text{ and } S_{\mathbf{I}}(location(i))S_{\mathbf{W}}(i) \geq 0 \\ -1 & \text{if } d(i) < T_0 \text{ and } S_{\mathbf{I}}(location(i))S_{\mathbf{W}}(i) < 0 \\ \frac{S_{\mathbf{I}}(location(i))}{S_{\mathbf{W}}(i)} & \text{if } d(i) \geq T_0 \end{cases} \quad (7.3)$$

The scaling factors are either +1 or -1 if difference d between the two nearest matched candidates is smaller than the threshold T_0 , depending on whether

they are of the same sign. If the candidates are of the same sign, *scale* equals 1; otherwise, *scale* is -1. Only when d is larger than the threshold T_0 , the scaling factor equals the ratio of these two candidates for the nearest match. Therefore, T_0 is employed as a control parameter for watermarked image's quality, and *scale* is an important refinement parameter for good watermark retrieval.

[Step 4] The *Key* of the proposed watermarking system includes both the watermark embedding location – *location* and the scaling factors – *scale*. Note that the dual key generated is image-dependent.

[Step 5] The watermarked image $\mathbf{X}$ is obtained by an inverse zig-zag scanning of the embedded sequence $S_{\mathbf{X}}$ followed by a 2-D inverse discrete cosine transform.

7.2.2 Automatic Embedding Range Selection

An automatic selection method of watermark embedding range of $[Initial, End]$ is proposed by determining the *End* index followed by an estimation of the *Initial* index.

1) Selection of *End* Index

JPEG is a standardized image compression mechanism. Lossy JPEG, which achieves much greater compression than lossless methods, is one of the most common signal processing attacks in digital watermarking technology. A nice property of the lossy JPEG compression is that the compression ratio can be varied by adjusting the *quality factor*. Therefore, it is possible to trade off the file size of the image against its quality factor varying from 100% (i.e., lossless) to 1% [52, 53].

In our proposed scheme, the *End* index of the embedding range is determined by

Figure 7.5: The *End* index selection for embedding range based on the relative changes of DCT coefficients with a reference of JPEG compression with quality factor 60%.

considering the robustness requirement of a watermark. One of the most important attacks, the effects of JPEG compression to the original image's DCT coefficients, is investigated. One would like to hide the watermark into the coefficients that are less sensitive to JPEG compressions. In particular, JPEG compression with quality factor 60% is used as a reference to analyze what portion of the DCT coefficients are more consistent against JPEG compression. The sensitivity of DCT coefficients against compression is measured by the relative change between the compressed coefficient $C_{compressed}$ and its original $C_{original}$, RC, defined by

$$RC = \frac{C_{compressed} - C_{original}}{C_{original}}. \quad (7.4)$$

As shown in Figure 7.5, it is observed that the frequency components of Lena image – roughly the first 8,000 most significant coefficients, are very robust against compression. Therefore, the *End* index of the embedding range for Lena image is set as 8,000 from the view point of the robustness of the embedded watermark. Specially, the sensitivity of DCT coefficients against JPEG compression with quality factor 60% is investigated to determine the *End* index of watermark embedding range experimen-

tally.

2) Selection of *Initial* Index

The *Initial* index of the embedding range is determined by considering the perceptual quality of the watermarked data. One may not want to touch those most significant coefficients of the original image to avoid the easily visualized changes caused by varying those coefficients. A weighted PSNR (wPSNR) [60] has been defined by Sviatoslav Voloshynovskiy *et al.* as an extension of the traditional quality metric – PSNR. In particular, a perceptual model of human vision - Noise Visibility Function (NVF) [60] - is used to weight each term of the PSNR. Thus, wPSNR is a HVS-based quality measurement, defined as

$$\text{wPSNR} = 10 \log_{10} \frac{\|255\|^2}{\|(\mathbf{I} - \mathbf{X}) \cdot \text{NVF}\|^2}, \quad (7.5)$$

where $\mathbf{I}$ is the original image, $\mathbf{X}$ is the watermarked image, and NVF denotes a visual masking function from (5.1). The level of NVF varies from 0 to 1. The magnitude of NVF approaches 1 in very flat region implying that the changes may be easily perceived for embedding in such regions.

The quality (in wPSNR) of the watermarked Lena image by selecting different *Initial* index is examined to estimate an optimal setting of *Initial* index. Once the *End* index is fixed, the DCT coefficients within the range of $[1, \text{End}]$ are divided into sub-bands with the octave-based division technique [102, 103], as illustrated in Figure 7.6. The coefficients of index $N \cdot 2^{-(s-1)}$ ($N = \text{End Index}$, $s \in \{1, 2, 3, \dots, \lfloor \log_2 N \rfloor\}$), are selected as *Initial* Index to test their corresponding qualities of the watermarked images. Figure 7.7 shows the visual quality for the watermarked image measured in wPSNR, with respect to the selection of the *Initial* index in the manner of the octave-band division. The optimal visual quality of the watermarked image is obtained when the *Initial*

Figure 7.6: *Initial* index selection by the technique of octave-band division. Figure 7.7: The watermarked image quality (in wPSNR) with respect to the selection of the *Initial* index.

index is 250 (see Figure 7.7). The embedding into the low frequency components below the index of 250 results in degradations in the visual quality. Therefore, for this watermarking example, the *Initial* index is set as 250 from the viewpoint of the visual quality of the watermarked image.

For embedding, the total number of DCT coefficients selected for the embedding range should be higher than the total number of DCT coefficients of the watermark image. For better nearest match, the embedding range needs to be chosen between 1.5 times to 3 times the length of the DCT sequence of the watermark. Because, if the embedding range is chosen too small as compared to the length of the watermark's DCT sequence, the number of selected coefficients would not be sufficient for good matching (see the lower wPSNR measures when $Initial > 250$). On the other hand, if the embedding range is too large, the robustness of the embedded watermark could

be degraded due to the embedding in the less significant coefficients. In addition, the perceptual quality of the marked content could be degraded as well due to the embedding in those most significant coefficients (see the lower wPSNR measure when $Initial < 250$). Moreover, it is rather a time-consuming process if the searching range is very large.

7.2.3 Threshold Setting

After fixing the watermark embedding range, the watermarked image quality is controlled by the threshold setting T_0 . One would like to ensure that the difference between the watermark coefficient and the original image's coefficient to be replaced is always lower than a tolerance (T_0) allowed, so that the good visual quality of a watermarked image can thus be guaranteed. The criterion to set the threshold T_0 is the total distortions allowed in a watermarked image. Generally, to make the watermarked image visually indistinguishable from the original image $\mathbf{I}$, the PSNR of $\mathbf{X}$ needs to be maintained above 40dB [2, 21]. The requirement of threshold setting T_0 can be derived from the following relationships:

$$\begin{aligned} \text{PSNR} &= 10 \log \left(\frac{255^2 \times \text{Total no. of pixels in the original image}}{\Delta E} \right) \geq 40\text{dB} \\ \text{and } T_0 &\leq \sqrt{\frac{\Delta E}{\text{Total no. of watermark samples}}} \end{aligned} \quad (7.6)$$

where ΔE is the total energy difference allowed between the watermarked image and the original image. If the original image is of 256×256 pixels and the watermark is of length 38×111 samples, the threshold T_0 is obtained as

$$\begin{aligned} T_0 &\leq \sqrt{\frac{\Delta E}{\text{Total no. of watermark samples}}} \\ &= \sqrt{\frac{255^2 \times 256^2}{10^4 \times (38 \times 111)}} \\ &\doteq 10.0514. \end{aligned} \quad (7.7)$$

For the given watermark embedding example, the threshold T_0 is to be set as 10.0514.

7.2.4 Dual Key Generation

The *Key* of the proposed watermarking system includes the watermark embedding positions – *location* as well as the scaling factors – *scale* (in (7.3)). Both location and scale information are working important roles during watermark extraction. Without anyone of them, the watermark extraction would fail. Note that the dual key of the algorithm is image-dependent, which implicates only the person, who knows the key, is able to do the ownership verification for the particular image content [97]. The dual-key watermarking enforces a higher watermarking security level, in the sense that even by knowing only the embedding location (or scale), the hacker is still unable to recover the watermark.

Most of the scaling factors are either +1 or -1 depending on whether the two nearest matches found are of the same signs; only when the difference d between the two nearest match is larger than the threshold T_0 , the scaling factor is the ratio between the original image's coefficient and the corresponding watermark's coefficient. Since positive ones and negative ones can be represented by 0 and 1 two binary codes, they can be coded by Run-Length Coding (RLC). The other scaling factors (rather than ± 1) could be encoded with Huffman or Arithmetic coding. Thus, the length of the key is altered by the total number, n , of the scaling factors with non-unit magnitude (see (7.3)). Therefore, the smaller n is, the less number of bits or payload are required to code the key. The value of n is determined by the threshold setting T_0 , such that the higher the threshold T_0 the lower the value of n . There exists a trade-off among the threshold T_0 , key length and watermarked image quality. The lower threshold setting ensures improvement of the watermarked image quality, while a higher value of n is required

Figure 7.8: (a) A watermarked Lena image (PSNR=61.20dB), (b) the noise-like difference image between the original Lena image and the watermarked Lena image in (a), (c) the enlarged eye region of the watermarked Lena image, and (d) the enlarged eye region of the original Lena image.

meaning that one has to allocate more number of bits to encode the secret key.

7.2.5 Watermark Embedding Results

A watermark embedding example is shown by using 256×256 Lena image (in Figure 7.1(a)) as the original image and 'NTU' (in Figure 7.1(b)) as the watermark. Set *Initial* index 250, *End* index 8,000 and threshold parameter T_0 as 10.0514. Figure 7.8 shows the embedding result – a watermarked Lena image with PSNR measure of 61.20dB. The enlarged eye region of the watermarked Lena image (in Figure 7.8(c)) and that of the original Lena image (in Figure 7.8(d)), are almost identical.

Table 7.1 shows the watermark embedding results for some standard test images

Table 7.1: The watermark embedding results: $[Initial, End]$: Embedding range; n : Total number of non-unit magnitude in $scale$ sequence (see (7.3)); PSNR and wPSNR: Quality of the watermarked image.

Test Image	Lena	Baboon	Pepper
<i>Initial</i> Index	250	40	265
<i>End</i> Index	8,000	10,000	8,500
n	85	100	127
PSNR (dB)	61.20	63.86	54.75
wPSNR (dB)	63.84	72.65	57.30

(a)

(b)

(c)

(d)

Figure 7.9: (a) The original Baboon image, (b) the watermarked Baboon image (PSNR=63.86dB), (c) the original Pepper image, and (d) the watermarked Pepper image (PSNR=54.75dB).

– Lena, Baboon and Pepper images. The selected range for watermark embedding, the number of scaling factors with non-one magnitude and the PSNR and wPSNR (in dB) measurements of the watermarked image are listed. Note that the watermarked images all preserve very good quality measurement, and we will further elaborate the excellent robustness of the proposed watermarking scheme.

It is found that the watermark embedding range for the Baboon image is much wider than that of Lena or Pepper image. The Baboon image contains very rich textures and edges such that it is much easier to hide information invisibly into such a channel of high capacity. Without employing the threshold T_0 , that is, simply replacing the original image's coefficients directly by the corresponding nearest matches of the watermark coefficients, it may result in an unacceptable perceptual quality of the watermarked data. On the other hand, without the scaling factors *scale*, the perfect reconstruction for the watermark during the extraction is impossible.

The payload to achieve good imperceptibility in the proposed adaptive replacement embedding scheme is the size of the key. For example, to obtain a watermarked Lena image of PSNR 61.20dB (in Figure 7.8(a)), the key length is supposed to be 59,647 bits if there is no any data compression. Here scaling factors with value of +1 or -1 are coded by 0 or 1, and each non-unit scaling factor is coded using 8 bits, where 7 bits are used to code the scaling factor's magnitude and 1 bit is used to code its sign. Each watermark embedding location is represented using 13 bits ($2^{13} = 8192$), since the maximum index of embedding range is 8000. Refer to Table 7.1, there are 85 scaling factors with non-unit magnitude for Lena image. Therefore the total number of bits for key is calculated as $(13 + 1) \times (4218 - 85) + (13 + 8) \times 85 = 59,647$. Note that because a watermark of 4,218 (38×111) pixels is used in our embedding example, the payload to code embedding locations for 4,218 watermark DCT coefficients is large. Please refer to Section 7.5.5 for discussion of applications for the proposed dual-key watermarking

system.

7.3 Proposed Blind Watermark Extraction Scheme

This section presents the *blind* watermark extraction scheme. The original image is not required, and only the key including the watermark embedding location and scaling factor is needed. Figure 7.10 shows the blind watermark extraction procedure.

Figure 7.10: Watermark extraction process.

[Step 1] Perform 2-D DCT transform for the received watermarked image $\mathbf{X}'$, which may or may not be the same as $\mathbf{X}$. The received watermarked image could be altered during the transmission through channels or manipulated deliberately due to attacks.

[Step 2] The DCT coefficients, $C_{\mathbf{X}}'$, are then re-arranged into a vector $S_{\mathbf{X}}'$ by zig-zag scanning.

[Step 3] Based on the key information of the watermark embedding locations (in vector *location*) and scaling factors (in vector *scale*), the vector form of the watermark's DCT coefficients, $S_{\mathbf{W}}'$, can be retrieved by

$$S_{\mathbf{W}}'(i) = \frac{S_{\mathbf{X}}'(location(i))}{scale(i)}, \quad i = 1, 2, \dots, M1 \times M2. \quad (7.8)$$

[Step 4] The watermark's DCT coefficients, $C_{\mathbf{W}}'$, in 2-D form, is obtained by rearrangement of $S_{\mathbf{W}}'$ in column-wise sequencing. The final extracted watermark,

$\mathbf{W}'$, is obtained by a 2-D inverse DCT transform of $S_{\mathbf{W}}'$. Figure 7.11 shows the extracted watermark from the watermarked Lena image in Figure 7.8(a).

Figure 7.11: Extracted watermark from the watermarked Lena image in Figure 7.8(a) (compare Figure 7.11 and Figure 7.1(b)).

7.4 Performance Evaluation

The performance of our proposed watermarking scheme is evaluated by using the Lena image of size 256×256 pixels (in Figure 7.1(a)) as the original image and a grayscale ‘NTU’ image of size 38×111 pixels (in Figure 7.1(b)) as watermark. The simulations are performed in the MATLAB 6.5 software environment. For a fair evaluation, performance is evaluated and compared with the other existing DCT-based watermarking schemes from two aspects – the imperceptibility of an embedded watermark and the robustness of this embedded watermark. The imperceptibility of a watermark is to be measured by the watermarked image’s quality in terms of PSNR as shown in Section 4.4.1. To investigate the robustness of an embedded watermark, the watermarked image (as shown in Figure 7.8(a)) is attacked by various signal processing techniques such as JPEG compression, quantization, additive noise, geometric distortions and filtering. The watermark extraction is performed on the attacked watermarked images and the extracted watermark is compared with the original watermark. The robustness performance of watermark extraction is evaluated by *normalized correlation coefficient*, r , of the extracted watermark $\mathbf{W}'$ and the original watermark $\mathbf{W}$ (see Section 4.4.1). A robust watermark would be able to survive various attacks even when the quality of the watermarked image is degraded heavily.

Table 7.2: The characteristics of Cox's, Piva's, Hsu's, Niu's and Tsai's schemes as well as our proposed scheme.

Scheme	Extraction Reference Data	Watermark	Embedding Domain	Modification Method
Cox's [46]	Non-blind	Binary	Whole DCT	Additive
Piva's [71]	Blind	Binary	Whole DCT	Additive
Hsu's [99]	Non-blind	Gray-level	8×8 block DCT	Additive
Niu's [100]	Non-blind	Gray-level	8×8 block DCT	Additive
Tsai's [101]	Blind	Gray-level	8×8 block DCT	Random replacement
Proposed	Blind	Gray-level	Whole DCT	Adaptive replacement

The characteristics of the five previously proposed DCT-based watermarking schemes as well as our proposed scheme are listed in Table 7.2. These five watermarking schemes, namely, Cox's scheme [46], Piva's scheme [71], Hsu's scheme [99], Niu's method [100] and Tsai's method [101], are used for performance comparisons with our proposed scheme. Note that the Cox's, Hsu's and Niu's schemes are non-blind, which could be a disadvantage in cases when the original data is not available for ownership verification. Tsai *et al.* [101] employed random replacement method to insert the watermark's DCT coefficient, whereas we use an adaptive replacement method to improve the quality of the watermarked image significantly.

7.4.1 Imperceptibility

We could not say which watermarking method can achieve better invisibility of an embedded watermark, as the watermarked image quality would be varied with watermark embedding strength. A lower watermark embedding strength would improve the imperceptibility. The purpose to investigate the imperceptibility is to include the transparent condition in the robustness evaluation. Although the PSNR measure may not be able to measure the perceptual quality of a watermarked image's quality accu-

Table 7.3: The PSNR of watermarked image for performance evaluation in our proposed, Cox's, Piva's, Hsu's, Niu's and Tsai's schemes (* denotes an approximate value).

Scheme	Proposed	Cox's	Piva's	Hsu's	Niu's	Tsai's
PSNR (dB)	61.20	40.00*	49.50*	40.83	30.70	39.65

rately, it is still an efficient objective measure used in most of the watermark schemes to evaluate the quality of a watermarked image. Table 7.3 lists the PSNR measure for the watermarked image being used for performance evaluation in our proposed scheme and in the above mentioned five DCT-based schemes, respectively (presented in papers [46, 71, 99, 100, 101]). Note that the watermarked image quality of our proposed scheme has the best PSNR measure, which is used for robustness tests in the following section.

7.4.2 Robustness Test

To enforce copyright protection through ownership verification, *robustness* is the primary concern of a proposed watermarking scheme. In the following, the experimental results on the robustness against various attacks – JPEG compression, color quantization, additive noise, geometric distortions and filtering, are shown and compared with the other watermarking schemes.

1) Robustness against JPEG Lossy Compression

As shown in Table 7.4, our proposed watermarking scheme has excellent resilience against JPEG compression attacks, since the watermark is embedded in the significant DCT coefficients. Figure 7.12(b) shows the extracted watermark with normalized correlation coefficient measure 0.9482, when the JPEG compression quality factor goes as low as 1% (the compression ratio is 27:1) as shown by Figure 7.12(a).

Table 7.4: The robustness results under JPEG lossy compression attacks, in terms of Normalized Correlation Coefficients, r .

Quality Factor (%)	100	90	80	70	60	50
PSNR (dB)	61.20	40.77	37.58	35.91	34.74	33.88
r	1.0000	1.0000	0.9999	0.9998	0.9997	0.9996
Quality Factor (%)	40	30	20	10	1	-
PSNR (dB)	33.03	32.05	30.69	28.26	22.79	-
r	0.9993	0.9988	0.9977	0.9932	0.9482	-

(a)

(b)

Figure 7.12: (a) The watermarked Lena image under JPEG compression with quality factor 1% (PSNR=22.79dB) and (b) the extracted watermark ($r=0.9482$).

Using Cox's method [46], the normalized correlation (NC) value of the extracted watermark with respect to its original data is around 0.9000 when the quality factor is 20%. Piva's method can survive JPEG compression up to 10% quality only [71]. For Hsu's method [99], the NC value of the extracted watermark is 0.4130 when the compression ratio goes to 11:1. The NC value of the extracted watermark using Niu's method [100] is 0.6030 when JPEG quality factor of the watermarked image becomes 50%. For Tsai's scheme [101], the NC value of the extracted watermark is 0.8311 when the JPEG compression ratio become 19:1. Comparing with above statistical figures, our proposed scheme has superior robustness against JPEG compression attack.

2) Robustness against Color Quantization

Figure 7.13(a) shows a binary version of watermarked Lena image, i.e., the color depth is reduced to 2 only. The NC value of the extracted watermark as shown in Figure 7.13(b) is 0.8889. The robustness of our proposed scheme outperforms most of the existing watermarking schemes.

3) Robustness against Additive Noise

The watermarked Lena image as shown in Figure 7.14(a) is corrupted by high additive Gaussian noise with resulted PSNR measure of 13.61dB. Though the image is heavily degraded, the extracted watermark has still good NC measure of 0.9123.

4) Robustness against Geometric Distortions

Figure 7.15 shows the good extraction result of a watermarked image under rotation, cropping and scaling transform. In this scenario, a synchronization operation would be needed to adjust the distorted image back to its original size and position. Many watermarking systems cannot cope with combinations of various attacks and especially the minor random geometric distortions. Figure 7.16(a) shows a watermarked Lena image of PSNR=17.48dB which has been attacked by some random geometric distortions as obtained using StirMark [20]. The image is slightly stretched, sheared, shifted and rotated by an unnoticeable random amount. The NC value of the extracted watermark from this random geometric distorted watermarked image is found as 0.6984.

Figure 7.13: (a) The watermarked Lena image after quantization to a binary image and (b) the extracted watermark ($r=0.8889$).

Figure 7.14: (a) The watermarked Lena image (PSNR=13.61dB) under attacks of heavy additive Gaussian noise and (b) the extracted watermark ($r=0.9123$).

Figure 7.15: (a) The watermarked Lena image under rotation, cropping and scaling transformations and (b) the extracted watermark ($r=0.8299$).

Figure 7.16: (a) The watermarked Lena image with random geometric distortions (PSNR=17.48dB) and (b) the extracted watermark ($r=0.6984$).

Figure 7.17: (a) The blurred watermarked Lena image (PSNR=25.66dB) and (b) the extracted watermark ($r=0.9858$).

Figure 7.18: (a) The sharpened watermarked Lena image (PSNR=9.59dB) and (b) the extracted watermark ($r=0.9339$).

5) Robustness against Filtering

The extraction results against low-pass and high-pass filtering attacks are shown in Figures 7.17 and 7.18, respectively. Our proposed scheme can successfully survive blurring as well as sharpening attacks. Comparing to Tsai's scheme, we can achieve better robustness performance against blurring attacks and comparable results against sharpening attacks.

This part has evaluated the robustness performance of our proposed scheme against various types of attacks. Experimental results have shown that our method provides excellent resilience against a wide range of attacks. It has been demonstrated that the proposed scheme has better robustness performance than Cox's scheme [46], Piva's scheme [71], Hsu's scheme [99], Niu's method [100] and Tsai's method [101].

7.5 Discussions

7.5.1 Binary versus Gray-level Watermark

Most watermarking methods are based on the same basic principle, that is, pseudo-random binary watermarks are used to make small changes to the selected coefficients in the spatial or transform domain. Later on, these changes are identified using a correlation or a correlation-like similarity measure. The presence of a watermark is determined by comparing the computed correlation with a pre-defined threshold. Recently, grayscale images such as some specially designed pattern, logo and signature images are also suggested as watermarks. This type of watermarks is very unique to identify the ownership. Rather than the objective similarity measurement, it could be verified easily by human eyes as well.

7.5.2 Additive versus Replacement Embedding

The embedding method of the most watermarking schemes is *additive* modification of the selected transformed coefficients or spatial pixels. As shown in our experimental results, the *replacement* embedding method, for example, in Tsai's scheme and our proposed schemes, can achieve higher robustness against the *additive* method. This is because a larger amount of information (watermark bits) can be hidden into cover data. For instance, the selected original image's DCT coefficients are replaced by the watermark's DCT coefficients, whereas additive method makes only small modifications to those selected coefficients. The conventional additive embedding technique is less robust against attacks comparing to our adaptive replacement embedding technique. However, the random replacement technique as proposed by Tsai *et al.* [101] can not guarantee to preserve good quality of the watermarked image. In our proposed scheme, the watermark's DCT coefficients *adaptively* replace the corresponding nearest matches of the original image's coefficients. Moreover, a threshold T_0 is employed to ensure the good quality of a watermarked image.

7.5.3 Whole DCT Domain versus Block DCT Domain Embedding

In our proposed scheme, the DCT transform of the whole image is employed as the embedding domain. The lower frequency components are selected for watermark embedding to achieve higher robustness against attacks. The DCT coefficients are zig-zag scanned and re-arranged in a way from the *most* significant one to the *least* significant one. The criterion of coefficients selection for embedding is then relatively simple. The embedding range is selected according to the investigation of the effects of JPEG compression in the DCT coefficients. Those less sensitive coefficients against compres-

sion – the low to middle frequency components – are selected. The problem of using whole DCT transform could be its heavy computational load in processing the DCT transform when an original image size is very large. The solution would be to divide the original image as well as the watermark into several blocks of smaller sizes. Each block of the DCT coefficients for the original image is to be adaptively replaced by a DCT block of the watermark image. Therefore, the computational load due to the large image size could be effectively reduced.

7.5.4 Security Issues

First of all, as mentioned earlier, the replacement embedding technique can achieve higher robustness against removal attacks comparing additive embedding technique as used in other methods [46, 71, 99, 100]. Because the embedded watermark by additive modification could be easily removed or altered by knowing the corresponding embedding locations. Replacement embedding technique is thus more secure against the removal attack even though the hacker knows the embedding locations of the watermark. Since the watermark is embedded in the coefficients that are the most significant, their removal will result serious degradations in the watermarked image. Furthermore, our robust watermarking scheme can provide a higher level of security due to use of dual-key generated during the embedding process. By knowing only the embedding locations, it is still not possible to recover the watermark in the absence of scaling or polarity information of the DCT coefficients.

There are many discussions on the uses of watermarking schemes in resolving ownership disputes. One interesting and well-known scenario is the inversion attacks addresses by Craver, Memon *et al.* [104]. The counterfeit watermarking schemes are performed on a watermarked image to allow multiple claims of the rightful ownership.

We may need further study on non-invertible watermarking scheme to overcome such attacks, by generating the watermark using a secure hash of the original image [104]. Another possible solution could be considered is to generate a watermark using a cryptographically secure pseudo-random number generator (CSPRNG) as proposed by Li and Chang [105].

7.5.5 Applications

In a practical watermarking system for ownership verification, there are at least two parties involved – the owner of the images and a legal authority [106]. The legal authority stores the watermark as well as one of the two keys. Only when the owner presents the other key, the final watermark detection can be performed by the legal authority and the ownership is thus verified. Consider the scenario that one Work may be created jointly by multiple owners, which arises the issue of joint ownership verification [107]. The use of dual key in a watermarking system could be an advantage for an ownership verification system. Suppose there are two owners sharing the ownership of the Work jointly. Then the secret key can be simply separated in two parts – one part contains only the watermark location information and the other part contains the scaling information. The rightful ownership can be thus verified only if both the owners present their secret keys. In some scenarios, a slight alternation of the digital content might affect the interpretation of the original data. For example, an accidental or intentional modification for an X-ray or CT scan image could result total misinterpretation in clinical diagnosis. In such case, the requirement of the watermarked data quality is needed to be quite high. By applying our *transparent* robust watermarking technique, one can always preserve a good quality of the watermarked data, while securing high robustness. The *transparent* embedding in our proposed scheme thus make it quite useful in applications, such as for ownership verification of biomedical images, where

the quality of the watermarked data is required to be high.

7.6 Chapter Summary and Conclusion

To enforce copyright protection through ownership verification, the essential requirement of a watermarking scheme would be its robustness against attacks. In this chapter, a novel transparent robust watermarking technique has been proposed based on the whole DCT transform of the original image. Meaningful gray-level images rather than binary sequences are used as watermarks. The DCT coefficients of the watermark image adaptively replace the DCT coefficients of the original image in the sense of a nearest match in terms of magnitude. The automatic selection of the optimal embedding range is achieved by setting the *End* index from the viewpoint of robustness against JPEG compressions, and the *Initial* index from the viewpoint of visual quality of the watermarked image. The key information consists of both the embedding locations and the scaling factors. Experimental results show that the proposed watermarking scheme can provide excellent resistance under various attacks, such as, JPEG compression, color quantization, additive noise, geometric distortions and filtering. Compared with the other DCT based robust watermarking techniques, our proposed scheme has higher robustness performance, while preserving better quality of the watermarked image.

The adaptive replacement embedding method in the DCT-domain can hide the gray level watermark transparently. The use of the replacement technique and the selection of low to middle frequency range for embedding enhances the robustness of the proposed scheme. A good perceptual quality of the watermarked data is achieved as well by our adaptive replacement with a threshold setting. Furthermore, the image-dependent dual key generation enhances the security level of the proposed watermarking scheme. The proposed transparent information hiding by adaptive replacement technique, has

partially compromised the trade-off between the robustness and the imperceptibility of the hidden watermark. All these good features make the proposed information hiding scheme very effective in ownership verification to enforce the intellectual property protection of digital media, especially when the quality of the watermarked data is required to be high. The disadvantage could be the heavy computational load due to the whole DCT transform when the image size is large. This could be compromised by dividing the original image and the watermark into smaller processing blocks. The practical applications of our proposed scheme, for example, for the ownership verification of biomedical images, will be further investigated.

Part III

Information Hiding for Content Authentication

Chapter 8

Introduction

The Part III of this thesis concentrates on the works of information hiding for content authentication of digital media. Specially, complex chirp signals [12] are embedded using multiresolution technique, as fragile watermarks [9] to assist in verifying the content integrity and assessing the quality of its associated cover digital Media. For ease of exposition, we assume that the contents being watermarked is digital image, however the technique proposed is, in principle, equally applicable to other types of digital media, such as, audio and video.

8.1 Authentication Watermarking

As a great volume of multimedia data is stored in the digital form, it makes very easy to modify and forge the digital content by widely available software editing tools. The content authentication has become extremely important, which is to guarantee that the content has not been altered, or at least that the semantic characteristics of the image is still maintained. The severe distortions, such as, heavy additive noises, visible block effects due to high compression ratio, cropping of the content and etc., could

be observed easily by human perception. However, some deliberate alterations could not be noticed. For example, the slight modifications of few pixels are difficult to be perceived. In many cases, the changes may be intentionally malicious or may inadvertently affect the interpretation of the digital content. For example, an inadvertently or intentionally change in an X-ray or CT scan image may result in a totally wrong diagnosis; the ease of tampering with photographs in the newspaper or the video object in video frames can provide wrong information to the public; or the malicious tampering of evidence in a criminal trial can even result in the wrong judgment. Content integrity in such applications is of high importance, thus, a secure/sensitive authentication tool is crucial to verify whether the content has not been altered.

8.1.1 Requirements of Authentication

Authentication tools are expected to answer the following questions [2]:

- Has the content been altered in any way whatsoever?
- Has the content been significantly altered?
- What parts of the content have been altered?
- Can an altered part be restored?

Information hiding terminology provides the capabilities to answer the above questions by hiding an authentication watermark or a digital signature in the protected content. Because the watermark/signature undergoes the same transforms as the cover content in which it is embedded, by analyzing the extracted watermark/signature it can detect an alteration occurred and might be possible to further infer what, when, and where changes happened. Some desirable features to construct an effective authentication scheme include the following [108]:

- Sensitivity – the authentication watermark/signature shall be able to detect any possible change in the content;
- Robustness – the authentication watermark/signature shall be able to sustain certain moderate distortions such as lossy compression, content-preserving manipulations;
- Security – the authentication watermark/signature itself shall be difficult to be forged or manipulated by an attacker;
- Portability – the authentication watermark integrated with the cover content invisibly has better portability than the authentication signature that is always stored as a separate file;
- Localization – the authentication watermark/signature shall be able to locate the altered areas and verify other areas as authentic;
- Recovery – the use of authentication watermark/signature provides the ability of reconstruction of the digital media that has been corrupted.

These are the essential requirements of “ideal” authentication system. The proposed authentication schemes shall partially or fully achieve above six requirements depending on the practical applications. According to the demands of practical applications, the content authentication has mainly two categories – *exact authentication* and *selective authentication* [2]. In some applications, the *sensitivity* of the authentication watermark is essential, for example, even single bit change in the content is considered as alterations. We refer this type of authentication as *exact authentication*, which is implemented by *fragile* watermarking. In contrast, *selective authentication* requires detecting only unacceptable changes such as intentionally malicious attacks by cropping and replacement of a part of the content, and tolerates certain acceptable changes such

as flipping of a single bit or low level JPEG compressions. Semi-fragile watermarking provides a mechanism for implementing selective authentication. A semi-fragile watermark is unaffected by legitimate distortions but destroyed by illegitimate distortions. Semi-fragile watermarking is beyond the scope of our concern.

The scope of Part III focuses on the fragile watermarking. The proposed information hiding scheme in Part III aims to fulfill the following four elements of the authentication requirements:

1. Sensitivity
2. Security
3. Portability
4. Localization

In Chapter 9, we present a multiresolution fragile watermarking scheme, which detects the alterations of the watermarked media with the localization capability.

8.1.2 Review of Authentication Techniques

There have been a large number of fragile/semi-fragile authentication watermarking schemes proposed for tamper detection and content integrity verification of multimedia content [109]. Firstly, these authentication algorithms can be divided into two broad categories, depending on the method to convey the authentication data [108, 110]:

1. Labelling-based approach:

The authentication data are stored in a separate file and transmitted as the overhead of the data bitstream, as proposed in [111, 112, 113]. The weakness of

this approach is that a label or the appended authentication data may be easily removed or lost during transmission. Furthermore, it requires extra bandwidth to be transmitted or stored.

2. Watermarking-based approach:

The authentication data are embedded in the digital media, which is a more secure and efficient way. The watermarking techniques for content authentication are usually referred to *fragile/semi-fragile* watermarking.

The existing fragile/semi-fragile watermarking techniques differ mainly in the way to embed the authentication data into the host data [9]. The embedding techniques proposed include spatial-domain embedding and transform-domain (DCT, wavelet and etc.) embedding.

The early spatial-domain fragile watermarking systems embedded the watermark in the least significant bit (LSB) plane for perceptual transparency [114, 115]. Wolfgang and Delp improved the robustness and the localization properties of schemes in [114, 115], by using VW2D technique [116]. In Wong's method [117], the image, image dimensions and marking key are hashed during embedding and used to modify the LSB plane of the original image. However, the attacker may attempt to modify the marked image itself without affecting the LSB plane. The technique proposed by Yeung and Mintzer [118] uses the marking key to generate several pseudo-random look-up-table (LUT) that control how subsequent modification of the pixel data will occur. The security of the technique depends on the difficulty of inferring the look-up tables [119].

Many fragile/semi-fragile watermarking systems benefit from the transform domain embedding, such that they preserve moderate robustness to compression attacks. Wu and Liu [120] proposed fragile watermarking scheme based on modifying the quantized

DCT coefficients according to a special lookup table. Xie and Arce's technique [121] firstly implemented the signature in DWT domain and then embedded the signature using SPIHT compression algorithm. Kundur and Hatzinakos [122, 123] embedded the pseudorandom sequences by modifying the quantization process of Harr wavelet coefficients. In particular, the Kundur's fragile watermarking algorithm is compared with our proposed scheme. More detailed information on Kundur's scheme is presented in Chapter 9.

There are other possible classifications of the tamper-detection and content-authentication watermarking techniques [9, 10, 124]:

- Visually-authenticated (semi-)fragile watermarking, where a typically visual pattern is hidden in the image (such as, replacing the LSB plane), and tamper detection is based on the visual assessment of perceived differences by an operator as in Marvel [125, 126].
- Statistically-authenticated (semi-)fragile watermarking, where an estimate of tampering likelihood is obtained based on correlation coefficient or measured mismatch between the original inserted and recovered authentication sequences. In these techniques the performance can typically be given in terms of probability of correct watermark sequence detection as a function of false alarm rate [111, 127, 128, 129, 130, 131].
- Self-embedding techniques for proof of authenticity and image protection as in Fridrich *et al.* [132], Lan *et al.* [133] and in the case of video in Robie *et al.* [134]. For example, Fridrich's scheme embeds the quantized 8×8 block DCT coefficients into the LSBs of the other distant blocks.

The authentication based on the visual pattern is rather insecure, for example, the substitution of the LSB plane or modification of the content while maintaining

the LSB plane unchanged. The self-embedding authentication technique would fail if large portions of the image were corrupted, since the authentication data are probably lost in such case. Many recent proposed schemes are under the second category, i.e., the statistically-authenticated watermarking. However, they are non-oblivious since the original watermark is still required during authentication. In this context, the embedded watermark should know previously or should be able to regenerate during authentication. Without knowing the previously hidden watermark, it is unable to perform the authentication.

In Chapter 9, we propose the wavelet-domain fragile watermarking scheme, which is not under above three categories, i.e., visually-, statistically-authenticated or self-embedded techniques. Specially, we make use of special signals – complex chirp signals [12] – as watermarks. The authentication does not require any *a priori* information on the original watermark as well as the original image; instead the special feature of the extracted complex chirp watermark signal is investigated for content authentication.

8.2 Complex Chirp Signals

Chirp signals are sinusoidal signals with time-varying frequencies (FM-frequency modulation) and/or magnitudes (AM-amplitude modulation) [12]. Chirp signal is also known as non-stationary signal [135, 136]. They are used in many engineering applications such as speech, seismic, sonar and radar applications. A generalized form of a complex chirp signal, $x(t)$, can be written as

$$x(t) = a(t)e^{j\vartheta(t)}, \quad (8.1)$$

where the magnitude $a(t)$ and the phase $\vartheta(t)$, are arbitrary functions of time.

The expression in (8.1) can be expanded in the terms of both the real and imaginary

parts:

$$x(t) = a(t)[\cos\vartheta(t) + j \sin\vartheta(t)]. \quad (8.2)$$

The cosine term is the real signal and the sine term is the imaginary signal. If the phase $\vartheta(t)$ is constant such that it is not varying with time, the chirp signal becomes a purely amplitude modulated (AM) signal. On the other hand, if the magnitude $a(t)$ is constant, the chirp signal is a frequency modulated (FM) signal [135, 136].

Time-frequency distributions [13, 137] have been suggested to characterize and analyze such non-stationary signals or (complex) chirp signals. Time-frequency representations (TFRs) of signals map a one-dimensional signal of time, $x(t)$, into a two-dimensional function of time and frequency, $T_x(t, f)$. A distribution that represents the energy or intensity of a signal simultaneously in time and frequency is devised. There are numerous distributions can be used to present a signal in time-frequency plane. For instance, one well-known distribution among them is *Wigner distribution* (WD) [13, 137], which is also known as *Wigner-Ville distribution*¹ [13, 137].

There are four features or descriptors to measure the relevance derived from the TFR of chirp signals [138].

- Instantaneous energy (IE): The IE is computed as the mean of the TFR along each time slice, which gives a measure of energy variation with time.
- Instantaneous energy spread (IES): IES measures the spread of energy over frequency for each time slice. It is computed as the standard deviation of the TFR along each time slice.
- Instantaneous mean frequency (IMF): IMF measures the frequency dynamics of

¹The Wigner-Ville distribution (WVD) function was originally proposed by E. Wigner in quantum mechanics and J. Ville applied it for a signal analysis, some 15 years after Wigner's paper. This method has made it possible to represent signal's power density spectrum in time-frequency domain as a natural extension of the Fourier transformation method.

the signal by computing as the first moment along each time slice.

- Instantaneous mean frequency spread (IMFS): IMFS measures the spread of frequency about the mean frequency for each time instant given by the second central moment along each time slice.

In Chapter 9, complex chirp signal with time-varying frequency and constant magnitude, that is, complex FM chirp signal, is used as the fragile watermarks. However, the complex chirp signal with constant frequency and time-varying magnitude (i.e., AM chirp signal) could be also applied. We use only the FM complex chirp signal to elaborate the proposed fragile watermarking scheme.

The complex FM chirp signal is written as

$$x(t) = a[\cos(\vartheta(t)) + j \sin(\vartheta(t))], \quad (8.3)$$

where a is an arbitrary real number, denoting a constant magnitude of the chirp signal. The frequency sweep can be linear, quadratic, cubic and etc.. In the following, some FM chirp signals with special patterns are shown. Note that the imaginary part in the expression of the complex chirp signal is not shown for simplicity in the following equations.

For instance, a non-stationary signal consisting of a linear frequency modulated signal is generated by [139]

$$x(t) = \prod_T(t - \frac{T}{2}) \cos[2\pi(f_0 t + \alpha \frac{t^2}{2})], \quad (8.4)$$

where $\prod_T(t - \frac{T}{2})$ is 1 for $|t| \leq \frac{T}{2}$ and zero elsewhere, and f_0 is the center frequency and α represents the rate of the frequency change. In discrete-time domain, this chirp signal can be defined as [139]

$$x(n) = \cos[2\pi(f_0 n + \alpha \frac{n^2}{2})], n = 0, 1, \dots, N_w - 1. \quad (8.5)$$

Here, we assume a unit sampling frequency and define N_w as the length of the signal. Applying WD, the time-frequency presentation of a linear FM signal of 256 samples and frequency range [0.1, 0.4] Hz, is shown in Figure 8.1 (a). A non-stationary signal consisting of a quadratic FM signal [140] is given by

$$x(t) = \prod_T \left(t - \frac{T}{2}\right) \cos[2\pi(a_0 t + a_1 t^2 + a_2 t^3)], \quad (8.6)$$

where a_0 , a_1 and a_2 are real coefficients. Similarly, in discrete domain, the chirp signal can be defined as [140]

$$x(n) = \cos[2\pi(a_0 n + a_1 n^2 + a_2 n^3)], n = 0, 1, \dots, N_w - 1. \quad (8.7)$$

A quadratic FM signal's TFR of 256 samples and frequency range [0.1, 0.4] Hz, is shown in Figure 8.1 (b).

Many other TFR patterns could be obtained, for example, hyperbolic (in Figure 8.1(c)), cubic (in Figure 8.1(d)) and so on. Infinite number of FM chirp signals can be generated by choosing different combinations of TFR patterns, values for the chirp rate, and the initial, end frequencies. The common feature for all the distinct FM complex chirp signals is that their magnitudes are always *constant*. Therefore, the special feature or descriptor – instantaneous energy or (instantaneous) magnitude – is to be used as the quantitative measure for the quality of the extracted authentication watermark.

8.3 Objective

As discussed in Section 8.1.2, the statistically-authenticated (semi-)fragile watermarking requires the original watermark. It could be a disadvantage to require a prior knowledge on the embedded watermark. Self-embedding authentication would fail because the hidden authentication data could be distorted or lost as well after tampering.

Figure 8.1: Time-frequency representations of (a) a linear FM chirp signal, (b) a quadratic FM chirp signal, (c) a hyperbolic FM chirp signal, and (d) a cubic FM chirp signal.

Therefore, we propose to hide the complex chirp signals rather than the pseudorandom sequences to cope with the above issues.

The objective of Part III is thus to propose a multiresolution fragile watermarking scheme using complex chirp signals for content authentication. The watermark is embedded in the wavelet domain. A Quantization-Index-Modulation (QIM) technique [123] is employed for information hiding. The proposed fragile watermarking scheme is applied for content integrity verification without resorting the original watermark and blind quality assessment for the distorted watermarked content.

Chapter 9

Multiresolution Fragile Watermarking Using Complex Chirp Signals for Content Authentication

In this Chapter, both the real and the imaginary parts of a complex chirp signal [12] are hidden into the digital media to be protected for content authentication. Specially, frequency modulated (FM) complex signal is used to demonstrate the proposed fragile watermarking scheme. The amplitude modulated (AM) complex chirp signal could also be applied, such that the feature used for authentication is its constant instantaneous frequency rather than the constant instantaneous magnitude for FM signal. The principle of using AM complex chirp signal would be the same as using FM complex chirp signal presented in chapter. The digital image watermarking is demonstrated to show our proposed fragile watermarking technique, but the technique and concept can be suitably generalized to the other media, such as, video.

9.1 Multiresolution Representation

9.1.1 Wavelet Decomposition of Host Image

In our proposed fragile watermarking scheme, the watermark is embedded in the wavelet domain of the original image, I . The wavelet transformation decomposes the host image hierarchically into a series of successively lower resolution reference images and their associated detail images. The low resolution image and the detail images including the horizontal, vertical and diagonal details contain the information to reconstruct the reference image of the next higher resolution level. A grayscale Lena image of 256×256 pixels, as shown in Figure 9.1(a), is used as the original image to demonstrate the proposed watermarking scheme in the following. In particular, the decomposition level, l , is set as three, as shown in 9.1(b).

Figure 9.1: (a) The original Lena image (256×256 pixels) and (b) the three level wavelet decomposition of (a).

9.1.2 Generation of Multiresolution Watermark

In our proposed method the arbitrary non-stationary FM signals, or the complex chirp signals, with particular signatures or patterns are worked as the authentication watermarks. The generated analog chirp signal, including both the real and imaginary

parts, is processed by Pulse Code Modulation (PCM) technique [141, 142]. It is an analog-to-digital (A/D) conversion technique to obtain the digital representation of an analog signal where the signal is sampled regularly at uniform intervals of duration T . Every sample is quantized to a series of symbols in a digital code, which is usually a binary code. The higher the sample rate and the greater the number of quantization levels, the more closely the digital codes will represent the original analog signal. PCM technique is widely used in digital telephone systems and for digital audio recording on compact discs.

For instance, a quadratic chirp signal of 256 samples, $\mathbf{s}_i$, $i = 1, 2, \dots, 256$, is generated as watermark from equation (8.6). It is written as

$$\mathbf{s}_i = a_i + jb_i, \quad (9.1)$$

where $\sqrt{a_i^2 + b_i^2} = 1$. Note that the FM complex chirp signals always preserve constant magnitudes while varying their phases. In particular, in this example the magnitude of $\mathbf{s}_i$ is set as one. The time plots for real signal a_i and the imaginary signal b_i are shown in Figures 9.2 (a) and (b), respectively, and the time-frequency representation (TFR) of this quadratic chirp signal $\mathbf{s}_i$ is displayed in Figure 9.2(c). From the TFR plane, the quadratic waveform with the frequency varying with the range of [0.1, 0.4] Hz, is clearly observed.

The watermark samples are then coded into the binary representations with N bits. There are 2^N quantization levels, and the quantization step, Δ , is thus

$$\Delta = \frac{V_{max} - V_{min}}{2^N}, \quad (9.2)$$

where V_{max} and V_{min} are the maximal and minimal values in a_i and b_i . If the magnitude of $\mathbf{s}_i$ is one, $V_{max}=1$ and $V_{min}=-1$. To represent the chirp watermark sample, $\mathbf{s}_i$, given in (9.1) using seven bits, the quantization step Δ is to be $\frac{2}{2^7}$ (≈ 0.0156).

Figure 9.2: The time plots of (a) the real part and (b) the imaginary part for a quadratic chirp watermark signal s_i ('o' denotes the samples of the watermark), (c) the TFR display of s_i .

To simplify the coding process, we can directly map the values of watermark samples to the non-negative integers ranging from 0 to 127 ($=2^7 - 1$), by

$$w_i = \text{round}\{(\mathbf{s}_i + 1) \times 63.5\}, \quad (9.3)$$

where $w_i \in [w_a(i) + jw_b(i)]$ denotes the i -th watermark sample after mapping. The binary representations of the real part w_a and the imaginary part w_b are denoted as a_{in} and b_{in} ($i = 1, 2, \dots, 256$, $n = 1, 2, \dots, 7$), respectively, which satisfy

$$\begin{aligned} w_a(i) &= \sum_{n=1}^7 a_{in} \cdot 2^{7-n}; \\ w_b(i) &= \sum_{n=1}^7 b_{in} \cdot 2^{7-n}. \end{aligned} \quad (9.4)$$

Figure 9.3 illustrates the generation of a multiresolution chirp watermark. Figure 9.3(a) gives an example on how to obtain the 7-bit binary representation for a decimal chirp sample with value of 0.8820. In this manner, the samples are successively sliced into seven bit-planes of different significance, as displayed in Figures 9.3 (b) and (c) for the real signal and the imaginary signal, respectively. The first bit-plane ($n=1$) is of most significance, and the seventh bit-plane ($n=7$) is the least significant one. To correspond with the 3-level wavelet decomposition of the host image, these seven bit-planes of the watermark are classified into 3 groups having different resolutions (see Figures 9.3 (b) and (c)):

1. Low resolution component – the most significant bit (MSB) plane comprising a_{i1} and b_{i1} ;
2. Intermediate resolution components – a_{i2} , a_{i3} , b_{i2} and b_{i3} ;
3. High resolution components – $a_{i4} - a_{i7}$ and $b_{i4} - b_{i7}$.

Figure 9.3: Multiresolution chirp watermark generation: (a) an illustration to generate binary representation for a sample of chirp signal; (b) the binary watermark bits of the real part and (c) the binary watermark bits of the imaginary part.

9.2 Proposed Multiresolution Watermark Embedding Scheme

9.2.1 Watermark Embedding Locations

In our proposed watermark embedding scheme, both the real part a_{in} and the imaginary part b_{in} of the complex chirp watermark are simultaneously embedded into the wavelet domain of the original data by a quantization-index-modulation (QIM) technique [123]. A 3-level DWT of the host image is conducted using Harr bases. As shown in Figure

Figure 9.4: Watermark embedding locations.

9.4, a wavelet coefficient in the 3rd level has successive correspondence with (2×2) higher frequency components in the 2nd level and (4×4) high frequency components in the 1st level, forming a *cluster* of the wavelet coefficients which is generated from the original 8×8 image block.

The rules for selection of the watermark embedding locations include the follows:

Rule 1: The sub-bands containing the horizontal details, i.e., HL3, HL2 and HL1, are selected for embedding the real part of complex chirp watermark samples - a_{in} ; and the sub-bands containing the vertical details, i.e., LH3, LH2 and LH1, are selected for embedding the imaginary part of complex chirp watermark samples - b_{in} , as illustrated in Figure 9.4.

Figure 9.5: A pair of clusters of wavelet coefficients for embedding a pair of i -th watermark samples of a_{in} and b_{in} , $n = 1, 2, \dots, 7$.

Rule 2: A cluster of wavelet coefficients consists of a single 3rd level coefficient, (2x2) 2nd level coefficients and (4x4) 1st level coefficients. One 7-bit watermark sample would be embedded into one cluster of wavelet coefficients as shown in Figure 9.5.

Rule 3: The 7 bits of a watermark sample is embedded in a multiresolution manner by grouping them into three different levels of resolution as described in Section 9.1.2. That is to embed the most significant bit of the watermark sample into the low frequency sub-bands of the host image, e.g., sub-band HL3 or LH3. While the high-resolution (i.e., less significant) watermark bits are embedded into higher frequency sub-bands of the host image. For instance, 2nd and 3rd bits are embedded into the HL2 or LH2 sub-band while 4th - 7th bits are embedded into the HL1 or LH1 sub-band.

Rule 4: In a given decomposition level and a given position, only one of the randomly chosen horizontal or vertical components is marked. The embedding of the imag-

inary part of a watermark sample is coupled with the embedding of its real part, such that if the embedding position for the 1st bit of the real part is (p, q) in sub-band HL3, the embedding position for the 1st bit of the imaginary part is then at $(p+1, q)$ in sub-band LH3.

Rule 5: To enhance the security of the watermarking system, a pseudo-random permutation is performed to produce a random embedding order to disperse the frequency modulation relationship between the consecutive samples of the chirp signal. This pseudo-random permutation is the *key* of the proposed fragile watermarking system. The random order can also be interpreted as the randomly selected embedding locations for watermark embedding.

Figure 9.5 demonstrates one possible solution to select the watermark embedding locations for a pair of watermark samples a_{in} and b_{in} following the above rules. Note that the embedding locations in the second level wavelet decomposition, are not overlapped for all four watermark bits a_{i2} , a_{i3} , b_{i2} and b_{i3} . The selection of the embedding locations in the first level wavelet decomposition also follows the same rules.

9.2.2 Quantization Process

The embedding of the binary watermark bits are based on the technique of Quantization-Index-Modulation (QIM) [122, 123, 143] of wavelet coefficients. Let denote the l th-level wavelet coefficient of the original image as $C_{k,l}(p, q)$, where $k = h$ or v (which stands for “horizontal” and “vertical” detail coefficients, respectively), $l = 1, \dots, L$ and (p, q) is a particular spatial location index at decomposition level l . These real-valued coefficients are quantized into two symbols of 0 and 1, as shown in Figure 9.6. The quantization function $Q(\cdot)$, which maps the wavelet coefficients to 0

Figure 9.6: The quantization function for the wavelet coefficients.

or 1, are given by

$$Q(C) = \begin{cases} 0, & \text{if } z\Delta \leq C < (z+1)\Delta \text{ for } z = 0, \pm 2, \pm 4, \dots \\ 1, & \text{if } z\Delta \leq C < (z+1)\Delta \text{ for } z = \pm 1, \pm 3, \pm 5, \dots \end{cases}, \quad (9.5)$$

where Δ is the quantization parameter or quantization step, which is a positive real number.

Let $\{w(i) \in (a_{in}, b_{in})\}$ denote a watermark bit to be embedded into a selected wavelet coefficient $C(i)$ based on the key. Differ from Kundur's quantization method [122, 123], a secure embedding zone, $\mathbb{Z}$, is defined in our proposed scheme. The reason of using the secure embedding zone is to further improve the watermark extraction performance of Kundur's method. Note that there still exists 0.1% false detection even if there is no any attacks in the Kundur's method [144]. We would improve the extraction performance by adjusting all the watermarked wavelet coefficients falling into the secure zone $\mathbb{Z}$, of width r ($r \leq \Delta$). As shown in Figure 9.7, the shadowed region denotes the zone $\mathbb{Z}$, and the mid point of the zone coincides with the mid point of one quantization step. The rules for embedding a watermark bit $w(i)$ are as follows.

Rule 1: If $Q(C(i)) = w(i)$ and the coefficient $C(i)$ falls into the zone $\mathbb{Z}$ (as illustrated in Figure 9.7(a)), then no change in this coefficient $C(i)$ is necessary such that the watermarked coefficient, $\tilde{C}(i)$, is

$$\tilde{C}(i) = C(i). \quad (9.6)$$

Rule 2: If $Q(C(i)) = w(i)$ but the coefficient $C(i)$ is not within the zone $\mathbb{Z}$, the following adjustment is done to shift the coefficient into the half portion of the

Figure 9.7: The rules for embedding of the watermark bits: (a) when $Q(C(i)) = w(i)$ and $C(i)$ falls within the zone $\mathbb{Z}$; (b) when $Q(C(i)) = w(i)$ but $C(i)$ falls outside the zone $\mathbb{Z}$; and (c) when $Q(C(i)) \neq w(i)$.

zone $\mathbb{Z}$ that is closer to this coefficient, as shown in Figure 9.7(b):

$$\tilde{C}(i) = \begin{cases} \Delta \text{floor}(\frac{C(i)}{\Delta}) + \frac{\Delta}{2} - \xi, & \text{if } \text{floor}(\frac{C(i)}{\Delta}) = \text{round}(\frac{C(i)}{\Delta}) \\ \Delta \text{floor}(\frac{C(i)}{\Delta}) + \frac{\Delta}{2} + \xi, & \text{if } \text{floor}(\frac{C(i)}{\Delta}) \neq \text{round}(\frac{C(i)}{\Delta}) \end{cases}, \quad (9.7)$$

where ξ is a non-negative random real number such that $\xi \leq \frac{r}{2}$, and operation ‘floor($\cdot$)’ is to round the element to the nearest integer towards minus infinity.

Rule 3: If $Q(C(i)) \neq w(i)$, the coefficient $C(i)$ is then shifted to the secure zone of its nearest neighboring quantization step, as illustrated in Figure 9.7(c). The

watermarked coefficient $\tilde{C}(i)$ in this scenario is given by

$$\tilde{C}(i) = \begin{cases} \Delta \text{floor}(\frac{C(i)}{\Delta}) - \frac{\Delta}{2} + \xi, & \text{if } \text{floor}(\frac{C(i)}{\Delta}) = \text{round}(\frac{C(i)}{\Delta}) \\ \Delta \text{round}(\frac{C(i)}{\Delta}) + \frac{\Delta}{2} - \xi, & \text{if } \text{floor}(\frac{C(i)}{\Delta}) \neq \text{round}(\frac{C(i)}{\Delta}) \end{cases}, \quad (9.8)$$

where ξ ($\leq \frac{\Delta}{2}$) is a non-negative random real number.

The random number ξ is used to adjust the coefficient $C(i)$ to a *randomly* selected position within the half-portion of the secure embedding zone that is closer to $C(i)$. As a result, the use of ξ enhances the security of the quantization embedding scheme, in the sense that it would be hard for a pirate to detect the embedding of watermark bits. Furthermore, watermark embedding is performed by quantization of $C(i)$ to its nearer half-portion of the embedding zone. Therefore, the change of a coefficient's value is always smaller than a quantization parameter Δ . Hence, a better visual quality of the watermarked data is achieved compared to Kundur's algorithm, where the change of the coefficient for embedding is done by the amount of Δ [123]. The use of a secure zone $\mathbb{Z}$ in our proposed scheme also reduces the bits error rate efficiently at the extraction stage. The narrower the width of the zone $\mathbb{Z}$, the higher possibility to decode the embedded watermark bit correctly. On the other hand, the change of the coefficient value to embed watermark bits would be higher.

9.2.3 Watermark Embedding Algorithm

To describe the proposed watermark embedding algorithm, the Lena image with the size of 256×256 pixels in Figure 9.1(a) is used as the original image; and a chirp signal of 256 samples in Figure 9.2 is used as the watermark. Figure 9.8 shows the flowchart of the proposed multiresolution watermark embedding scheme, which includes the following steps:

Figure 9.8: The watermark embedding process.

[Step 1] Discrete wavelet transform of the host image:

A three level DWT of the original image $\mathbf{I}$ is performed using Harr bases. The obtained wavelet coefficients are denoted as C . In the given watermarking example, the size of the sub-band HL3 or LH3 is of $32 \times 32 = 1,024$ pixels.

[Step 2] Generate the multiresolution watermark:

Both the real and the imaginary parts of the complex chirp signal $\mathbf{s}$ are sampled and coded into binary sequences, using the technique described in Section 9.1.2. The same chirp watermark $\mathbf{s}$ is to be embedded twice. So the total number of watermark samples (including both real and imaginary samples) to be embedded is $256 \times 2 \times 2 = 1,024$. The total number of watermark bits is thus $1,024 \times 7 = 7,168$. Note that the number of watermark samples is the same as the number of pixels of a sub-band in the third level wavelet decomposition of $\mathbf{I}$. In this context, every position (either from the horizontal or vertical sub-band) of the third level decomposition is to be embedded with one authentication watermark bit. To authenticate the content of the protected media, one would like to embed at least one authentication bit into every non-overlapping partition of the image. Here the minimum authentication unit is a 8×8 image block, as the 3-level wavelet decomposition is used and every position of the 3-rd level coefficient is embedded

with one watermark bit.

[Step 3] Key generation:

A random sequencing order is generated in order to disperse the embedding locations of the watermark samples. The watermark bits are to be embedded based on the randomly selected locations.

[Step 4] Multiresolution watermark embedding:

The detailed multiresolution embedding technique is presented in the Section 9.2.1. The rules to embed 7 bits of a watermark sample in a multiresolution manner are explained. The embedding of the watermark bits 0 or 1 is based on the quantization technique elaborated in Section 9.2.2. The equations for watermark embedding under three different scenarios are given in (9.6), (9.7) and (9.8). The marked wavelet coefficients are denoted as $\tilde{C}$.

[Step 5] Inverse wavelet transform:

The final watermarked image $\mathbf{X}$ is obtained by an inverse DWT of $\tilde{C}$ using Harr bases.

The watermark embedding locations and the quantization function $Q(\cdot)$ mapping the coefficients to odd or even multiples of the quantization coefficients, would be randomized via a key. In practice, the quantization step Δ as well as the width r of the embedding zone $\mathbb{Z}$ need to adjust according to the requirements of the watermarked image quality and the watermark extraction accuracy.

9.2.4 Watermark Embedding Results

Two watermark embedding results following the above presented embedding algorithm are shown in Figures 9.9 and 9.10, respectively. In the first example (Figure 9.9), the

Figure 9.9: Watermark embedding example 1: (a) a watermarked Lena image (PSNR=45.97dB), (b) the difference image between the original Lena image and the watermarked Lena image in (a), (c) the enlarged eye region of the watermarked Lena image, and (d) the enlarged eye region of the original Lena image.

quantization step Δ is set as 16, 8, 4 for $l=3, 2, 1$, respectively; and the width r of the embedding zone $\mathbb{Z}$ is set as $\frac{\Delta}{2}$, i.e., 8, 4, 2 for $l=3, 2, 1$, respectively. In the second example (Figure 9.10), the quantization step Δ is set as 32, 16, 8 for $l=3, 2, 1$, respectively; and the width r of the embedding zone $\mathbb{Z}$ is 16, 8, 4 for $l=3, 2, 1$, respectively.

The quality of a watermarked image is altered with the different settings of the quantization step Δ . The smaller value of Δ results in higher PSNR measure of the watermarked image. For instance, the PSNR of the watermarked Lena image in Figure 9.9(a) is 45.97dB, whereas the PSNR of the watermarked Lena image in Figure 9.10(a) is 40.20dB. The differences between the watermarked images and their original images have random image patterns as shown in Figures 9.9(b) and 9.10(b), respectively. No

Figure 9.10: Watermark embedding example 2: (a) another watermarked Lena image (PSNR=40.20dB), (b) the difference image between the original Lena image and the watermarked Lena image in (a), (c) the enlarged eye region of the watermarked Lena image, and (d) the enlarged eye region of the original Lena image.

visual difference is noticed from the enlarged images of the eye zone in Figures 9.9 (c) and (d). It is also difficult to perceive the difference between Figures 9.10 (c) and (d).

Here, we have shown two watermark embedding results. A watermarked image of different qualities can be attained as required by adjusting the quantization parameter Δ .

9.3 Watermark Extraction and Performance Evaluation

9.3.1 Watermark Extraction and Verification Scheme

This section presents the proposed *blind* watermark extraction scheme. Neither the original image nor the original watermark is required during the extraction and verification stage. The key information includes the random permutation for the random watermark embedding locations, the wavelet type, the values of the quantization parameter Δ and the quantization function $Q(\cdot)$. Figure 9.11 shows the flowchart of the blind watermark extraction and verification procedure.

Figure 9.11: The watermark extraction process.

[Step 1] Discrete wavelet transform of the received image:

The received image denoted as $\mathbf{X}'$, which could be the same as the watermarked image $\mathbf{X}$ or altered by attackers. Similarly, a three level DWT of the received image $\mathbf{X}'$ is performed using Harr bases. The resulted wavelet coefficients are denoted as $\tilde{C}'$.

[Step 2] Extraction of the embedded watermark bits:

Based on the watermark embedding locations provided by the key, the selected wavelet coefficients are quantized into symbol 0 or 1 using the same quantization function during embedding in (9.5). The extracted watermark bits $\{w'(i) \in (a_{in}', b_{in}')\}$ are estimated from odd or even quantization of the selected wavelet coefficients by

$$w'(i) = Q(\tilde{C}'(i)). \quad (9.9)$$

[Step 3] Content authentication:

In the case when the original watermark w is available at authentication stage, the authentication is performed by simple comparisons between the extracted watermark bits w' and the original watermark bits w .

The good feature of our proposed fragile watermarking scheme is that the original watermark is not necessarily required during authentication. In the absence of the original watermark, the extracted watermark bits are then converted back to the watermark sample, $\mathbf{s}'_i (= a'_i + jb'_i)$, with the values within the range of $[-1, 1]$, in the following way:

$$\begin{aligned} w'_a(i) &= \sum_{n=1}^7 a'_{in} \cdot 2^{7-n}, & w'_b(i) &= \sum_{n=1}^7 b'_{in} \cdot 2^{7-n}; \\ a'_i &= \frac{w'_a(i)}{63.5} - 1, & b'_i &= \frac{w'_b(i)}{63.5} - 1. \end{aligned} \quad (9.10)$$

Content authentication is thus based on the verification of the magnitude of the extracted complex chirp watermark $\mathbf{s}'$, which is computed as

$$mag(i) = \sqrt{a_i'^2 + b_i'^2}. \quad (9.11)$$

Without resorting the original watermark, the content can be still verified using the quantitative measurement in (9.11) – the magnitude of the extracted chirp

watermark signal. The parameter $mag(i)$ should be a constant, specially, with value of 1, in the absence of any attacks on the protected content. Note that we make use of the FM complex chirp signal with the magnitude of one as the watermark (see (9.1)).

Since the authentication process does not require the original watermark, it is possible to embed different chirp authentication watermarks into different images to be protected, but all of them can share the same authentication process without knowing the embedded watermark previously. The applications of the proposed scheme for content integrity verification and blind quality assessment in the absence of the original watermark, will be presented in Section 9.4.

9.3.2 Performance Evaluation: Robustness versus Tamper-Sensitivity

1) Performance against Signal Processing Attacks

The watermark extraction performance of the proposed watermarking scheme against common signal processing is evaluated. The sensitivity or the robustness of the proposed scheme is investigated under various attacks in the following scenarios:

- JPEG compression of quality factors 90%, 80%, 70%, 60%, 50%, 40%;
- Histogram equalization (uniform distortion);
- Sharpening (low-pass filtering) - processed by Adobe Photoshop 7.0;
- Blurring (high-pass filtering) - processed by Adobe Photoshop 7.0;
- Additive Gaussian noise;

Table 9.1: The BERs of the extracted watermarks under JPEG compressions of quality factor (QF) of 90%, 80%, 70%, 60%, 50% and 40%.

Example 1				Example 2			
QF	$l=3$	$l=2$	$l=1$	QF	$l=3$	$l=2$	$l=1$
90%	0	0.0728	0.4587	90%	0	0.0010	0.2629
80%	0.0068	0.1880	0.4802	80%	0	0.0405	0.4248
70%	0.0313	0.2847	0.4832	70%	0.0020	0.0835	0.4656
60%	0.0781	0.3413	0.5034	60%	0.0020	0.1274	0.4678
50%	0.1357	0.4214	0.4995	50%	0.0098	0.1865	0.4805
40%	0.2168	0.4634	0.4978	40%	0.0186	0.2529	0.4783

- Salt-and-pepper noise (value set to 255 or 0, respectively).

The watermark extraction of the two watermarked Lena images in Figures 9.9(a) and 9.10(a) are tested under the above mentioned manipulations. The performance is measured in terms of the bit error rate (BER) of the extracted watermark bits, which is the ratio of bits in errors, N_e , relative to the total number of the hidden watermark bits, N_w , given by

$$\text{BER} = \frac{N_e}{N_w}. \quad (9.12)$$

The BER of the extracted watermark of all the three wavelet decomposition levels are analyzed.

Table 9.1 shows the BERs of the extracted watermarks under attacks of JPEG compressions. Table 9.2 shows the extraction results when the watermarked image is attacked by various signal processing attacks. The corresponding PSNR measures of the distorted watermarked images are given in Table 9.3. Note that the BERs in both given watermarking examples are zeros when there is no attack. The watermark embedded in a higher decomposition level (low frequency band) has better resistance against distortions. This characteristic is resulted by using the multiresolution embedding technique. Comparing the BERs for examples 1 and 2, it is observed that the embedding using a larger quantization step (in example 2) is less sensitive against

Table 9.2: The BERs of the extracted watermarks under various signal processing attacks.

Attacks	Example 1			Example 2		
	$l=3$	$l=2$	$l=1$	$l=3$	$l=2$	$l=1$
No attacks	0	0	0	0	0	0
Histogram equalization	0	0.0728	0.4587	0	0.0010	0.2629
Sharpening	0.0068	0.1880	0.4802	0	0.0405	0.4248
Blurring	0.0313	0.2847	0.4832	0.0020	0.0835	0.4656
Gaussian noise	0.0781	0.3413	0.5034	0.0020	0.1274	0.4678
Salt-and-pepper noise	0.1357	0.4214	0.4995	0.0098	0.1865	0.4805

Table 9.3: The PSNR measures (in dB) of the watermarked image under various signal processing attacks.

Attacks	Example 1	Example 2
JPEG comp. 90%	38.83	37.02
JPEG comp. 80%	35.94	34.96
JPEG comp. 70%	34.41	33.74
JPEG comp. 60%	33.34	32.81
JPEG comp. 50%	32.56	32.14
JPEG comp. 40%	31.82	31.45
Histogram equalization	16.69	16.66
Sharpening	29.31	28.81
Blurring	31.53	31.30
Gaussian noise	32.77	32.26
Salt-and-pepper noise	24.85	24.97

tampering.

2) Performance against Width of the Embedding Zone $\mathbb{Z}$

Even though the quantization parameter Δ is fixed, the quality of the watermarked image as well as the BER of the extracted watermark may still be varied by adjusting the width r of the secure embedding zone $\mathbb{Z}$. The two examples in Figures 9.9 and 9.10 are obtained by setting the width $r = \frac{\Delta}{2}$ with the varying quantization step of Δ . In this part, the experiments are performed by fixing the value of Δ as 16, 8, 4 for

Figure 9.12: The quality of the watermarked Lena image with respect to the width r of the secure embedding zone $\mathbb{Z}$.

$l = 3, 2, 1$, with the varying width r . The effects of the choice of r on the quality of the watermarked image as well as the BER of the extracted watermark have also been investigated.

Figure 9.12 shows the quality of watermarked image measured in terms of PSNR with respect to the width r . The PSNR value is gradually increased when the width r is getting wider. It implies that the less number of selected coefficients are falling outside of the secure embedding zone $\mathbb{Z}$ when the embedding zone is wider, and therefore the less number of coefficients are required to be adjusted for embedding.

The BER of the extracted watermark with respect to the width r is shown in Figure 9.13. The three curves correspond to the BERs of the extracted watermark for decomposition level $l = 3, 2, 1$, respectively. When $r \leq \frac{\Delta}{2}$ ($8 \times \frac{\Delta}{16}$), i.e., the width of the embedding zone is no larger than half of the quantization step Δ , the error-free decoding process of the watermark bits is achieved. The embedding in a higher decomposition level, i.e., in the lower resolution level, can achieve lower BER.

Figure 9.13: The BERs of the extracted watermark with respect to embedding using different width r of the secure embedding zone $\mathbb{Z}$.

3) Performance Comparisons with Kundur's Scheme

The main difference of our proposed scheme from Kundur's scheme [123] is on the watermark embedding algorithm. In Kundur's scheme, if $C(i) = w(i)$, no change in the coefficient is necessary; otherwise the change of the coefficient for embedding follows the rule:

$$\tilde{C}(i) = \begin{cases} C(i) + \Delta & \text{if } C(i) \leq 0 \\ C(i) - \Delta & \text{if } C(i) > 0 \end{cases}. \quad (9.13)$$

Specially, a secure embedding $\mathbb{Z}$ is employed in our proposed scheme. The embedding equations are given in (9.6), (9.7) and (9.8). To elaborate the benefits of using such a secure embedding zone, the extraction results of Kundur's scheme are shown to compare with the results of our proposed scheme.

A watermarked Lena image using Kundur's embedding technique is shown in Figure 9.14. Here we use the same quantization setting Δ as those used for the watermarked image in Figure 9.9, i.e., $\Delta = 16, 8, 4$ for $l = 3, 2, 1$, respectively. Note that Kundur's scheme results a relatively worse PSNR measure of the watermarked image under

Figure 9.14: A watermarked Lena image (PSNR=42.63dB), by using the Kundur's watermarking scheme.

the same embedding conditions. As discussed in the Section 9.2.2, the change of the coefficient value in our embedding scheme is always less than the change required for Kundur's embedding scheme, which is the amount of Δ .

Both the embedded watermarks for the two watermarking examples using proposed scheme in Figures 9.9(a) and 9.10(a) can be fully recovered without any bit error, when there is no attack, i.e, $\mathbf{X}' = \mathbf{X}$. However, the experiments show that using the Kundur's scheme the extracted watermark contains some error bits even if the watermarked image is untampered, as shown in Table 9.4. This could be explained as follows. Some of the marked coefficients could be very close to the boundary of two neighboring quantization steps. Therefore, a very small error introduced during DWT and IDWT processing could make the marked coefficient falling into the neighboring quantization step and thus invert the symbol of a embedded watermark bit. The use of a secure embedding zone as in our scheme successfully brings such a coefficient into a more 'secure' position against the processing errors due to transformations.

Compare the figures in Tables 9.1 and 9.4, it is observed that the proposed watermarking scheme has slightly higher robustness than Kundur's scheme due to the use of a secure embedding zone. For authentication applications, it is desirable that the extracted watermark has no error bits if there is no attack at all. Our proposed scheme

Table 9.4: The BERs of the extracted watermarks under JPEG compressions for Kundur's watermarking scheme.

QF	$l=3$	$l=2$	$l=1$
No attacks	0.0059	0.0103	0.0371
90%	0.0488	0.1099	0.4277
80%	0.0732	0.1787	0.4739
70%	0.1201	0.2813	0.4875
60%	0.1592	0.3721	0.4656
50%	0.1924	0.4150	0.4897
40%	0.2598	0.4551	0.4880

improves the Kundur's scheme such that it gives no false alarms if the received data is untampered.

9.4 Applications

Fragile watermarking is designed to be a powerful content authentication tool for detection of every possible change in the protected digital content, no matter what position the attack occurs and how small the modification is in the content. A multiresolution fragile watermarking scheme by using complex chip signals is proposed in the previous section. This section demonstrates the applications for content authentication through the proposed multiresolution fragile watermarking scheme. The first application is for the content integrity verification with localization capability. Unlike the traditional statistically-authenticated fragile watermarking schemes [144], our scheme does not require the original watermark for content integrity verification. The second application is to assess the quality of the watermarked content by investigating the extracted complex chirp watermark in the absence of the original watermark.

9.4.1 Content Integrity Verification Without Resorting the Original Watermark

The promising feature for the authentication applications by using complex chirp signals is that the authentication does not require any a priori information about the hidden watermark. This allows the embedding of arbitrary chirp watermarks into different images. Here we present how to use the proposed scheme for content integrity verification with localization without resorting the original watermark. Especially we aim to detect and locate the malicious changes, such as feature adding, cropping and replacement.

Recall that the real and the imaginary parts of a sample for the given complex chirp watermark are embedded simultaneously into two vertically neighboring clusters of wavelet coefficients, as shown in Figures 9.4 and 9.5. Let the embedding position for the real part is (p, q) in sub-band HL3, then the embedding position for imaginary part is $(p+1, q)$ in sub-band LH3. For three level DWT, every wavelet coefficient at 3rd level corresponds to a 8×8 block of the original image. The decimal values of real and imaginary samples of the extracted watermark bits are obtained by (9.10). The detector is performed based on examining the magnitudes of the extracted complex chirps signal computed using (9.11). The magnitudes of frequency modulated chirp signal should be a constant when the hidden watermark bits can be extracted perfectly. In particular, in our watermarking example, the constant magnitude is one. Consider the scenario that the received content is altered. The tampered area can be detected by the non-one values in the magnitude measure of the extracted complex chirp signal. Note that the authentication unit in this case is a block of 16×8 pixels, which is the region corresponding to a pair of watermark samples – one real and one imaginary samples of a complex chirp watermark sample. For example, a Lena image of 256×256

pixels is partitioned into 512 such 16×8 image blocks, and they are labelled from 1 to 512 in column-wise order, as illustrated in Figure 9.15. As shown in Figure 9.16, the magnitude of the extracted complex chirp signal is almost constant around the value 1 if no any change occurs in the watermarked data.

Figure 9.15: The column-wise location indices using 16×8 blocks.

Figure 9.16: The detector response when there is no attacks.

1) Tamper in Single Authentication Block

Assume the watermarked Lena image is tampered by one pixel's value at location of (135, 138), by changing the value from 196 to 0 (as shown in Figure 9.17). This pixel falls into the 16×8 authentication block with index of 281. The detector response is displayed in Figure 9.18(a). A peak at block index of 281 is detected. The index 281 corresponds to a 16×8 image block with four corner locations: (129, 137), (129, 144), (144, 137) and (144, 144), as illustrated in Figure 9.18(b).

As the same complex chirp signal is embedded repeatedly twice, it is possible to further locate the error pixel within a smaller authentication unit – the upper or lower 8×8 block. The other copy of the same sample in the extracted watermark is used as a reference to verify which 8×8 block is tampered. If the real parts of these two extracted complex watermark samples differ, the tampered pixel is in the upper 8×8 block. On the other hand, the tampered pixel falls in the lower 8×8 block if the imaginary parts are not matched. The minimum authentication unit achievable by the proposed scheme is 8×8 block, however, it is true only when there is at least one copy of the watermark samples decoded correctly without tampering. Otherwise the authentication unit is 16×8 block only.

Figure 9.17: A tampered Lena image by setting the pixel's value at (135, 138) to 0 (A black dot in the left eye region is observed).

(a)

(b)

Figure 9.18: (a) The detector response when there is one pixel's value altered, (b) localization of the alteration.

The tampered region is investigated for the example in Figure 9.17. The extracted complex chirp sample at image block of index 281 is found to be $(0.7638 + j0.9685)$, which is tampered as its magnitude is not one. Comparing this error sample with the other correct sample $(-0.2441 + j0.9685)$, the real parts of the two complex samples do not match, therefore, it draws the conclusion that the tampered pixel is within the upper half 8×8 block. In order to achieve better localization capability, that is, the authentication block of 8×8 pixels, it is suggested to embed the repeated copy of the

watermark sample is the other distant block avoiding the lost of both copies of the sample at the same time.

2) Tamperers in Multiple Authentication Blocks

One more example is illustrated where more than one authentication block are tampered. As shown in Figure 9.19, the mouth region of the watermarked Lena image is replaced deliberately by a different mouth image. From the image itself without comparing with the original image, it would be difficult to tell the alterations in the content. In the following, we demonstrate the detection of alterations without the help of the original watermark signal during authentication. The detector detects ten authentication blocks in error, as shown in Figure 9.20(a). These ten tampered blocks are of index 267, 268, 283, 284, 299, 300, 315, 316, 331 and 332. They form a larger rectangular block corresponding to the marked Lena image with four corner locations: (161, 129), (161, 168), (192, 129) and (192, 168), as shown in Figure 9.20(b). The tampered region can be further located in a smaller region by comparing the error samples with the corresponding correct copies embedded in the other randomly selected locations. The alteration in the Lena image is finally located within the shadowed areas as shown in Figure 9.20(b). The real parts of the chirp watermark samples in block with index 267, 283, 299, 315 and 331 are untampered; and the imaginary parts of the chirp watermark samples in block with index 268, 316 and 332 has also no alterations. This shadowed region matches the deliberately tampered mouth region of the watermarked Lena image.

This part has shown one application of our proposed fragile watermarking scheme in content integrity verification. The use of the complex chirp watermark enables the authentication without resorting the original watermark. The authentication unit is 16×8 block. The localization accuracy for alterations detection is able to be improved

Figure 9.19: (a) A tampered watermarked Lena image, and (b) the tampered region of (a) – the mouth area of the watermark Lena image.

(a)

(b)

Figure 9.20: (a) The detector response for the tampered watermarked Lena image in 9.19(a) ('o' denotes the samples that are of non-one magnitude), (b) localization of the tampered blocks.

to 8×8 block if the other copy of a watermark sample can be recovered correctly.

9.4.2 Blind Quality Assessment Using Multiresolution Fragile Watermarking Technique

The *quality assessment* of the marked content could be another potential application of our proposed multiresolution fragile watermarking. The chirp watermark is represented in 7-bit binary code, and embedded using multiresolution technique. The MSB is inserted in the highest level of wavelet decomposition and the less significant bits are in the lower decomposition level. In this context, the more significant bits of the watermark samples have better resistance against tampering. Such multiresolution watermark is able to sustain a certain degree of attacks before it is distorted or lost totally. The embedded watermark is distorted gradually with respect to the strength of attacks. Therefore, we are able to evaluate the strength of quality degradation of the marked content by assessing the distortion of the extracted fragile watermark. In [145], campisi *et al.* presents a ‘blind’ quality estimation of a multimedia communication link by means of digital fragile watermarking. Specially, a spread-spectrum based fragile watermark is inserted in the video sequence to evaluate the coder-channel quality by quality of the extracted watermark from the received video after a coding or transmission process. The quality is performed based on the mean-square-error (MSE) between the extracted and the original watermarks. Note that the original watermark is still needed for the quality assessment in this presented method.

In this section, we employ our proposed fragile watermarking scheme as means of blind quality assessment for the received marked content. The quality of the watermarked image under JPEG compression and various common signal processing attacks is assessed. Comparing to campisi’s method [145], the use of complex chirp watermark

Figure 9.21: The actual and the “ideal” curves for the magnitude of the extracted complex chirp signal.

signal has good feature working as a tool of quality assessment – the original watermark is not required, thus a ‘true’ blind quality assessment is achieved. The evaluation is performed directly by investigating the magnitudes of the extracted complex chirp watermark signal.

Ideally the instantaneous magnitudes of the extracted complex chirp signal should be a constant (a flat curve), as illustrated in Figure 9.21. However, the actual curve is distorted due to the possible manipulations of the digital content. The distortion of the actual curve with respect to the ideal curve is measured for the blind quality assessment of the received digital media. The MSE between the actual and the ideal curves is computed as

$$\begin{aligned}
 \text{MSE} &= \frac{1}{N_w} \sum_{i=1}^{N_w} \left[\text{mag}'(i) - \text{mag}'_{ideal}(i) \right]^2 \\
 &= \frac{1}{N_w} \sum_{i=1}^{N_w} \left[\text{mag}'(i) - 1 \right]^2,
 \end{aligned} \tag{9.14}$$

where mag' and mag'_{ideal} denote the actual and the ideal magnitudes of the complex chirp signal, respectively, and N_w is the length of the watermark. Here, mag'_{ideal} is constant one. A more sensitive quality measure, signal-to-noise ratio (SNR) in dB, is

defined as

$$\text{SNR} = 10 \log_{10} \left(\frac{1}{\text{MSE}} \right). \quad (9.15)$$

Note that the numerator within the ‘ $\log_{10}$ ’ operation is 1, meaning that the average energy of the “ideal” curve of the extracted complex chirp signal is expected to be one. A higher MSE or a lower SNR measure implies a worse quality for the received marked image.

Table 9.5 shows the quality assessments in terms of MSE in (9.14) and SNR in (9.15) when the watermarked Lena image in Figure 9.9(a) is processed by JPEG compression operations with various quality factor. It is observed that the MSE increases and the SNR drops gradually as the quality factor decreases from 100% to 50%. The change of SNR measure follows the same trend as that of the actual PSNR quality measure for the compressed watermarked image. Note that the MSE for the marked image with JPEG compression quality factor 100% is non-zero. This is due to the quantization noise, which becomes smaller for the smaller quantization step.

More quality assessment results for the watermarked image under various signal processing attacks are shown in Table 9.6. The actual PSNR measures (in dB) of the corresponding altered marked images are given in Table 9.3. By estimating the distortions of the magnitude of the extracted complex chirp watermark, it can approximately indicate the amount of degradation of the watermarked content.

The proposed quality assessment technique does not require the original watermark signal as well as the original watermark signal. The assessment is performed according to the distortions of the extracted watermark. Specially, the magnitude of the extracted complex chip watermark is tested to estimate the distortions of the watermark. In this way, the original watermark is not required and *blind* quality assessment can be achieved. This multiresolution embedding technique allows the gradual degradation of

Table 9.5: The quality assessment based on the MSE (in (9.14)) and SNR (in (9.15)) measures of the extracted watermark when the watermarked Lena image is under various JPEG compression attacks.

JPEG Comp. (QF)	Actual PSNR (dB)	Quality Assessment	
		MSE	SNR (dB)
100%	45.97	2.0025×10^{-5}	46.98
99%	45.35	5.1460×10^{-4}	32.89
98%	44.41	0.0014	28.69
95%	41.53	0.0101	19.95
90%	38.83	0.0267	15.74
80%	35.94	0.0641	11.93
70%	34.41	0.0956	10.19
60%	33.34	0.1061	9.74
50%	32.56	0.1227	9.11

Table 9.6: The quality assessment based on the MSE (in (9.14)) and SNR (in (9.15)) measures of the extracted watermark when the watermarked Lena image is under various signal processing attacks.

Attacks	Quality Assessment	
	MSE	SNR (dB)
Histogram equalization	0.1282	8.92
Sharpening	0.1038	9.84
Blurring	0.1062	9.74
Gaussian noise	0.1202	9.20
Salt-and-pepper noise	0.0926	10.34

the hidden watermark when the attack is getting heavier. Hence, before the watermark is totally distorted or lost, the distortion of the embedded watermark provides efficient information on the trend of how serious the attack is on the marked content. However, the fragile watermark is rather brittle. As we can see from Table 9.5, the SNR measure degrades rapidly with slight compression. Therefore, a ‘more robust’ or a semi-fragile watermarking scheme is desirable to perform the blind quality assessment. This would be one of the open research areas for the future work.

9.5 Chapter Summary and Conclusion

This chapter presents a wavelet-domain multiresolution fragile watermarking scheme. The FM complex chirp signal is employed as watermark, which is encoded into binary watermark bit sequence by PCM technique. Both the real part and the imaginary part of the chirp signal are embedded simultaneously into the HL and LH sub-bands, respectively, in a hierarchical manner. More significant bits are embedded in the lower frequency sub-bands of the wavelet decomposition of the host image. Our embedding algorithm is improved from the Quantization-Index-Modulation (QIM) technique proposed by Kundur *et al.*. In particular, a secure embedding zone is exploited such that all the embedded wavelet coefficients are adjusted to fall into this secure zone. It has shown that the proposed scheme successfully improves the Kundur's scheme in terms of the quality of the watermarked image and the bit error rates (BERs) for the extracted watermark. The sensitivity of the hidden fragile watermark under various attacks is also demonstrated.

The excellent feature by using complex chirp watermark signal in authentication is that no original watermark is required. Thus arbitrary FM chirp watermark can be used for different images. Based on the embedding locations provided by the key, the embedded watermark can be extracted by the quantization process of the wavelet coefficients for the received content. In the absence of the original watermark, the magnitude of the extracted FM complex chirp signal is tested to detect and locate the possible alterations in the content. Two potential applications of our proposed multiresolution fragile watermarking scheme are presented. The first one is for content integrity verification without resorting the original watermark. The second one is the blind quality assessment for the received watermarked image. The more challenging problem in content authentication is to propose techniques used to authenticate

the content, rather than representations of the content in various file formats, such as, JPEG, GIF [109]. The semi-fragile authentication watermarking that can sustain moderate degree of compressions needs to pay attention for further research.

Chapter 10

Final Remarks

This chapter highlights some open research areas (which are not addressed in this thesis) and concludes with a brief summary of research contributions and concluding remarks.

10.1 Future Works

1) Issue of Attacks and Counterattacks with ICA

Collusion attack is a cost effective attack in digital watermarking where several copies with the same content but different watermarks are combined. The attackers try to remove the watermark by averaging these images or by taking small parts of all images and reassembling them. In Part I of the thesis, we have discussed the multimedia distribution applications, where multiple copies of the same content are available. A challenging issue is raised that is worthy of further investigation: ICA method itself could be used as an attack tool by attackers or illegal users. The attacker who has two or more copies (mixtures) of the same content, could try to extract/recover the

original image as well as the hidden watermarks by the means of blind source separation techniques. The issue of attacks and counterattacks with ICA in digital watermarking would be investigated. How good is the extracted original image in the collusion attack by using ICA methods? Could it be as good as the original un-watermarked content? The embedding of watermarks could be carefully designed such that even the attackers know a few copies of the content, it is still not possible in practice to recover the original and the embedded watermark without knowing the key or the original content.

2) Fragile Authentication Watermarking with Recovery Capability

We have proposed multiresolution fragile watermarking scheme in Part III of this thesis. It enables blind content authentication with localization capability. However, the proposed scheme lacks recovery property for error bits, which could be another promising feature. The embedding of recovery information has been proposed to enable the recovery of the damaged or lost blocks [146, 147, 148]. In [146], the lost information in a self-authentication-and-recovery (SARI) image can be approximately recovered based on the embedded watermark, which includes the content-based authentication information and recovery information. The recovery information bits are generated from the scaled down version of the original image through DCT transform and quantization. The paper [147] proposed a secure and flexible fragile digital video authentication watermark which enables the self-recovery of video content after malicious manipulations. The recovery packet contains a highly compressed version of a spatio-temporally distant block. The embedding of recovery information could be investigated to enable the recovery capability of an authentication system.

3) Semi-fragile Authentication Watermarking

Semi-fragile authentication watermarking [10] would be carried out. It is a desirable feature for an authentication system to distinguish malicious attacks, such as, cropping-and-replacement, from the non-malicious attacks, such as, compressions. Furthermore, by the means of semi-fragile watermarking, the extracted watermark is not so brittle against attacks. It would achieve better performance for quality assessment of the distorted watermarked content, since the embedded watermark can resist a moderate degree of attacks and would not be removed easily in the presence of slight attack.

4) Watermark Synchronization

To evaluate the extraction performance of a proposed scheme, various signal processing operations are applied to degrade the quality of the watermarked content for testing. The schemes in this thesis deal with the degradations on the watermarked signal as the nature of fading noise. When the noise level is quite high, the watermark could be undetectable. There are a number of other types of distortions such as geometric transformations, which cannot be modelled as fading noise. These include scaling, rotation or general affine transformations. Slight modifications of this type would make it difficult to extract the hidden watermark due to the loss of synchronization. It should be emphasized that the watermark still exists within the geometrically distorted multimedia signal, but it is not easy to recover.

Works have been presented that attempt to overcome the geometric distortions by embedding the watermark in RST-invariant domain [79, 149]. Lin *et al.* proposed a watermarking algorithm that the watermark is embedded into a one-dimensional (1-D) signal obtained by taking the Fourier transform of the image, resampling the Fourier magnitudes into log-polar coordinates, and then summing a function of those

magnitudes along the log-radius axis [149]. In the paper [79], Fourier-Mellin transform is applied in watermarking to produce watermarks that are resistant to RST distortions. Brute-force re-synchronization would be attempted for the watermarking schemes, which embed the watermark in a non-affine invariant domain. Some algorithms are proposed to invert affine transformations by tuning a particular set of parameters that has been used to warp the watermarked content [92, 93, 94, 150]. In [150], a fast automatic resynchronization scheme has been developed. It incorporates multi-resolution matching and coarse-fine searching, achieving low computation complexity.

Further investigation could be involved in the issue of watermark synchronization. The RST-invariant embedding techniques could be adopted in our proposed watermarking schemes to improve their performance against RST attacks.

5) Applications

- The techniques presented in this thesis are all demonstrated using the media of digital images. The proposed watermarking schemes and watermark detection/extraction technique could be generalized to the other medias, such as, audio and video. The applications of the works in this thesis would be investigated for audio and video, by considering the practical requirements of those media.
- The applications of the proposed transparent robust information hiding scheme in Part II of the thesis will be investigated, such as, for the security of biomedical images, where the quality of the watermarked content is required to be high.
- The practical implementations of the works presented in the thesis will be investigated in the media security systems. These would lead to the solutions for Digital Rights Management (DRM) of digital multimedia data [151, 152, 153].

- More topics on the multimedia security such as *copy control and access control* [2], would be also studied.

10.2 Summary and Concluding Remarks

This thesis introduced and addressed several technical challenges of information hiding technique for intellectual property protection of digital media. The information hiding technique is also known as digital watermarking in the context of hiding information for the protection of cover media. We concentrated on three tasks to enforce the IP rights of digital media – transaction tracking, ownership verification and content authentication. We firstly proposed a new watermark detection and extraction method by using Independent Component Analysis technique, which is not previously proposed. The blind watermarking scheme by using ICA method for extraction is proposed and efficiently applied for transaction tracking in multimedia distribution applications. We then dedicated in designing and developing a transparent robust information hiding scheme for ownership verification and a muliresolution fragile watermarking scheme by using complex chirp signal for content authentication.

Cryptography techniques are developed to enhance the secure transmission of multimedia data, while information hiding techniques provide an efficient means to enhance the intellectual property protection of the digital media. A great volume of information hiding or watermarking schemes are thus proposed for copyright protection, content verification and so on. However, the future role of watermarking within the multimedia communication and processing framework is not clear yet. There has been no rush yet to embrace any of the current schemes. A period of inspection and appraisal is still needed. As researchers propose solutions and users test them and debate for their merits, then some methods of watermarking will emerge as useful and widely used

tools. In the meantime, the challenge for the researchers is to develop more invisible, decodable, and secure watermarking or information hiding methods, and perhaps to meet even more specifications to meet the demands.

Appendices

Appendix A:

A Statistic-Based ICA Algorithm

– JADE

Cardoso *et al.* proposed a statistic-based blind identification algorithm by Joint Approximate Diagonalization of Eigen-matrices (JADE) [42, 43]. Consider exploiting second order information by whitening the observed mixture signals $\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t)$ ($\mathbf{s}$ is the source signals, $\mathbf{A}$ is the mixing matrix). This is done via a whitening matrix $\mathbf{W}$, i.e., a $n \times m$ matrix such that $\mathbf{W}\mathbf{y}(t)$ is spatially white. The whiteness condition is

$$\mathbf{I}_n = \mathbf{W}\mathbf{R}_y\mathbf{W}^H = \mathbf{W}\mathbf{A}\mathbf{A}^H\mathbf{W}^H, \quad (\text{A.1})$$

where $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. Equation (A.1) implies that $\mathbf{W}\mathbf{A}$ is a unitary matrix: for any whitening matrix $\mathbf{W}$, it then exists a unitary matrix $\mathbf{U}$ such that $\mathbf{W}\mathbf{A} = \mathbf{U}$. As a consequence, matrix $\mathbf{A}$ can be factored as

$$\mathbf{A} = \mathbf{W}^\sharp \mathbf{U} = \mathbf{W}^\sharp [\mathbf{u}_1, \dots, \mathbf{u}_n], \quad (\text{A.2})$$

where $\mathbf{U}$ is unitary, and $^\sharp$ denotes the pseudoinverse. The use of second-order information – in the form of an estimate of $\mathbf{R}_y$ which is used to solve for $\mathbf{W}$ in (A.2) – reduces the determination of the $m \times n$ mixing matrix $\mathbf{A}$ to the determination of a unitary $n \times n$ matrix $\mathbf{U}$. The whitened process $\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t)$ (see Figure A.1) still obeys a linear model:

$$\mathbf{z}(t) \stackrel{\text{def}}{=} \mathbf{W}\mathbf{x}(t) = \mathbf{W}(\mathbf{A}\mathbf{s}(t) + \mathbf{n}(t)) = \mathbf{U}\mathbf{s}(t) + \mathbf{W}\mathbf{n}(t). \quad (\text{A.3})$$

This signal part of the whitened process is a unitary mixture of the source signals. Note that all the information contained in the covariance is ‘exhausted’ after the whitening, in the sense that changing $\mathbf{U}$ in (A.3) to any other unitary matrix leaves unchanged the covariance of $\mathbf{z}(t)$.

Let $\mathbf{V}$ denotes a $n \times n$ unitary matrix and further define $\mathbf{e}(t)$, as in Figure A.1,

$$\mathbf{e}(t) \stackrel{\text{def}}{=} \mathbf{V}^H \mathbf{z}(t) = \mathbf{V}^H \mathbf{U} \mathbf{s}(t) + \mathbf{V}^H \mathbf{W} \mathbf{n}(t). \quad (\text{A.4})$$

If $\mathbf{V} = \mathbf{U}$ then $\mathbf{V}^H \mathbf{U} = \mathbf{I}_n$ and the coordinates of $\mathbf{e}(t)$ are the (noise corrupted) source signals.

Figure A.1: Inverting $\mathbf{A}$ by chaining a whitener and a unitary matrix.

More generally, if $\mathbf{V}$ is essential equal to $\mathbf{U}$, the coordinates of $\mathbf{e}(t)$ are the phase-shifted source signals, possibly permuted and corrupted by additive Gaussian noise, so that higher-order *cross*-cumulants are zero.

Cardoso *et al.* proposed the approach to blind identification by exploiting the fourth-order cumulants of the array output. Let say, for $\mathbf{v}$ a complex d -dimensional random vector with coordinates $v_1, \dots, v_d$ and finite 4th-order cumulants, a cumulant set denoted as $\mathcal{Q}_{\mathbf{v}}$ is defined as

$$\mathcal{Q}_{\mathbf{v}} \stackrel{\text{def}}{=} \{Cum(v_i, v_j^*, v_k, v_l^*) \mid 1 \leq i, j, k, l \leq d\}. \quad (\text{A.5})$$

In [43], Cardoso *et al.* proposed to determine $\mathbf{U}$ as the unitary maximizer of the criterion $c(\mathbf{V})$:

$$c(\mathbf{V}) \stackrel{\text{def}}{=} \sum_{i,k,l=1,n} |Cum(e_i, e_i^*, e_k, e_l^*)|^2, \quad (\text{A.6})$$

which is equivalent to minimizing the sum of the squared cross-cumulants with distinct first and second indices. The main reason for considering criterion (A.6) is its link to underlying eigenstructures which allows for an efficient optimization of it by the mean of ‘joint diagonalization’.

- **Joint Diagonalization**

Let $\mathcal{N} = \{N_r | 1 \leq r \leq s\}$ be a set of $\mathbf{s}$ matrices with common size $n \times n$. A *joint diagonalizer* of the set $\mathcal{N}$ is defined as a unitary maximizer of the criterion

$$C(\mathbf{V}, \mathcal{N}) \stackrel{\text{def}}{=} \sum_{r=1,s} |\text{diag}(\mathbf{V}^H N_r \mathbf{V})|^2, \quad (\text{A.7})$$

where $|\text{diag}(\cdot)|$ is the norm of the vector build from the diagonal of the matrix argument. When the set $\mathcal{N}$ contains only one hermitian matrix, joint diagonalization is equivalent to usual unitary diagonalization. If the set $\mathcal{N}$ cannot be exactly jointly diagonalized (this is the case when sample cumulants are processed), the unitary maximization of (A.7) defines a somewhat arbitrary but quite natural ‘joint approximate diagonalization’.

- **Representation of the 4th-order cumulants by eigenmatrices**

The computational efficiency of joint diagonalization can be further increased by downsizing $\mathcal{N}_p$ to a smaller set made of the significant ‘eigenmatrices’ of Q_z . Let say, for any d -dimensional complex random vector $\mathbf{v}$ with 4th-order cumulants, there exist d^2 real numbers $\lambda_1, \dots, \lambda_{d^2}$ and d^2 matrices $\mathbf{M}_1, \dots, \mathbf{M}_{d^2}$, called eigenmatrices verifying

$$Q_v(\mathbf{M}_r) = \lambda_r \mathbf{M}_r \quad \text{Tr}(\mathbf{M}_r \mathbf{M}_s^H) = \delta(r, s) \quad 1 \leq r, s \leq d^2. \quad (\text{A.8})$$

After ordering the eigenvalues by decreasing order of magnitude, the eigen-set of Q_z , which contains the relevant 4th-order information, is defined as

$$\mathcal{N}^e \stackrel{\text{def}}{=} \{\lambda_r \mathbf{M}_r | 1 \leq r \leq n\}. \quad (\text{A.9})$$

This reduced set of n matrices (rather than n^2 in $\mathcal{N}^p$) together with the extended Jacobi technique makes the maximization of $c(\mathbf{V})$ computationally attractive.

- **The JADE Algorithm**

A blind identification algorithm by Joint Approximate Diagonalization of Eigenmatrices (JADE) can be described by the following steps.

-
- Step 1.** Form the sample covariance $\hat{\mathbf{R}}_x$ and compute a whitening matrix $\hat{\mathbf{W}}$.
- Step 2.** Form the sample 4th-order cumulants $\mathcal{Q}_z$ of the whitened process $\hat{\mathbf{z}}(t) = \hat{\mathbf{W}}\mathbf{x}(t)$; compute the n most significant eigenpairs $\{\hat{\lambda}_r, \hat{\mathbf{M}}_r | 1 \leq r \leq n\}$.
- Step 3.** Jointly diagonalize the set $\hat{\mathcal{N}}^e = \{\hat{\lambda}_r \hat{\mathbf{M}}_r | 1 \leq r \leq n\}$ by a unitary matrix $\hat{\mathbf{U}}$.
- Step 4.** An estimate of $\mathbf{A}$ is $\hat{\mathbf{A}} = \hat{\mathbf{W}}^\# \hat{\mathbf{U}}$.

Step 1 is implemented via eigendecomposition of $\hat{\mathbf{R}}_x$. With the assumption of white noise, an estimate $\hat{\sigma}$ of the noise variance is the average of the $m-n$ smallest eigenvalues of $\hat{\mathbf{R}}_x$. Denote $\mu_1, \dots, \mu_n$ the n largest eigenvalues and $\mathbf{h}_1, \dots, \mathbf{h}_n$ the corresponding eigenvectors of $\hat{\mathbf{R}}_x$. A whitener is $\hat{\mathbf{W}} = [(\mu_1 - \hat{\sigma})^{-1/2} \mathbf{h}_1, \dots, (\mu_n - \hat{\sigma})^{-1/2} \mathbf{h}_n]^H$.

In step 2, computation of the eigenmatrices amounts to diagonalizing a $n^2 \times n^2$ matrix made from the elements of $\hat{\mathcal{Q}}_z$. A standard algorithm for eigendecomposition of hermitian matrices will perfectly do, but more efficient implementations can also be devised, by taking into account additional cumulant symmetries or the fact that only the n most significant eigenpairs are needed [154]. Recall that computation of the eigenmatrices may be bypassed if, for simplicity, joint diagonalization is performed on the parallel set $\mathcal{N}^p$. An even simpler implementation is to form a set $\mathcal{N} = \{Q_z(C_r) | 1 \leq r \leq s\}$, where C_r 's are s arbitrary matrices in arbitrary number. Of course, identifiability cannot be guaranteed *a priori* and performance may be significantly lower than when $\mathcal{N}^e$ or $\mathcal{N}^p$ are used.

Step 3 is implemented by extending the single-matrix Jacobi technique. Note that when $n=2$, the Jacobi technique is not iterative: a unique Given rotation achieves (joint) diagonalization. Also recall that joint diagonalization may be initialized with the (usual) diagonalizer of a single cumulant matrix.

In step 4, the pseudo-inverse of $\hat{\mathbf{W}}$ needs not be explicitly computed: the eigendecomposition of $\hat{\mathbf{R}}_x$ may be recycled by $\hat{\mathbf{W}}^\# = [(\mu_1 - \hat{\sigma})^{-1/2} \mathbf{h}_1, \dots, (\mu_n - \hat{\sigma})^{-1/2} \mathbf{h}_n]$.

Appendix B:

A Non-linear Information

Maximization ICA Algorithm

– Infomax

The problem of blind separation is illustrated in Figure B.1 [155]. A set of sources, $s_1, \dots, s_n$, are presumed to be mixed approximately linearly so that all we receive is N superpositions of them, $x_1, \dots, x_n$, which are input to our single-layer information maximization network. Providing the mixing matrix, $\mathbf{A}$, is non-singular then the original sources can be recovered if our weight matrix, $\mathbf{W}$, is such that $\mathbf{WA}$ is a ‘permutation’ matrix containing only one high value in each row and column. If the sources are statistically independent and non-gaussian, then the information in the output nodes will be maximized when each output transmits one independent component only. Maximizing the information contained in a layer of units involves maximizing the entropy of the individual units while minimizing the mutual information (*the redundancy*) between them [44, 155]. Considering two units:

$$H(y_1, y_2) = H(y_1) + H(y_2) - I(y_1, y_2). \quad (\text{B.1})$$

For $I(y_1, y_2)$ to be zero, y_1 and y_2 must be statistically independent of each other, so that $f_{y_1 y_2}(y_1, y_2) = f_{y_1}(y_1)f_{y_2}(y_2)$.

Consider a network with an input vector $\mathbf{x}$, a weight matrix $\mathbf{W}$ and a monotonically transformed output vector $\mathbf{y} = g(\mathbf{W}\mathbf{x} + \mathbf{w}_0)$. The multivariate probability density function of $\mathbf{y}$ can be written [31]:

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{f_{\mathbf{x}}(\mathbf{x})}{|J|}, \quad (\text{B.2})$$

where $|J|$ is the absolute value of the Jacobian of the transformation. The Jacobian

Figure B.1: Illustration for the problem of blind source separation.

is the determinant of the matrix of partial derivatives:

$$J = \det \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_n} \end{bmatrix}. \quad (\text{B.3})$$

For sigmoidal units, $\mathbf{y} = (1 + e^{-\mathbf{u}})^{-1}$, $\mathbf{u} = \mathbf{W}\mathbf{x} + \mathbf{w}_0$, the resulting rules by maximizing $\ln|J|$, are

$$\Delta \mathbf{W} \propto [\mathbf{W}^T]^{-1} + \mathbf{x}(\mathbf{1} - 2\mathbf{y})^T; \quad \Delta \mathbf{w}_0 \propto \mathbf{1} - 2\mathbf{y}, \quad (\text{B.4})$$

except that now $\mathbf{x}$, $\mathbf{y}$, $\mathbf{w}_0$ and $\mathbf{1}$ are vectors ($\mathbf{1}$ is a vector of ones), $\mathbf{W}$ is a matrix, and the anti-Hebbian term has become an outer product. The anti-decay term has generalized to the inverse of the transpose of the weight matrix. For an individual weight, w_{ij} , this rule amounts to [155]:

$$\Delta w_{ij} \propto \frac{\text{cof } w_{ij}}{\det \mathbf{W}} + x_j(1 - 2y_i), \quad (\text{B.5})$$

where $\text{cof } w_{ij}$, the cofactor of w_{ij} , is $(-1)^{i+j}$ times the determinant of the matrix of the matrix obtained by removing the i th row and the j th column from $\mathbf{W}$. When the weight vectors entering two output units become too similar, $\det \mathbf{W}$ becomes small and the natural dynamic of learning causes these weight vectors to diverge. This effect is

that there is a degeneracy in the weight matrix of the *rest* of the layer (ie: those weights not associated with input x_j or output y_i). In this case, any degeneracy in $\mathbf{W}$ is less to do with the specific weight w_{ij} that we are adjusting.

Through mutual information, ICA can be tackled directly (in the sense of (B.5)) through a stochastic gradient approach. Sigmoidal units, being bounded, are limited in their ‘channel capacity’. Weights transmitting information try, by following (B.4), to project their inputs to where they can make a lot of difference to the output, as measured by the log of the Jacobian of the transformation. In the process, each set of statistically ‘dependent’ information is channelled to the same output unit [155].

The non-linearity is crucial. If the network were just linear, the weight matrix would grow without bound since the learning rule would be:

$$\Delta \mathbf{W} \propto [\mathbf{W}^T]^{-1}. \quad (\text{B.6})$$

This reflects the fact that the information in the outputs grows with their variance. The non-linearity also supplies the high-order cross-moments necessary to maximize the infinite-order expansion of the information. For example, when $\mathbf{y} = \tanh(\mathbf{u})$, the learning rule has the form $\Delta \mathbf{W} \propto [\mathbf{W}^T]^{-1} - 2\mathbf{y}\mathbf{x}^T$, from which we can write that the weights stabilize (or $\langle \Delta \mathbf{W} \rangle = 0$) when $\mathbf{I} = 2\langle \tanh(\mathbf{u})\mathbf{u}^T \rangle$. Since $\tanh$ is an odd function, its series expansion is of the form $\tanh(u) = \sum_j b_j u^{2j+1}$, the b_j being coefficients, and thus this convergence criterion amounts to the condition $\sum_{i,j} b_{ijp} \langle u_i^{2j+1} u_j \rangle = 0$ for all output unit pairs $i \neq j$, for $p=0,1,2,\dots$, and for the coefficients b_{ijp} coming from the Taylor series expansion of the $\tanh$ function [44, 155].

Appendix C:

A Fast Fixed-Point Algorithm for Independent Component Analysis

– FastICA

To avoid the problems encountered with the preceding objective functions, new one-unit contrast functions for ICA were developed in [45, 156, 157]. Such contrast functions try to combine the positive properties of the preceding contrast functions, i.e. have statistically appealing properties (in contrast to cumulants), require no prior knowledge of the densities of the independent components (in contrast to basic maximum likelihood estimation), allow a simple algorithmic implementation (in contrast to maximum likelihood approach with simultaneous estimation of the densities), and be simple to analyze (in contrast to non-linear cross-correlation and non-linear PCA approaches). The generalized contrast functions (introduced in [156]), which can be considered generalizations of kurtosis, seem to fulfill these requirements.

To begin with, note that one intuitive interpretation of contrast functions is that they are measures of non-normality [158]. A family of such measures of non-normality could be constructed using practically any functions $\mathbf{G}$, and considering the difference of the expectation of $\mathbf{G}$ for the actual data and the expectation of $\mathbf{G}$ for Gaussian data. In other words, we can define a contrast function $\mathbf{J}$ that measures the non-normality of a zero-mean random variable using any even, non-quadratic, sufficiently smooth function $\mathbf{G}$ as follows

$$\mathbf{J}_{\mathbf{G}}(y) = |E_y\{\mathbf{G}(y)\} - E_{\nu}\{\mathbf{G}(\nu)\}|^p, \quad (\text{C.1})$$

where ν is a standardized Gaussian random variable, y is assumed to be normalized to unit variance, and the exponent $p=1,2$ typically. The subscripts denote expectation with respect to y and ν . (The notation $\mathbf{J}_{\mathbf{G}}$ should not be confused with the notation for negentropy, $\mathbf{J}$.)

Adaptive algorithms based on stochastic gradient descent may be problematic when used in an environment where no adaptation is needed. This is the case in many practical situations. The convergence is often slow, and depends crucially on the choice of the learning rate sequence. As a remedy for this problem, one can use batch (block) algorithms based on fixed-point iteration [38, 45, 156]. In [38], a fixed-point algorithm, named FastICA, was introduced using kurtosis, and in [45, 156], the FastICA algorithm was generalized for general contrast functions. For sphered data, the one-unit FastICA algorithm has the following form:

$$\mathbf{w}(k) = E\{\mathbf{x}_g(k-1)^T \mathbf{x}\} - E\{g'(\mathbf{w}(k-1)^T)\}\mathbf{w}(k-1) \quad (\text{C.2})$$

where the weight vector $\mathbf{w}$ is also normalized to unit norm after every iteration, and the function g is the derivative of the function $\mathbf{G}$ used in the general contrast function in (C.1). The expectations are estimated, in practice, using sample averages over a sufficiently large sample of the input data. Units using this FastICA algorithm can then be combined, just as in the case of neural learning rules, into systems that estimate several independent components. Such systems may either estimate the independent component one-by-one using hierarchical decorrelation (deflation), or they may estimate all the independent components in parallel, with symmetric decorrelation [38, 45]. Versions of (C.2) that need no preliminary sphering were also introduced in [156]. The FastICA algorithm is neural in that it is parallel and distributed, but it is not adaptive. Instead of using every data point immediately for learning, FastICA uses sample averages computed over larger samples of the data. The convergence speed of the fixed-point algorithms is clearly superior to those of the more neural algorithms. Speed-up factors in the range from 10 to 100 are usually observed [159].

An interesting point proven in [160] is that when FastICA is used with symmetric decorrelation, it is essentially equivalent to a Newton method for maximum likelihood estimation. This means that FastICA is a general algorithm that can be used to optimize both one-unit and multi-unit contrast functions.

The Fixed-Point Algorithm

1) Estimating one independent component

Assume that we have collected a sample of the sphered (or prewhitened) random vector $\mathbf{x}$, which in the case of blind source separation is a collection of linear mixtures of independent source signals. The *fixed-point algorithm* for ICA is as follows:

1. Take a random initial vector $\mathbf{w}(0)$ of norm 1. let $k = 1$.
2. Let $\mathbf{w}(k) = E\{\mathbf{x}(\mathbf{w}(k-1)^T \mathbf{x})^3\} - 3\mathbf{w}(k-1)$. The expectation can be estimated using a large sample of $\mathbf{x}$ vectors (say, 1,000 points).
3. Divide $\mathbf{w}(k)$ by its norm.
4. If $|\mathbf{w}(k)^T \mathbf{w}(k-1)|$ is not close enough to 1, let $k = k + 1$ and go back to step 2. Otherwise, output the vectors $\mathbf{w}(k)$.

The final vector $\mathbf{w}(k)$ given by the algorithm equals one of the (orthogonal) mixing matrix $\mathbf{B}$. In the case of blind source separation, this means that $\mathbf{w}(k)$ separates one of the non-Gaussian source signals in the sense that $\mathbf{w}(k)^T \mathbf{x}(t)$, $t = 1, 2, \dots$ equals one of the source signals. A remarkable property of fixed-pointed algorithm is that a very small number of iterations, usually 5-10, seems to be enough to obtain the maximal accuracy allowed by the sampled data.

2) Estimating several independent components

To estimate n independent components, we run this algorithm n times. To ensure that we estimate each time a different independent component, we only need to add a simple *orthogonalizing projection* inside the loop. The columns of the mixing matrix $\mathbf{B}$ are orthonormal because of sphering. Thus the independent components can be estimated one by one by projecting the current solution $\mathbf{w}(k)$ on the space orthogonal to the columns of the mixing matrix $\mathbf{B}$ previously found. Define the matrix $\bar{\mathbf{B}}$ as a matrix whose columns are the previously found columns of $\mathbf{B}$. Then the projection operation in the beginning of step 3:

3. Let $\mathbf{w}(k) = \mathbf{w}(k) - \bar{\mathbf{B}}\bar{\mathbf{B}}^T \mathbf{w}(k)$. Divide $\mathbf{w}(k)$ by its norm.

Also the initial random vector should be projected this way before starting the iterations. To prevent estimation errors in $\bar{\mathbf{B}}$ from deteriorating the estimate $\mathbf{w}(k)$, this projection step can be omitted after the first few iterations: once the solution $\mathbf{w}(k)$ has entered the basin of attraction of one of the fixed points, it will stay there and converge to that fixed point.

In addition to the hierarchical (or sequential) orthogonalization described above, any other method of orthogonalizing the weight vectors could also be used. In some applications, a symmetric orthogonalization might be useful. This means that the fixed-point step is first performed for all the n weight vectors, and then the matrix $\mathbf{W}(k) = (\mathbf{w}_1(k), \dots, \mathbf{w}_n(k))$ of the weight vectors is orthogonalized, e.g., using the well-known formula

$$\mathbf{W}(k) = \mathbf{W}(k)(\mathbf{W}(k)^T \mathbf{W}(k))^{-1/2}, \quad (\text{C.3})$$

where $(\mathbf{W}(k)^T \mathbf{W}(k))^{-1/2}$ is obtained from the eigenvalue decomposition of $\mathbf{W}(k)^T \mathbf{W}(k) = \mathbf{E} \mathbf{D} \mathbf{E}^T$ as $(\mathbf{W}(k)^T \mathbf{W}(k))^{-1/2} = \mathbf{E} \mathbf{D}^{-1/2} \mathbf{E}^T$.

3) A semi-adaptive version

A disadvantage of many batch algorithms is that large amounts of data must be stored simultaneously in working memory. The fixed-point algorithm, however, can be used in *semi-adaptive* manner so as to avoid this problem. This can simply be accomplished by computing the expectation $E\{\mathbf{x}(\mathbf{w}(\mathbf{k}-1)^T \mathbf{x})^3\}$ by an on-line algorithm for N sample points has been computed. This semi-adaptive version also makes adaptation to non-stationary data possible. Thus the semi-adaptive algorithm combines many of the advantages usually attributed to either on-line or batch algorithms.

Author's Publications

Book Chapter

1. D. Yu and F. Sattar. Intellectual property protection for multimedia distribution through digital watermarking. Invited book chapter in *Digital Watermarking for Digital Media*, Idea Group Publishing, March 2005.

Journal Paper

1. D. Yu, F. Sattar, and K.-K. Ma. Watermark detection and extraction using independent component analysis method. *EURASIP Journal on Applied Signal Processing - Special Issue on Nonlinear Signal and Image Processing (Part II)*, 2002(1):92-104, Jan. 2002.
2. D. Yu, F. Sattar, and K.-K. Ma. Image watermarking and dewatermarking based on independent component analysis method. *EEE Research Journal*, Nanyang Technological University, Singapore, Jan. 2002.
3. D. Yu and F. Sattar. A new blind image watermarking technique based on independent component analysis. *Springer-Verlag Lecture Notes in Computer Science*, 2613:51-63, May 2003.
4. F. Sattar and D. Yu. On copyright protection through digital watermarking. Accepted in *Encyclopedia of Virtual Communities and Technologies*, Dec. 2004.
5. D. Yu, F. Sattar, and B. Barkat. Multiresolution fragile watermarking using complex chirp signals for content authentication. Submitted to *Pattern Recognition*, Dec. 2004.
6. D. Yu, F. Sattar, and S. Gulam Razul. Transparent robust information hiding for ownership verification. Re-submitted to *IEEE Transactions on Multimedia*, Jan. 2005.

Conference Paper

1. D. Yu, F. Sattar, and K.-K. Ma. Watermark detection and extraction using independent component analysis method. In *Proceedings of the IEEE-EURASIP Nonlinear Signal and Image Processing Workshop 2001 (NSIP'01)*, Baltimore, USA, June 2001.

Author's Publications

2. D. Yu and F. Sattar. Robust image watermarking based on independent component analysis. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing 2002 (ICASSP'02)*, May 2002.
3. D. Yu and F. Sattar. A new blind image watermarking technique based on independent component analysis. In *Proceedings of 1st International Workshop of Digital Watermarking*, Nov. 2002.
4. D. Yu, F. Sattar, and S. Gulam Razul. Transparent robust information hiding for ownership verification. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing 2004 (ICASSP'04)*, May 2004.
5. F. Sattar, D. Yu, S. Gulam Razul, and S. M. Krishnan. Transparent robust information hiding with automatic embedding range selection for ownership verification. In *Proceedings of IEEE International Conference on Image Processing 2004 (ICIP'04)*, Jan. 2004.
6. D. Yu, F. Sattar, and B. Barkat. Multiresolution fragile watermarking using complex chirp signal for content integrity verification. In *Proceedings of IEEE International Conference on Image Processing 2004 (ICIP'04)*, Jan. 2004.

Bibliography

- [1] S. Katzenbeisser and F. A. P. Petitcolas. *Information Hiding Techniques for Steganography and Digital Watermarking*. Boston: Artech House, 2000.
- [2] I. J. Cox. *Digital Watermarking*. San Francisco, Calif.: Morgan Kaufmann; London: International Thomson, 2001.
- [3] B. Schneier. *Applied Cryptography*. John Wiley and Sons Inc., 2nd edition, 1995.
- [4] Neil F. Johnson, Zoran Duric, and Sushil Jajodia. *Information Hiding: Steganography and Watermarking: Attacks and Countermeasures*. Boston, MA: Kluwer Academic, 2001.
- [5] Peter Wayner. *Disappearing Cryptography : Information Hiding : Steganography & Watermarking*. Amsterdam; Boston: MK/Morgan Kaufmann Publishers, 2002.
- [6] M. Arnold, M. Schmucker, and S. D. Wolthusen. *Techniques and Applications of Digital Watermarking and Content Protection*. Boston: Artech House, 2003.
- [7] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn. Information hiding – a survey. In *Proc. of IEEE, Special Issue on Protection of Multimedia Content*, volume 87, pages 1062–1078, July 1999.
- [8] W. Trappe, M. Wu, J. Wang, and K. J. R. Liu. Anti-collusion fingerprinting for multimedia. *IEEE Trans. on Signal Processing*, 51(4):1069–1087, April 2003.
- [9] E. T. Lin and E. J. Delp. A review of fragile image watermarks. In *Proc. of the Multimedia and Security Workshop (ACM Multimedia'99) Multimedia Contents*, pages 25–29, Orlando, FL, Oct. 1999.
- [10] Ö. Ekici, B. Coskun, U. Naci, and B. Sankur. A comparative assessment of semi-fragile watermarking methods. In *SPIE Conf. 4518, Multimedia Systems and Applications IV*, Denver, USA, Aug. 2001.

- [11] J. J. Eggers, R. Bäuml, R. Tzschoppe, and J. Huber. Applications of information hiding and digital watermarking. *ECDL WS Generalized Documents*, 2001.
- [12] A. S. Kayhan. Representation and analysis of complex chirp signals. *Signal Processing*, 66(1998):111–116, 1998.
- [13] F. Hlawatsch and G. F. Boudreaux-Bartels. Linear and quadratic time-frequency signal representations. *IEEE Signal Processing Magazine*, 9(2):21–67, April 1992.
- [14] K. K. Parhi and T. Nishitani. *Digital Signal Processing for Multimedia Systems*. New York: M. Dekker, 1999.
- [15] F. Hartung and M. Kutter. Multimedia watermarking technique. In *Proc. IEEE*, volume 87, pages 1079–1106, July 1999.
- [16] M. Kutter and F. A. P. Petitcolas. A fair benchmark for image watermarking systems. *Electronic Imaging*, 3657:87–90, Jan. 1999.
- [17] F. A. P. Petitcolas and R. J. Anderson. Evaluation of copyright marking systems. In *Proc. of IEEE Multimedia Systems'99*, volume 1, pages 574–579, June 1999.
- [18] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn. Attacks on copyright marking system, second workshop on information hiding. *Lecture Notes in Computer Science*, 1525:218–238, 14-17 April 1999.
- [19] S. Voloshynovskiy, S. Pereira, V. Iquise, and T. Pun. Attack modelling: towards a second generation watermarking benchmark. *Signal Processing*, 81:1177–1214, 2001.
- [20] F. A. P. Petitcolas. In stirmark benchmark 4.0. Available: <http://www.petitcolas.net/fabien/watermarking/stirmark/>, 2003.
- [21] M. Ramkumar and A. Akansu. Information theoretic bounds for data hiding in compressed images. In *Proc. of IEEE Multimedia Signal Processing Workshop*, 1998.
- [22] J. R. Smith and B. O. Comiskey. Modulation and information hiding in images. In *Proc. of 1st Information Hiding Workshop*, 1996.
- [23] S. D. Servetto, C. I. Podilchuk, and K. Ramchandran. Capacity issues in digital image watermarking. In *Proc. of Int. Conf. Image Processing (ICIP'98)*, 1998.
- [24] P. Moulin and J. A. O'Sullivan. Information-theoretic analysis of information hiding. *IEEE Trans. on Information Theory*, 49:563–593, March 2003.

- [25] P. Loo and N. Kingsbury. Digital watermarking with complex wavelets. In *IEE Seminar on Secure Images and Image Authentication (Ref. No. 2000/039)*, pages 10/1 – 10/7, April 2000.
- [26] M. Barni, F. Bartolini, A. De Rosa, and A. Piva. Capacity of the watermark channel: how many bits can be hidden within a digital image? In *Proc. SPIE*, volume 3657, pages 437–448, Jan. 1999.
- [27] C. E. Shannon. The zero-error capacity of a noisy channel. *IRE Trans. on Info. Theory*, IT-2:8–19, 1956.
- [28] A. Hyvärinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999.
- [29] A. Hyvärinen and E. Oja. Independent component analysis: A tutorial. <http://www.cis.hut.fi/projects/ica/>, April 1999.
- [30] T.-W. Lee. *Independent Component Analysis: Theory and Applications*. Kluwer Academic Publishers, 1998.
- [31] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, second edition, 1987.
- [32] J.-F. Cardoso. Blind signal separation: statistical principles. In *Proc. of the IEEE*, pages 2009–2025, 1998.
- [33] T.-W. Lee. Blind source separation of more sources than mixtures using overcomplete representations. *IEEE Signal Processing Letters*, 6(4):87–90, April 1999.
- [34] J. Karhunen, E. Oja, L. Wang, R. Vigário, and J. Joutsensalo. A class of neural networks for independent component analysis. *IEEE Trans. on Neural Networks*, 8(3):486–504, May 1997.
- [35] A. Cichocki and A. K. Barros. Robust batch algorithm for sequential blind extraction of noisy biomedical signals. In *Proc. Int. Symposium on Signal Processing and its Applications (ISSPA'99)*, pages 363–366, 1999.
- [36] A. Cichocki and S. Andrzej. *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons, Ltd, 2002.
- [37] W. Kasprzak and A. Cichocki. Hidden image separation from incomplete image mixtures by independent component analysis. *Proc. Int. Conf. on Pattern Recognition (ICPR'96)*, pages 394–398, 1996.

-
- [38] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9:1483–1492, 1997.
 - [39] Michael S. Lewicki and Terrence J. Sejnowski. Learning overcomplete representation. <http://citeseer.nj.nec.com/lewicki98learning.html>, 1998.
 - [40] T.-W. Lee. Nonlinear approaches to independent component analysis. *Proceedings of the American Institute of Physics*, Oct. 1999.
 - [41] T.-W. Lee. Blind source separation of nonlinear mixing models. *Proceedings of IEEE International Workshop on Neural Networks for Signal Processing*, pages 406–415, Sept. 1997.
 - [42] J.-F. Cardoso. High-order contrasts for independent component analysis. *Neural Comput.*, 11:157–192, 1999.
 - [43] J.-F. Cardoso and A. Soulomiac. Blind beamforming for non gaussian signals. In *IEE-Proceedings-F*, volume 140, pages 362–370, Dec. 1993.
 - [44] A. Bell and T. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.*, 7:1129–1159, 1995.
 - [45] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Networks*, 10(3):626–634, May 1999.
 - [46] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Trans. on Image Processing*, 6(12):1673–1687, Dec. 1997.
 - [47] D. Yu, F. Sattar, and K.-K. Ma. Watermark detection and extraction using independent component analysis method. In *Proc. of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, June 2001.
 - [48] D. Yu, F. Sattar, and K.-K. Ma. Watermark detection and extraction using independent component analysis method. *EURASIP Journal on Applied Signal Processing - Special Issue on Nonlinear Signal and Image Processing (Part II)*, 2002(1), Jan. 2002.
 - [49] R. L. Kashyap and Kie-Bum Eom. Robust image models and their applications. *Advances in Electronic and Electron Physics*, 1988.
 - [50] Andreas S. Weigend and Neil A. Gershenfeld. *Time Series Prediction: Forecasting the Future and Understanding the Past*. Addison-Wesley Publishing Company, 1993.

- [51] Wayne A. Fuller. *Introduction to Statistical Time Series*. A Wiley-Interscience Publication, 1996.
- [52] Anil K. Jain. *Fundamentals of Digital Image processing*. Prentice-Hall, Inc., 1989.
- [53] V. Bhaskaran and K. Konstantinides. *Image and Video Compression Standards: Algorithms and Architectures*. Kluwer Academic Publishers, second edition, 1997.
- [54] J. G. Proakis. *Digital Communications*. McGraw-Hill, 2nd edition, 1989.
- [55] H.-J. M. Wang, P.-C. Su, and C.-C. J. Kuo. Wavelet-based digital image watermarking. *Optics Express*, 3(12):491–496, 1998.
- [56] H. Inoue, A. Miyazaki, A. Yamamoto, and T. Katsura. A digital watermark based on the wavelet transform and its robustness on image compression. *Proc. of IEEE Int. Conf. Image Processing (ICIP'98)*, 1998.
- [57] R. Dugad, K. Ratakonda, and N. Ahuja. A new wavelet-based scheme for watermarking images. *Proc. of IEEE Int. Conf. Image Processing (ICIP'98)*, 1998.
- [58] D. Kundur and D. Hatzinakos. Digital watermarking using multiresolution wavelet decomposition. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'98)*, 5:2969–2972, 1998.
- [59] L. Xie and G. R. Arce. Joint wavelet compression and authentication watermarking. *Proc. of IEEE Int. Conf. Image Processing (ICIP'98)*, 1998.
- [60] S. Voloshynovskiy, A. Herrigel, N. Baumgaertner, and T. Pun. A stochastic approach to content adaptive digital image watermarking. In *Proc. of International Workshop on Information hiding*, Dresden, Germany, Sept. 1999.
- [61] Jae S. Lim. *Two-Dimensional Signal and Image Processing*. Prentice-Hall, Inc., 1990.
- [62] J. F. Delaigle, C. De Vleeschouwer, and B. Macq. Watermarking algorithm based on human visual model. *Signal Processing*, 66:319–335, 1998.
- [63] M. S. Kankanhalli, Rajmohan, and K. R. Ramakrishnan. Content based watermarking of images. In *Proc. of The Sixth ACM International Multimedia Conference*, England, Sept. 1998.
- [64] S. Efstratiadis and A. Katsaggelos. Adaptive iterative image restoration with reduced computational load. *Optical Engineering*, 29(12):1458–1468, Dec. 1990.

-
- [65] P. Peter. Digital image watermarking in the wavelet transform domain. Master's thesis, 2001.
 - [66] F. Mintzer and G. W. Braudaway. If one watermark is good, are more better? In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'99)*, volume 4, pages 2067–2069, March 1999.
 - [67] P. H. W. Wong, O. C. Au, and Y. M. Yeung. A novel blind multiple watermarking technique for images. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(8):813–830, Aug. 2003.
 - [68] N. Nikolaidis and I. Pitas. Robust image watermarking in the spatial domain. *Signal Processing*, 66:385–403, 1998.
 - [69] M. Barni, F. Bartolini, V. Cappellini, and A. Piva. A DCT-domain system for robust image watermarking. *Signal Processing*, 66:357–372, 1998.
 - [70] M. Barni, F. Bartolini, V. Cappellini, A. Piva, and F. Rigacci. A M.A.P. identification criterion for DCT-based watermarking. In *Proc. Europ. Signal Processing Conf. (EUSIPCO'98)*, Sept. 1998.
 - [71] A. Piva, M. Barni, E. Bartoloni, and V. Cappellini. DCT-based watermark recovering without resorting to the uncorrupted original image. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, volume 1, Santa Barbara, CA, 1997.
 - [72] D. Benham, N. Memon, B.-L. Yeo, and M. Yeung. Fast watermarking of DCT-based compressed images. In *Proc. Int. Conf. Image Science, System, and Technology (CISST'97)*, pages 243–253, June 1997.
 - [73] F. Hartung and B. Girod. Digital watermarking of raw and compressed video. In *Proc. SPIE digital compression technologies and systems for video commun.*, volume 2952, pages 205–213, Oct. 1996.
 - [74] E. Koch and J. Zhao. Towards robust and hidden image copyright labelling. In *Proc. Workshop Nonlinear Signal and Image Processing*, Marmaros, Greece, June 1995.
 - [75] C. Langelaar, J. C. A. van der Lubbe, and R. L. Lagendijk. Robust labelling methods for copy protection of images. In *Proc. Electronic Imaging*, volume 3022, pages 298–309, San Jose, CA, Feb. 1997.
 - [76] C. Podilchuk and W. Zeng. Watermarking of the JPEG bitstream. In *Proc. Int. Conf. Imaging Science, Systems, and Applications (CISST'97)*, pages 253–260, Las Vegas, NV, June 1997.

- [77] X. Xia, C. Boncelet, and G. Arce. A multiresolution watermark for digital images. In *Proc. Int. Conf. Image Processing (ICIP'97)*, volume 1, pages 548–551, Santa Barbara, CA, Oct. 1997.
- [78] J. J. K. Ó Ruanaidh, W. J. Dowling, and F. M. Boland. Phase watermarking of digital images. In *Proc. Int. Conf. Image Processing*, volume 3, pages 239–242, Sept. 1996.
- [79] J. J. K. Ó Ruanaidh and T. Pun. Rotation, scale and translation invariant digital watermarking. In *Proc. Int. Conf. Image Processing (ICIP'97)*, volume 1, pages 536–539, Santa Barbara, CA, Oct. 1997.
- [80] J. J. K. Ó Ruanaidh and T. Pun. Rotation, scale and translation invariant digital watermarking. *Signal Processing (Special Issue on Watermarking)*, 66(3):303–318, May 1998.
- [81] H. Liu, X.-W. Kong, X.-D. Kong, and Y. Liu. Content based color image adaptive watermarking scheme. In *Proc. of IEEE International Symposium on Circuits and Systems*, volume 2, pages 41–44, May 2001.
- [82] D. Yu and F. Sattar. A new blind image watermarking technique based on independent component analysis. *Springer-Verlag Lecture Notes in Computer Science*, 2613:51–63, May 2003.
- [83] W. Bender, D. Gruhl, N. Morimoto, and A. Lu. Techniques for data hiding. *IBM Systems Journal*, 35(3&4):313–336, 1996.
- [84] G. Voyatzis and I. Pitas. Digital image watermarking using mixing systems. *IEEE Computer & Graphics*, 22(3), 1998.
- [85] S. A. M. Gilani, I. Kostopoulos, and A. N. Skodras. Color image-adaptive watermarking. In *Proc. of IEEE 14th International Conference on Digital Signal Processing*, volume 2, pages 721–724, July 2002.
- [86] J. J. Chae, D. Mukherjee, and B. S. Manjunath. Color image embedding using multidimensional lattice structures. In *Proc. IEEE Int. Conf. Image Processing (ICIP'98)*, volume 1, pages 460–464, Chicago, Oct. 1998.
- [87] M. Kutter. Digital signature of color images using amplitude modulation. In *SPIE*, volume 3022, pages 518–525, San Jose, CA, Feb. 1997.
- [88] D. J. Fleet and D. J. Heeger. Embedding invisible information in color images. In *Proc. IEEE Int. Conf. Image Processing (ICIP'97)*, volume 1, pages 532–535, Santa Barbara, CA, 1997.

-
- [89] A. Piva, M. Barni, F. Bartolini, and V. Cappellini. Exploiting the cross-correlation of RGB-channels for robust watermarking of color images. In *Proc. IEEE Int. Conf. Image Processing (ICIP'99)*, volume 1, pages 532–535, 1999.
 - [90] J.-B. Zheng, Y.-N. Zhang, D.-G. Feng, and R.-C. Zhao. Color image watermarking based on DCT-domains of color channels. In *Proc. of IEEE TENCON'02*, volume 1, pages 281–284, Oct. 2002.
 - [91] J.-L. Dugelay and F. A. P. Petitcolas. Possible counter-attacks against random geometric distortions. In *Proc. SPIE Security and Watermarking of Multimedia Contents II*, volume 3971, pages 518–525, California, USA, 2000.
 - [92] F. Divoine, P. Bas, P.-A. Hébert, and J.-M. Chassery. Watermarking et résistance aux déformations géométriques. In *Cinquièmes journées d'études et d'échanges sur la compression et la représentation des signaux audiovisuels (CORESA'99)*, France, June 1999.
 - [93] N. F. Johnson, Z. Duric, and S. Jajodia. Recovery of watermarks from distorted images. In *Preliminary Proc. of the Third Int. Information Hiding Workshop*, pages 361–375, Dresden, Germany, 29 Sept.-1 Oct. 1999.
 - [94] M. Kutter. Watermarking resisting to translation, rotation and scaling. In *Proc. of SPIE Int. Symposium on Voice, Video, and Data Communications - Multimedia Systems and Applications*, volume 3528, pages 423–431, Boston, Massachusetts, U.S.A., Nov. 1998.
 - [95] M. Holliman and N. Memon. Counterfeiting attacks and blockwise independent watermarking techniques. *IEEE Transactions on Image Processing*, 9(3):432–441, March 2000.
 - [96] M. Kutter, S. Voloshynovskiy, and A. Herrigel. The watermar copy attack. In *Proc. of SPIE's 12th Annual Symposium, Electronic Imaging 2000: Security and Watermarking of Multimedia Content II*, volume 3971, pages 23–28, January 2000.
 - [97] M. Holliman, N. Memon, and M. Yeung. On the need for image dependent keys in watermarking. In *Proc. of the Second Workshop on Multimedia*, March 1999.
 - [98] V. Saravanan, P. K. Bora, and D. Ghosh. Oblivious image-adaptive watermarking using quantization index modulation. In *The Eighth National Conference on Communications*, Powai, Mumbai, India, Jan. 2002.
 - [99] C.-T. Hsu and J.-L. Wu. Hidden digital watermarks in images. *IEEE Trans. on Image Processing*, 8(1):58–68, Jan. 1999.

- [100] X.-M. Niu, Z.-M. Lu, and S.-H. Sun. Digital watermarking of still images with gray-level digital watermarks. *IEEE Trans. on Consumer Electronics*, 46(1):137–145, 2000.
- [101] C.-S. Tsai and C.-C. Chang. *Distributed Multimedia Databases: Techniques and Applications*, chapter Embedding robust graylevel watermark in an image using discrete cosine transformation. Timothy Shih (Ed.). Idea Group Publishing, Jan. 2002.
- [102] Y. J. Jung, M. Hahn, and Y. M. Ro. Spatial frequency band division in human visual system based-watermarking. *Springer-Verlag Lecture Notes in Computer Science*, 2613:224 – 234, May 2003.
- [103] J. G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(11), 1993.
- [104] S. Craver, N. Memon, B. Yeo, and M. Yeung. Resolving rightful ownership with invisible watermarking techniques: limitations, attacks, and implications. *IEEE Journal on Selected Areas in Communications*, 16(4):573–586, May 1998.
- [105] Q. Li and E.-C. Chang. On the possibility of non-invertible watermarking schemes. In *Proc. of Information Hiding Workshop*, 2004.
- [106] Y. Wang, J. F. Doherty, and R. E. V. Dyck. A wavelet-based watermarking algorithm for ownership verification of digital images. *IEEE Trans. on Image Processing*, 11(2):77–88, Feb. 2002.
- [107] H. Guo and N. D. Georganas. Digital image watermarking for joint ownership verification without a trusted dealer. In *Proc. IEEE Int. Conf. on Multimedia and Expo (ICME'03)*, volume 2, pages 497–500, July 2003.
- [108] M. Wu. *Multimedia Data Hiding*. Phd thesis, Princeton University, June 2001.
- [109] N. Memon and P. Vora. Authentication techniques for multimedia content. In *Proc. of Multimedia Systems and Applications*, Oct. 1998.
- [110] M. P. Queluz. Authentication of digital images and video: Generic models and a new contribution. *Signal Processing: Image Communication*, 16(2001):461–475, 2001.
- [111] C.-Y. Lin and S.-F. Chang. A robust image authentication method distinguishing JPEG compression from malicious manipulation. *IEEE Trans. on Circuits and Systems of Video Technology*, 11(2):153–168, Feb. 2001.

- [112] C.-Y. Lin and S.-F. Chang. A robust image authentication method surviving JPEG lossy compression. In *Proc. of IS&T/SPIE Symposium on Electronic Imaging*, San Jose, USA, Jan. 1998.
- [113] M. Schneider and S.-F. Chang. A robust content-based digital signature for image authentication. In *Proc. IEEE Int. Conf. Image Processing (ICIP'96)*, pages 227–230, Sept. 1996.
- [114] S. Walton. Information authentication for a slippery new age. *Dr. Dobbs Journal*, 20(4):18–26, April 1995.
- [115] R. van Schyndel, A. Trikel, and C. Osborne. A digital watermark. In *Proc. IEEE Int. Conf. Image Processing (ICIP'94)*, volume 2, pages 86–90, Austin, Texas, Nov. 1994.
- [116] R. Wolfgang and E. Delp. A watermark for digital images. In *Proc. IEEE Int. Conf. Image Processing (ICIP'96)*, volume 3, pages 219–222, 1996.
- [117] P. Wong. A watermark for image integrity and ownership verification. In *Final Program and Proc. of the IS&T PICS 99*, pages 374–379, Savanna, Georgia, April 1999.
- [118] M. Yeung and F. Mintzer. Invisible watermarking for image authentication. *Journal of Electronic Imaging*, 7(3):578–591, July 1998.
- [119] N. Memon, S. Shende, and P. Wong. On the security of the yeung-mintzer authentication watermark. In *Final Program and Proc. of the IS&T PICS 99*, pages 301–306, Savanna, Georgia, April 1999.
- [120] M. Wu and B. Liu. Watermarking for image authentication. In *Proc. IEEE Int. Conf. Image Processing (ICIP'98)*, volume 2, pages 437–441, Chicago, Illinois, Oct. 1998.
- [121] L. Xie and G. Arce. Joint wavelet compression and authentication watermarking. In *Proc. IEEE Int. Conf. Image Processing (ICIP'98)*, volume 2, pages 427–431, Chicago, Illinois, Oct. 1998.
- [122] D. Kundur and D. Hatzinakos. Towards a telltale watermarking technique for tamper-proofing. In *Proc. IEEE Int. Conf. Image Processing (ICIP'98)*, volume 2, pages 409–413, Chicago, Illinois, Oct. 1998.
- [123] D. Kundur and D. Hatzinakos. Digital watermarking for telltale tamper proofing and authentication. In *Proc. of IEEE*, volume 87, pages 1167–1180, July 1999.

Bibliography

- [124] J. Fridrich. Methods for tamper detection in digital images. In *Proc. ACM Workshop on Multimedia and Security*, pages 19–23, Orlando, FL, Oct. 1999.
- [125] L. M. Marvel, Jr. G. W. Hartwig, and Jr. C. Boncelet. Compression compatible fragile and semi-fragile tamper detection. In *Proc. of the SPIE International Conference on Security and Watermarking of Multimedia Contents II*, volume 3971, San Jose, CA, Jan. 2000.
- [126] L. M. Marvel, Jr. C. Boncelet, and C. T. Retter. Spread spectrum image steganography. *IEEE Trans. on Image processing*, 8(8):1075–1083, Aug. 1999.
- [127] J. J. Eggers, J. K. Su, and B. Girod. A blind watermarking scheme based on structured codebooks. In *IEE Conf. Secure Images and Image Authentication*, London, April 2000.
- [128] J. Fridrich. Robust digital watermarking based on key-dependent basis functions. In *The 2nd Information Hiding Workshop*, Oregon, Portland, April 1998.
- [129] C. Y. Lin and S. F. Chang. Semi-fragile watermarking for authentication JPEG visual content. In *Proc. of the SPIE International Conference on Security and Watermarking of Multimedia Contents II*, volume 3971, San Jose, CA, Jan. 2000.
- [130] E. T. Lin, C. I. Podilchuk, and E. J. Delp. Detection of image alterations using semi-fragile watermarks. In *Proc. of the SPIE International Conference on Security and Watermarking of Multimedia Contents II*, volume 3971, San Jose, CA, Jan. 2000.
- [131] M. P. Queluz. Content-based integrity protection of digital images. In *Proc. of the SPIE International Conference on Security and Watermarking of Multimedia Contents*, volume 3657, pages 85–93, 1999.
- [132] J. Fridrich and M. Goljan. Protection of digital images using self embedding. In *Symposium on Content Security and Data Hiding in Digital Media*, New Jersey Institute of Technology, May 1999.
- [133] T.-H. Lan and A. H. Tewfik. Fraud detection and self embedding. In *ACM Multimedia (2)*, pages 33–36, 1999.
- [134] D. L. Robie and R. M. Mersereau. Video error correction using steganography. *Applied Signal Processing*, 2002(2):164–173, Feb. 2002.
- [135] John G. Proakis and Dimitris G. Manolakis. *Introduction to Digital Signal Processing*. New York: Macmillan; London: Collier Macmillan, 1988.

- [136] Alfred Martins. *Signal Analysis: Wavelets, Filter Banks, Time-Frequency Transforms and Applications*. New York: John Wiley & Sons, 1999.
- [137] Leon Cohen. *Time-Frequency Analysis*. Englewood Cliffs, N.J: Prentice Hall PTR, 1995.
- [138] S. Krishnan, R. M. Rangayyan, F. D. Bell, and C. B. Frank. Time-frequency signal feature extraction and screening of keen joint vibroarthrographic signals using the matching pursuit method. In *Proc. of IEEE EMBS*, pages 1309–1312, Chicago, IL. USA, 1997.
- [139] F. Sattar and B. Barkat. A new time-frequency based fragile watermarking scheme for image authentication. In *Proc. IEEE Pacific Rim Conf. on Communications, Computers and Signal Processing (PACRIM'03)*, volume 2, Aug. 2003.
- [140] B. Barkat and F. Sattar. A new time-frequency based private fragile watermarking scheme for image authentication. In *Proc. Int. Symposium on Signal Processing and its Applications (ISSPA'03)*, Paris, France, July 2003.
- [141] M. S. Roden. *Analog and Digital Communication Systems*. Englewood Cliffs, N.J.: Prentice-Hall, 1996.
- [142] R. E. Ziemer and R. L. Peterson. *Introduction to Digital Communication*. Upper Saddle River, New Jersey: Prentice Hall, 2001.
- [143] Mohsen Ashourian and Yo-Sung Ho. Analysis of quantization watermarking in the wavelet transform. In *Proc. of 7th Int. Signal Processing and Its Applications*, July 2003.
- [144] Ö. Ekici, B. Sankur, and M. Akcay. Comparative assessment of semifragile watermarking methods. *Journal of Electronic Imaging*, 2002.
- [145] P. Campisi, M. Carli, G. Giunta, and A. Neri. Blind quality assessment system for multimedia communications using tracing watermarking. *IEEE Trans. on Signal Processing*, 51(4):996–1002, April 2003.
- [146] C.-Y. Lin and S.-F. Chang. SARI: Self-authentication-and-recovery image watermarking system. In *ACM Multimedia*, Sep. 2001.
- [147] M. Celik, G. Sharma, A. M. Tekalp, and E. Saber. Video authentication with self-recovery. In *Proc. of SPIE Security and Watermarking of Multimedia Contents IV*, volume 4675, Jan 2002.

-
- [148] B. Mobasseri and A. Evans. Content-dependent video authentication by self-watermarking in color space. In *Proc. of SPIE Security and Watermarking of Multimedia Contents III*, Jan. 2001.
 - [149] C.-Y. Lin, M. Wu, J. A. Bloom, I. J. Cox, M. L. Miller, and Y. M. Lui. Rotation, scale, and translation resilient watermarking for images. *IEEE Transactions on Image Processing*, 10(5):767–782, May 2001.
 - [150] X. Kang, J. Huang, and Yun Q. Shi. An image watermarking algorithm robust to geometric distortions. *Springer-Verlag Lecture Notes in Computer Science*, 2613:212–223, May 2003.
 - [151] F. Hartung and F. Ramme. Digital rights management and watermarking of multimedia content for m-commerce applications. *IEEE Communications Magazine*, 38(11):78–84, Nov. 2000.
 - [152] A. O. Waller, G. Jones, T. Whitley, J. Edwards, D. Kaleshi, A. Munro, B. MacFarlane, and A. Wood. Securing the delivery of digital content over the internet. *Electronics & Communication Engineering Journal*, 14(5):239–248, Oct. 2002.
 - [153] M. Lesk. The good, the bad, and the ugly: what might change if we had good drm. *IEEE Security & Privacy Magazine*, 1(3):63–66, May-June 2003.
 - [154] J.-F. Cardoso and P. Comon. Tensor based independent component analysis. In *Proc. Europ. Signal Processing Conf. (EUSIPCO'90)*, pages 673–676, 1990.
 - [155] A. Bell and T. Sejnowski. A non-linear information maximization algorithm that performs blind separation. In G. Tesauro, D. Touretzky, and T. Leen, editors, *Advances in Neural Information Processing Systems*, volume 7, pages 467–474. The MIT Press, 1995.
 - [156] A. Hyvärinen. A family of fixed-point algorithms for independent component analysis. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'97)*, pages 3917–3920, Munich, Germany, 1997.
 - [157] A. Hyvärinen. New approximations of differential entropy for independent component analysis and projection pursuit. *Advances in Neural Information Processing Systems*, pages 273–279, 1998.
 - [158] P. Comon. Independent component analysis - a new concept? *Signal Processing*, 36:287–314, 1994.
 - [159] X. Giannakopoulos, J. Karhunen, and E. Oja. Experimental comparison of neural ICA algorithms. In *Proc. Int. Conf. on Artificial Neural Networks (ICANN'98)*, pages 651–656, Skövde, Sweden, 1998.

Bibliography

- [160] A. Hyvärinen. The fixed-point algorithm and maximum likelihood estimation for independent component analysis. *Neural Processing Letters*, 10(1):1–5, 1999.