

RESILIENCE OF DIGITAL WATERMARKS UNDER TRANSFORMATION ATTACKS

Srilal Buddika T.M.

Registration No. : 138204V



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk
Degree of Master of Science

Department of Computer Science & Engineering

University of Moratuwa
Sri Lanka

April 2015

RESILIENCE OF DIGITAL WATERMARKS UNDER TRANSFORMATION ATTACKS

Srilal Buddika T.M.

Registration No. : 138204V



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Dissertation submitted in partial fulfillment of the requirements for the
degree Master of Science in Computer Science

Department of Computer Science & Engineering

University of Moratuwa
Sri Lanka

April 2015

Declaration

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Candidate

.....
Date *T.M.S. Buddika*

The above candidate has carried out research for the Masters Dissertation under my supervision.

Supervisor

.....
Date *Dr. Chandana Gamage*

Acknowledgments

I offer my sincerest gratitude to my research supervisor Dr. Chandana Gamage, who gave endless support and encouragement to me during the research work and thesis preparation. His supervision has always inspired me to improve my writing skills and to produce quality work on this thesis.

I would like to extend my gratitude to my wife and family for providing assistance, support, encouragement and inspiration during the writing of this dissertation. I greatly appreciate their support and commitment in making this thesis a success.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Abstract

Digital watermarking of electronic data has become a popular research area over last several years. The research work presented in this thesis blends a more practical perspective of digital watermarking schemes against transformation/geometric attacks with theoretical studies presented in past literature. Main objective of this dissertation is to come up with an attack success/failure evaluation metric against chosen watermarking schemes and with set of attacks. “Success” or “Failure” is defined as per given criteria: If there is a legitimate owner for the image, *success* of image attack would be determined if recovery of the watermark is denied given the fact that attacked image lies within the perceptual range. Otherwise it would be depicted as a *failure* of image attack.

In this thesis, limits that preserve perceptual quality of subjected images were analyzed for different types of attacks. Both PSNR and Histogram Analysis provide a foundation for meaningful measurement. We have shown that coupling such measurements with a perceptual quality measurement provide meaningful values for effectiveness of a watermarking scheme.

Table of Contents

Declaration	i
Acknowledgments	ii
Abstract	iii
List of Figures	vi
List of Tables	x
List of Equations	xi
1 Introduction	1
1.1 Origin of Information Hiding	2
1.2 Watermarking, Steganography and Cryptography	3
1.3 Digital Image Watermarking Applications	3
1.4 Digital Image Watermarking Techniques	4
1.5 Attacks on Digital Image Watermarks	6
1.6 Research Objective	13
2 Literature Review	14
2.1 Generic Watermarking System	15
2.1.1 Steps of Image Watermark Encoding	16
2.1.2 Steps of Image Watermark Decoding	16
2.2 Spatial Domain based Image Watermarking	17
2.2.1 LSB Watermarking Schemes	17
2.2.2 Patchwork based Watermarking Schemes	19
2.2.3 Correlation based Watermarking Schemes	21
2.2.4 Other Spatial Domain Watermarking Schemes	23
2.3 Transform Domain based Image Watermarking	24
2.3.1 Discrete Fourier Transform (DFT) based Schemes	24
2.3.2 Discrete Cosine Transform (DCT) based Schemes	25
2.3.3 Discrete Wavelet Transform (DWT) based Schemes	29
2.3.4 Hybrid Watermarking Schemes	34
2.4 Summary of Transform Domain	37
3 Research Methodology	39
3.1 Selection of Watermarking Schemes	39

TABLE OF CONTENTS

3.2	Selection of Transformation and Geometric Attacks	40
3.3	Selection of Images	48
3.4	Development of an Attack Success/Failure Evaluation Metric	49
4	Testing and Evaluation	50
4.1	Post Attack Evaluation on Watermarking Schemes	50
4.2	Summary of Test and Evaluation	59
5	Analysis and Conclusion	60
5.1	Distortion Analysis	60
5.2	Conclusion	69
	References	72
	Appendix	78



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

List of Figures

1.1	History of Steganography 5 th Century	2
1.2	Digital Image Watermark Classification	5
1.3	Generic Watermark Embedding Process	5
1.4	Watermark Recovery/Detection process	6
1.5	Two-Dimensional True Image Compression	9
1.6	Image Rotation Attack	10
1.7	Image Cropping Attack	10
1.8	Image Resize Attack	11
1.9	Image Enhancement Techniques: Filters and Histogram Equalization . .	12
1.10	Image Noise Attack	13
2.1	Digital Watermarking Hierarchy	15
2.2	Image Watermark Encoding	16
2.3	Image Watermark Decoding	17
2.4	Image Comparison-LSB Watermarking	18
2.5	Recovered Message Image	19
2.6	Image Comparison-LSB Watermarking	20
2.7	Histogram Comparison-LSB Watermarking	20
2.8	Watermark Encoding-Correlation Scheme	21
2.9	Image converted into DCT domain	26
2.10	Watermark Insert and Extract in Transform Domain	26
2.11	Original Image and Histogram-DCT	27
2.12	DCT Domain Watermarked Image and Histogram	27
2.14	Two-Dimensional Discrete Wavelet Transform	29
2.15	DWT watermark embedding block diagram	31
2.16	DWT watermark decoding block diagram	31
2.17	Original Image and Histogram-DWT	32
2.18	DWT Domain Watermarked Image and Histogram	32
2.19	Image Differences w.r.t Original Image-DWT	32
2.20	Resulting DWT Decomposed Image: Peppers	33
2.21	DWT block tree organization	33
2.22	Block-based hybrid method	35
2.23	Hybrid watermarked still image: Lena	37

LIST OF FIGURES

3.1	Block Diagram of Visible Differences Predictor	41
3.2	Cropped Lena Image	42
3.3	Cropped Peppers Image	42
3.4	Cropped OldLady Image	42
3.5	Image Flip Attack on Peppers Image	43
3.6	Image Resize Attack on Lena Image:ScaleUp	44
3.7	Image Resize Attack on OldLady:ScaleDown	44
3.8	Image JPEG Compression StirMark Attack on OldLady	45
3.9	Image Noise StirMark Attack on OldLady	46
3.10	Removal Attack on Peppers	46
3.11	Lowpass Filter Attack on OldLady	47
3.12	Lena	48
3.13	Peppers	48
3.14	OldLady	48
4.1	Watermark Detection after Crop Attack	51
4.2	Rotation:COX Scheme (Method:Bicubic)	52
4.3	Rotation:COX Scheme (Method:Nearest)	52
4.4	Watermark Detection after Flip Attack	53
4.5	Resize:WANG Scheme (Scale Down)	54
4.6	Resize:WANG Scheme (Scale Up)	54
4.7	JPEG Compression:KUCH Scheme	55
4.8	JPEG Compression:KUNDUR Scheme	56
4.9	Noise Attack:COX Scheme	56
4.10	Noise Attack:KUNDUR Scheme	57
4.11	Removal Attack:WANG Scheme	57
4.12	Removal Attack:KUNDUR Scheme	58
4.13	Filter Attack:XIA Scheme	59
4.14	Filter Attack:KUNDUR Scheme	59
5.1	Histogram Comparison of Lena (Flip Attack)	61
5.2	Histogram Comparison of Peppers (Flip Attack)	61
5.3	Histogram Comparison of OldLady (Flip Attack)	61
5.4	Histogram Comparison of Lena:Kundur (Filter Attack)	62
5.5	Histogram Comparison of Peppers:Kundur (Filter Attack)	62
5.6	Histogram Comparison of OldLady:Kundur (Filter Attack)	62
5.7	Histogram Comparison of Lena:Xia (Filter Attack)	63
5.8	Histogram Comparison of Peppers:Xia (Filter Attack)	63
5.9	Histogram Comparison of OldLady:Xia (Filter Attack)	63
5.10	Histogram Comparison of Lena:Kundur (StirMark JPEG Attack)	64
5.11	Histogram Comparison of Peppers:Kundur (StirMark JPEG Attack)	64

LIST OF FIGURES

5.12	Histogram Comparison of OldLady:Kundur (StirMark JPEG Attack) . .	64
5.13	Histogram Comparison of Lena:Kundur (StirMark Noise Attack)	65
5.14	Histogram Comparison of Peppers:Kundur (StirMark Noise Attack) . .	65
5.15	Histogram Comparison of OldLady:Kundur (StirMark Noise Attack) . .	65
5.16	Histogram Comparison of OldLady:Kundur (StirMark Removal Attack)	66
5.17	Histogram Comparison of Peppers:Kundur (Scale Down Attack)	66
5.18	Histogram Comparison of Lena:Kundur (Scale Up Attack)	66
5.19	Histogram Comparison of OldLady:Kundur (Rotate Attack)	67
5.20	PSNR Comparison:Kundur (Rotate Attack)	68
5.21	PSNR Comparison (Flip Attack)	68
5.22	Rotation:COX Scheme(Method:Bicubic)	78
5.23	Rotation:COX Scheme(Method:Nearest)	79
5.24	Rotation:WANG Scheme(Method:Bicubic)	79
5.25	Rotation:WANG Scheme(Method:Nearest)	79
5.26	Rotation:KUNDUR Scheme(Method:Bicubic)	80
5.27	Rotation:KUNDUR Scheme(Method:Nearest)	80
5.28	Rotation:XIA Scheme(Method:Bicubic)	80
5.29	Rotation:XIA Scheme(Method:Nearest)	81
5.30	Rotation:KOCH Scheme(Method:Bicubic)	81
5.31	Rotation:KOCH Scheme(Method:Nearest)	81
5.32	Resize:COX Scheme(Scale Down)	82
5.33	Resize:COX Scheme(Scale Up)	82
5.34	Resize:WANG Scheme(Scale Down)	83
5.35	Resize:WANG Scheme(Scale Up)	83
5.36	Resize:KUNDUR Scheme(Scale Down)	83
5.37	Resize:KUNDUR Scheme(Scale Up)	84
5.38	Resize:XIA Scheme(Scale Up)	84
5.39	Resize:XIA Scheme(Scale Up)	84
5.40	Resize:KOCH Scheme(Scale Down)	85
5.41	Resize:KOCH Scheme(Scale Up)	85
5.42	JPEG Compression:COX Scheme	86
5.43	JPEG Compression:WANG Scheme	86
5.44	JPEG Compression:KUNDUR Scheme	87
5.45	JPEG Compression:XIA Scheme	87
5.46	JPEG Compression:KOCH Scheme	87
5.47	Noise Attack:COX Scheme	88
5.48	Noise Attack:KUNDUR Scheme	88
5.49	Noise Attack:WANG Scheme	89
5.50	Noise Attack:XIA Scheme	89
5.51	Noise Attack:KOCH Scheme	89

LIST OF FIGURES

5.52 Removal Attack:COX Scheme	90
5.53 Removal Attack:WANG Scheme	90
5.54 Removal Attack:KUNDUR Scheme	91
5.55 Removal Attack:XIA Scheme	91
5.56 Removal Attack:KOCH Scheme	91
5.57 Filter Attack:COX Scheme	92
5.58 Filter Attack:WANG Scheme	92
5.59 Filter Attack:KUNDUR Scheme	93
5.60 Filter Attack:XIA Scheme	93
5.61 Filter Attack:KOCH Scheme	93
5.62 Histogram Comparison of Lena:Kundur(StirMark Removal Attack) . . .	94
5.63 Histogram Comparison of Peppers:Kundur(StirMark Removal Attack) .	94
5.64 Histogram Comparison of OldLady:Kundur(StirMark Removal Attack) .	94
5.65 Histogram Comparison of Lena:Kundur(Scale Down Attack)	95
5.66 Histogram Comparison of Peppers:Kundur(Scale Down Attack)	95
5.67 Histogram Comparison of OldLady:Kundur(Scale Down Attack)	95
5.68 Histogram Comparison of Lena:Kundur(Scale Up Attack)	96
5.69 Histogram Comparison of Peppers:Kundur(Scale Up Attack)	96
5.70 Histogram Comparison of OldLady:Kundur(Scale Up Attack)	96
5.71 Histogram Comparison of Lena:Kundur(Rotate Attack)	97
5.72 Histogram Comparison of Peppers:Kundur(Rotate Attack)	97
5.73 Histogram Comparison of OldLady:Kundur(Rotate Attack)	97
5.74 PSNR Comparison(Flip Attack)	98
5.75 PSNR Comparison:Kundur(JPEG Attack)	98
5.76 PSNR Comparison:Kundur(Filter Attack)	99
5.77 PSNR Comparison:Xia(Filter Attack)	99
5.78 PSNR Comparison:Kundur(Noise Attack)	100
5.79 PSNR Comparison:Kundur(Removal Attack)	100
5.80 PSNR Comparison:Kundur(Scale Down Attack)	101
5.81 PSNR Comparison:Kundur(Scale Up Attack)	101
5.82 PSNR Comparison:Kundur(Rotate Attack)	102

List of Tables

3.1	Selected Transformation and Geometric Attacks	41
3.2	Rotation Angles for Rotate Attack	43
3.3	Scale Factors for Resize Attack	43
3.4	Quality Levels of JPEG Compression Attack	45
3.5	Selected Images for Evaluation	48
5.1	Attack Success/Failure Summary	69



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

List of Equations

2.1	Watermarking Encoding Process	16
2.2	Watermarking Decoding Process with Original Image	16
2.3	Watermarking Decoding Process without Original Image	17
2.4	Watermark Encoding - Correlation Scheme	21
2.5	Edge-enhancing finite impulse response (FIR) filter	22
2.6	Spatial Domain Watermark Detection - Extended Method	23
2.7	DFT-1D Transformation	24
2.8	DFT-2D Transformation	24
2.9	DFT-2D Transformation-Magnitude and Phase	25
2.10	Discrete Cosine Transformation	28
2.11	Inverse of Discrete Cosine Transformation	28
2.12	DWT Embedding Procedure	30
2.13	DWT Extraction Procedure	30
2.14	Signature Function	30
2.15	DWT signal decomposition Level-1	34
2.16	DWT Signal Reconstruction Level-1	34
2.17	DWT and IDWT Orthogonality	34
2.18	Equation for modified DCT coefficients	38
2.19	Calculate Correlation Factor	38
2.20	PSNR Calculation	38
2.21	Mean Square Error Calculation	38

Chapter 1

Introduction

Modern information technology development has taken the world towards a digital era. It has become a common practice to copy, store, transmit and distribute content as digital data over various communication channels. This wide distribution has created significant challenges to creators and owners of the content to enforce their rights as owners of digitized content. One way of applying a digital image watermark is for authentication where it identifies ownership of copyright information. In this context digital watermarks are seen as a possible solution. Hence, it provides copyright protection by hiding valid ownership information in digital images. The ownership information may be in the form of audio, pictures, texts or 3D models. Since a digital copy of data is same as original, digital watermarking is a passive protection tool. It just marks data, but does not degrade it nor controls access to data.

Just as digital watermarks have been used to ensure ownership rights to content, many different attacks have been developed to counter and invalidate properties of digital watermarks. More importantly recent studies focus on particular transformation attacks and subjected watermark scheme behaviors. Voloshynovskiy et al. [1], Kutter and Martin [2] and Collberg et al. [3] describe comprehensive attacks on digital watermarks: classification, based attacks, and benchmarks in this area of study as a whole. Usually digital water marking's first priority is to control robustness. If a digital watermark distorts due to some image transformation attack, it is no longer usable.

Nearly a decade from 1990's to 1999, there were number of researches on "*robust digital watermarking schemes*" proposed but none of them uses same criteria of robustness. The disadvantage of it is, the watermarking scheme resilient and its robustness cannot be compared. This issue has addressed in multiple survey researches by Potdar et al [4], Cox et al [5] and Jeng-Shyang et al [6]. Specially, Kutter and Petitcolas [7] have completed a comprehensive research on benchmarking image watermarking systems. Digital watermarking is an essential technology as advances are made in the

field of communication in 20th century and it became necessary to cipher and decipher information/data. Significant number of applications for digital watermarking includes copyright protection, image authentication, copy control, fingerprinting, tamper detection and data hiding applications (E.g. Broadcast Monitoring).

1.1 Origin of Information Hiding

Data or Information hiding is also known as *Steganography*. The history of steganography goes back nearly to the 5th century. The earliest known uses of steganography can be traced back to 440 BC when Herodotus mentions two examples in his Histories. First one, Demaratus sent a warning about a forthcoming attack to Greece by writing it directly on the wooden backing of a wax tablet shown in Figure 1.1 before applying its beeswax surface. Wax tablets were in common use then as reusable writing surfaces. Second one, is “Hidden messages on messenger’s body” which was used in ancient Greece depicted in Figure 1.1. Herodotus tells the story of a message tattooed on the shaved head of a slave of Histiaeus. Once the slaves’ hair was long enough to cover the message he was sent to Aristagoras in the city of Miletus. The message allegedly carried a warning to Greece about Persian invasion plans.



Figure 1.1: History of Steganography 5th Century [8], [9]

There are number of ways in which steganography can be used in secret communication. One way is, the use of invisible inks which helps a person could draft a secret message and send it to another person without anyone else knowing what’s written and later reveal the message by applying lime juice or other technique as there are various forms of invisible ink were used to conceal messages.

Generally, steganography comes with techniques of embedding data inside a medium usually called “*cover medium*” or “*host*”, in a way that hidden data is not visible

(Imperceptibility) to human eyes. Cover medium could be any form of digital data like audio, video, images and text documents etc. Information hiding theory defines steganography as the art and science of writing hidden messages in the way that no one apart from sender and intended recipient, suspects the existence of the message. In other words it is a form of security achieved through obscurity.

1.2 Watermarking, Steganography and Cryptography

In Steganography, hidden data has no relationship with cover medium and seems innocent. In Watermarking, data which is hidden has a relationship with cover medium because it carries cover medium or host ownership information which provides copyright data for digital content. Unlike steganography, watermarking systems are known to be robust against malicious attacks. Compared to watermarking and steganography, cryptography is focused on protection of message content. The protection is provided in a way that message content is scrambled also known as cipher text and cannot be understood by anyone unless the person is authorized. This does not happen in digital watermarking or steganography.

1.3 Digital Image Watermarking Applications



Electronic Theses & Dissertations

www.lib.mru.ac.in

The book of Information Hiding [10] and Cox et al [5] describe an overview of watermarking applications and their properties. In digital watermarking there is no universal watermarking method is used. Hence different level of robustness can be seen depending on the watermarking application needs.

Copyright Protection

Copyright for digital artifacts is being considered in this topic under watermarking application. Embedding content represents ownership information of digital data. Organizations and individuals are interested in securing their works by copyright protection nowadays. It happens mainly because, embedded copyright ownership information prevent other parties from claiming copyright on same data of individual or organization owned. These types of watermarks help to identify rightful origin of the source. Fundamental property should have in this type of watermarking application is “*Robustness*”.

Fingerprinting

This technique is used to back track illegal copy distributors for particular digital data content such as music CDs, video CDs, images, commercial software version etc. It is similar to software license registration process. The concept behind fingerprinting technique is that, it uses different types of watermarks but unique to embed on each

distributed copy. It helps to trace back and track illegal users who consume copy of data. Why not single watermark is used? Because, there have been few research studies [1], [11]–[13] published, single watermarked copies allow “*collusion attacks*” and those watermarks should be unique to be collusion-secure.

Copy Protection

This technique is widely used in CD/DVD producing companies and their intention is to prevent illegal copying of sold CD/DVDs. The objective of this technique is to embed watermark on original content and restrict further copies (Need to use a special hardware DVD/CD players or software) or although somebody gets a copy of it, the copy will not get the watermark data. Hence, a duplicate can be easily found.

Image Authentication

Authentication watermark has the lowest level of robustness to detect slightest modification of data. In watermarking theory, it is called “*fragile watermarks*”. Its primary objective is to determine modification of watermarked data.

Tamper Detection

This scheme is mainly used for securing images in photographic forensics applications and devices. Main objective is, if an attacker tries to tamper or alter original data, watermark will get destroyed and indicates that it has been tampered as well. Such mechanism proposed by Friedman [14] in 1993 for digital camera.



University of Moratuwa, Sri Lanka
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

1.4 Digital Image Watermarking Techniques

Digital image watermarks can be classified into two categories: visible and invisible watermarks. In Information hiding theory and practice, perceptible and imperceptible words are used for these two categories as illustrated in Figure 1.2.

Imperceptible watermarking techniques can be fragile, semi-fragile or robust. These attributes define robustness of a watermarking scheme.

- **Fragile:** Modification to original work that clearly is detectable and commonly used for tamper detection (Integrity Proof).
- **Semi-fragile:** This kind of watermarked electronic data, resists benign transformation attacks, but fails detection after malignant transformations. Hence, it is primarily used to detect malignant transformations.
- **Robust:** This type of watermark resists a designated class of transformation attacks. Sometime they are used in copyright protection applications.

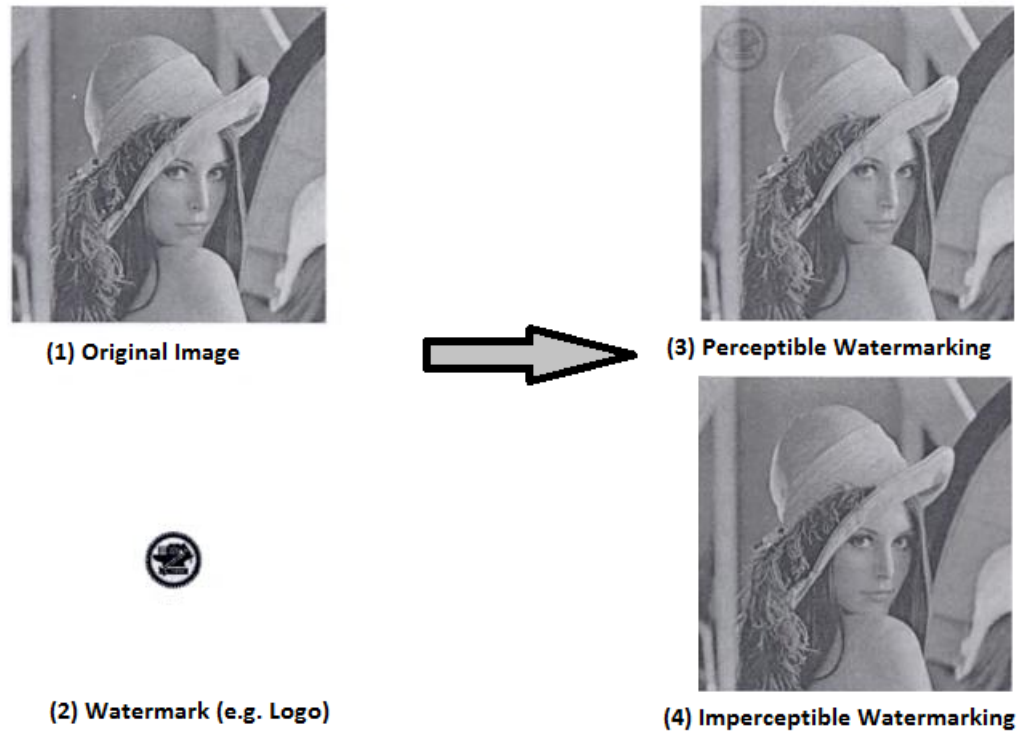


Figure 1.2: Digital Image Watermark Classification

Watermarking scheme classification falls into three categories. Each scheme has its own properties and qualities.

Private Watermarking. This kind of systems are known to be robust and they have two types of watermarking detection methods as in Figure 1.4, which is *Original-Image*, and/or *Watermark* and *Secret key*. Furthermore, watermark embedding process depicted in Figure 1.3.

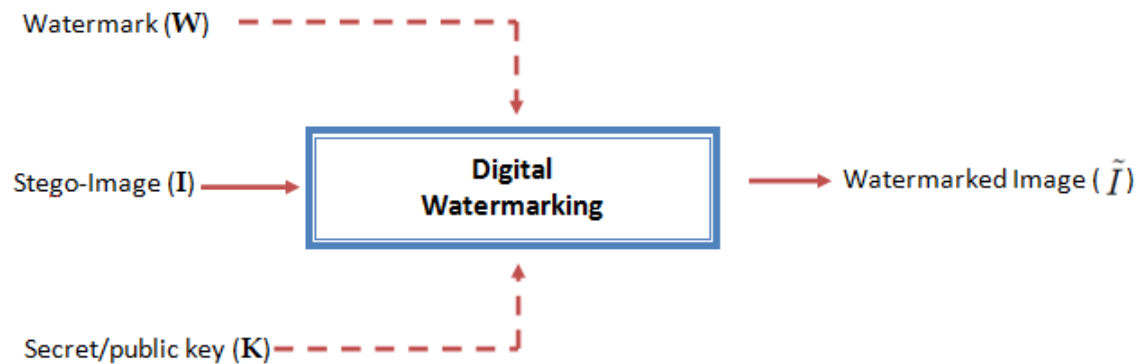


Figure 1.3: Generic Watermark Embedding Process

Given an image I , a watermark W and a key K (Usually the seed of a random number generator) the embedding process can be defined as a mapping of the form:

$I \times K \times W \rightarrow \tilde{I}$ and is common to all watermarking methods.

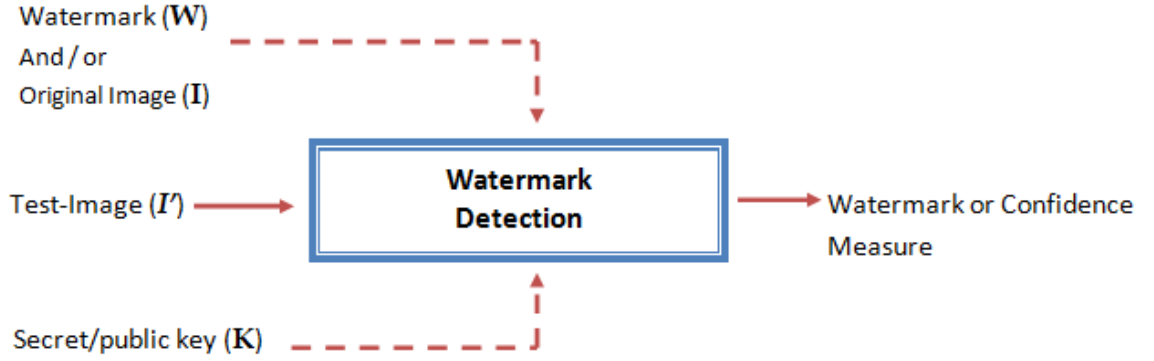


Figure 1.4: Watermark Recovery/Detection process

Semi-Private Watermarking: In contrast to private watermarking, this scheme does not use *original-image*, and/or *watermark* but *secret key*. Based on semi-private watermarking researches published by Barni et al [15], Kundur et al [16] and Schyndel et al [17], significant number of currently proposed schemes come under this category.

Public Watermarking: This technique does not rely on previously explained detection methods but it use a technique which extract m bits information (the watermark) from the watermarked image. Multiple research studies [2], [18], [19] have been published on public watermarking schemes and there are many application widely use in applications of watermarking. In fact, this algorithm can be used in private and semi-private watermarking systems as well.

1.5 Attacks on Digital Image Watermarks

This section describes most common attacks perform on digital watermarking systems. This thesis focuses on geometric transformation type of attacks on imperceptible watermarking schemes and their robustness. Geometric transformation type of attacks are important to study because, they're the hardest to counter but easiest to mount. A detail explanation of popular attacks described in the book of “*Watermark Systems Engineering*” [20] and few research studies published by Voloshynovskiy et al [1], Kutter et al [2], [7] and Collberg et al [3].

A good watermarking system embeds watermark information into host and it guarantees that watermark is resistant to format conversion that can occur when watermarked content is stored, transmitted etc. If watermarked data is coupled to asset's header or to a separate file, no resilience can be guaranteed. The problem is, in the event of

transformations like “*lossy compression-JPEG*” can cause a loss of watermarked information in the asset. Hence it is important to figure out limits of watermarked data can survive on particular transformations. One objective of this section is to analyze and model types of impairments that can cause imperceptible watermarked data. Usually impairments are also known as attacks. It is important to understand what type of attacks work on different watermarking schemes and their scope.

Classification of Attacks

This topic depicts a possible classification of various attacks. In the beginning, it is convenient to distinguish terms between malicious and non malicious attacks:

- **Non Malicious:**

Attacks that can occur during normal use of asset fall in to this category. Attack nature and strength are strongly dependent on application for which watermarking system is devised. Among these, lossy compression, geometric and temporal manipulations, digital to analogue conversion, extraction of asset fragments (cropping), processing aimed at enhancing asset quality (e.g. noise reduction) etc are common in nature.

- **Malicious:**

Attack is malicious if its main objective is just to remove or make watermark unrecoverable. According to digital watermarking theory, malicious attacks can be further classified into two distinct classes:

1. **Blind:** A malicious attack is said to be blind if it tries to remove or make watermark unrecoverable without exploiting any knowledge of particular algorithm use for watermarking asset. For instance, watermark copy attack which estimates watermark signal with the aim of adding it to another asset can be depicted as a blind attack.
2. **Informed:** Malicious attacks are said to be informed if they attempt to remove or make watermark unrecoverable by exploiting the knowledge of particular algorithm use for watermarking asset. These attacks first try to extract some secret information about algorithm from publicly available data and then based on this information, nullify the effectiveness of watermarking system.

Watermark classification has some limitations due to distinction between malicious, non malicious, blind and informed attack terms, because sometimes these attacks cannot be defined clearly. For example, some noise reduction algorithms which are commonly used for enhancing image or audio quality, can also be applied with explicit aim of removing the watermark treated as noise. Similarly, malicious attacks which perform

some random local geometric manipulation could not be considered really blind attacks as they exploit some knowledge of the structure of most common watermarking algorithms.

Hartung et al. [21] have described a way of classifying attacks on transform domain watermarks. The terms that authors have been used in classification are common wordings in digital watermarking. They distinguish attacks into four groups:

1. **Simple Attacks:** This kind of attacks are called ‘waveform attacks’ or ‘noise attacks’. These attacks are conceptually simple attacks that attempt to impair embedded watermark by manipulations of whole watermarked data (Host or asset data together with watermark information), without an attempt to identify and isolate the watermark. Some of examples include linear and general non-linear filtering, waveform-base compression (JPEG, MPEG), addition of noise, addition of offset, cropping, quantization in the pixel domain, conversion to analog, and GAMMA correction.
2. **Detection-disabling attacks:** Other possible names include ‘synchronization attacks’. This kind of attacks that attempt to break the correlation and make the recovery of the watermark impossible or infeasible for a watermark decoder, mainly by geometric distortion like shift in spatial or temporal (for video) direction, resizing, rotation, shear, cropping, pixel permutations, sub-sampling, removal or insertion of pixels or pixel clusters, or any other geometric transformation of the host data as published by Licks et al [22]. This type of attacks have a typical property that the watermark remains intact in the attacked data and can usually be recovered by applying some additional knowledge of the watermark decoder.
3. **Ambiguity attacks:** This attack type is also known as ‘confusion attacks’, ‘dead-lock attacks’, ‘inversion attacks’, ‘fake-watermark attacks’, ‘fake-original attacks’. This attack attempts to confuse by producing fake original data or fake watermarked data. For example, Craver et al. [23] described an inversion attack (Sometimes referred as ‘IBM attack’) that attempts to discredit authority of the watermark by embedding one or several additional watermarks. Hence, it is unclear which was first made authoritative watermark of “intellectual property right” owner.
4. **Removal attacks:** This type of attacks do attempt to analyze watermarked data, estimate the watermark or host data, separate watermarked data into host and watermark data and finally discard only watermark. Collusion attacks [11]–[13], denoising attacks, certain non-linear filter operations, or compression attacks using synthetic modeling of the image (e.g. using texture models or 3D models) are examples of such attacks. Idea is, if attacker knows the position of single

watermark bits in watermarked data (pixel that host one specific watermark bit), which gives rise to a certain collusion attack where different fingerprinted copies of data are interleaved on a watermark-bit by watermark-bit basis.

It is a complex task to separate these classified attacks with a clear line. Because some attacks do not clearly belong in one group. For example, cropping can be taken as either simple attack or a detection-disabling attack. MATLAB implemented examples with other common attacks are explained further by Kutter et al [7].

1. **JPEG compression:** JPEG is most widely used compression algorithm for images. There should be at least significant resilient to some degree of compression on any image watermarking system. Also there are other compression methods like “Two-dimensional true compression” schemes. An example of compression with histogram visualization illustrated in Figure 1.5. Image parameters are:

- *Compression Scheme: Two-Dimensional True Compression*
- *Compression Method: “GBL_MMC_H”*
- *Compression Ratio: 3.46 %*
- *Image: Mandrill.tiff (a.k.a. Baboon)*

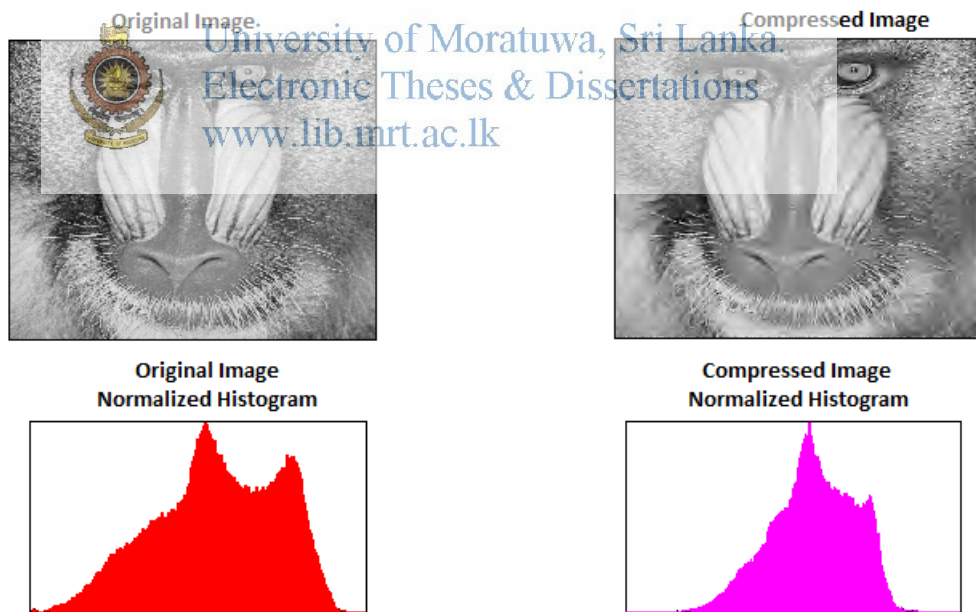


Figure 1.5: Two-Dimensional True Image Compression

2. **Geometric transformation:**

There are multiple types of transformations illustrated with a standard image Mandrill.tiff (a.k.a. Baboon):

- **Rotation & Image flipping:** An example is depicted in Figure 1.6. Usually, small angle rotation, often perform in combination with cropping. It does not usually change commercial value of image but can make watermark undetectable. Rotations are used to realign horizontal features of an image and it is certainly the first modification applied to an image after it has been scanned.

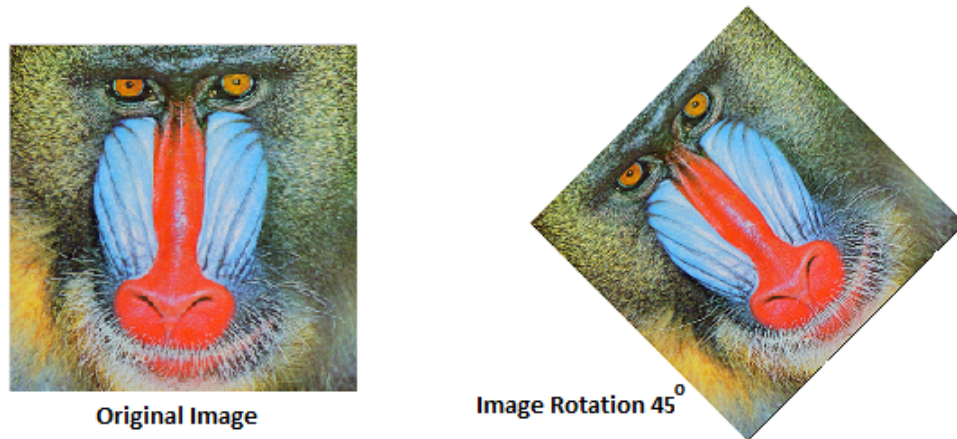


Figure 1.6: Image Rotation Attack

- **Cropping:** In some cases, infringers are just interested by the “central” part of copyrighted material, moreover more and more web sites use image segmentation, which is basis of the “Mosaic attack”. This is of course an extreme case of cropping. If the algorithm is non-blind, it is better to bring back those parts from the original watermarked image for a better recovery of message. An example illustrated in Figure 1.7.

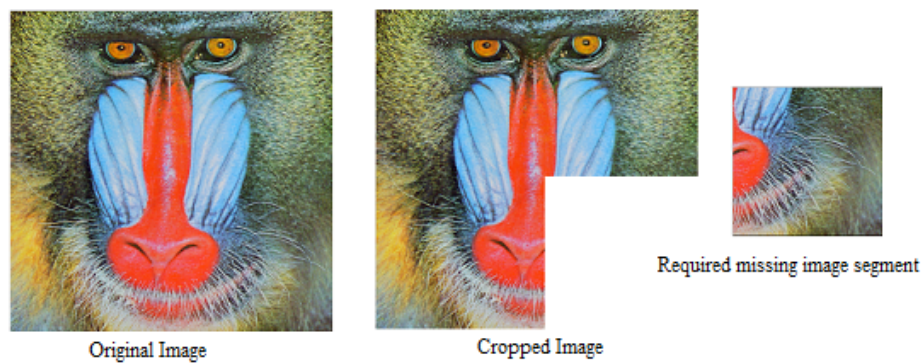


Figure 1.7: Image Cropping Attack

- **Scaling:** This attack initiates when a printed image is scanned or a high resolution digital image is used for electronic applications such as web publishing. Scaling can be divided into two groups; “uniform” and “non-uniform” scaling. Under uniform scaling, it is defined equally in horizontal and vertical

direction. Non-uniform scaling uses different scaling factors in horizontal and vertical direction (change of aspect ratio). Very often, digital watermarking methods are resilient to uniform scaling. For example, scaling down an image is shown in Figure 1.8.

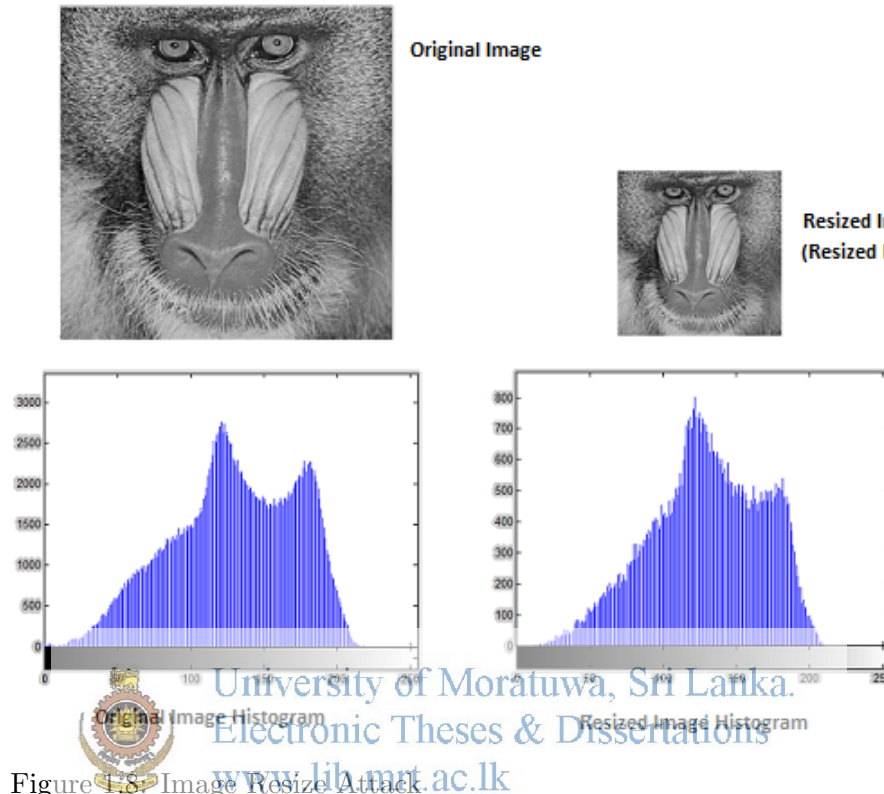


Figure 1.8: Image Resize Attack

- Generalized geometrical transformations: A generalized geometrical transformation is a combination of non-uniform scaling, rotation, and shearing.
- Random geometric distortions (StirMark): StirMark [24], [25] is an image-watermarking and processing tool that measure watermarked data robustness [7]. Images which are not survived during StirMark attacks are considered unacceptably easy to break.
- Geometric distortion with JPEG: Rotation and scaling alone are not enough and they should be also tested in combination with JPEG compression. Since most artists will first apply geometric transformation and then save image in a compressed format. It makes sense to test robustness of watermarking system to geometric transformation followed by compression. It is difficult to choose a minimal “quality factor” for JPEG as artifact quickly appear. Most research studies show that, “quality factors” down to 70 % are reasonable.

3. Enhancement Techniques

- Low pass filtering: This attack type includes linear and non-linear filters.

Frequently use filters include median, gaussian, and standard average filters. An example of image filtering with morphological operators and noise removal using a Wiener filter is shown in Figure 1.9.

- **Histogram modification:** This attack includes histogram stretching or equalization which is sometimes used to compensate poor lighting conditions by means of contrast enhancement techniques. The end results have been exposed many of previously hidden features. An example of histogram equalization is shown in Figure 1.9.



Figure 1.9: Image Enhancement Techniques: Filters and Histogram Equalization (Source-Standard Grayscale Images: rice.tiff, cameramen.tiff, child.tiff)

- **Gamma correction:** Very frequently used operation to enhance images or adapt images for display. This is typically done after scanning of digital images.
4. **Noise Addition:** Additive noise and uncorrelated multiplicative noise have been largely addressed in communication theory and signal processing theory literature. Many authors often claim that their copyright marking techniques survive this kind of noise but many forget to mention the maximum level of acceptable noise. Any noise could be gaussian, poisson, salt & pepper, and speckle. In Figure 1.10 shows “salt & pepper” noise attacked image.
 5. **Statistical Averaging and Collusion:** This is performed by having two or more copies of same image but with different watermarks. It should not be possible to remove the marks by averaging these images or by taking small parts of all images and reassembling them. Some of experiments for collusion attacks were explained by Wolfgang et al [11], Wang et al [12] and Stone Harold [13].

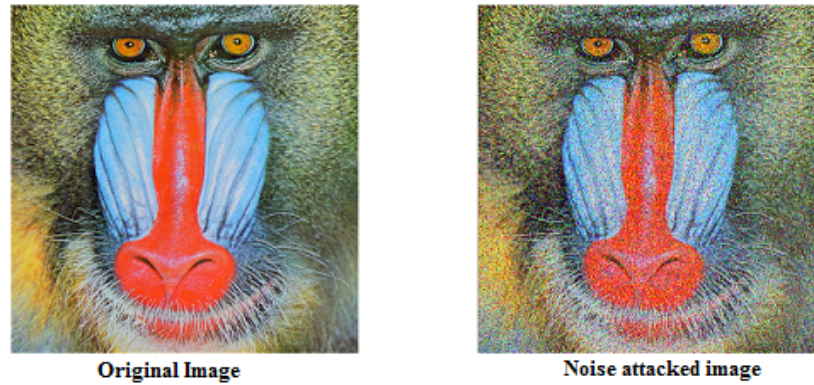


Figure 1.10: Image Noise Attack

6. **Oracle attack:** When a public decoder is available, an attacker can remove a mark by applying small changes to image until decoder cannot find it anymore. A theoretical analysis of this attack and a possible countermeasure (randomizing the detection process) have been presented recently in [26]. Apart from that, an inventor could make decoding process computationally expensive. However neither approach is really satisfactory in the absence of tamper-resistant hardware.

1.6 Research Objective



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

The research focuses on reliability and robustness of an invisible (imperceptible) watermarking scheme under various transformation/geometric attacks. The initial aim of this research is to complete a comprehensive study of selected digital watermarking schemes and related attacks shown in chapter 2. Main objective of this dissertation is to come up with an attack success or failure evaluation metric against chosen watermarking schemes depicted in section 3.1 and with set of attacks as in section 3.2. “Success” or “Failure” is defined as per given criteria: If there is a legitimate owner for the image, *success* of the image attack would be determined if recovery of watermark is denied given the fact that attacked image lies within the perceptual range. Otherwise it would be depicted as a *failure* of image attack. This study also reviews several different invisible watermarking schemes under different transformation schemes and is used to evaluate different properties identified in this research.

Chapter 2

Literature Review

Rapid growth of Internet during last two decades, immensely increase the usability and availability of digital content such as images, audio and video. The Internet has its own features such as inexpensiveness and popularity. So it has become very easy to distribute digital media contents among people. Digital data which is shared or manipulated or transmitted without consent of original author is an intellectual property violation. It is very important to prevent such activities and protect intellectual property of creators or inventors. This is where *digital watermark* come in to play a huge role on digital content in order to secure it. In past research studies, digital multimedia watermarking has significantly improved and various copyright protection methods, algorithms were proposed. In the field of watermarking, majority of watermarking research studies are done on *image watermarking* area rather compared to video and audio digital data. There could be several reasons for that, such as sample images are available all the time; Image carries redundant information which helps the watermark information to embed with less distortions and so on.

Comprehensive studies of watermarking areas and categorization have been done by Petitcolas et al [27], Mohanty et al [28] and Hartung et al [29]. Typically, watermarking schemes can be classified in to two different sets based on digital watermark embedding domain.

1. **Spatial Domain:** In spatial domain, image is represented in terms of *Pixels* values. Hence watermarking system directly modifies subjected image data elements such as intensities of image pixels to hide/embed the watermark.
2. **Transformed Domain:** In transformed domain, image is represented in terms of *Frequency* values. Hence watermarking system uses DCT (Discrete Cosine Transform), DFT (Discrete Fourier Transform) etc techniques to modify “*transform domain coefficients*” to hide/embed watermark.

In addition, a significant number of research work in past decade have been studied on hybrid methods for image watermarking which use both spatial and transform domain techniques. The digital watermarking hierarchy illustrated in Figure 2.1 is based on survey published by Mohanty [28].

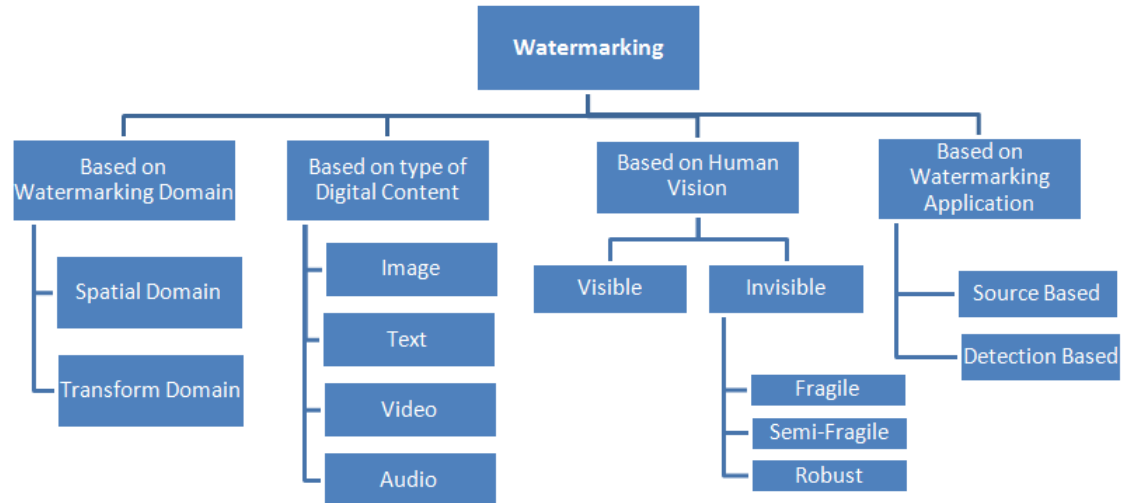


Figure 2.1: Digital Watermarking Hierarchy

As shown in Figure 2.1, watermarking techniques can be divided into four categories based on document type, image, text, video and audio. Due to human perception, again watermarks are divided into invisible and visible categories. Finally, watermarking applications fall into these categories:

- **Source-based Application:** Used for ownership identification for digital content. It also supports for tamper detection on host content type. (E.g. image, text etc)
- **Destination-based Application:** This kind of application is used to embed unique watermark information on distributed digital contents. It helps to back-track end user identification as well as preserves copyright ownership information of distributor.

2.1 Generic Watermarking System

A typical image watermarking scheme has two main components. Those are: watermark encoder and watermark decoder. Sometimes encoding process and decoding process also known as embedding and detection/extraction of the watermark. Craver et al [23] have explained a structure of a general image watermarking scheme as illustrated in Figure 2.2 and Figure 2.3.

2.1.1 Steps of Image Watermark Encoding

Let's take the original image I , the watermark S feed in to respective encoder function ϵ and the output watermarked image as $\hat{I}$.

Hence, equation can be derived as:

$$\epsilon(I, S) = \hat{I} \quad (2.1)$$

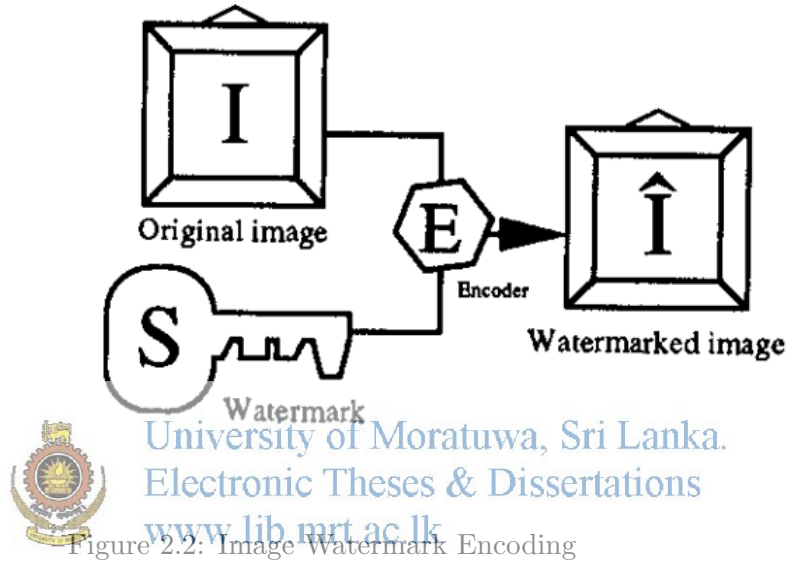


Figure 2.2: Image Watermark Encoding

2.1.2 Steps of Image Watermark Decoding

For decoding process, D denotes decoder function and it takes input image J which could be either watermarked or not and distorted image which needs to be verified the ownership. Hence, decoder recovers a watermark S' or output a confidence measure value for existence of original watermark in J . Sometimes original image also include in decoding process for a particular watermarking system.

If decoding process has original image I taken as reference image, equation can be derived as:

$$D(J, I) = P(T) \quad (2.2)$$

The extracted watermark T in subjected image J is expressed by function P . It is also known as “*Private Watermarking Schemes*”. If decoding process does not need original

image I , equation would be:

$$D(J) = P(T) \quad (2.3)$$

After that, extracted watermark T and original watermark S is compared by a comparator function C_δ and output a binary value to make the decision of watermark comparison.

$$f(n) = \begin{cases} 1 & \text{if } c \geq \delta \\ 0 & \text{otherwise} \end{cases}$$

Where c represents correlation of two watermarks.

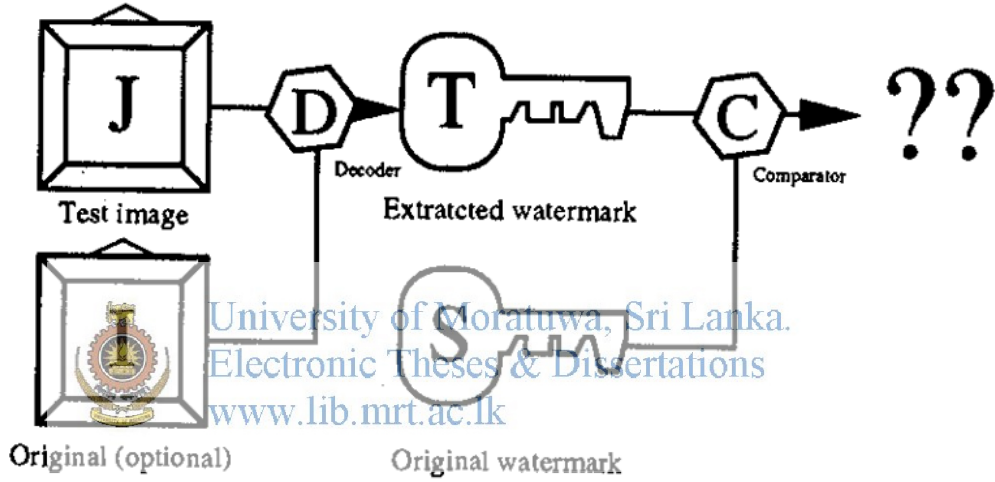


Figure 2.3: Image Watermark Decoding

This chapter describes different image watermarking schemes along with imperceptible watermarking techniques. Primary focus is here to elaborate problem statement of the thesis with objectives mentioned in section 1.6.

2.2 Spatial Domain based Image Watermarking

2.2.1 LSB Watermarking Schemes

In spatial domain, it's the simplest way of embedding digital watermark to an image. Image is represented as two dimensional array of pixels and hide watermark data using *least significant bit* (LSB) method. Small change in LSB will not detectable as human eyes are not sensitive to notice a tiny variance in color of images. According to Darshana

Mistry [30] and Chopra et al [31], a typical algorithm of LSB watermarking can be explained as:

LSB Algorithm Steps:

1. Convert RGB image to 8-bit gray scale image
2. Make double precision of image – All values will be mapped between 0-255
3. Shift Most significant bits (MSBs) to Low significant bits of watermark image
4. Set all LSBs of host image to zero ($n=0$)
5. In step-3, shifted version of watermarked image is added to modified host image in step 4

Significant number of spatial domain techniques is used in fragile watermarking systems and they are primarily focused on tamper detection of images. Such methods are described by Walton [32] and Van Schyndel et al [17]. As per author's point of view, they embed watermark in least significant bit pane for perceptual transparency. It is a known fact that LSB techniques cannot withstand to lossy compressions attacks and end up with damaging the watermark. By passing security of a LSB watermarked images are also explained by Fridrich [33] and Minerva et al [34].

Typical LSB watermarking uses a technique of hiding message (can be another image) in host image bit planes. The *Lena.tif* is used as the host image (cover image) in the experiment. The representation of all 8 bit plane of host image is shown in left of Figure 2.4.

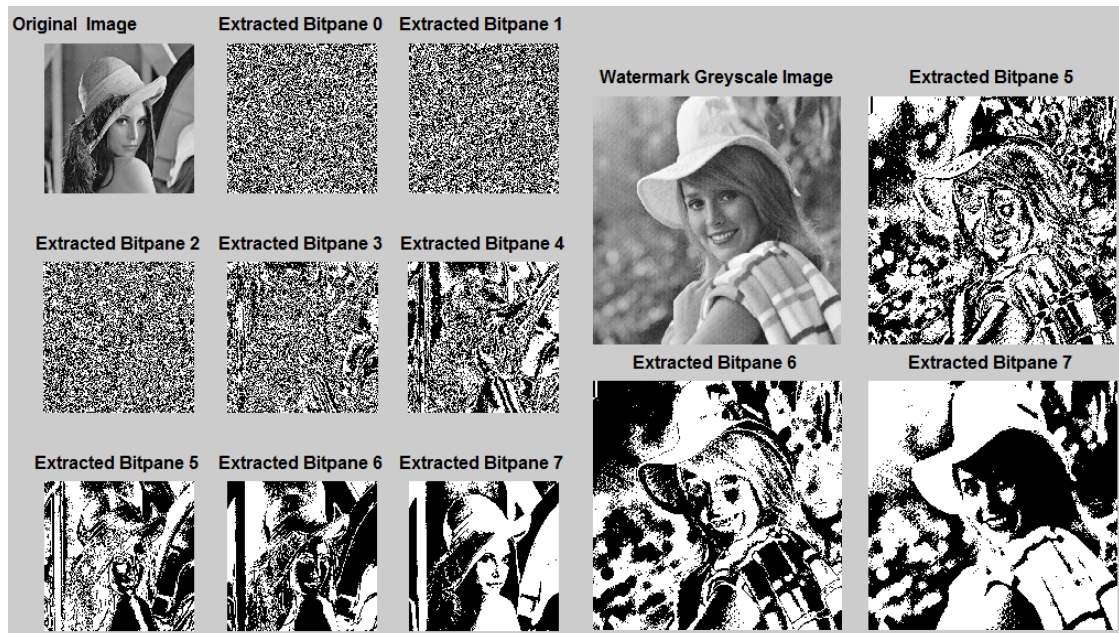


Figure 2.4: Image Comparison-LSB Watermarking

First bit-plane (Bitplane 0), is least significant one (LSB) and it does contain noise most of the time. Second and third bit-planes are hardly related to original shapes of the image. Last bit-plane (Bitplane 7) is most significant one (MSB) and consists significant lines and edges of the image. The message which is used to hide in host image is an image (*Elaine.tif*). Similarly important bitpane can be extracted from message image as shown in right of Figure 2.4 and replaced those bit-planes instead of insignificant bit-planes of host image. Finally all bit-planes are encoded and LSB watermarking is completed. LSB watermarking algorithm and bit-plane extraction method are implemented using MATLAB software.

It is observed that output watermarked image has a good perceptual quality and watermarked message is imperceptible. In LSB decoding process, message image can be recovered by extracting insignificant bitplanes of watermarked image and then reconstructed it as in Figure 2.5.

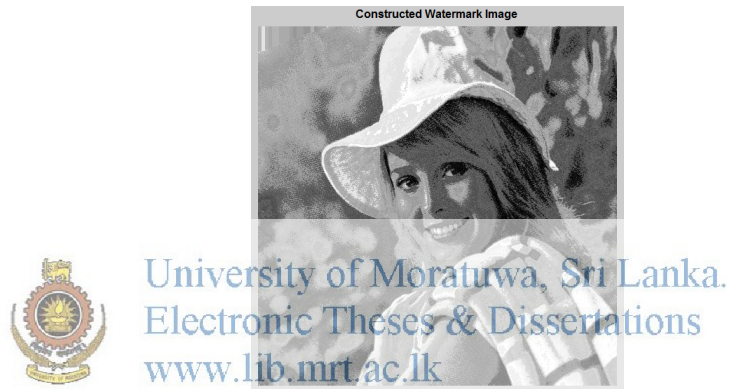


Figure 2.5: Recovered Message Image

Among LSB watermarking surveys explained so far, fair amount of schemes can survive simple attacks such as cropping and noise addition but not other attacks like lossy compression etc. Original image and watermarked image with relevant histograms are illustrated in Figure 2.6 and Figure 2.7.

However Celik et al. [35] have proposed a modified LSB scheme to have lossless feature. Authors present a high-capacity, low-distortion, lossless data-embedding algorithm which is a generalization of well-known least significant bit (LSB). This technique modifies lowest levels of bit planes of image.

2.2.2 Patchwork based Watermarking Schemes

Another spatial domain based watermarking method is patchwork-based watermarking technique. Bender et al. [36] have proposed two watermarking techniques. They are:



Figure 2.6: Image Comparison-LSB Watermarking

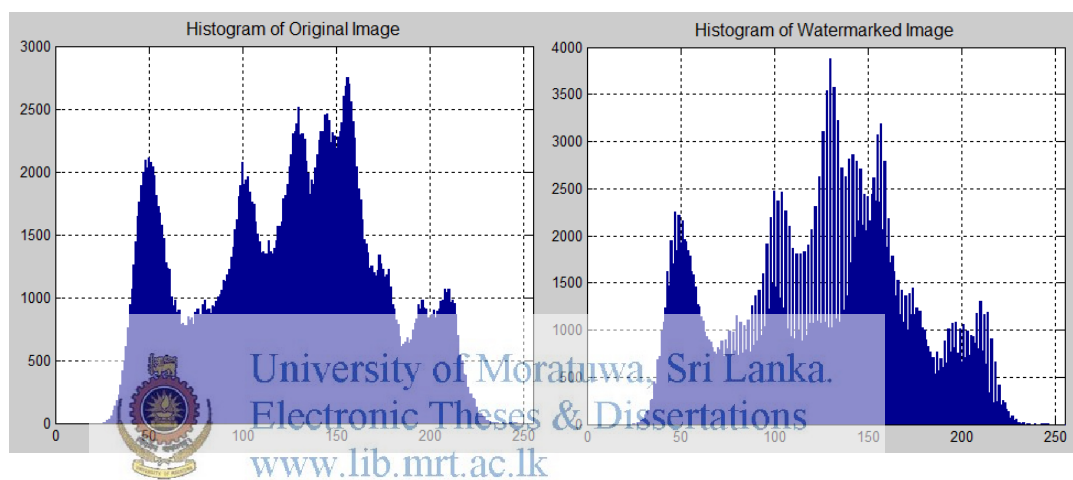


Figure 2.7: Histogram Comparison-LSB Watermarking

1. *Patchwork* method based on a statistical analysis

In this patchwork method, it randomly chooses pair of image points and increases brightness at a point by unit value whilst other point decreases that value by same amount.

2. *Texture Block Coding* Method

In this method, random texture pattern regions are selected from watermark and copied into an area of host image with similar texture. At recovery of each region, they used a technique called “*Autocorrelation*”. This method fails to process with images that have large areas of random textures such as images of texts etc. Authors have explained it as a drawback which they couldn’t avoid in their research.

However, other patchwork based algorithms invented for image watermarking were explained by In-Kwon et al [37] and ShaoWei et al [38].

2.2.3 Correlation based Watermarking Schemes

In spatial domain watermarking, these schemes use a method of adding a pseudorandom noise pattern (PN) to luminance values of each image pixels during the process of adding watermark in to host image. Many methods [17], [39]–[48] have been proposed in past for this type of watermarking schemes. The book of “*Watermarking digital image and video Data*” published by Langelaar et al. [49] and also Neil Johnson [50] have been described few types of schemes.

- **Scheme with 1-PN sequence**

Pseudorandom noise pattern consists of integers $-1, 0, 1$. This pattern is generated based on various techniques. For example; it is based on key or seeds, linear shift registers, randomly shuffled binary images etc. The constraints need to be satisfied are: energy in the pattern should uniformly distribute (more or less amount) and that pattern is not correlated with host image content. To create a watermarked image $I_w(x,y)$, pseudorandom pattern $W(x,y)$ is multiplied by a small gain factor k and added to host image $I(x,y)$, as depicted in Figure 2.8.

$$I_w(x, y) = I(x, y) + k \cdot W(x, y) \quad (2.4)$$

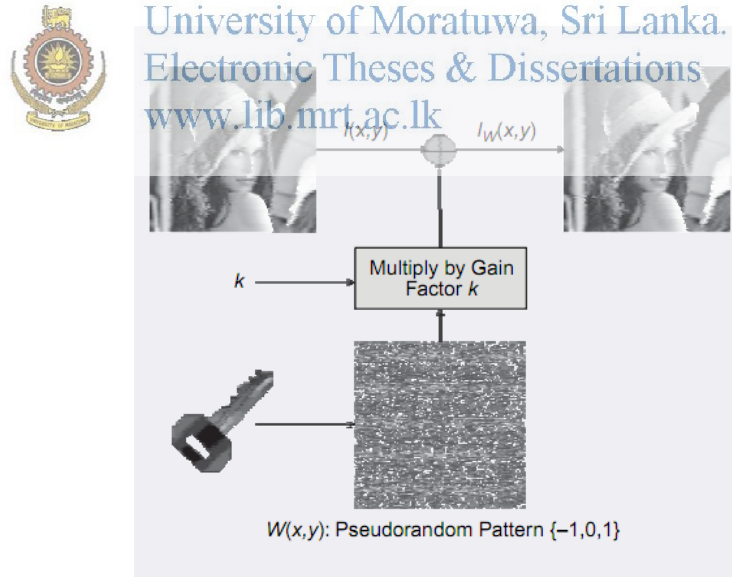


Figure 2.8: Watermark Encoding-Correlation Scheme

Detection of the watermark existence on a scheme is done by calculating correlation between possibly watermarked image $I'_w(x,y)$ and pseudorandom noise pattern $W(x,y)$. In general, $W(x,y)$ is normalized to a zero mean before correlation. Pattern is generated using different keys which have very low correlation

with each other. Hence, during decoding process there could be very high correlation value or otherwise for a pattern generated with correct key. Usually it is a common practice to set a threshold T to decide whether watermark detected or not. If correlation exceeds a particular threshold T , the watermark decoder determines image $I'_w(x,y)$ has watermark $W(x,y)$. It can be simply depicted as:

$$R_{I'_w(x,y)W(x,y)} = \begin{cases} > T & W(x,y) \text{ Detected} \\ < T & \text{No } W(x,y) \text{ Detected} \end{cases}$$

- **Scheme with 2-PN sequence**

This is an improved version of previously stated algorithm. Use of a threshold can be evaded by adding two different pseudorandom patterns RP_0 and RP_1 for watermark bit 0 and 1. Hence, decoder now calculates subimage and two patterns. Same technique also described by Smith et al [51]. There, two patterns are chosen such that $RP_0 = -RP_1$ and then decoder only has to calculate correlation among subimage and one of pattern. Sign of correlation would depicts watermark bit value. In simple terms, one pattern is designated a logical “1” and the other a logical “0”, then same procedure is continued once for each pattern and finally, pattern with higher resulting correlation is used. This is more efficient than 1-PN sequence scheme as it increases the probability of correct watermark detection.

- **Scheme with Pre-filtering**

Before applying the watermark, image is pre-filtered by improved version of same algorithm described in subsection 2.2.3. As stated in 1-PN sequence scheme, reducing correlation between cover image and PN sequence is mandatory. This can be achieved by applying pre-filtering via a simple edge-enhancing finite impulse response (FIR) filter F_{edge} . Hence, the scheme is resistant to noise. F_{edge} is given by following convolution kernel as explained by Langelaar et al [49]:

$$F_{edge} = \frac{1}{2} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 10 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad (2.5)$$

- **CDMA spread spectrum scheme**

This scheme has been introduced to be resilient against cropping attacks as well as in order to achieve high robustness than other schemes. If CDMA (*code division multiple access*) techniques are used, probability that all bits can be recovered after cropping a particular image is high. However, many multiplications are required to compute correlations, since each bit is completely spread over the

image.

⇒ Embedding Procedure

Watermark is formatted in to large string instead of 2-D image. For each value of watermark, a PN sequence is generated using an independent seed. Watermark is represented by summation of all PN sequences. Then it is scaled and embedded in to host image.

⇒ Extraction Procedure

Each and every seed is correlated with whole image as they are used to generate its PN sequence. If a high correlation is detected, bit in watermark is set to “1” else to “0”. Repeat this process for all values of watermark and extract it.

2.2.4 Other Spatial Domain Watermarking Schemes

Apart from previously mentioned watermarking schemes, Ioannis Pitas [52] has proposed a method that total number of intensity levels of host image can be split into two subsets: A and B . Then, intensity level of one subset (A) of image is increased by a constant value called embedding factor k . Luminance values are small enough to be undetected by human eye. In watermark detection, they evaluated difference of mean values of pixels in both subsets. If difference is equal to factor k , then image is watermarked and if it is zero then it is not watermarked. Hypothetical testing is used for the determination of difference of two mean values. Although watermarking scheme is designed to be robust, yet it is vulnerable to attacks like low-pass filtering, clipping, line or column removal etc.

Nikolaïdis [53] has proposed an extension to stated algorithm in subsection 2.2.4. In order to achieve robustness against compression attacks, marked pixels in subset A are grouped into blocks. These blocks are composed of certain size $b_1 \times b_2$, e.g. 2×2 or 2×4 . Such grouping of pixels can be implemented by selecting $P/(b_1.b_2)$ randomly positioned block of size $b_1 \times b_2$ where P is a total number of pixels in subset A . In fact authors have put energy of new watermark signal in to low frequencies in order to achieve higher JPEG compression. When block size increases, robustness against compression also increases. During watermark embedding part, technique which is used to select pixels that belong to subset A is the only difference between this variation and Pitas’s work [52]. Detection algorithm will output same results if, instead of modifying intensity of all pixels in A by same constant k , authors use a different value k_{nm} for each pixel in A , provided that sum of all k_{nm} is equal to Pk :

$$\sum_{n,m \in A} k_{nm} = Pk \quad (2.6)$$

2.3 Transform Domain based Image Watermarking

It is a well known fact that transform domain watermarking schemes are highly robust against attacks compared to spatial domain watermarking schemes. Many algorithms are known to be robust against simple attacks such as low pass filtering, brightness, JPEG compression, blurring and noise image transformations etc whereas spatial domain watermarking systems failed to survive. A well known transform base scheme is published by Cox et al [54] and also Podtar et al [4] described a survey of generic transform domain watermarking techniques in detail.

2.3.1 Discrete Fourier Transform (DFT) based Schemes

DFT domain watermarking algorithm modifies DFT magnitude and phase coefficients to embed watermark into host images. A method of DFT based scheme is described by Wolfgang et al [45]. In the algorithm, they modify phase coefficients and embed watermark in order to achieve robustness against contrast operation transformation attacks. Another watermarking method is proposed by Ramkumar et al [55]. In that algorithm, authors modify DFT magnitudes components and embed watermark into host image. Additionally, they show that their scheme is resilient to compression attacks also. In general, many watermarking-compression schemes use this technique.

Mathematical Definition of Fourier Transform

The Fourier transform is a representation of an image as a sum of complex exponentials of varying magnitudes, frequencies, and phases. Fourier transform plays a critical role in a broad range of image processing applications, including enhancement, analysis, restoration, and compression.

For a length- M One-Dimensional (1-D) DFT, the relationship between spatial or temporal domain signals $f[n]$, and their corresponding transform in frequency domain $F[k]$ is:

$$F[k] = \sum_{n=0}^{M-1} f(n) \cdot W_m^{kn} \quad (2.7)$$

where ; $W_m^r \equiv e^{-j2\pi r/M}$

For a digital image, two-dimensional (2D) DFT can be defined as follows.

$$Y[u, v] = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} X(i, j) \cdot W_M^{iu} \cdot W_N^{jv} \quad (2.8)$$

Calculated DFT value is a complex value. Hence it can be represented by **Magnitude** and **Phase** as in Equation 2.9.

$$M(u, v) = |Y(u, v)| \quad \text{and} \quad \phi(u, v) = \angle Y(u, v) \quad (2.9)$$

2.3.2 Discrete Cosine Transform (DCT) based Schemes

DCT is a technique of converting signal into frequency components. Host image is represented as a sum of different magnitudes and frequencies. Number of DCT schemes are published by Chiou-Ting et al [56], Cox et al [54] and Jiansheng et al [57]. In DCT algorithm, initially image is divided into square block of size 8 by 8. Then using block selection criteria and coefficient selection criteria, DCT is applied to each block. After that by modifying selected coefficients, digital watermark is embedded. When DCT transform segments apply to an image; it will give three frequency sub-bands: low frequency sub-band, mid-frequency sub-band, and high frequency sub-band. Each sub-band properties are:

- **Low frequency sub-band:** Large amount of signal energy lies in this band which contain most significant visual parts of host image
- **High frequency sub-band:** Less amount of signal energy lies in this band and easily can be removed through compression and noise attacks
- **Mid-frequency sub-band:** Average impact of Low and High bands are applied

Hence watermark is embedded into middle frequency sub-band so that visibility of image will not be affected and watermark will not be removed by compression and noise attacks. In light of that concept, one good publication on DCT watermarking is presented by Cox et al [54]. In this method, discrete cosines transform (DCT) is applied on host image as illustrated in Figure 2.9 , absolute values of DCT coefficients belonging to low frequencies are higher and locate in upper left corner of output image. Highest frequency coefficients are appeared in bottom right corner with lower absolute values. Authors suggest not changing some coefficients near to DC term in low frequency sub-band and suggested area is shaded in Figure 2.9.

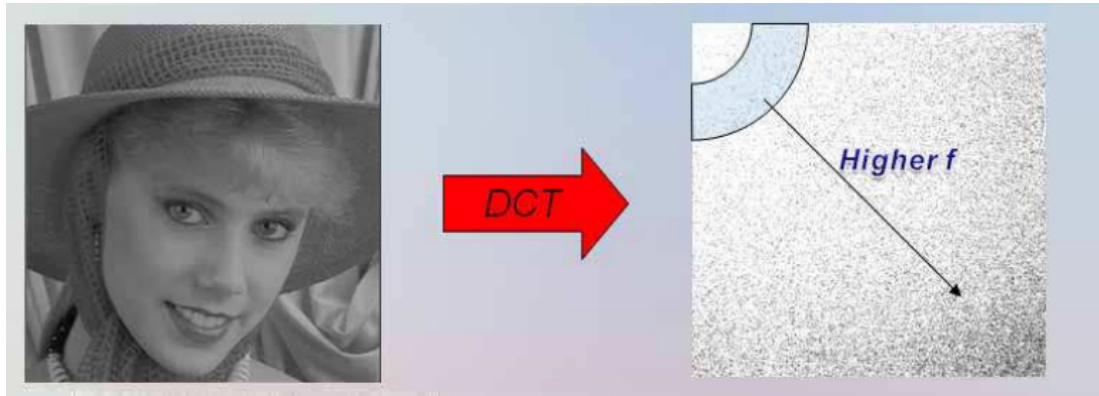


Figure 2.9: Image converted into DCT domain

Secure Spread Spectrum Watermarking (Cox et al. 1997)

Cox et al. use 1000 largest coefficients to embed a watermark sequence of length 1000. The only exception is that, the DC term which is located in (0,0) of DCT matrix should not be changed due to its perceptible change of picture. Beside, highest frequencies are vulnerable to common transformation attacks. Watermark insert/extract process is illustrated in Figure 2.10.

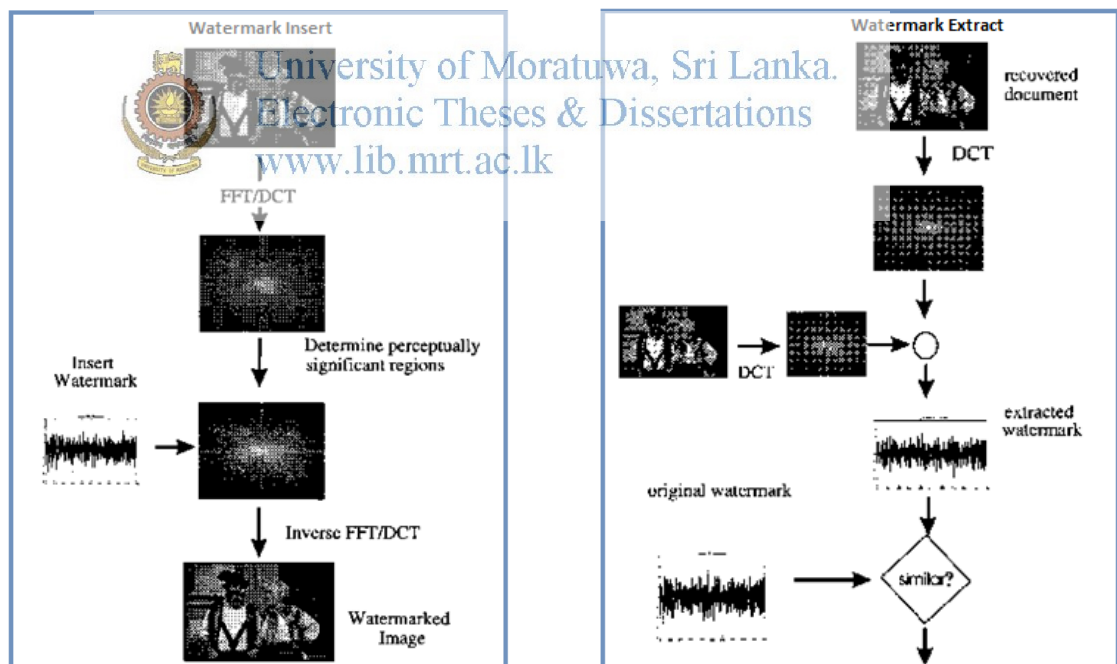


Figure 2.10: Watermark Insert and Extract in Transform Domain (Cox et al. 1997)

DCT domain scheme advantages over spatial domain schemes

- Comparatively much better than spatial domain watermarking because, DCT domain watermarking can survive against attacks such as noising, compression,

sharpening, and filtering

- Since JPEG compression method is a core component of DCT watermarking, an expert may use different parameters related to image processing, and come up with equal or even stronger robustness against various attacks
- By using pseudorandom sequences, such as M-sequences, can be added to DCT middle frequencies as signatures

Discrete Cosine Transform Watermarking Results (Cox et al. 1997)

Watermarking results are generated as per implementation of algorithm explained by Cox et al. 1997. Original image with histogram is illustrated in Figure 2.11.

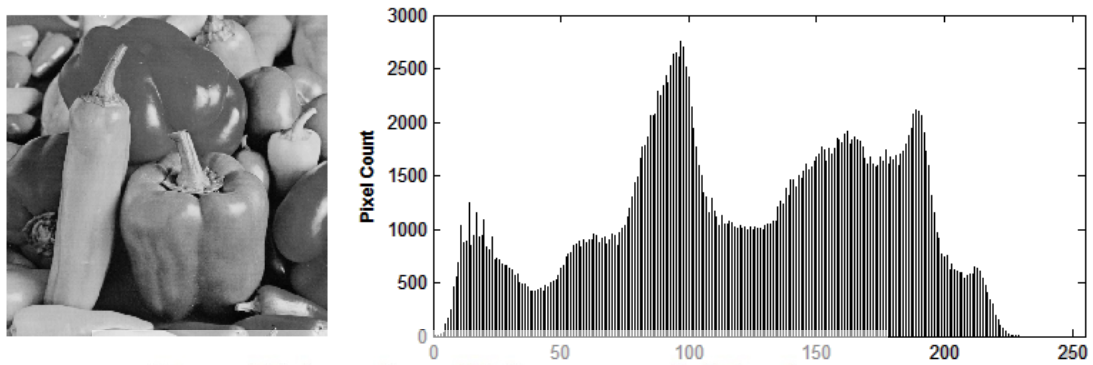


Figure 2.11: Original Image and Histogram (Image:Peppers)

After watermark embedding, output watermarked image and relevant histogram is shown in Figure 2.12.

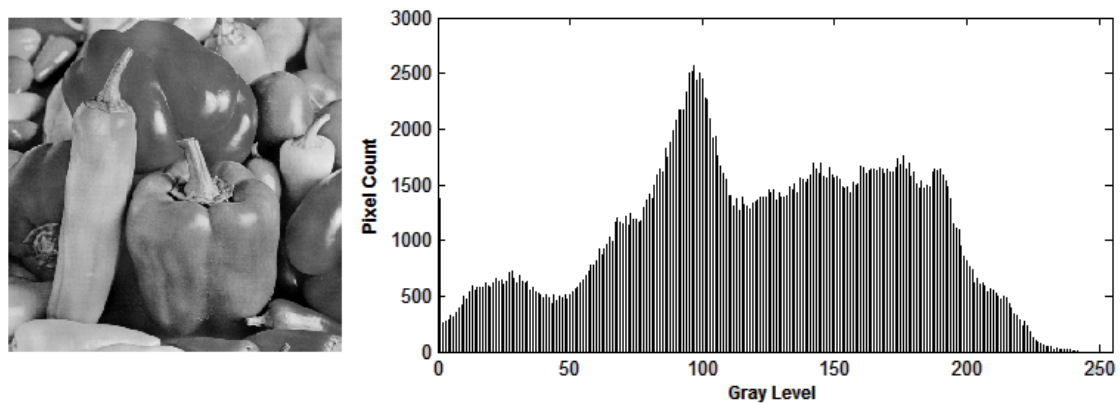


Figure 2.12: Transform Domain Watermarked Image and Histogram (Image:Peppers)

The Figure 2.13 represents image differences in watermarked image (A), noise added watermarked image (B) and cropped and scaled watermarked image (C) with original image. In image differences of “B” and “C” show loss of original image data and added

noises during watermarking process. Hence watermarking process degrades original image upto some extent.

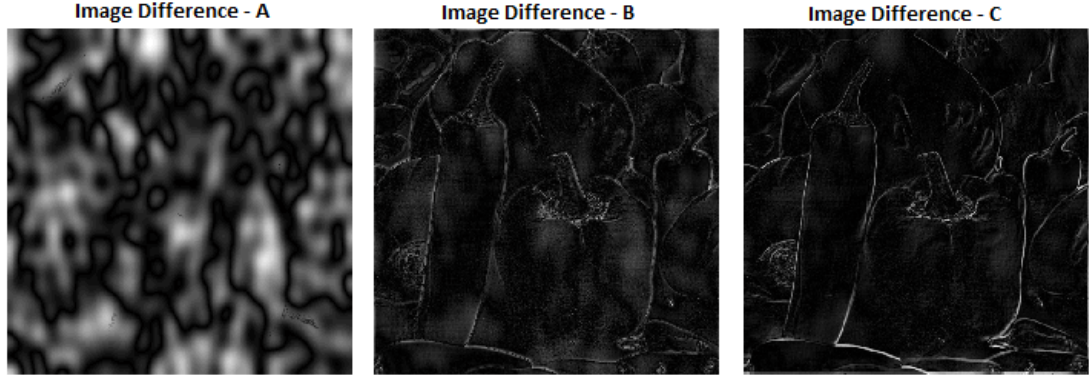


Figure 2.13: Image Differences w.r.t Original Image-DCT

Mathematical Definition of Discrete Cosine Transform

Discrete cosine transform (DCT) represents an image as a sum of sinusoids of varying magnitudes and frequencies. The `dct2` function in MATLAB computes two-dimensional discrete cosine transform (DCT) of an image. Most of visual significant information about image is concentrated in just a few coefficients of DCT and it is a key property of DCT. The DCT is the core standard component of international standard lossy image compression algorithm a.k.a JPEG. For DCT with block size $(M \times N)$, the connection between spatial domain image pixels $x(i, j)$ and transform domain coefficients $Y(u, v)$ is given by

$$Y(u, v) = \frac{2c(u).c(v)}{\sqrt{MN}} \cdot \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} x(i, j) \cdot \cos \left[\frac{(2i+1)u\pi}{2M} \right] \cos \left[\frac{(2j+1)v\pi}{2N} \right] \quad (2.10)$$

where $u = 0, 1, \dots, M-1$, $v = 0, 1, \dots, N-1$ and

$$c(k) = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } k = 0 \\ 1 & \text{otherwise} \end{cases}$$

Image is reconstructed by applying inverse DCT as shown in Equation 2.11

$$X(i, j) = \frac{2c(u).c(v)}{\sqrt{MN}} \cdot \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} y(u, v) \cdot \cos \left[\frac{(2i+1)u\pi}{2M} \right] \cos \left[\frac{(2j+1)v\pi}{2N} \right] \quad (2.11)$$

2.3.3 Discrete Wavelet Transform (DWT) based Schemes

Significant number of approaches and algorithms operating in wavelet domain has been proposed during last few years. Some DWT schemes are published by Jiansheng et al [57] and Terzija et al [58]. Basic theory of discrete wavelet transform (DWT) in image process is to decompose image into sub-images of high frequencies and low frequencies. Thus, it transforms coefficient of sub-image. Once original image has been DWT transformed, it is decomposed into four frequency areas which is called “districts” which is one low-frequency district (LL) and three high-frequency districts (LH,HL,HH) as is illustrated in left part of Figure 2.14. If information of low-frequency district is DWT transformed, the sub-level frequency district information will be obtained. A two-dimensional image after three-times DWT decomposed can be illustrated in right part of Figure 2.14. “L” represents low-pass filter, “H” represents high-pass filter. Now original image can be decomposed of frequency districts of HL1, LH1, HH1 and low-frequency district (LL) information further decomposed into sub-level frequency district information of LL2, HL2, LH2 and HH2. This process is continued an arbitrary number of times and by doing this original image can be decomposed for “ n -level” wavelet transformation.

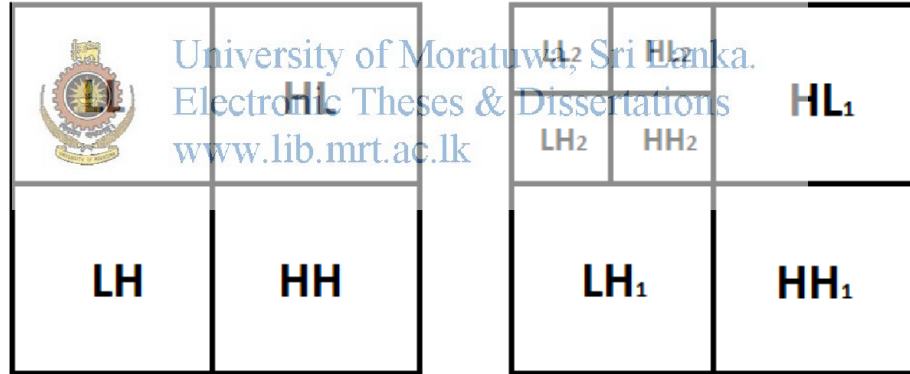


Figure 2.14: Two-Dimensional Discrete Wavelet Transform (1-scale and 2-scale)

Discrete Wavelet Transform Watermarking Embedding and Extraction

Similar to DCT watermarking process, DWT also has two procedures in watermark embedding and extraction. As explained by Terzija et al. [58], for embedding a bit stream is transformed into a sequence $w(1) \dots w(L)$ by replacing 0 by -1 , where L is the length of bit stream and $w(k) \in \{-1, 1\} (k = 1, \dots, L)$. This sequence is used as watermark.

Initially, original image is decomposed into several bands using discrete wavelet transformation as depicted in Figure 2.14. Let's assume, decomposition is performed through

two decomposition levels using “Haar filter”. Watermark is added to largest coefficients in all bands of details which represent higher and middle frequencies of image. Let $f(m, n)$ denote DWT coefficients which are not located at approximation band LL_2 of image. The embedding procedure is performed by Equation 2.12:

$$f'(m, n) = f(m, n) + \alpha \cdot f(m, n) \cdot w(k) \quad (2.12)$$

where α is watermark strength which controls the level of watermark $w(1) \dots w(L)$. In this method, DWT coefficients at lowest resolution band are not modified. Hence, watermarked image is obtained by applying inverse discrete wavelet transform(IDWT).

Both received image and original image are decomposed into two levels during extraction procedure. It is assumed that original image is known for extraction. The extraction procedure is performed by Equation 2.13:

$$w_r(k) = \frac{f'_r(m, n) - f(m, n)}{\alpha \cdot f(m, n)} \quad (2.13)$$

Where $f'_r(m, n)$ are DWT coefficients of received image. Sometimes output image may include noise due to transmission or attacks over communication channel. Hence, extracted sequence $w_r(1) \dots w_r(L)$ consists of positive and negative values. Therefore it is better to take extracted watermark which is depicted in Equation 2.14 by using a signum function. Signum function returns -1, zero and one, when $w_r(k)$ is less than, equal to and greater than zero, respectively.

$$w_e(k) = \text{sgn}(w_r(k)) \quad (2.14)$$

After watermark signal extraction, obtained bit stream is reconstructed as per replacement rule use at initial step. Scenario is illustrated in block diagram of DWT watermark embedding in Figure 2.15 and DWT watermark extraction in Figure 2.16 respectively. This explains non-blind/informed watermarking technique as detection process requires both original watermark and original image.

DWT domain scheme advantages

- According to the character of HVS, human eyes are sensitive to change of smooth parts or districts of image, but not sensitive to tiny change of edges and textures. The edges and textures are usually belongs to high frequency sub bands, such as HH, LH, HL etc. Therefore, putting watermarking signal into big amplitude coefficient of high-frequency band of discrete wavelet transformed image is a better approach. Then it can carry more watermarking signal and has good concealing effect.

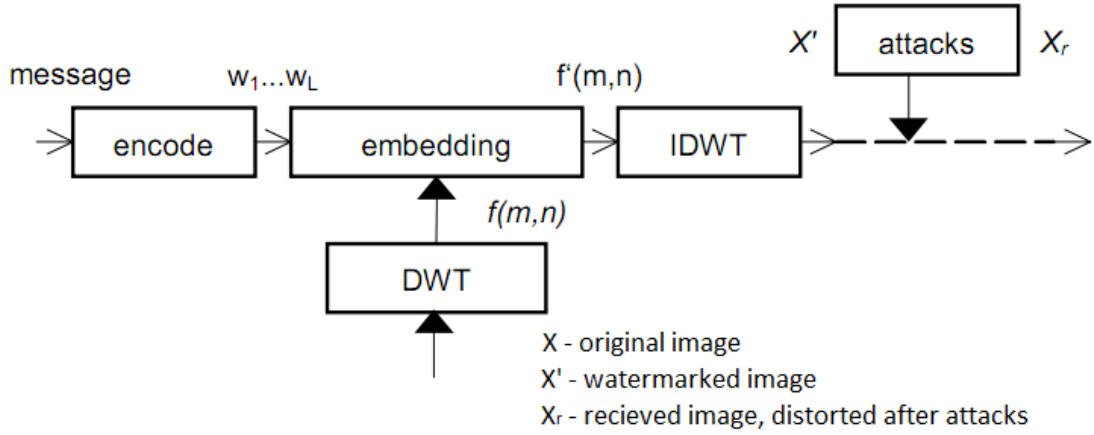


Figure 2.15: DWT watermark embedding block diagram

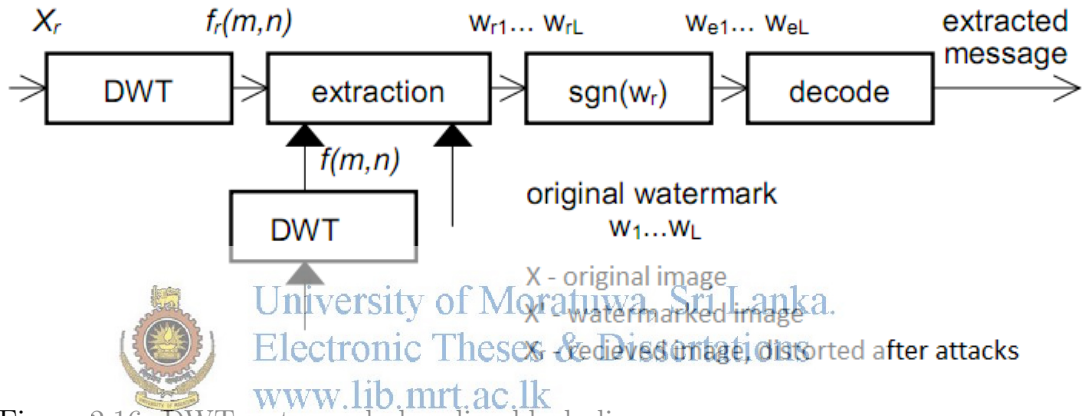


Figure 2.16: DWT watermark decoding block diagram

- This watermarking method robust to wavelet transform base image compressions, such as “*Embedded Zero-tree Wavelet (EZW)*” image compression scheme, and as well as to other common image transformations, such as additive noise, rescaling/stretching, and half toning. This is an advantage over DCT based watermarking.
- Advantages over spatial domain will remain same as DCT in subsection 2.3.2.

Discrete Wavelet Transform Watermarking Results (Wang et al,1998)

Watermarking results are generated as per implementation of algorithm explained by Wang et al. [59]. Original image with histogram is illustrated in Figure 2.17. After embedding watermark, output watermarked image is depicted in Figure 2.18 with histogram.

The Figure 2.19 represents image differences in watermarked image (A), noise added watermarked image (B) and cropped and scaled watermarked image (C) with original image. In image differences of “B” and “C” show original image degradation during

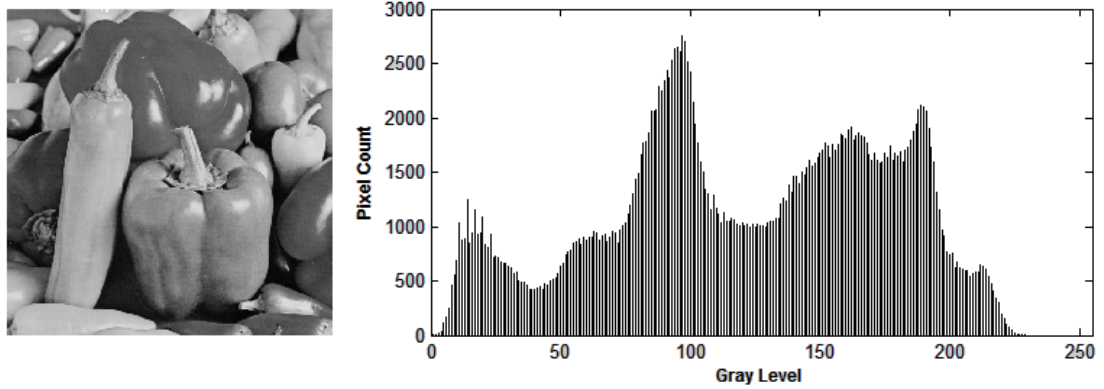


Figure 2.17: Original Image and Histogram-DWT

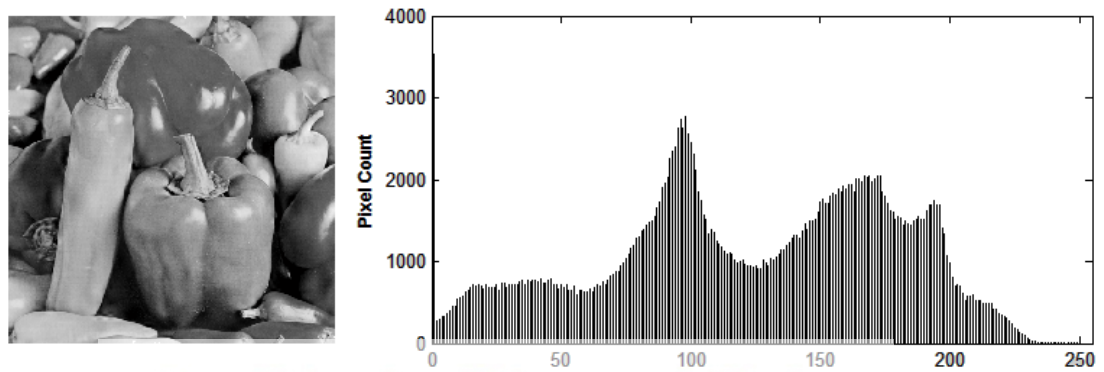


Figure 2.18: Transform Domain Watermarked Image and Histogram

watermarking process.



Figure 2.19: Image Differences w.r.t Original Image-DWT

In Figure 2.20 shows a level 2 decomposed image which is filtered by “haar” filter. Underline structure of this image in block level is depicted in Figure 2.21.

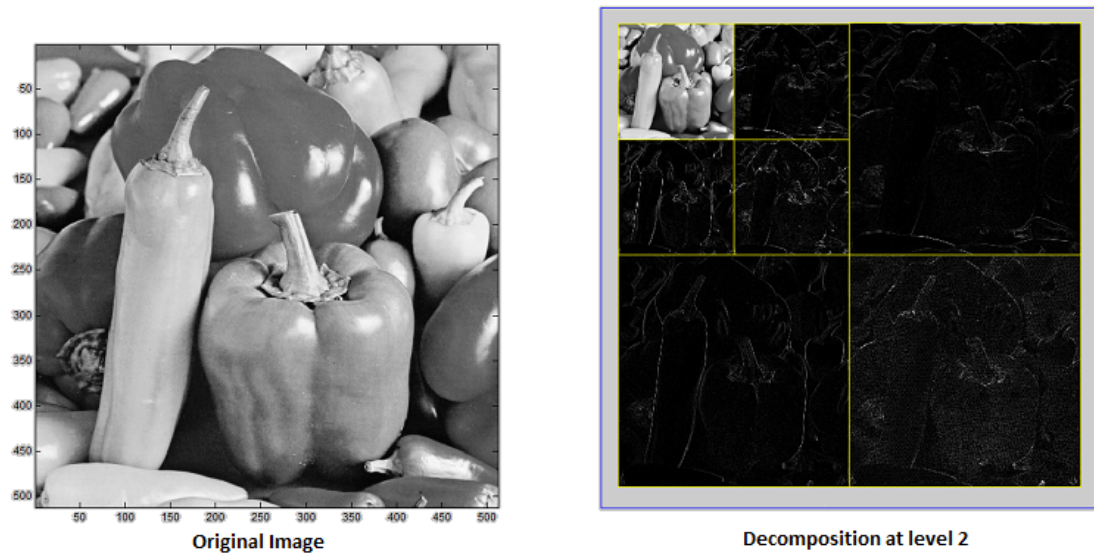


Figure 2.20: Resulting DWT Decomposed Image “Peppers” (Level = 2)

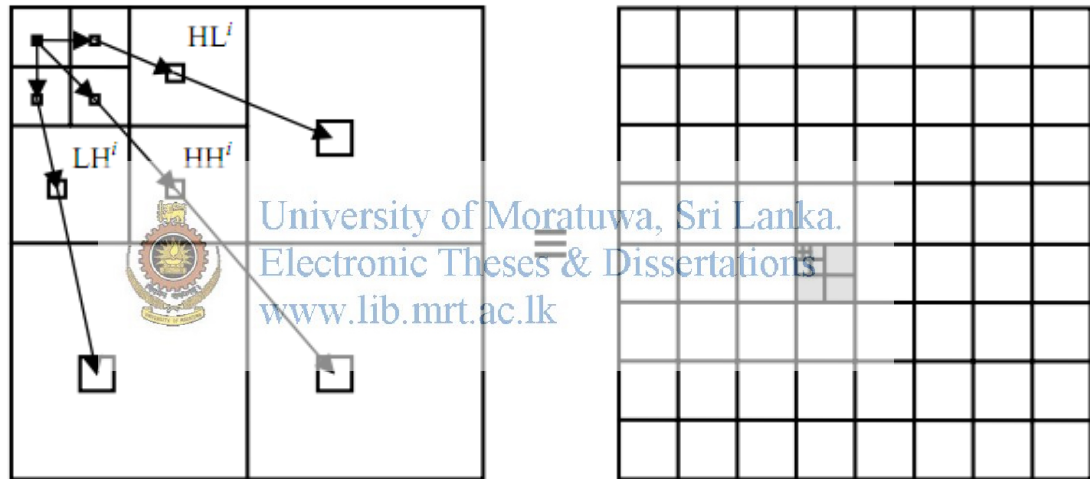


Figure 2.21: DWT block tree organization (Level = 3)

Mathematcal Definition of Wavelet Transform

DWT signal decomposition (Level 1)

Wavelet techniques provide excellent space and frequency energy compaction. Basic idea of DWT for one-dimensional signal has been explained. A signal is split into two parts known as “*High Frequency*” and “*Low Frequency*”. The technique is called **De-composition**. Signal is passing through a number of high pass filters and low pass filters to analyze high frequencies and low frequencies. Suppose $x[n]$ is the original signal, spanning a frequency band of 0 to $\pi rad/s$.

Original signal $x[n]$ is first pass through a halfband high pass filter $g[n]$ and a low pass filter $h[n]$. Then signal is subsampled by two and continues one level of decomposition

and equations are listed in Equation 2.15.

$$Y_{high}[k] = \sum_n x[k].g[2k - n] \quad \text{and} \quad Y_{low}[k] = \sum_n x[k].h[2k - n] \quad (2.15)$$

where $Y_{high}[k]$ and $Y_{low}[k]$ are outputs of highpass and lowpass filters after subsampling by two (Repeat same for each layer).

Those are called **DWT Coefficients** and watermark is added to these components. These DWT coefficients can be used to recover original watermark. Signal reconstruction process is called “*Inverse Discrete Wavelet Transform (IDWT)*”.

DWT Signal Reconstruction (Level 1)

The signals at every level are upsampled by two, passed through synthesis filters $g'[n]$ and $h'[n]$ (highpass and lowpass, respectively), and then added. The analysis and synthesis filters are identical to each other and reconstruction equation becomes (for each layer),

$$X[n] = \sum_n (y_{high}[k].g[-n + 2k] + y_{low}[k].h[-n + 2k]) \quad (2.16)$$

In addition, orthogonality condition on the filters $H(\omega)$ and $G(\omega)$ must hold to prove the IDWT and DWT relationship. Hence,

$$|H(\omega)|^2 + |G(\omega)|^2 = 1$$

where

$$H(\omega) = \sum_n h[n]e^{-jn\omega} \quad \text{and} \quad G(\omega) = \sum_n g[n]e^{-jn\omega} \quad (2.17)$$

2.3.4 Hybrid Watermarking Schemes

Many hybrid watermarking techniques have been proposed due to matter of trade off among advantages of spatial domain and transformed domain techniques. Number of watermarking schemes [20], [26], [60]–[65] were published in years 2004 to 2013 based on DCT, DWT and mix of spatial domain techniques. Qualities of localization of watermark disturbance in spatial domain techniques and good resilience against attacks of transformed domain techniques are considered in hybrid techniques. Although variety of techniques proposed, many of them follow same basic property suggested by Barni et al [20]: they keep trace of spatial/temporal characterization of host signal, while at the same time exploiting richness of frequency interpretation. Among hybrid techniques, those based on block DCT/DFT and those relying on wavelet de-composition of signals have been received a particular attention. However, it is said that hybrid techniques

inherit drawbacks of both approaches as well as advantages. But, still no proven superiority of a set of host features with respect to others has ever been demonstrated as described by Barni & Bartolini 2004 [20]. Hence, digital watermarking in hybrid domain means modifying image in both spatial and spectral specifications.

Block-based Hybrid Method Technique proposed by “Barni & Bartolini” 2004

Block-based DCT is one of most popular choices of image watermarking. Because it is a basic component of image and video compression standards such as JPEG and MPEG. By choosing a framework that matches compression standards, for example 8×8 blocks ideally match JPEG compression to provide least distort to message facing with JPEG compression attacks. In Figure 2.22 pixels that are in light-red color represent intensity of middle frequencies in host image. Authors suggest that middle frequencies are good for carrying watermark data. Rest of the work is to simply performing spread spectrum algorithm in separate smaller blocks.

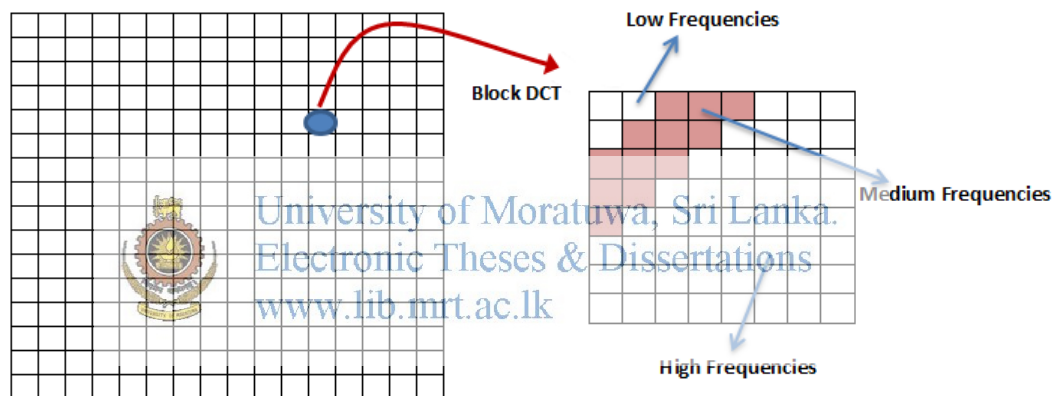


Figure 2.22: Block-based hybrid method proposed by Barni et al [20]

This section presents hybrid watermarking techniques published in last few years. More or less recent researches are combinations of DFT, DWT and DCT concepts and modeled algorithm using mathematics.

A reversible watermarking scheme for 2D-vector data (point coordinates) proposed by Michael et al [61] is widely used in geographical information related applications. This reversible watermarking scheme exploits high correlation among points in same polygon in a map and achieves reversibility of whole scheme by an 8-point integer DCT. Furthermore, this ensures that original 2D-vector data can be watermarked during watermark embedding process and then perfectly restored during watermark extraction process, with the watermark itself. Authors have used an efficient highest frequency coefficient modification technique in integer DCT domain to modulate watermark bit “0” or “1”. In watermark extraction process, this technique is used to determine watermark without using any additional information to indicate. Watermarked map visual distortions

occur due to coefficient modification were alleviated by proposed improved reversible watermarking scheme based on original coefficient modification technique. By combining this improved scheme, embedding capacity is greatly increased whilst watermarking distortion is reduced as compared to original coefficient modification scheme. In fact, a novel distortion limitation scheme presented to control distortion between watermarked 2D-vector data and its original version.

A new zero-knowledge watermark robust to sensitivity attacks is presented by Troncoso-Pastoriza et al [26]. Generic zero-knowledge watermark detectors are vulnerable to sensitivity attacks [66] because of their linear correlation between host image (or asset) features and a given secret sequence. Authors have selected Generalized Gaussian Maximum Likelihood (ML) detector as basis. For detection, they used three algorithms in derived transformed domain and spatial domain, namely Additive Spread Spectrum with a correlation based detector (SS), Spread-Transform Dither Modulation without distortion compensation (ST-DM), and Additive Spread Spectrum with a Generalized Gaussian maximum likelihood detector (GG).

Piper et al [62] proposed a method of obtaining a watermark scheme which provides both resolution and quality scalability. Though there are tradeoffs between resolution and quality scalability, authors have managed to take advantage of human visual system characteristics to increase quality scalability without compromising resolution scalability of watermarked image. They have focused on providing a spread spectrum watermarking algorithm used with JPEG2000 wavelet-based image compression standard for their work. However, designed texture algorithm is applied in DWT domain but applies only a single resolution of multi-resolution wavelet decomposition components.

Another hybrid watermarking scheme is proposed by Schlauweg et al [63] and it has high resilience to transformation attacks. The method is a dual watermarking scheme where two different watermarks are embedded in the same host data to aim at different applications. Also authors claim that this is a new meaning of dual watermarking due to the fact that, all previous approaches have in common that second embedding process slightly affects robustness of firstly embedded bits but their algorithm is not. Original image and watermarked image difference is illustrated in Figure 2.23.

An improved, invariant wavelet discrete cosine transform blind watermarking algorithm against Rotation-and Scaling-and Translation (RST) attacks is proposed by Huang et al [64] based on Xiong's invariant wavelet algorithm. Since this has done in mix domain, it is called as "RSTXIW Domain". In face of geometrical attacks, all shortcomings of almost all digital watermarking algorithms have been exposed. The experiments show that this novel watermarking algorithm is robust against filter, noise and arbitrary RST geometrical attacks, however, sensitive to local crop attacks.



Figure 2.23: Hybrid watermarked still image “Lena” with 18 bits embedded (left), difference to original (right)

Another robust watermarking scheme is implemented in DCT domain with help of spatial domain colour channel techniques proposed by Li Xiaoqiang et al [65]. Authors have proposed a new DCT domain watermarking expressly devised for RGB color images based on diversity technique in communication system. By modifying a subset of block DCT coefficients of each color channel, watermark signal coefficients are hidden in same sequence. Watermark detection process is based upon a combination method by considering the information conveyed by three color channels. Advantage of this scheme is, even particular channel is severely faded, still there is a high probability of recover a reliable estimate of transmitted watermark through other propagation channel. However, due to algorithm design, this novel approach is operating on image luminance only. The main shortcoming of new algorithm is that it fails to resist synchronization attack on blind watermarking. Cox et al [67] also explained that this is a difficult problem.

2.4 Summary of Transform Domain

The transform domain watermarking is composed of;

1. *Image transformation*
2. *Watermarking embedding and*
3. *Watermark recovery*

Based on watermarking application, image transformation can be applied to whole image [54] or to block by block pattern [56]. Generic coefficients modification can be

represented as,

$$I'_i = I_i + \alpha \cdot I_i \cdot W_i \quad (2.18)$$

Where I and I' represent original and watermarked image and W depicts watermark and i represents embedding position and α is watermarking embedding strength factor.

Watermark detection is done by “*Correlation*”. Correlation factor measures similarity between original watermark and extracted watermark of image (robustness is measured). The correlation factor may take values between 0 (random relationship) to 1 (perfect linear relationship). Mathematical expression is given in Equation 2.19

$$C_{(I',W)} = \frac{1}{L} \sum_{i=0}^{L-1} I'_i \cdot W_i \quad (2.19)$$

Where, The correlation $C_{(I',W)}$ between possible attacked image I' and watermark W . Based on threshold T , watermark existence is determined. It can be expressed as:

$$C_{(I',W)} = \begin{cases} \geq T & \text{watermark is present} \\ < T & \text{watermark is not present} \end{cases}$$

Watermark verification is done by quality of watermark extracted. Image quality is verified using Peak Signal to Noise Ratio (PSNR). It should be always greater than one.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

PSNR: The Peak Signal to Noise Ratio is utilized to evaluate image quality. Visual distortion of watermarked image is determined by PSNR from Equation 2.20. The MAX_I is maximum possible pixel value of image. In this thesis, all image represented using 8-bit grayscale mode and $MAX_I = 255$. PSNR is measured in decibels (dB).

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad (2.20)$$

Hence final value can be depicted as:

$$PSNR = 20 \log_{10} \left(\frac{255}{\sqrt{MSE}} \right)$$

MSE: Mean Squared Error (MSE) is one of the earliest tests which were performed to test if two pictures are similar. A function could be simply written according to Equation 2.21.

$$MSE = \frac{1}{mn} \cdot \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \left[I(i, j) - K(i, j) \right]^2 \quad (2.21)$$

Chapter 3

Research Methodology

A review of several surveys [1], [28], [29] and significant number of research studies reveal that there are no generic methods or schemes available for digital watermarking. While, a particular algorithm is suitable for one type of image, the same methodology may not work for other images. The type of image in this context is determined based on perceptual requirements. Majority of digital watermarking schemes are designed to overcome specific issues. In digital watermarking, grayscale image types are widely used. In general, color image (RGB Image) watermarking use “blue color channel” for watermark embedding because of human eye has less sensitivity of detecting blue color in images. Also it is a known fact that transform domain watermarking has been proved better qualities than spatial domain watermarking schemes. Main objective of this dissertation is to come up with an attack success/failure evaluation metric against chosen watermarking schemes depicted in section 3.1 by means of attack set described in section 3.2. This study also reviews several invisible watermarking schemes under different transformation attacks and their different properties which contribute for effectiveness of a watermarking scheme.

Research methodology is organized by selection criteria depicted in this chapter. Initially, five digital watermarking algorithms are selected based on the attack resilience judged by the authors. However, watermark imperceptibility and domain category which is transform domain are mandatory factors for all watermarking algorithms. After that, the attack methods are selected as in section 1.5. Finally, an attack success/failure metric is developed to prove the state of the art of the research proposed.

3.1 Selection of Watermarking Schemes

Five image watermarking algorithms has been selected in transform domain watermarking area. The common property of that image watermarking scheme is robustness

against transformation attacks. In addition, imperceptible property, public source code availability and attack resilience judged by the authors are also considered during the selection. The narrowed down research area is depicted in chapter 2, Figure 2.1.

Selected Algorithms are:

1. Secure spread spectrum watermarking for multimedia [54] - Transform Domain (DCT based)
 - **Cox**, refer to Cox et al research published in 1997.
2. Wavelet-based digital image watermarking [59] - Transform Domain (DWT based)
 - **Wang**, refer to Wang et al research published in 1998.
3. Digital watermarking using multiresolution wavelet decomposition [68] - Transform Domain (DWT based)
 - **Kundur**, refer to Deepa Kundur et al et al research published in 1998.
4. Wavelet transform based watermark for digital images [69] - Transform Domain (DWT based)
 - **Xia**, refer to Xiang-Gen Xia et al research published in 1998.
5. Towards robust and hidden image copyright labeling [70] - Transform Domain (DCT based)
 - **Koch**, refer to Eckhard Koch et al research published in 1995.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

3.2 Selection of Transformation and Geometric Attacks

Transformation and geometric attacks have been depicted in Table 3.1 and each attack has explained in detail under section 1.5. However, attacks are restricted in to certain limits such as target image perceptual quality, image significant areas etc in order to cope with research objectives. Attack simulations are performed using software tools: StirMark [24], [25] and MATLAB R2012 [71].

Image Perceptual Difference Criteria:

A software program developed by Hector Yee and Newman [72] is used to measure perceptual quality between attacked and original images. Authors have described a perceptual image comparison technique which can be used to determine when images are perceptually identical even though they contain imperceptible numerical differences. Furthermore, this technique has proven much utility in the production testing of rendering software.

Table 3.1: Selected Transformation and Geometric Attacks

Attack Type	Summary
Image Cropping	Image is cropped without degrading the image perceptual quality and details. After that, it is reconstructed to original size
Image Rotation	Smaller degree of image rotations are considered along with image cropping
Image Flip	Horizontal flip is considered
Image Resizing	Image Scale up and Scale Down in smaller factors are considered
StirMark JPEG Compression	JPEG quality factors are selected using StirMark tool
StirMark Noise	Noise levels are selected using StirMark tool
StirMark Line and Column Removal	Line/Column of pixels removed at selected rates using StirMark tool
Low pass filtering	Average and Gaussian filters are selected using MATLAB tool

The software tool [72] is designed to use a perceptual error metric method and calculated output of the automated test suite to determine if two images are perceptually identical. In its latest version, the tool allows perceptual thresholds to be tweaked for individual tests or for entire suites of tests. Images that fail the perceptual-metric test are then presented as error images highlighting the pixel difference along with other parameters.

Moreover, perceptual metric component is based on the “visible differences predictor” (VDP) proposed by Scott Daly [73]. Scott used VDP tool to map RGB image color scheme to XYZ. The underline block diagram of the input of two images to the VDP algorithm to create output image of visible differences is shown in Figure 3.1.

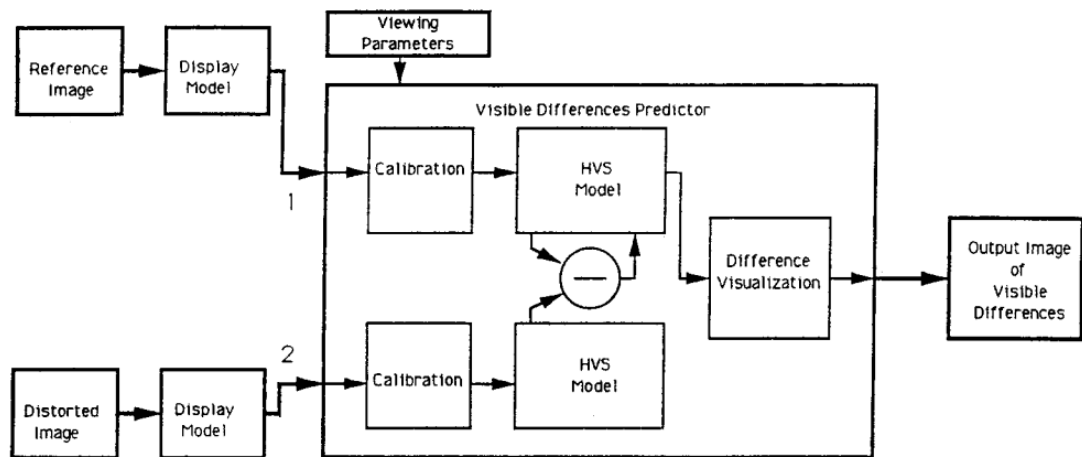


Figure 3.1: Block Diagram of Visible Differences Predictor (Calculation of Visible Differences Only)

Example output:

Command line argument is: “perceptualdiff original-image attacked-image”

Example Output:

Field of view is 45.000000 degrees

Threshold pixels is 100 pixels

The Gamma is 2.200000

The Display's luminance is 100.000000 candela per meter squared

Converting RGB to XYZ

Constructing Laplacian Pyramids

Performing test

PASS: Images are perceptually indistinguishable

0 pixels are different

Transformation Attaks Criteria:

1. **Crop Attack:** In this attack, watermarked image is cropped 10-15% of its original image size from the boarder as illustrated in Figure 3.2. In this way, attacked image has almost similar features compared to its watermarked image. However, the attacked image will be reconstructed with missing parts obtained from the original image in order to perform watermark detection. This methodology is described by Cox et al [54].



Figure 3.2: Cropped Lena Image



Figure 3.3: Cropped Peppers Image

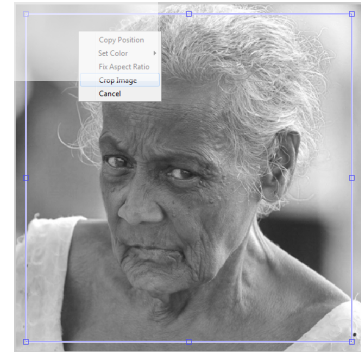


Figure 3.4: Cropped Old-Lady Image

2. **Rotate Attack:** MATLAB tool is used to perform the rotate attack on watermarked images. It is a known fact that the MATLAB function outputs image identical to the source image for smaller angles and images sizes. In this attack, series of tests were conducted and rotate angles were identified while preserving image perceptual quality. Consequently, images were cropped to its original dimensions (E.g. 512x512 pixels) by means of two interpolation methods: nearest and bicubic by same MATLAB function. Numerical values are depicted

in Table 3.2.

Table 3.2: Rotation Angles for Rotate Attack

Interpolation		Angle of Rotation/Degree						
bicubic	0.01	0.02	0.03	0.04	0.060	0.061	0.062	0.063
nearest	0.01	0.05	0.10	0.1201	0.1205	0.1207	-	-

3. **Flip Attack:** Attacked image is generated by mirroring (or Horizontal Flip) the watermarked image. Example is depicted in Figure 3.5.

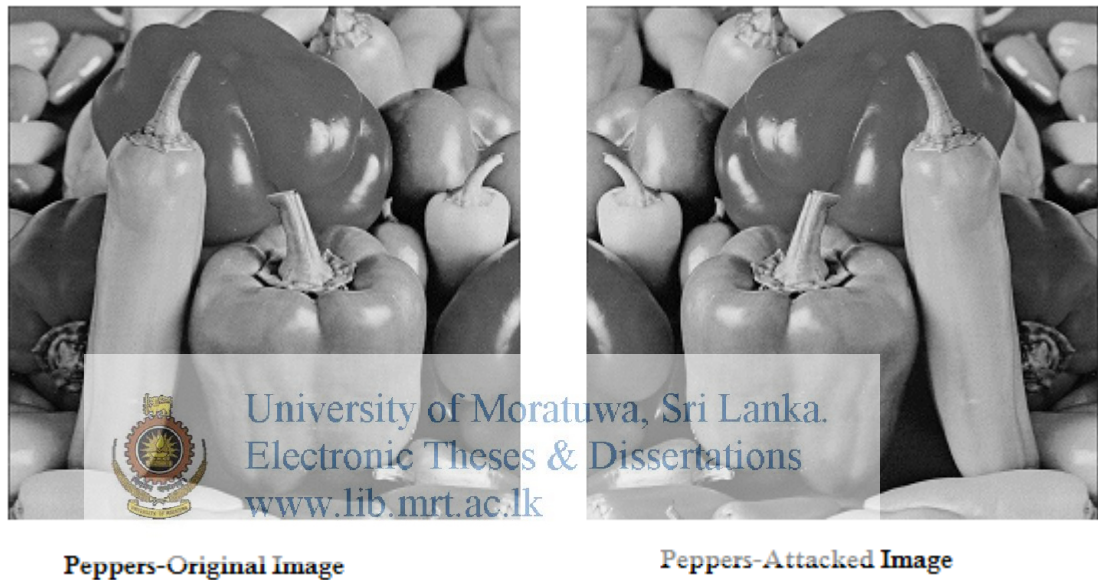


Figure 3.5: Image Flip Attack on Peppers Image

4. **Resize Attack:** Image resize attack is categorized in to two parts. According to the scale factor, watermarked image is Scale up or Scale down during the attack. Scale factors are determined after number of experiments and the resultant image is resized to original dimensions accordingly. Selected scale factors are shown in Table 3.3 and attack is illustrated in Figure 3.6 and Figure 3.7.

Table 3.3: Scale Factors for Resize Attack

Scaling Method	Scaling Factors				
ScaleUp	0.75	0.990	0.997	0.998	0.999
ScaleDown	1.005	1.009	1.15	1.25	1.5



Figure 3.6: Image Resize Attack on Lena Image:ScaleUp



Figure 3.7: Image Resize Attack on OldLady:ScaleDown

5. **JPEG Compression Attack:** This attack is performed using StirMark image processing tool. Compression quality values are determined by observing perceptual quality output. Selected JPEG compression values are shown in Table 3.4. Some compression quality levels are illustrated in Figure 3.8.
6. **Noise Attack:** StirMark image processing software is used in the Noise attack. Noise level is normalised from 0 to 100 and 0 gives identity and 100 a completely

Table 3.4: Quality Levels of JPEG Compression Attack

Image	Quality Level		
Lena	80	90	100
Peppers	80	90	100
OldLady	80	90	100

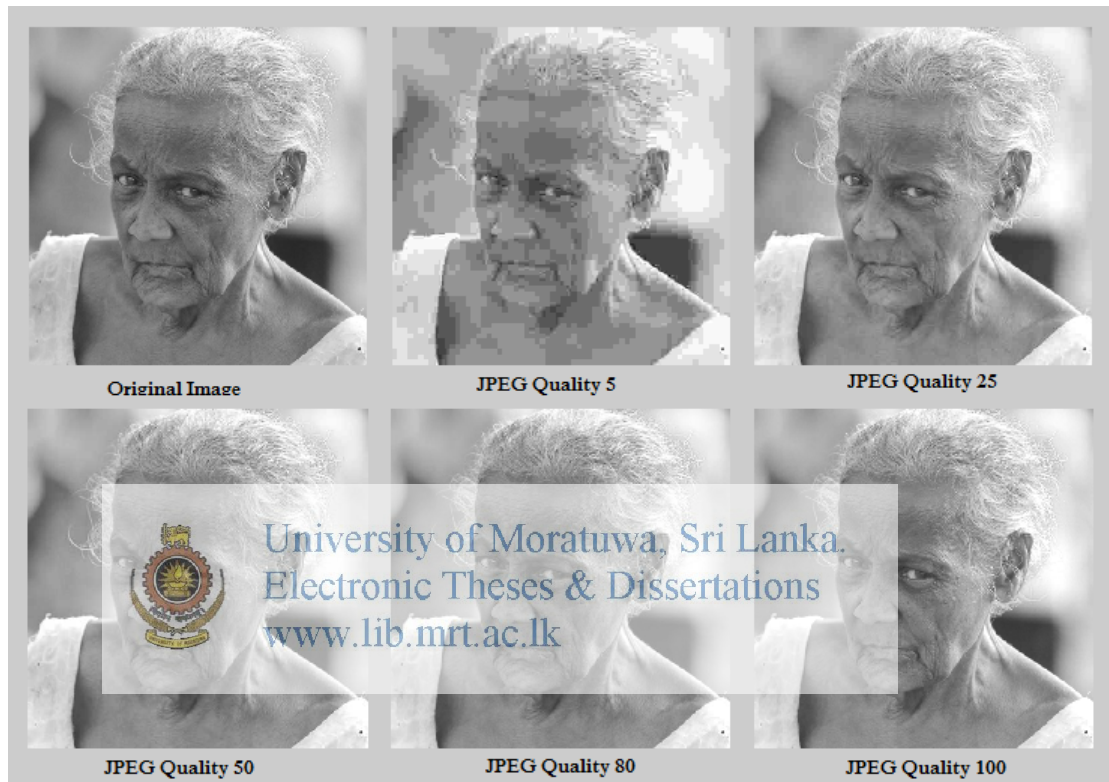


Figure 3.8: Image JPEG Compression StirMark Attack on OldLady

random image. In the experiment, noise levels 1.000, 1.005 and 1.010 are used and some noise levels are depicted in Figure 3.9 to understand the visual differences.

7. **Removal Attack:** This attack also simulated by using StirMark tool. The lines and columns are removed based on a rate. Removal rate one (1) in 90 to 100 is considered for the testing purpose. Some StirMark removal attacked images are illustrated in Figure 3.10 along with histograms. In fact the pixel loss due to this attack is addressed by scaling up the attacked image to original image dimensions during the watermark detection.
8. **Filter Attack:** In this attack, “Lowpass Filter” is used for the evaluation. In Figure 3.11 has different types of filter attacks depicted. Based on the perceptual selection utility program [72], *AVERAGE* and *GAUSSIAN* Filters are used

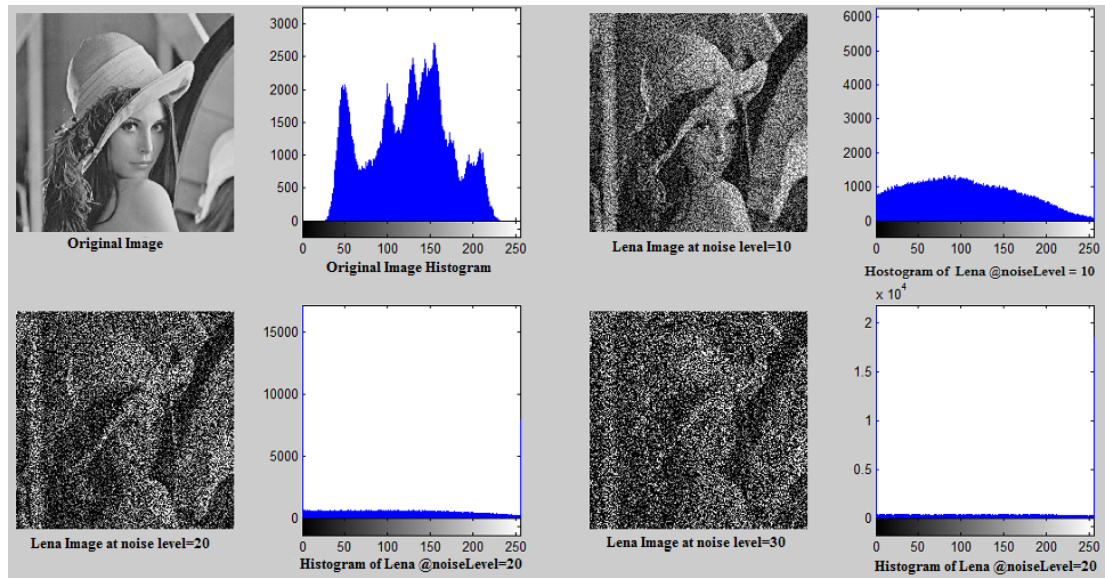


Figure 3.9: Image Noise StirMark Attack on OldLady

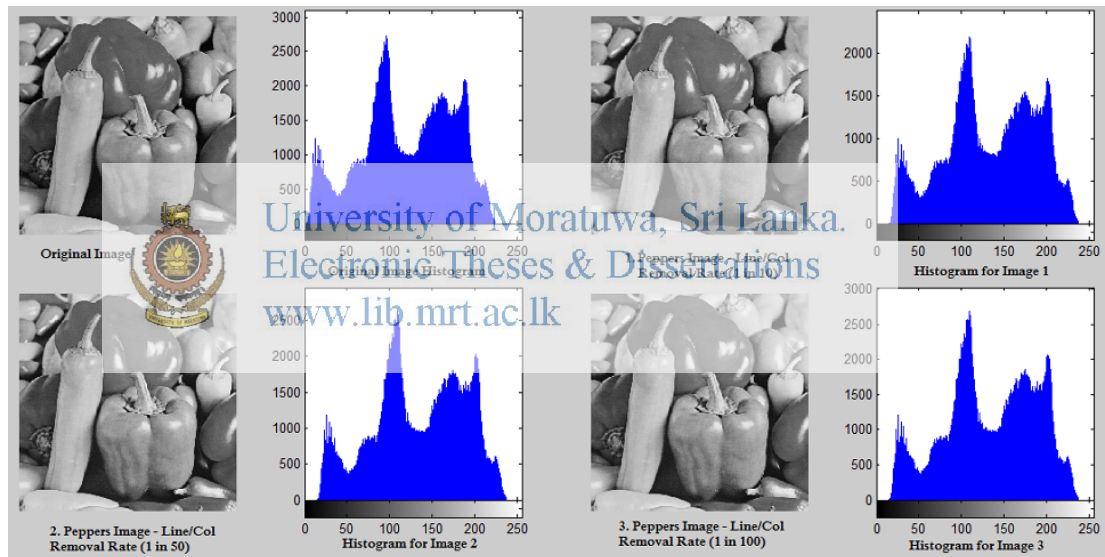


Figure 3.10: Removal Attack on Peppers

in the experiments.

Filter Parameters: MATLAB function - fspecial()

Average Filter:

$h = \text{fspecial}(\text{'average'}, \text{hsize})$ returns an averaging filter h of size hsize . The argument hsize can be a vector specifying the number of rows and columns in h , or it can be a scalar, in which case h is a square matrix. The default value for hsize is $[3 \ 3]$.

Gaussian Filter:

$h = \text{fspecial}(\text{'gaussian'}, \text{hsize}, \text{sigma})$ returns a rotationally symmetric Gaussian lowpass filter of size hsize with standard deviation sigma (positive). hsize can be a vector specifying the number of rows and columns in h , or it can be a scalar, in which case h is a square matrix. The default value for hsize is $[3\ 3]$; the default value for sigma is 0.5.



Figure 3.11: Lowpass Filter Attack on OldLady

3.3 Selection of Images

Three images are selected for image watermarking and evaluation in this dissertation. All images comply NetPBM [74] format and the attacks have been tested with 8-bit grayscale images of size 512×512 in pixels. Two out of those are well known images in watermarking Peppers and Lena. In addition to those two well known images, one image resembling the life of Negombo, Sri Lanka [75] was chosen as the third test image. All selected images are shown in Figure 3.12, Figure 3.13 and Figure 3.14.

Table 3.5: Selected Images for Evaluation

Image(Image Format:PGM)	Description
Lena	Standard grayscale test image
OldLady	Image is taken from Flickr Photo Sharing [75]
Pepeprs	Standard grayscale test image



Figure 3.12: Lena



Figure 3.13: Peppers



Figure 3.14: OldLady

3.4 Development of an Attack Success/Failure Evaluation Metric

Final outcome of this dissertation is to come up with an attack success/failure evaluation metric against selected watermarking schemes and transformation attacks by conducting a series of experiments. Thesis primary objective is to determine criteria explained in this section.

If there is a legitimate owner for the image, *success* of image attack would be determined if recovery of the watermark is denied given the fact that attacked image lies within the perceptual range. Otherwise it would be depicted as a *failure* of image attack.

Another possible variation of previously stated objective can be expressed as a scenario where the attacker's intention is to impose the ownership of the subjected image. Hence, preserving the image perceptual quality, if a person can carry out the attack while retaining the watermark and at the same time the owner cannot prove that the watermark's existence. A typical example can be observed from these test results is "*Image Flip Attack*" Figure 3.5. In addition some algorithms are failing to detect the watermark for some attacks executed.

Example: Watermark detection confidence level is very low with StirMark Noise attack and Removal attack on Kunchin [68] watermarking scheme.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Chapter 4

Testing and Evaluation

According to the proposed methodology explained in the chapter 3, different attacks are performed on two standard images and one custom selected image [75] with respect to five watermarking schemes built from the source [76]. Test results are illustrated in graphs and categorized by each attack name performed on each image.

4.1 Post Attack Evaluation on Watermarking Schemes

This section represents the watermark existence of each algorithm after image attack performed on all three images of Figure 3.12, Figure 3.13 and Figure 3.14. Confidence value is recorded after running the watermark detection program and output values are plotted against the watermarking schemes mentioned in section 3.1. In some instances, confidence values were plotted against attack type properties (*Example: Rotate attack has set of angles of rotation*) of each watermarking schemes as well.

For each watermarking scheme, watermark embedding is performed using a signature file and it is parsed to obtain the particular watermark sequence and the embedding strength. The signature file is generated randomly for watermark length ‘n’ of each watermarking algorithm with α factor. The watermarking capacity does not depend on a particular algorithm used in the watermark. The ‘n’ is a constant value for all tests were performed in the thesis and the embedding strength factor (α) is taken as 0.1. The reason for that is to keep perceptual quality of the image after watermarking and it is observed that, when increasing the embedding strength conversely the perceptual quality is lost.

Finally, the extracted signature of the watermarked image is compared with original signature. The output value is called “*Correlation Factor*” or “*Confidence Value*”. Watermark is detected if output values around ‘1’ and values around ‘0’ indicate that

the watermark is not detected. Sometimes the value could be a negative depending on the algorithm you choose. However, it is necessary to observe the output of the detector for many different keys to justify a reasonable threshold for detection. As per Peter Meerwald [76] and Petitcolas et al [77]–[79] research studies, a good value to go with this evaluation is ‘0.2’. Hence the watermark is detected if the correlation factor is < 0.2 .

Watermark Detection after “Crop Attack”

Images are manipulated as shown in the crop attack in section 3.2. Watermark detection of all five algorithms are illustrated in Figure 4.1. By inspection of confidence values, all algorithms show resilience to crop attack.

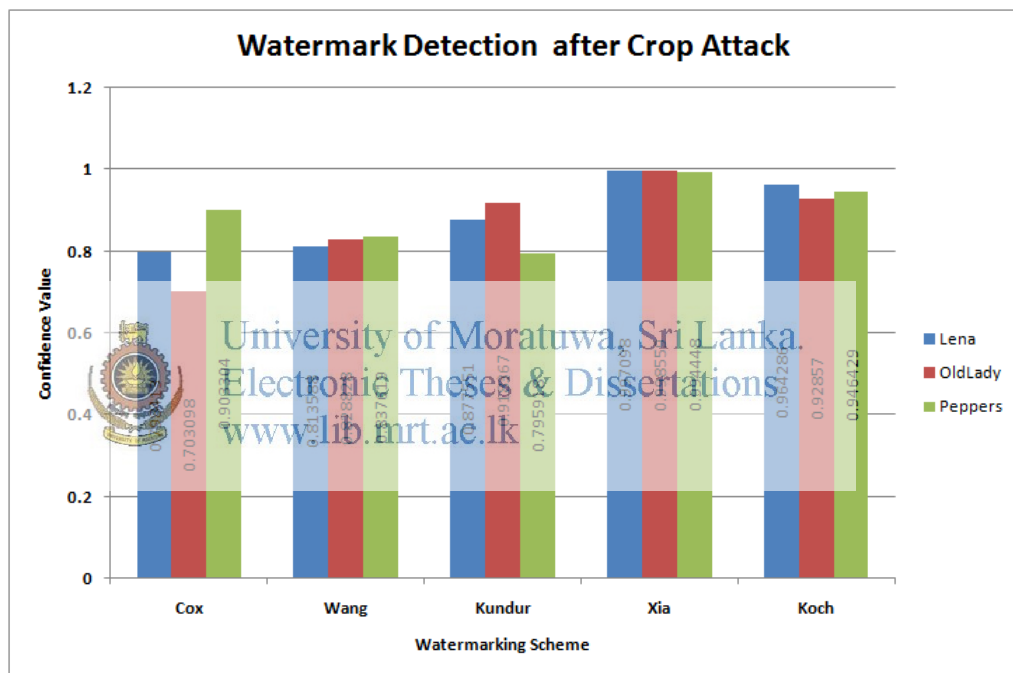


Figure 4.1: Watermark Detection after Crop Attack

Watermark Detection after “Rotate Attack”

Rotate attack is a “combined attack”, because it performs both rotate and crop techniques. When a rotation happens, it uses either “nearest” or “bicubic” interpolation method and perform the crop at the end. The crop technique outputs image same size as the original image. These two graphs Figure 4.2 and Figure 4.3 represent watermark detection for COX’s scheme.

Rotate attack is performed on each watermarking scheme and results are posted in Appendix A section 5.2 for further reference. Apart from COX’s scheme, it is observed

that there are certain occurrences where the watermark detector failed to determine the watermark existence after rotate attack. Rotate angle values were selected by performing series of experiments while preserving the perceptual visibility of the watermarked image. It is observed that when the angle value increases the confidence value is still high but perceptual image difference is very low. This happens due to smallest degree changes and they are not affecting the number of pixels in an image. Furthermore, rounding off keeps the pixels and intensity levels unchanged while using MATLAB rotate function.

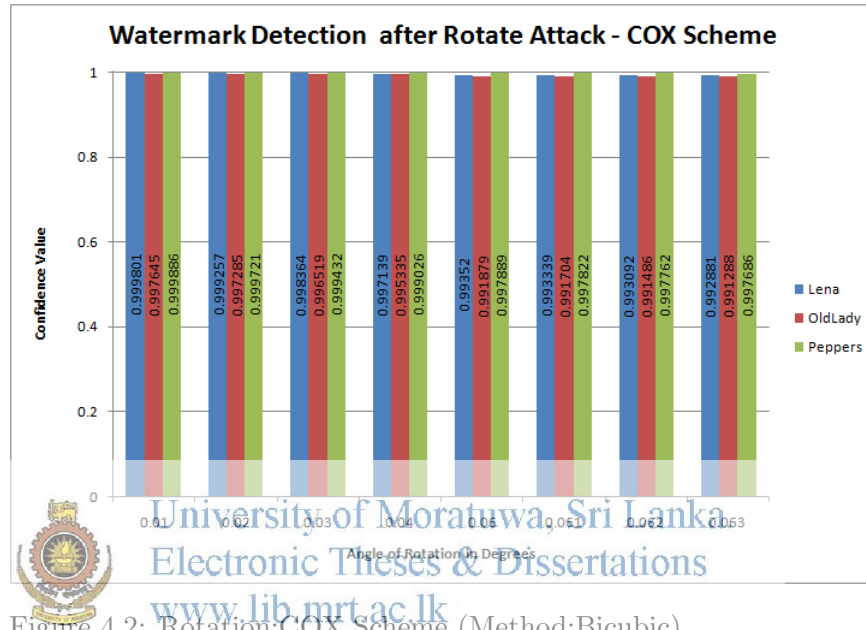


Figure 4.2: Rotation:COX Scheme (Method:Bicubic)

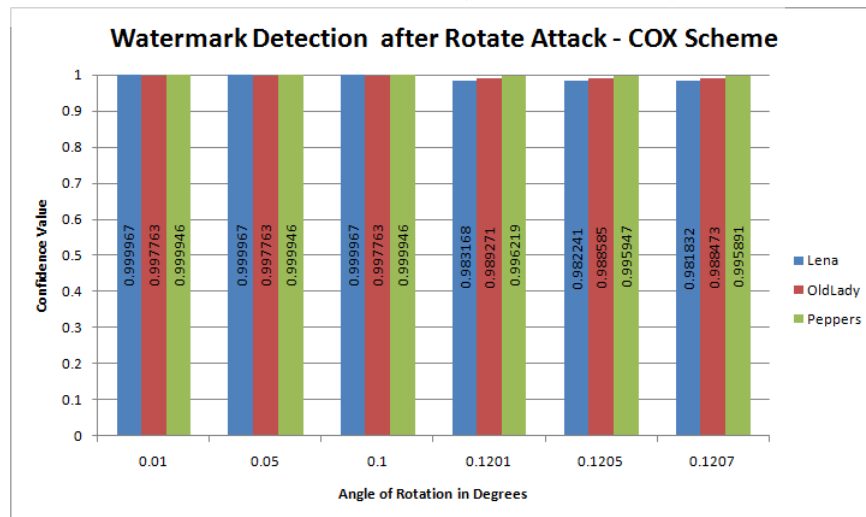


Figure 4.3: Rotation:COX Scheme (Method:Nearest)

Watermark Detection after “Flip Attack”

Basically, flip attack generates the mirror of original image. But none of watermarking schemes were able to detect the watermark after this image manipulation attack. Obtained results are depicted in Figure 4.4. This falls into a scenario where the attacker’s intention is to impose the ownership of the subjected image. Hence it is called an attack success case. Because attacker is preserving the image perceptual quality and carry out the attack while retaining watermark at the same time. Then the owner cannot prove that original watermark’s existence.

It is a known fact that, the watermarking algorithms are not properly working if the original image and watermarked image are out of synchronization. The “*Out of Synchronization*” in this context is, watermark detection algorithm cannot extract the hidden information in the watermarked image. There are number of attack types which output some amount of degradation and out of synchronization between watermarked and original image such as StirMark. But flip attack is different as it produces out of synchronized image with no perceptual degradation. Flipping attack properties are:

1. Resultant image is out of synchronized and hard to detect watermark for most of algorithms
2. No perceptual degradation
3. Without knowing the original images it cannot be judged whether the image was flipped or not
4. Perform a flip attack is a trivial operation on digital image processing

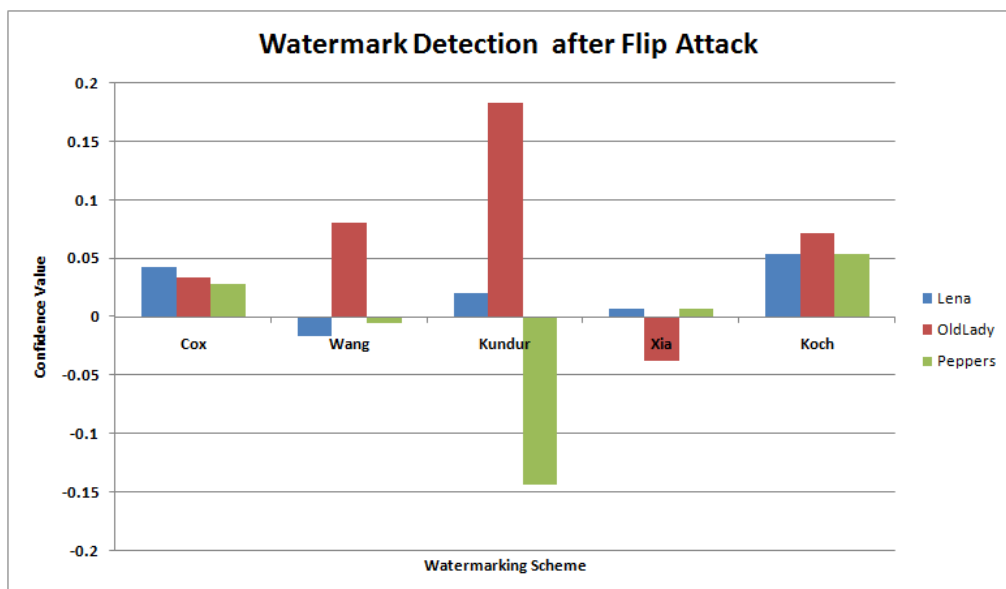


Figure 4.4: Watermark Detection after Flip Attack

It is observed that in Figure 4.4, all confidence values are around '0'. This means watermark is not detected as per the definition of the watermark detection threshold explained in section 4.1. Hence negative values highlighted in some instances during test depict that watermark is never recovered.

Watermark Detection after “Resize Attack”

Watermark resize attack is treated in two different ways. Two factors are defined for resize attack: scale down and scale up. These values are obtained by performing significant number of experiments while satisfying image criteria of perceptual validation. For instance, it is noticed that some watermarking schemes were failed to detect the watermark whereas some performed well during watermark detection. In Figure 4.5 and Figure 4.6 show WANG’s watermark detection and the rest is illustrated at Appendix B section 5.2.

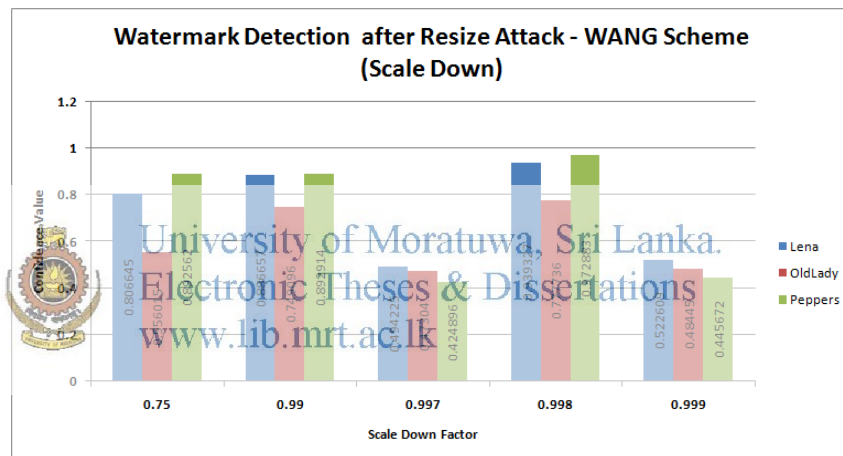


Figure 4.5: Resize:WANG Scheme (Scale Down)

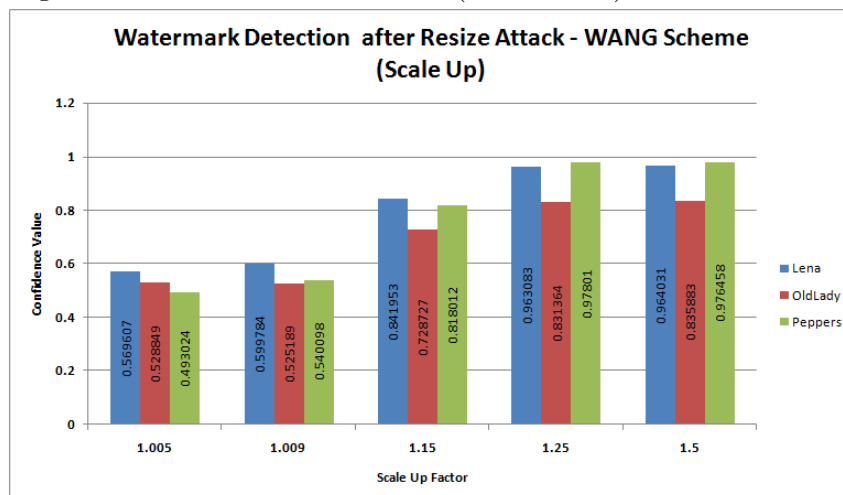


Figure 4.6: Resize:WANG Scheme (Scale Up)

Image resize has been tested using MATLAB tool. Matlab resize function uses bicubic interpolation during the scale down and up. In Matlab bicubic interpolation, the output image may have some values slightly outside the range of pixel values in input image. Kundur's scheme shows such confidence value differences when resize attack has applied. The behavior is not further investigated on specific ranges as it deviates from the primary objective of this research.

Watermark Detection after “JPEG Compression Attack”

JPEG stands for “Joint Photographic Experts Group” which is a standardization committee and they invented the JPEG compression algorithm. JPEG is a lossy compression algorithm and it reduces image file size naturally. Human eye is sensitive to changes in hue (chrominance) than changes in luminance (brightness). Hence brightness is more important than chrominance; the algorithm retains more of the luminance in the compressed file to achieve the file compression mentioned earlier.

This attack is processed using StirMark image tool. As per selected jpeg quality levels, it is observed that selected watermarking schemes are robust to jpeg attacks. However, Kundur's scheme is failed to prove watermark existence for some JPEG quality levels depicted in Figure 4.8 and Koch's scheme has illustrated absolute resistance to all jpeg quality levels shown in Figure 4.7. Further experiment results are posted in Appendix C section 5.2 for reference.

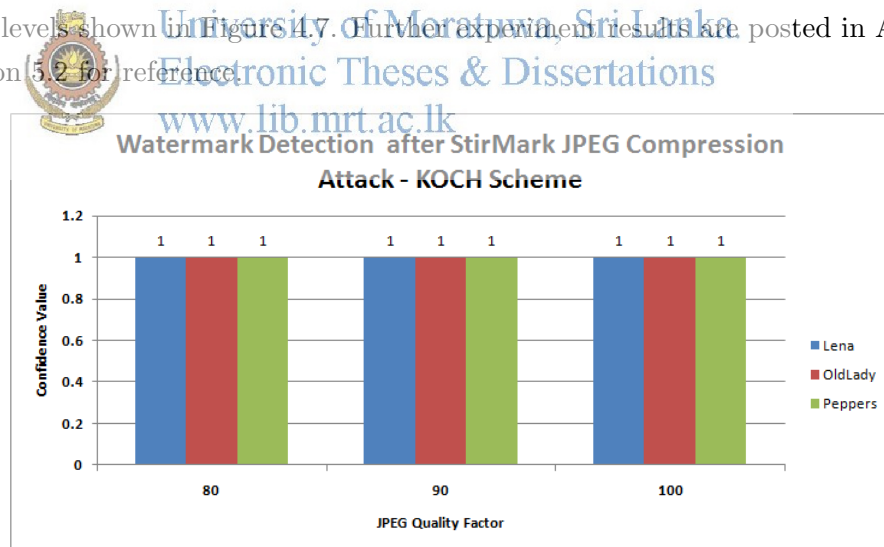


Figure 4.7: JPEG Compression:Koch Scheme

StirMark JPEG attack on Koch's and Kudur's schemes are totally different but only at JPEG quality factor 100 shows valid confidence value for both schemes. The JPEG quality factors 80 and 90 made watermark undetectable for kundur's scheme. This because of image is subjected to JPEG compression and it causes output image to lose data. It is clear that, the watermark detection is also depending on the watermarking

algorithm as well. The reason for all images has same perceptual similarity with original image is jpeg compression scheme rounding off on pixel intensity calculations.

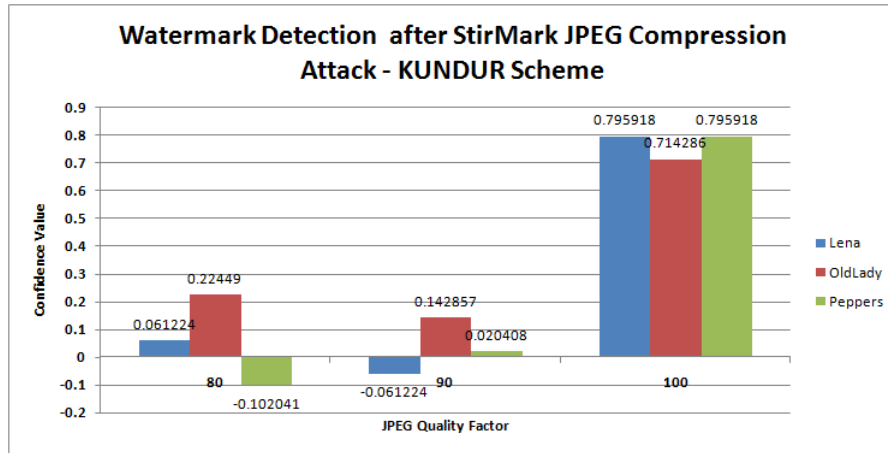


Figure 4.8: JPEG Compression:KUNDUR Scheme

Watermark Detection after “Noise Attack”

Noise levels are defined from noise attack criteria in section 3.2. Apart from Kundur’s watermarking scheme, all the other schemes are resilient to this attack. StirMark software tool is used to produce noise attacks upon each image.

Appendix D in section 5.2 shows the results for remaining watermarking algorithms. In this section, Figure 4.9 and Figure 4.10 represents noise attack on Cox’s and Kundur’s schemes respectively.

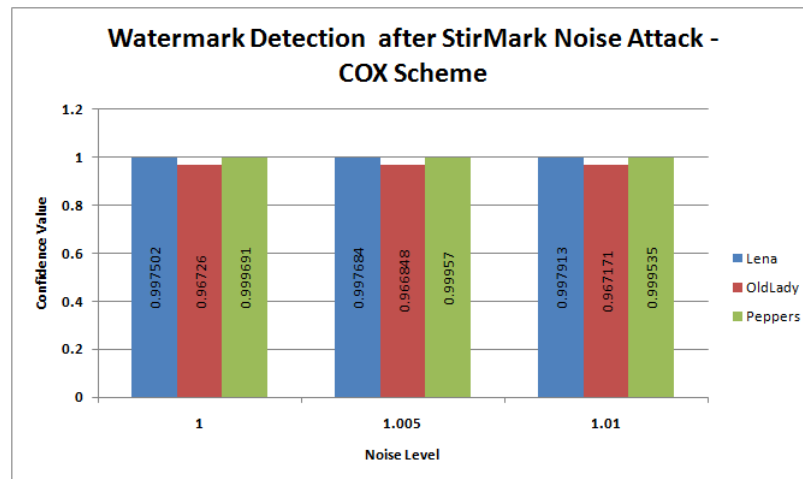


Figure 4.9: Noise Attack:COX Scheme

StirMark has its own proprietary noise generation algorithm. When noise level is increased image quality degrades severely. Only smallest noise variations are considered

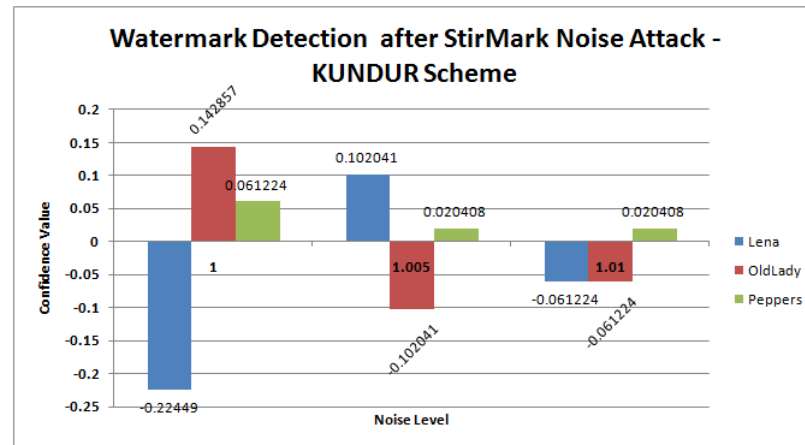


Figure 4.10: Noise Attack:KUNDUR Scheme

in the thesis. According to Figure 4.10, Kundur’s scheme is failed to detect watermark on each noise level but all other watermarking algorithms were succeeded at same noise levels. Since all confidence values are around ‘0’, the watermark is not detected.

Watermark Detection after “Removal Attack”

StirMark image processing tool is used to simulate removal attack. However, this attack has scaling method applied in order to adjust watermarked image size same as original image. This also a combined attack and similar kind of situation is observed in section 4.1. Appendix E in section 5.2 contains the rest of the watermark detection results.

As illustrated in Figure 4.11 and Figure 4.12, it is observed that many algorithms show low detection confidence values for this attack.

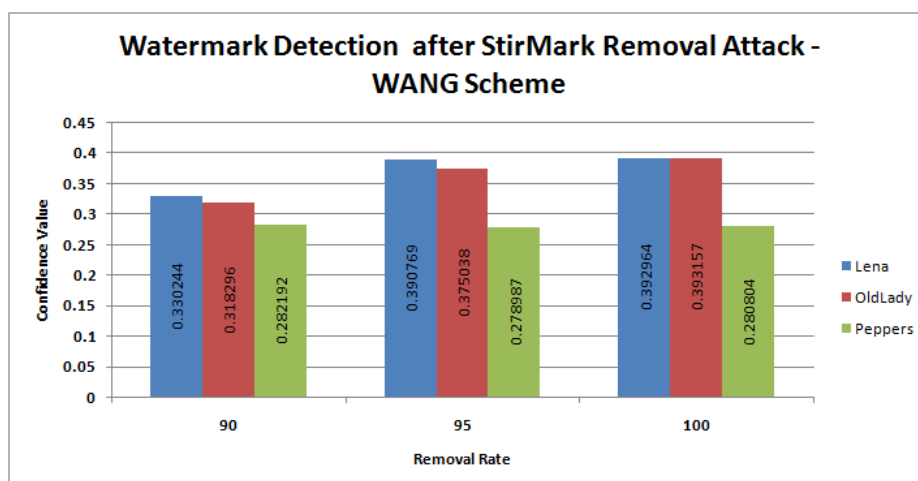


Figure 4.11: Removal Attack:WANG Scheme

StirMark removal attack is designed in a way that rate of row and column removal one in ‘x’ amount of pixel lines. Concealing perceptual quality between attacked and original images, removal rates have been decided. All watermarking schemes were resilient to this attack except Kundur’s scheme. This anomaly has been detected in other attacks as well. The negative confidence values are also possible but still output values are near ‘0’. As per conditions stated in watermark detection threshold, it claims that watermark is not detected. Further investigation was not conducted as focus here is to detect watermark existence under various transformation attacks.

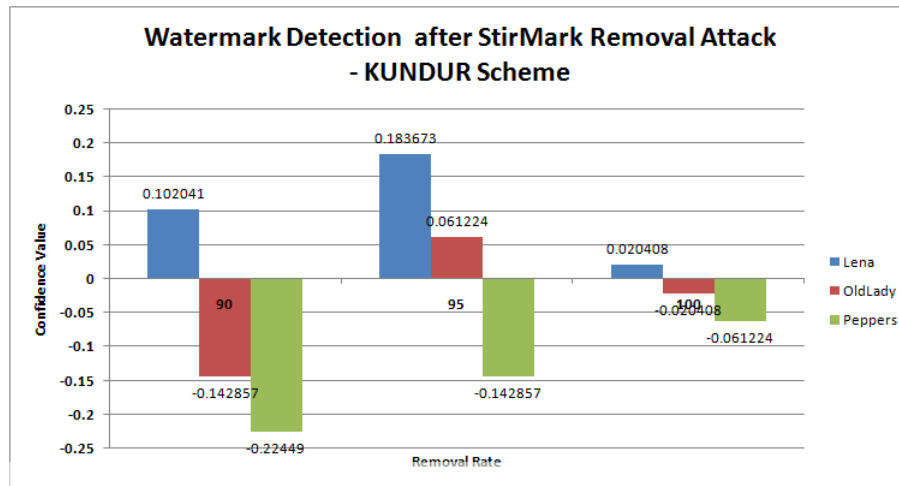


Figure 4.12: Removal Attack:KUNDUR Scheme

Watermark Detection after “Filter Attack”

Based on observed results, all watermarking schemes are resilient to “gaussian” filter attacks but “average” filter attacks. In Figure 4.13 and Figure 4.14 are shown confidence values for Xia’s and Kundur’s watermarking schemes. Remaining results for other schemes are posted at Appendix F section 5.2.

When an image is subjected to a similar low-pass filter attack, it undergoes a smoothen operation of image pixels. The reason behind that is “low pass filter” gives the average of ‘even’ and ‘odd’ (the consecutive) pixels which outputs a smoother low pass signal. Matlab filter specifications for an averaging filter is [3 3] and gaussian filter is [3 3] with standard deviation sigma equal to 0.5. The Kundur’s algorithm is not tested using such low-pass filter attacks as explained in [68].

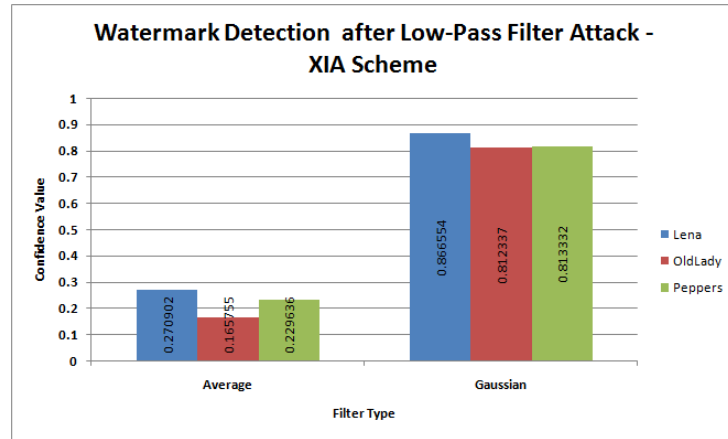


Figure 4.13: Filter Attack:XIA Scheme

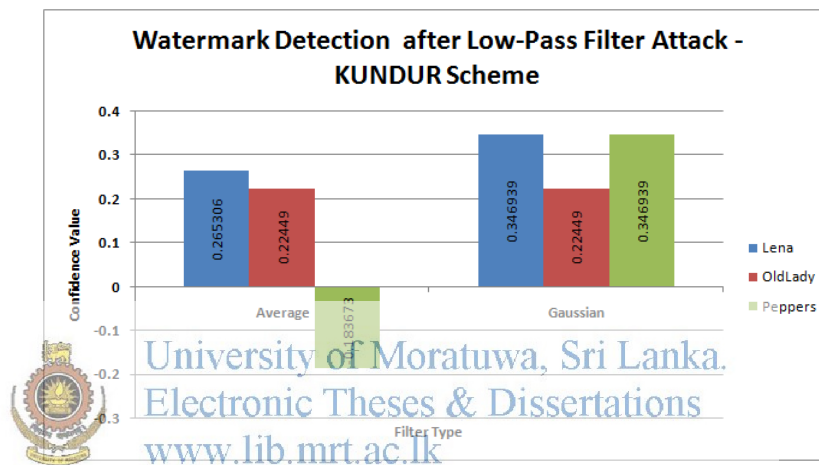


Figure 4.14: Filter Attack:KUNDUR Scheme

4.2 Summary of Test and Evaluation

In this evaluation, it is observed that Kundur’s watermarking is more vulnerable to significant number of transformation attacks. Furthermore, Xia’s watermarking scheme is not resilient for “filter attack” when the filter type is “average”. However, all watermarking schemes have failed to resist “flip attack” as shown in section 4.1. These attacked images are carefully selected along with subjected attacks while maintaining perceptual quality of attacked image against original image. Next step is to determine these attack failures further by distortion analysis. It is illustrated based on PSNR (Peak Signal-to-Noise Ratio) and Histogram comparison of original, watermarked images against attacked images. Analyzed results are depicted in chapter 5.

Chapter 5

Analysis and Conclusion

Digital watermarking schemes are introduced in order to resist transformation attacks and/or to resolve ownership proofs over digital objects. Apart from ownership proof cases, majority of watermarking schemes are not able to withstand against transformation attacks as they do not possess the property of robustness. In this chapter, image Histograms and PSNR values are analyzed under section 5.1 based upon key anomalies which encountered during watermark detection process.

5.1 Distortion Analysis



University of Moratuwa, Sri Lanka.
E-theses & Dissertations
www.lib.mrt.ac.lk

In this section, image Histogram and PSNR are compared and illustrated at each key incidents observed in test results.

- Analysis of Flip Attack Results: All watermarking schemes were failed to detect watermark
- Analysis of FilterAttack Results: Xia and Kundur Schemes were failed to detect watermark (Filter Type=Average)
- Analysis of JPEGCompression Attack: Kundur Scheme is failed to detect watermark (JPEG Quality=80,90)
- Analysis of Noise Attack: Kundur Scheme is failed to detect watermark
- Analysis of Removal Attack: Kundur Scheme is failed to detect watermark
- Analysis of Resize ScaleDown/ScaleUp Attack: Kundur Scheme is failed to detect watermark on certain scale factors
- Analysis of Rotate Attack: Kundur Scheme is failed to detect watermark at certain rotation angles (Method: Bicubic)

Histogram Analysis of Attack Success Incidents

In a grayscale image, a histogram represents tonal variation values per each pixel of image. The horizontal axis of graph represents “tonal variations”, while vertical axis represents “number of pixels” in that particular tone.

Analysis of Flip Attack Results:

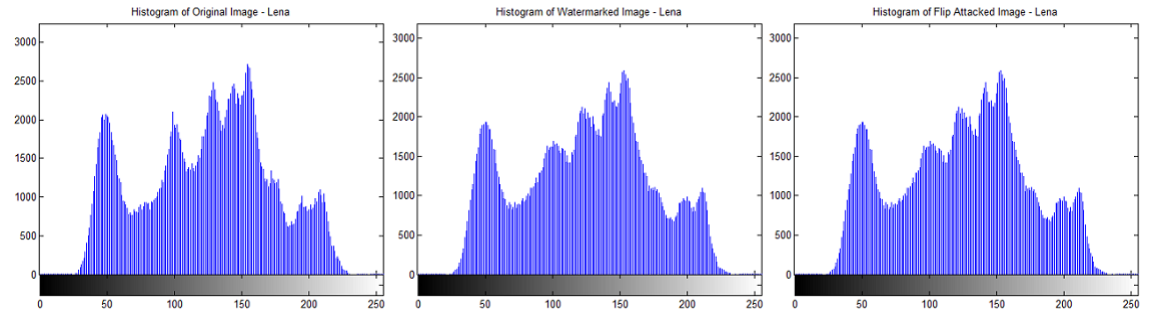


Figure 5.1: Histogram Comparison of Lena (Flip Attack)

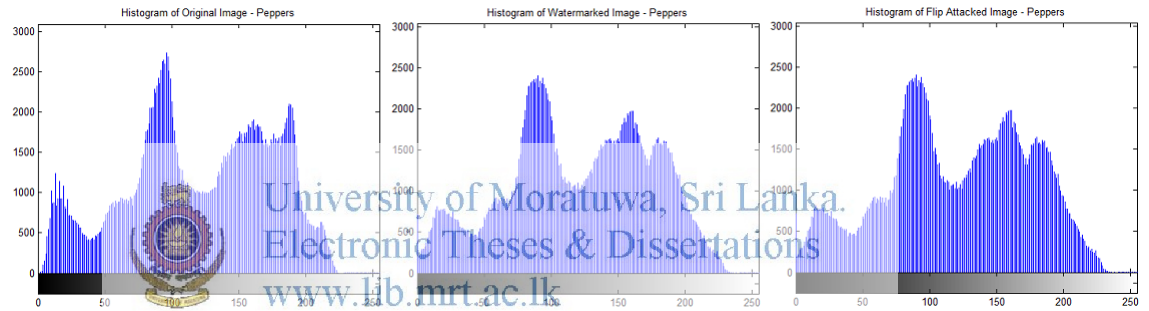


Figure 5.2: Histogram Comparison of Peppers (Flip Attack)

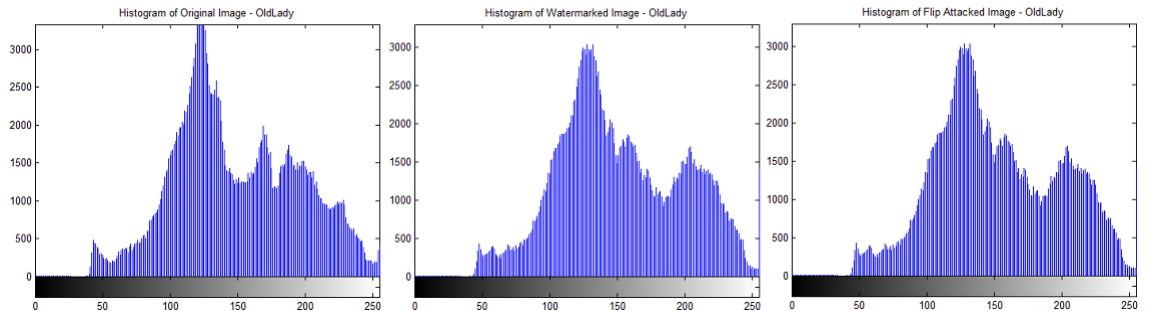


Figure 5.3: Histogram Comparison of OldLady (Flip Attack)

Histograms of flip attack show similar shape for original, watermarked and attacked images. Histogram of Figure 5.1 represents COX’s watermarking scheme for Lena image.

Analysis of FilterAttack Results:

It is observed that histograms are very much similar depicted in Figure 5.4 for Lena image. It is also noticed that perceptual difference of attacked image and original image are almost negligible. Hence, perceptual quality test is passed. However, watermark detection is denied for both attacks. According to Kundur's et al [68], authors have not tested watermarking algorithm for "average" filter type but median and mean filters which come under Spatial Filter category. In Xia's watermarking scheme, authors have not consider filtering attacks in their research [69].

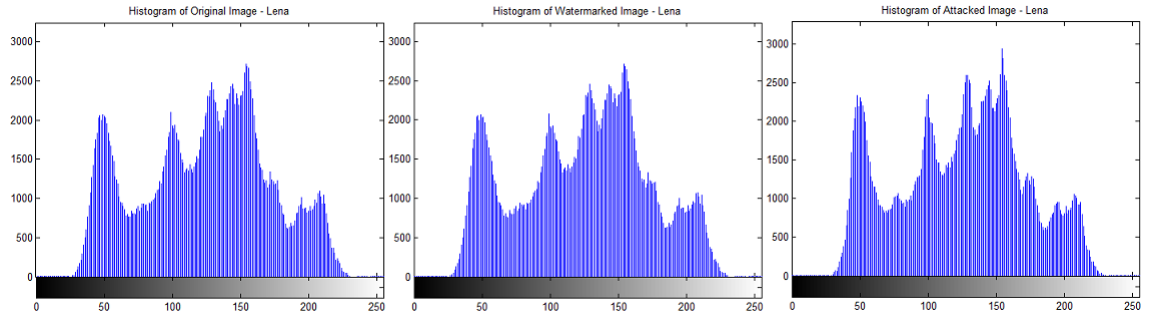


Figure 5.4: Histogram Comparison of Lena:Kundur (Filter Attack)

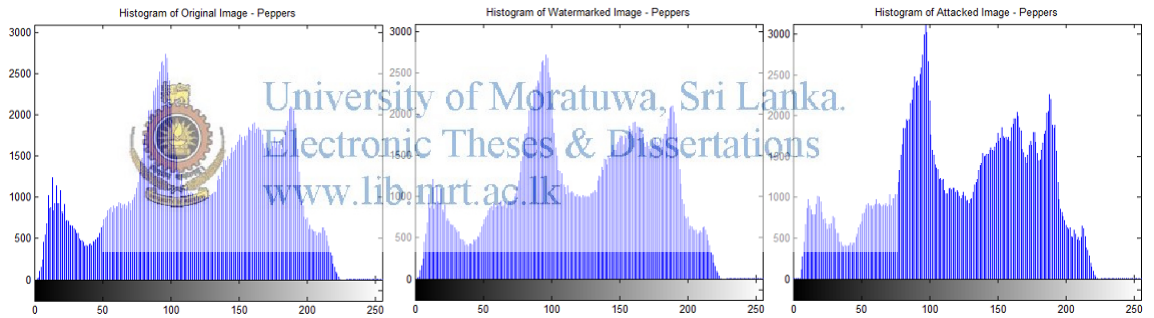


Figure 5.5: Histogram Comparison of Peppers:Kundur (Filter Attack)

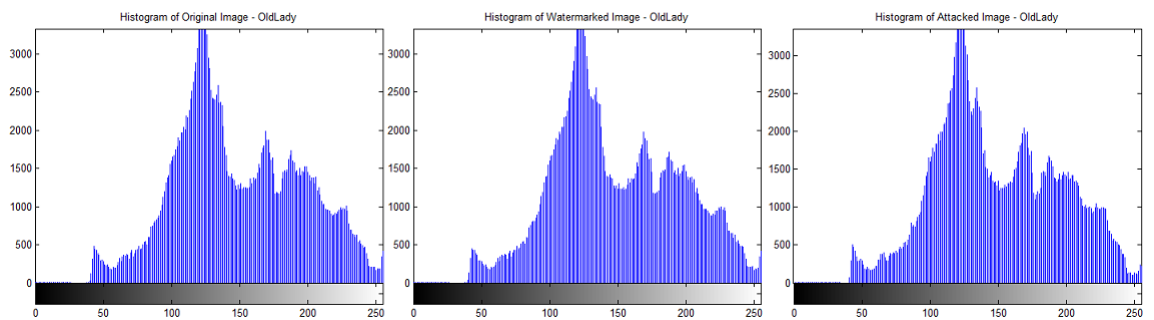


Figure 5.6: Histogram Comparison of OldLady:Kundur (Filter Attack)

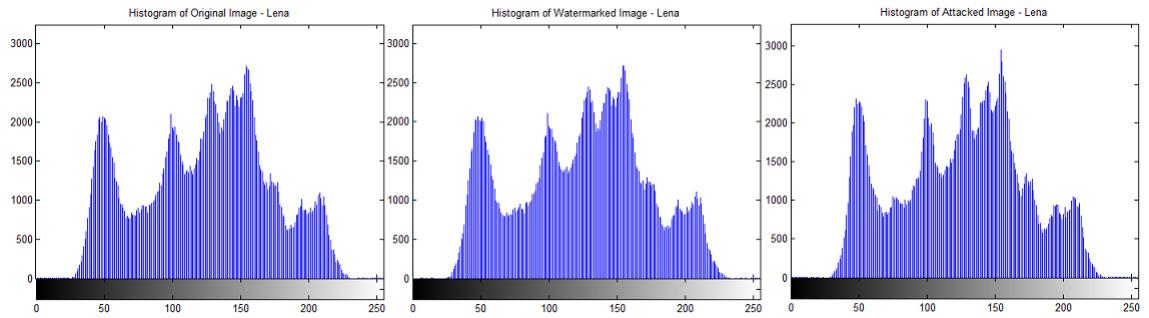


Figure 5.7: Histogram Comparison of Lena: Xia (Filter Attack)

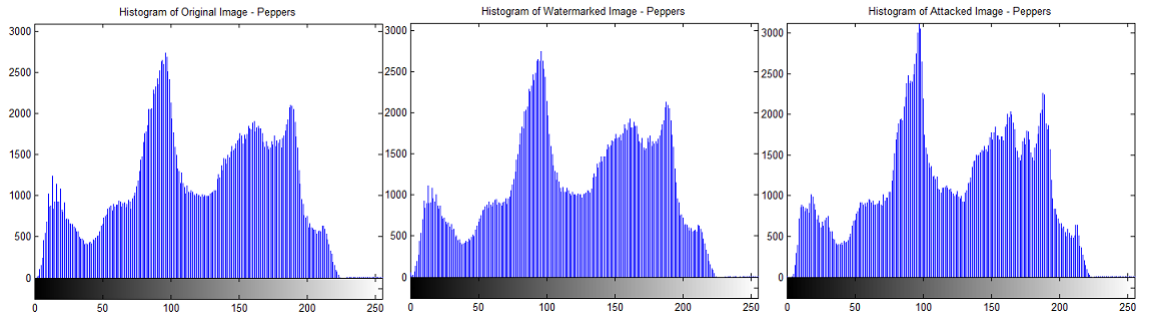


Figure 5.8: Histogram Comparison of Peppers: Xia (Filter Attack)

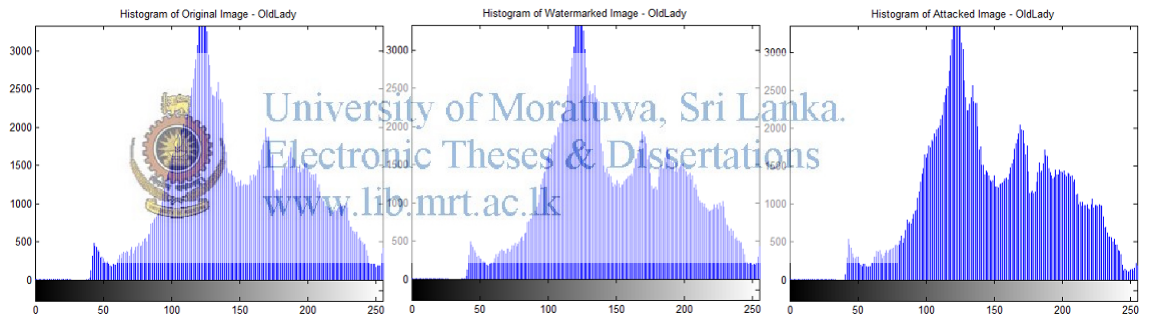


Figure 5.9: Histogram Comparison of OldLady: Xia (Filter Attack)

Analysis of JPEGCompression Attack Results:

In this context, original image histogram is compared with attacked image histogram with JPEG quality factors 90 and 100. Kundur's scheme is vulnerable to JPEG Stir-Mark attack compared with other watermarking schemes. It is observed that perceptual quality of attacked images shows high quality compared to original image. Furthermore, histograms are likely equal as shown in Figure 5.10, Figure 5.11 and Figure 5.12.

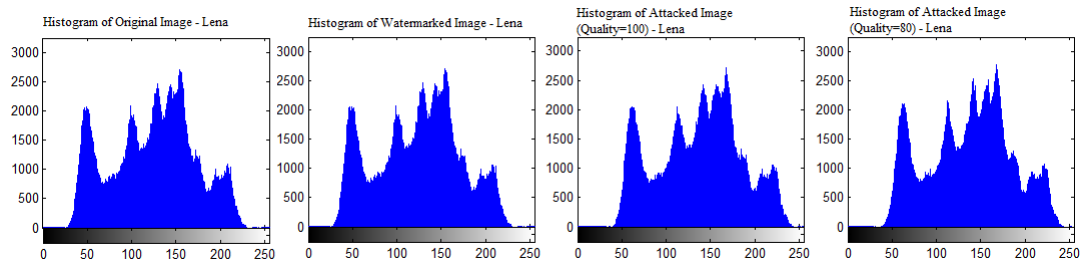


Figure 5.10: Histogram Comparison of Lena:Kundur (StirMark JPEG Attack)

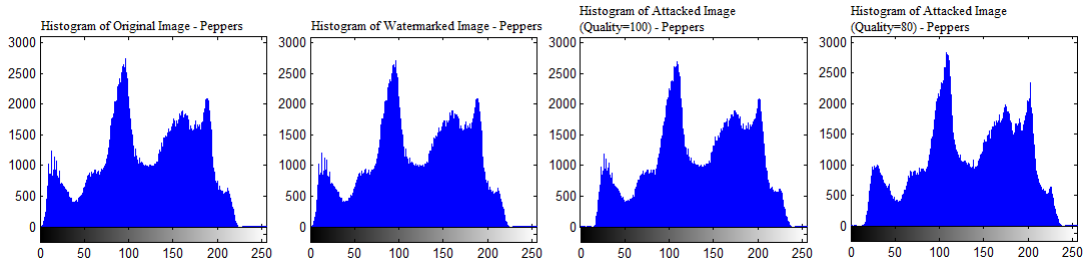


Figure 5.11: Histogram Comparison of Peppers:Kundur (StirMark JPEG Attack)

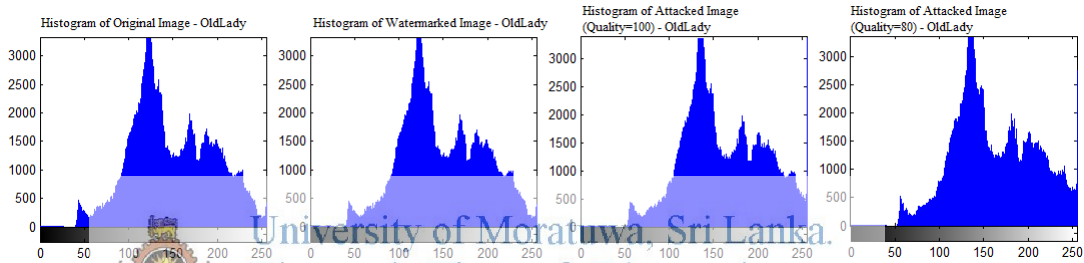


Figure 5.12: Histogram Comparison of OldLady:Kundur (StirMark JPEG Attack)

Analysis of Analysis of Noise Attack Results:

StirMark image processing tool is used to perform noise attack. Out of selected watermarking schemes, kundur's scheme has shown low resilience during this attack. Hence, it can be concluded that watermark detection has failed. However perceptual quality of attacked image and original image are same. Because, selected noise levels are too small and number of pixels changed is almost same. Histograms also prove same fact which are illustrated in Figure 5.13, Figure 5.14 and Figure 5.15.

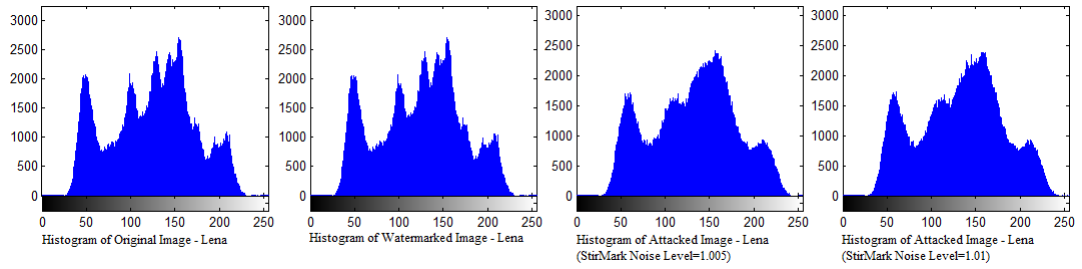


Figure 5.13: Histogram Comparison of Lena:Kundur (StirMark Noise Attack)

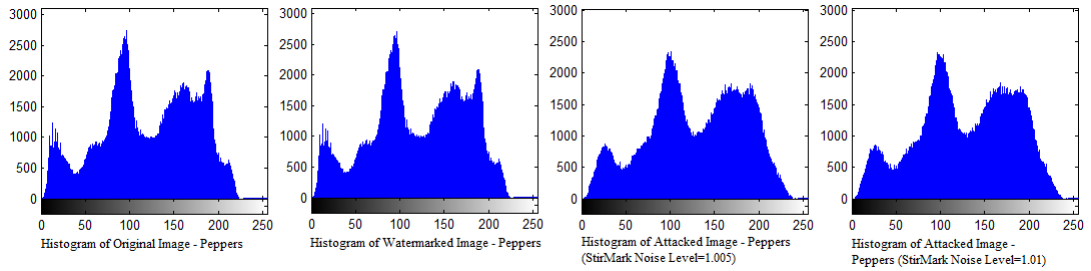


Figure 5.14: Histogram Comparison of Peppers:Kundur (StirMark Noise Attack)

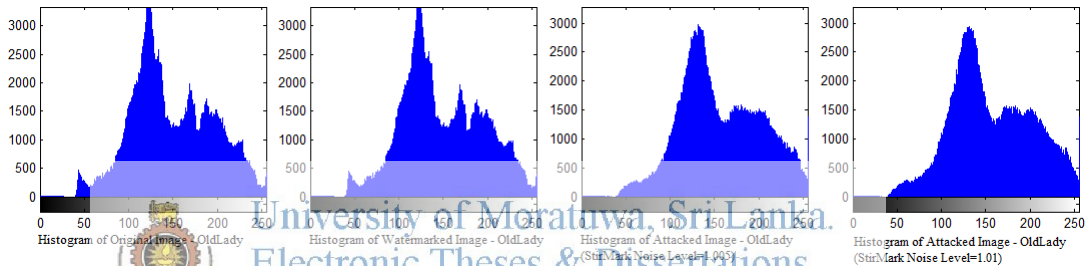


Figure 5.15: Histogram Comparison of OldLady:Kundur (StirMark Noise Attack)

Analysis of Removal Attack Results:

Removal attack is generated using StirMark image processing tool. Every other watermarking scheme was resilient during removal attack but kundur's scheme. Example of OldLady image histogram is shown in Figure 5.16 and histograms for other images are shown in Appendix G section 5.2. Perceptual image comparison and histograms justify same conclusion in this attack incident too.

Analysis of Resize ScaleDown/ScaleUp Attack:

Scale down attack is performed by MALAB tool. It is observed that Kundur's scheme is vulnerable to this attack. The image which subjected to scale down attack is perceptually similar to original. Furthermore, histogram diagrams also prove the same fact. Attacked and original image histograms are shown in Appendix H section 5.2. Example of attacked image histogram is depicted in Figure 5.17.

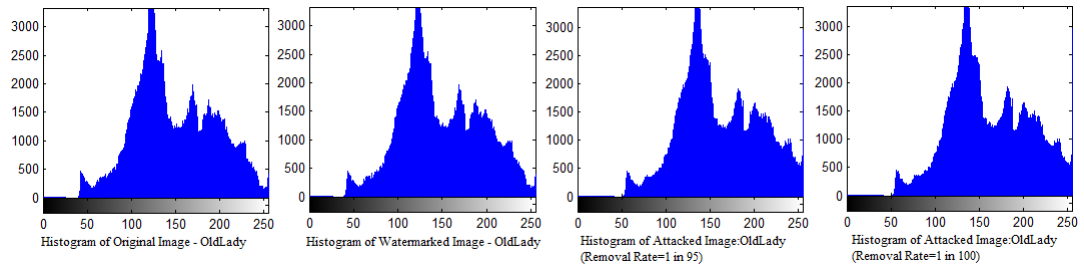


Figure 5.16: Histogram Comparison of OldLady:Kundur (StirMark Removal Attack)

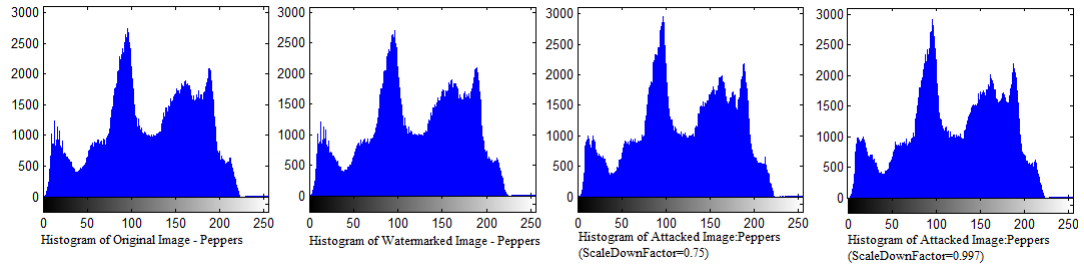


Figure 5.17: Histogram Comparison of Peppers:Kundur (Scale Down Attack)

Scale up attack is generated by MATLAB tool and similar behavior is observed in kundur's scheme. By inspection of histogram in Figure 5.18 and Appendix I section 5.2, almost all histograms are similar to each other.

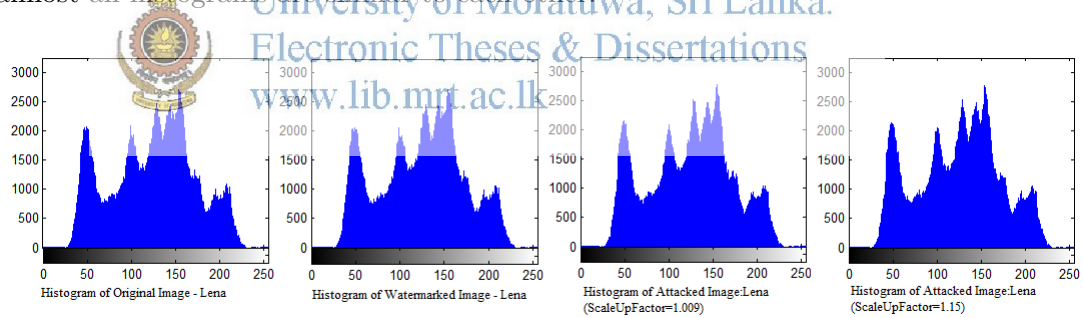


Figure 5.18: Histogram Comparison of Lena:Kundur (Scale Up Attack)

Analysis of Rotate Attack:

Rotate attack is used with bicubic interpolation method on watermarked images. As observed in previous instances, kundur's scheme has failed to detect watermark after this attack. The confidence level is very low and similar image perceptual quality can be observed in attacked and original images. Histograms are looklike as shown in Appendix J section 5.2 and image in Figure 5.19.

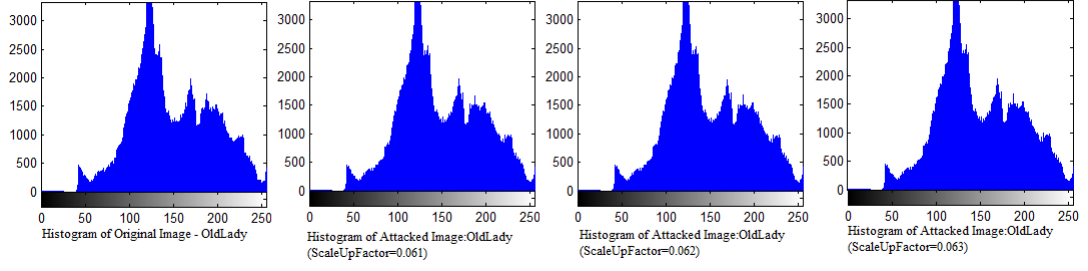


Figure 5.19: Histogram Comparison of OldLady:Kundur (Rotate Attack)

PSNR Analysis of Attack Success Incidents

Peak-Signal-to-Noise-Ratio (PSNR) calculation is a measurement of image processing to quantify the quality of two different images coming from the same source. For example, original image and the distorted image. The signal is original data and noise is the error introduced by the attack. Many authors have been highlighted the fact that, PSNR is a *approximation* to human perception of image quality. Higher the PSNR value is higher the quality of distorted image. However, it cannot be used to compare two different images. Mean squared error (MSE) is one of the earliest tests that were performed to test if two pictures are similar. Mean square error is calculated by:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - \hat{I}(i,j)]^2$$

Hence, PSNR is deried by equation:

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right)$$

Where MAX_I is the maximum possible pixel value of the image. In this thesis all images are represented using 8-bit grayscale mode and $MAX_I = 255$. PSNR is measured in decibels (dB). As thesis stated previously, PSNR is a better way of comparing image restoration results for same image but not for different images.

Remaining plots with PSNR values and graphs are posted in Appendix K section 5.2. Some of the attack experiments are explained for short reference.

PSNR Analysis of Rotate Attack:

In Figure 5.20, PSNR values prove output of histogram diagram illustrated in Figure 5.19. Hence image perceptual quality does not degrade over small changes of the angle as explained in section 4.1.

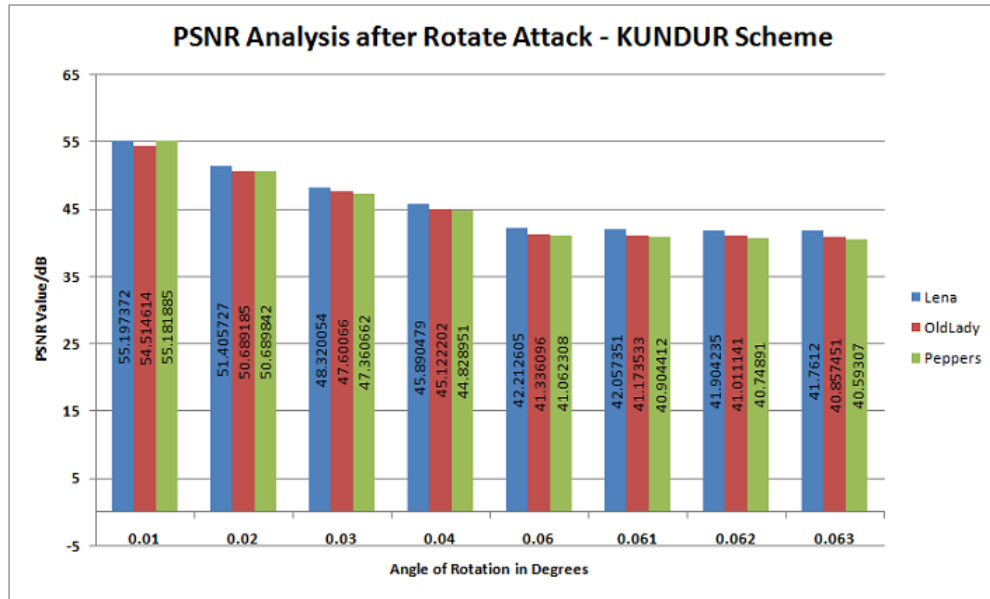


Figure 5.20: PSNR Comparison:Kundur (Rotate Attack)

PSNR Analysis of Flip Results:

According to depicted analysis in the thesis, it is known that similar histograms found for original and attacked images produce high PSNR values for particular image. By inspection of flip attack's histograms, it can be assumed that PSNR result values to be high. However, the Figure 5.21 shows low PSNR values. This does not mean that the PSNR is a not good measure. As explained in section 4.1, original and flipped images are not synchronized and that leads these two images are to be different from one another. Hence, calculating PSNR is not meaningful for such occurrence as depicted in section 5.1.

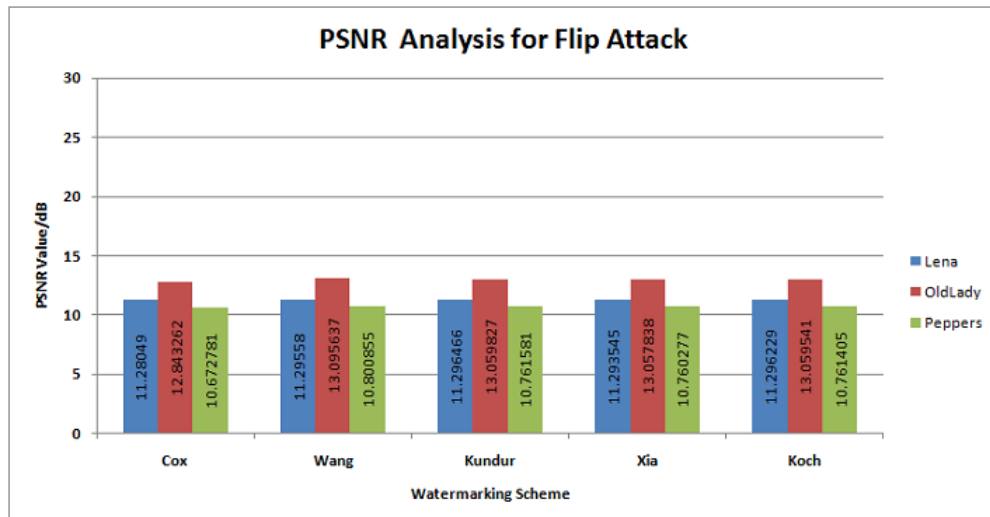


Figure 5.21: PSNR Comparison (Flip Attack)

5.2 Conclusion

The main use of digital watermarking is to prevent theft of intellectual property rights on multimedia objects. It is obtained using a hidden signature which contains proof of owner's data and is embedded into original data without degrading original image. This process perform in an imperceptible manner as primary intention is not to do a visible damage to original data. To prove the authenticity, owner can extract the signature whenever necessary. The outcome of watermark recovery does not depend on original image as none of algorithms uses original image's properties during their watermark embedding process. In this thesis, scope of watermarking multimedia data is limited to digital images. This section represents final outcome of attack success/failure metric with watermarking schemes and related transformation attacks.

Table 5.1: Attack Success/Failure Summary

Attack Type	Selected Algorithms				
	Cox	Wang	Kundur	Xia	Koch
Flip	Success	Success	Success	Success	Success
Crop	Failure	Failure	Failure	Failure	Failure
Scale Down	Failure	Failure	Success	Failure	Failure
Scale Up	Failure	Failure	Success	Failure	Failure
Filter(Average)	Failure	Success	Success	Success	Failure
Filter(Gaussian)	Failure	Failure	Success	Failure	Failure
StirMarkNoise	Failure	Failure	Success	Failure	Failure
StirMarkRemoval	Failure	Failure	Success	Failure	Failure
Rotation (bicubic)	Failure	Failure	Success	Failure	Failure
Rotation (nearest)	Failure	Failure	Failure	Failure	Failure
StirMark JPEG	Failure	Failure	Success	Failure	Failure

It is not possible to use an analytical scheme to detect watermark recovery probabilities by statistical means alone. An analytical scheme would not require image dependent values or set of values such as image perceptual quality, significant area of image (based on the image description) etc used in this thesis. Hence, an experimental analysis would work towards to achieve the research objectives.

Each scheme was tested against three test images depicted in section 3.3. Two standard images were selected from well known images in watermarking and one custom image was taken locally. The reason behind for selection of two standard images was that, results obtained in this thesis can also be tested by other researchers. The other non-standard image was taken to prove same results for different images.

According to attack evaluation metric with related concepts described in section 3.4

and experimental results taken from chapter 4 on watermark embedding and detection Table 5.1, a summary of results is created. This thesis has discussed the “attack success” instances in distortion analysis in section 5.1. Important facts are determined during the analysis of attacks with respect to selected watermarking schemes.

1. Kundur’s Scheme is vulnerable to many attacks performed in the thesis

This watermarking scheme is designed to detect watermark without referring to original image. This is known as “Blind watermark detection”. Authors have claimed that this scheme is robust to variety of signal distortions. However they have only tested “cropping” and “filtering”. In this thesis, filtering attacks are different from what Kundur scheme authors used in their paper. This research used average and gaussian filtering methods while kundur’s work has only used meadian and mean filtering methods. Hence it is observed that this scheme cannot withstand stronger filtering attacks.

2. All schemes failed to detect the watermark when subjected to the “Flip Attack”

In the section 4.1, the flip attack is preserving the image perceptual quality and carry out attack while retaining the watermark at the same time.

3. PSNR analysis of the Flip Attack shows low quality

In the Histogram analysis of attacked images histograms were almost identical with original histogram images. However, PSNR calculations are taken from Figure 5.21 shows counter expectation of what is observed in the previous attack analysis of other schemes. Observation result depicts that the obtained PSNR values are very low.

4. Highest attack resilient watermarking schemes are COX and KOCH

Cox’s scheme use non-blind watermarking whereas Koch’s scheme use blind watermarking technique to detect the watermark out of attacked images. It is noticed that these two watermarking schemes are resilient to all attacks performed in thesis except the flip attack.

The importance of work done in the thesis depicts in this section of the chapter. Christine et al. [80] worked on “Digital Watermarking: Algorithms and Application” paper and it has concluded with unanswered two questions. In this thesis, two questions mentioned in the paper have been answered:

Q1. What are reasonable distortions for particular applications that the watermark is expected to survive?

In this thesis, for different types of attacks, the limits that preserve the perceptual quality of the subjected images were analyzed.

Q2. What is a meaningful measure of distortion that can be used to determine the effectiveness of a watermarking scheme?

Both PSNR and Histogram analysis techniques provide a foundation for meaningful measurement. We have shown that coupling such measurements with a perceptual quality measurement provide meaningful values for effectiveness.

In conclusion, the research presented in this thesis confirms the findings from previous researches [4], [54] that DCT/DFT outperforms DWT when used in the creation of attack resilient digital watermarking schemes. More importantly, the research work in this thesis extends the known results to a series of new attacks.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

References

- [1] S. Voloshynovskiy, S. Pereira, T. Pun, J. J. Eggers, and J. K. Su, "Attacks on digital watermarks: classification, estimation based attacks, and benchmarks," *Communications Magazine, IEEE*, vol. 39, no. 8, pp. 118–126, 2001.
- [2] M. Kutter, "Watermarking resistance to translation, rotation, and scaling," in *Photonics East (ISAM, VVDC, IEMB)*, 1999, pp. 423–431.
- [3] C. S. Collberg and C. Thomborson, "Watermarking, tamper-proofing, and obfuscation-tools for software protection," *Software Engineering, IEEE Transactions on*, vol. 28, no. 8, pp. 735–746, 2002.
- [4] V. M. Potdar, S. Han, and E. Chang, "A survey of digital image watermarking techniques," in *Industrial Informatics, 2005. INDIN'05. 2005 3rd IEEE International Conference on*, 2005, pp. 709–716.
- [5] I. J. Cox, M. L. Miller, and J. A. Bloom, "Watermarking applications and their properties," in *Information Technology: Coding and Computing, 2000. Proceedings. International Conference on*, 2000, pp. 15–20.
- [6] J.-S. Pan, H.-C. Huang, and L.-C. Jain, *Intelligent watermarking techniques*. World scientific, 2004, vol. 7.
- [7] M. Kutter and F. A. Petitcolas, "A fair benchmark for image watermarking systems," *Electronic Imaging*, vol. 99, pp. 219–239, 1999.
- [8] G. Poupeau, *Wax tablet*, 2006. [Online]. Available: <http://www.flickr.com/photos/24266175@N00/256751204> (visited on 08/26/2014).
- [9] Google, *Tattooed head*. [Online]. Available: <http://5010.mathed.usu.edu/Fall2014/KKing/headtattoo.jpg> (visited on 08/26/2014).
- [10] S. Katzenbeisser and F. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking*. EDPACS: The EDP Audit, Control, and Security, 2006, vol. 28.
- [11] R. B. Wolfgang, C. I. Podilchuk, and E. J. Delp, "Perceptual watermarks for digital images and video," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1108–1126, 1999.
- [12] M. Wu, W. Trappe, Z. J. Wang, and K. R. Liu, "Collusion-resistant fingerprinting for multimedia," *Signal Processing Magazine, IEEE*, vol. 21, no. 2, pp. 15–27, 2004.
- [13] H. S. Stone, *Analysis of attacks on image watermarks with randomized coefficients*. NEC Research Institute, 1996.

- [14] G. L. Friedman, "The trustworthy digital camera: restoring credibility to the photographic image," *Consumer Electronics, IEEE Transactions on*, vol. 39, no. 4, pp. 905–910, 1993.
- [15] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, "A DCT-domain system for robust image watermarking," *Signal processing*, vol. 66, no. 3, pp. 357–372, 1998.
- [16] D. Kundur and D. Hatzinakos, "Digital watermarking using multiresolution wavelet decomposition," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, vol. 5, 1998, pp. 2969–2972.
- [17] R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne, "A digital watermark," in *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, vol. 2, 1994, pp. 86–90.
- [18] A. Herrigel, J. ó Ruanaidh, H. Petersen, S. Pereira, and T. Pun, "Secure copyright protection techniques for digital images," in *Information Hiding*, 1998, pp. 169–190.
- [19] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Robust data hiding for images," in *Digital Signal Processing Workshop Proceedings, 1996., IEEE*, 1996, pp. 37–40.
- [20] M. Barni and F. Bartolini, *Watermark Systems Engineering*. Marcel Dekker New York, 2004.
- [21] F. H. Hartung, J. K. Su, and B. Girod, "Spread spectrum watermarking: malicious attacks and counterattacks," in *Electronic Imaging'99*, International Society for Optics and Photonics, 1999, pp. 147–158.
- [22] V. Licks and R. J. Jordan, "Geometric attacks on image watermarking systems," *IEEE multimedia*, vol. 12, no. 3, pp. 68–78, 2005.
- [23] S. Craver, N. Memon, B.-L. Yeo, and M. M. Yeung, "Resolving rightful ownerships with invisible watermarking techniques: limitations, attacks, and implications," *Selected Areas in Communications, IEEE Journal on*, vol. 16, no. 4, pp. 573–586, 1998.
- [24] F. A. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Attacks on copyright marking systems," in *Information Hiding*, Springer, 1998, pp. 218–238.
- [25] F. A. Petitcolas, "Watermarking schemes evaluation," *Signal Processing Magazine, IEEE*, vol. 17, no. 5, pp. 58–64, 2000.
- [26] J. R. Troncoso-Pastoriza and F. Pérez-González, "Zero-knowledge watermark detector robust to sensitivity attacks," in *Proceedings of the 8th workshop on Multimedia and security*, ACM, 2006, pp. 97–107.
- [27] F. A. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding-a survey," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1062–1078, 1999.
- [28] S. P. Mohanty, "Digital watermarking: a tutorial review," URL: <http://www.csee.usf.edu/~smohanty/research/Reports/WMSurvey1999Mohanty.pdf>, 1999.
- [29] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1079–1107, 1999.
- [30] D. Mistry, "Comparison of digital water marking methods," *International Journal on Computer Science & Engineering*, 2010.

- [31] D. Chopra, P. Gupta, and A. Gupta, "Lsb based digital image watermarking for gray scale image," *IOSR Journal of Computer Engineering (IOSRJCE) ISSN*, pp. 2278-0661, 2012.
- [32] W. S., "Information authentication for a slippery new age," *Dr.Dobbs Journal*, vol. 20, no. 4, pp. 18-26, 1995.
- [33] J. Fridrich, "Image watermarking for tamper detection," in *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, IEEE, vol. 2, 1998, pp. 404-408.
- [34] M. M. Yeung and F. Mintzer, "An invisible watermarking technique for image verification," in *Image Processing, 1997. Proceedings., International Conference on*, IEEE, vol. 2, 1997, pp. 680-683.
- [35] M. U. Celik, G. Sharma, A. M. Tekalp, and E. Saber, "Lossless generalized-lsb data embedding," *Image Processing, IEEE Transactions on*, vol. 14, no. 2, pp. 253-266, 2005.
- [36] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM systems journal*, vol. 35, no. 3.4, pp. 313-336, 1996.
- [37] I.-K. Yeo and H. J. Kim, "Generalized patchwork algorithm for image watermarking," *Multimedia systems*, vol. 9, no. 3, pp. 261-265, 2003.
- [38] S. Weng, Y. Zhao, and J.-S. Pan, "Reversible watermarking based on improved patchwork algorithm and symmetric modulo operation," in *Knowledge-Based Intelligent Information and Engineering Systems*, Springer, 2005, pp. 317-323.
- [39] G. Caronni, "Assuring ownership rights for digital images," in *Verlässliche IT-Systeme*, Springer, 1995, pp. 251-263.
- [40] J. Fridrich, "Robust bit extraction from images," in *Multimedia Computing and Systems, 1999. IEEE International Conference on*, IEEE, vol. 2, 1999, pp. 536-540.
- [41] A. Hanjalic, G. Langelaar, P. Van Roosmalen, J. Biemond, and R. Lagendijk, *Image and Video Databases: Restoration, Watermarking and Retrieval: Restoration, Watermarking and Retrieval*. Elsevier, 2000, vol. 8.
- [42] G. C. Langelaar, J. C. van der Lubbe, and R. L. Lagendijk, "Robust labeling methods for copy protection of images," in *Electronic Imaging'97*, International Society for Optics and Photonics, 1997, pp. 298-309.
- [43] I. Pitas, "A method for signature casting on digital images," in *Image Processing, 1996. Proceedings., International Conference on*, IEEE, vol. 3, 1996, pp. 215-218.
- [44] J. R. Smith and B. O. Comiskey, "Modulation and information hiding in images," in *Information Hiding*, Springer, 1996, pp. 207-226.
- [45] R. B. Wolfgang Edward J and Delp, "A watermarking technique for digital imagery: further studies," in *International Conference on Imaging, Systems, and Technology*, 1997, pp. 279-287.
- [46] E. J. Wolfgang Raymond B and Delp, "Fragile watermarking using the vw2d watermark," in *Electronic Imaging'99*, International Society for Optics and Photonics, 1999, pp. 204-213.

- [47] R. B. Wolfgang Edward J and Delp, "Overview of image security techniques with applications in multimedia systems," in *Voice, Video, and Data Communications*, International Society for Optics and Photonics, 1998, pp. 297–308.
- [48] E. J. Wolfgang Delp and R. B., "A watermark for digital images," in *Image Processing, 1996. Proceedings., International Conference on*, IEEE, vol. 3, 1996, pp. 219–222.
- [49] G. C. Langelaar, I. Setyawan, and R. L. Lagendijk, "Watermarking digital image and video data," *IEEE Signal processing magazine*, 2000.
- [50] N. F. Johnson and S. Katzenbeisser, "A survey of steganographic techniques," in *Information hiding*, Norwood, MA: Artech House, 2000, pp. 43–78.
- [51] J. R. Smith and B. O. Comiskey, "Modulation and information hiding in images," in *Information Hiding*, Springer, 1996, pp. 207–226.
- [52] I. Pitas, "A method for watermark casting on digital image," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 8, no. 6, pp. 775–780, 1998.
- [53] N. Nikolaidis and I. Pitas, "Robust image watermarking in the spatial domain," *Signal processing*, vol. 66, no. 3, pp. 385–403, 1998.
- [54] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *Image Processing, IEEE Transactions on*, vol. 6, no. 12, pp. 1673–1687, 1997.
- [55] M. Ramkumar, A. N. Akansu, and A. A. Alatan, "A robust data hiding scheme for images using dft," in *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, IEEE, vol. 2, 1999, pp. 211–215.
- [56] C.-T. Hsu and J.-L. Wu, "Hidden digital watermarks in images," *Image Processing, IEEE Transactions on*, vol. 8, no. 1, pp. 58–68, 1999.
- [57] M. Jiansheng, L. Sukang, and T. Xiaomei, "A digital watermarking algorithm based on dct and dwt," in *Proceedings of the 2009 International Symposium on Web Information Systems and Applications (WISA '09)*, vol. 8, 2009, pp. 104–107.
- [58] N. Terzija, M. Repges, K. Luck, and W. Geisselhardt, "Digital image watermarking using discrete wavelet transform: performance comparison of error correction codes," in *Visualization, Imaging, and Image Processing*, 2002.
- [59] H.-J. Wang, P.-C. Su, and C.-C. J. Kuo, "Wavelet-based digital image watermarking," *Optics Express*, vol. 3, no. 12, pp. 491–496, 1998.
- [60] N. Chaturvedi and S. Basha, "A novel svd based digital watermarking scheme using dwt and a comparative study with dwt-arnold, svd-dct and svd-dft based watermarking," 2013.
- [61] M. Voigt, B. Yang, and C. Busch, "Reversible watermarking of 2d-vector data," in *Proceedings of the 2004 workshop on Multimedia and security*, ACM, 2004, pp. 160–165.
- [62] A. Piper, R. Safavi-Naini, and A. Mertins, "Resolution and quality scalable spread spectrum image watermarking," in *Proceedings of the 7th workshop on Multimedia and security*, ACM, 2005, pp. 79–90.

- [63] M. Schlauweg, D. Pröfrock, B. Zeibich, and E. Müller, "Dual watermarking for protection of rightful ownership and secure image authentication," in *Proceedings of the 4th ACM international workshop on Contents protection and security*, ACM, 2006, pp. 59–66.
- [64] X.-Y. Huang, M.-S. Tan, Y. Luo, and D.-Z. Lin, "An image digital watermarking based on dct in invariant wavelet domain," in *Wavelet Analysis and Pattern Recognition, 2007. ICWAPR'07. International Conference on*, IEEE, vol. 1, 2007, pp. 458–463.
- [65] X. Li and X. Xue, "Improved robust watermarking in dct domain for color images," in *Advanced Information Networking and Applications, 2004. AINA 2004. 18th International Conference on*, IEEE, vol. 1, 2004, pp. 53–58.
- [66] P. Comesana, L. Pérez-Freire, and F. Pérez-González, "Blind newton sensitivity attack," in *Information Security, IEE Proceedings, IET*, vol. 153, 2006, pp. 115–125.
- [67] I. J. Cox and M. L. Miller, "Electronic watermarking: the first 50 years," in *Multimedia Signal Processing, 2001 IEEE Fourth Workshop on*, IEEE, 2001, pp. 225–230.
- [68] D. Kundur and D. Hatzinakos, "Digital watermarking using multiresolution wavelet decomposition," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, IEEE, vol. 5, 1998, pp. 2969–2972.
- [69] X. G. Xia, C. Boncelet, and G. Arce, "Wavelet transform based watermark for digital images," *Optics Express*, vol. 3, no. 12, pp. 497–511, 1998.
- [70] E. Koch and J. Zhao, "Towards robust and hidden image copyright labeling," in *IEEE Workshop on Nonlinear Signal and Image Processing*, Neos Marmaras, Greece, 1995, pp. 452–455.
- [71] MATLAB, version 7.14 (R2012a). Natick, Massachusetts: The MathWorks Inc., 2012.
- [72] Y. H. Yee and A. Newman, "A perceptual metric for production testing," in *ACM SIGGRAPH 2004 Sketches*, ser. SIGGRAPH '04, Los Angeles, California: ACM, 2004, pp. 121–.
- [73] S. J. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," in *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*, International Society for Optics and Photonics, 1992, pp. 2–15.
- [74] B. Davidsen, *Netpbm*, 1993. [Online]. Available: <ftp://ftp.wustl.edu/graphics/graphics/packages/NetPBM> (visited on 01/26/2015).
- [75] A. D. Panditha, *Life at negombo*, 2013. [Online]. Available: <https://www.flickr.com/photos/akiladarshana/10066006836/?rb=1> (visited on 01/26/2015).
- [76] P. Meerwald, "Watermarking source code," Dept. of Computer Sciences, University of Salzburg, Salzburg, Austria, Tech. Rep., 2005, Available at <http://www.cosy.sbg.ac.at/~pmeerw/Watermarking/source>.
- [77] F. A. A. P. Petitcolas, R. J. J. Anderson, and M. G. G. Kuhn, "Attacks on copyright marking systems," in *Information Hiding: Second International Workshop*, D. Aucsmith, Ed., ser. Lecture Notes in Computer Science, vol. 1525, Portland, OR, USA: Springer Verlag, Berlin, Germany, Apr. 1998, pp. 218–238.

- [78] F. A. A. P. Petitcolas and R. J. J. Anderson, "Weaknesses of copyright marking systems," in *Multimedia and Security Workshop at the 6th ACM International Multimedia Conference*, Bristol, England, 1998, pp. 55–61.
- [79] F. A. P. Petitcolas and R. J. Anderson, "Evaluation of copyright marking systems," in *Proceedings of IEEE International Conference on Multimedia Computing and Systems '99*, vol. 1, Florence, Italy, 1999, pp. 574–579.
- [80] C. I. Podilchuk and E. J. Delp, "Digital watermarking: algorithms and applications," *Signal Processing Magazine, IEEE*, vol. 18, no. 4, pp. 33–46, 2001.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Appendix

Appendix A

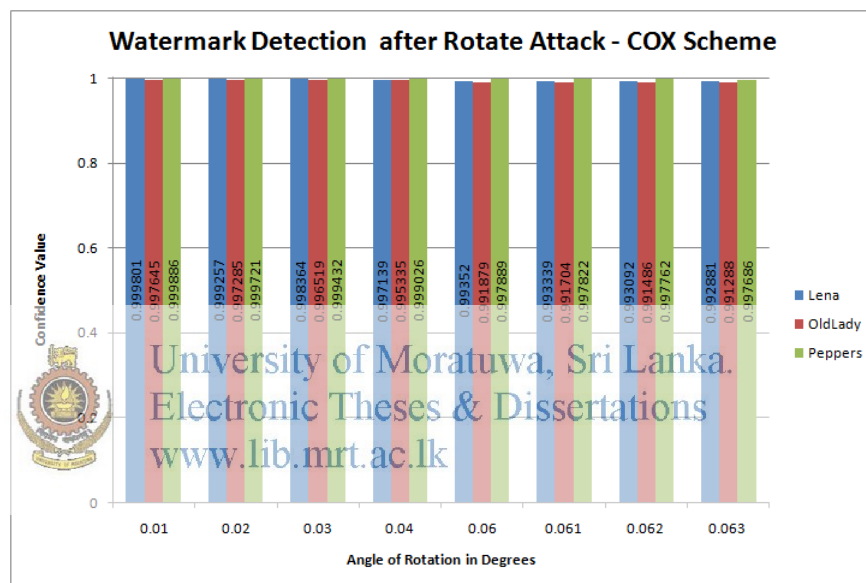


Figure 5.22: Rotation:COX Scheme(Method:Bicubic)

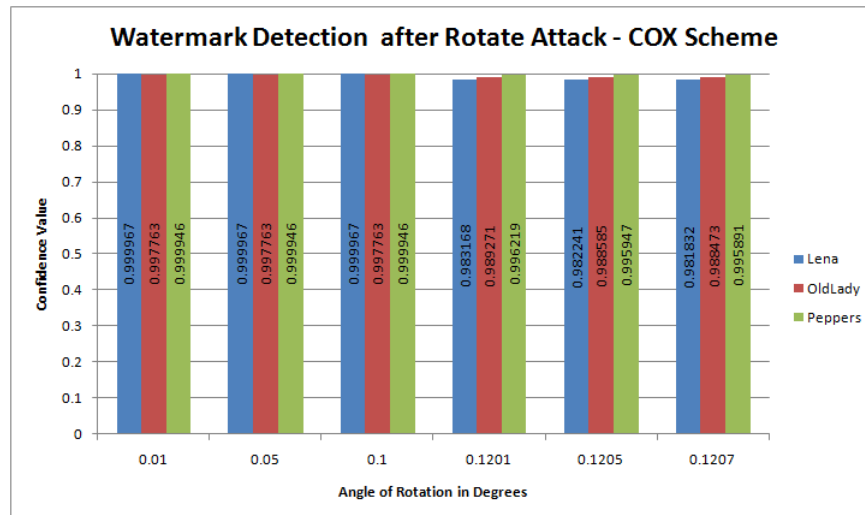


Figure 5.23: Rotation:COX Scheme(Method:Nearest)

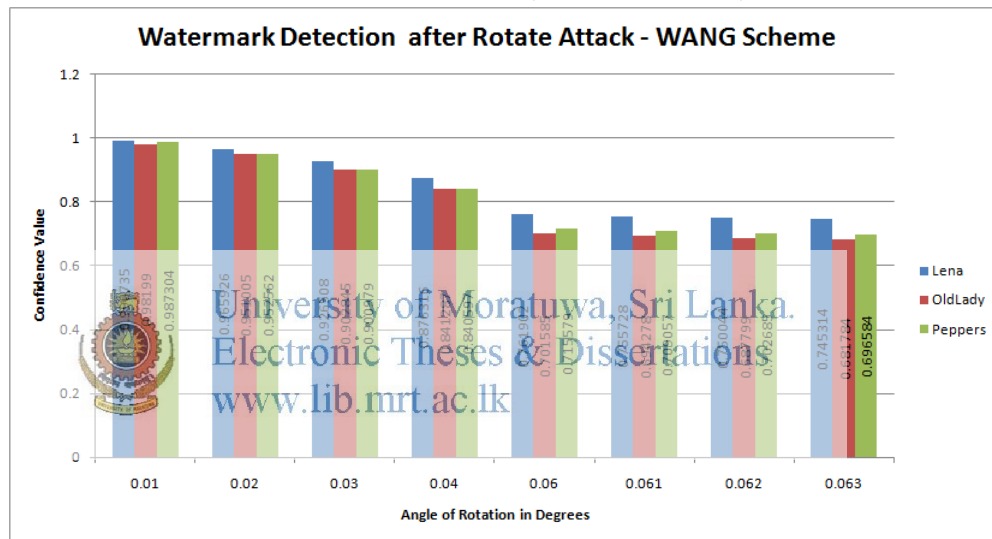


Figure 5.24: Rotation:WANG Scheme(Method:Bicubic)

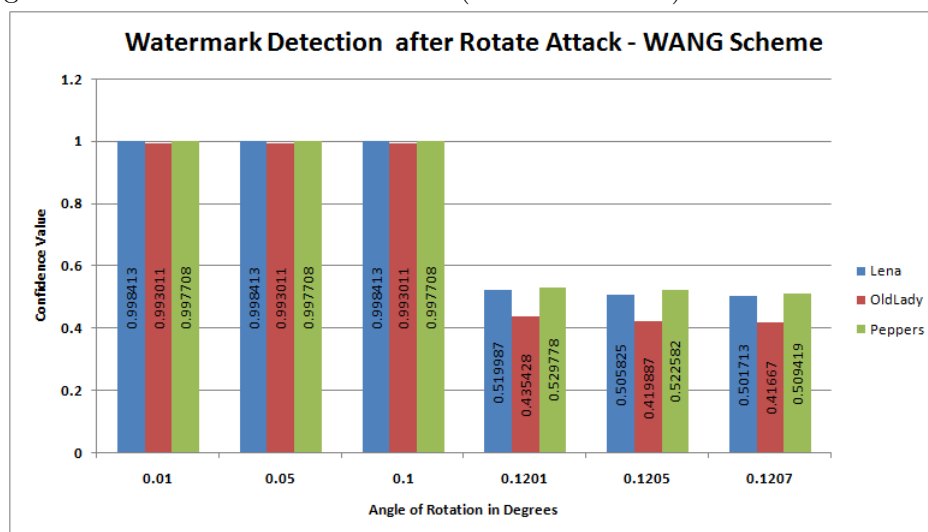


Figure 5.25: Rotation:WANG Scheme(Method:Nearest)

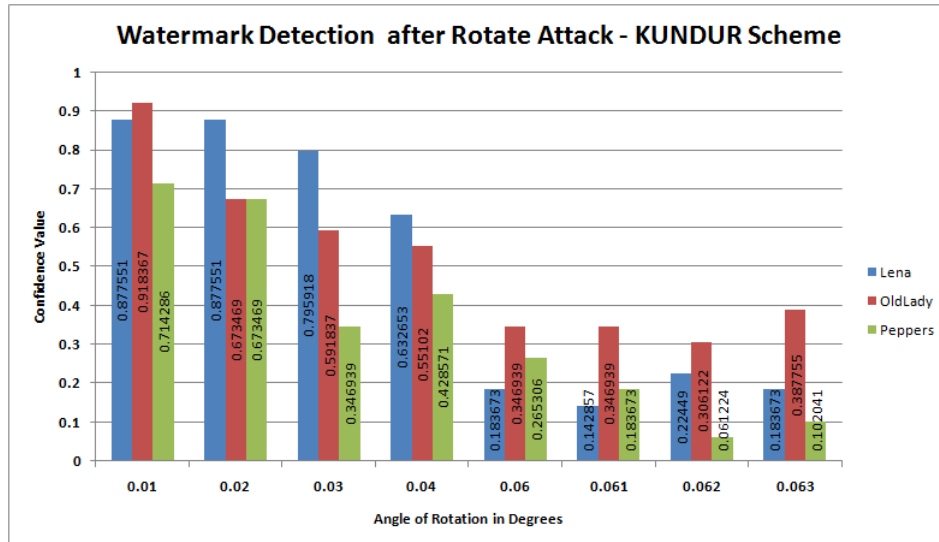


Figure 5.26: Rotation:KUNDUR Scheme(Method:Bicubic)

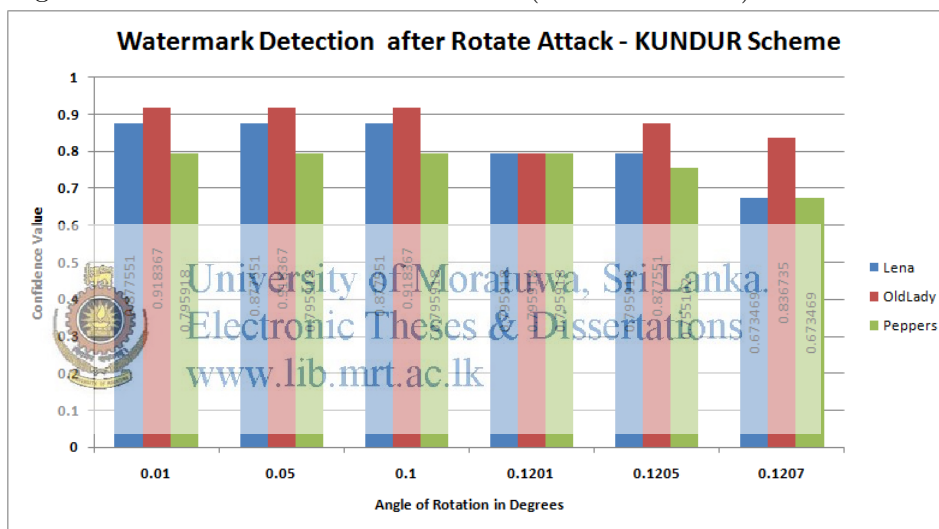


Figure 5.27: Rotation:KUNDUR Scheme(Method:Nearest)

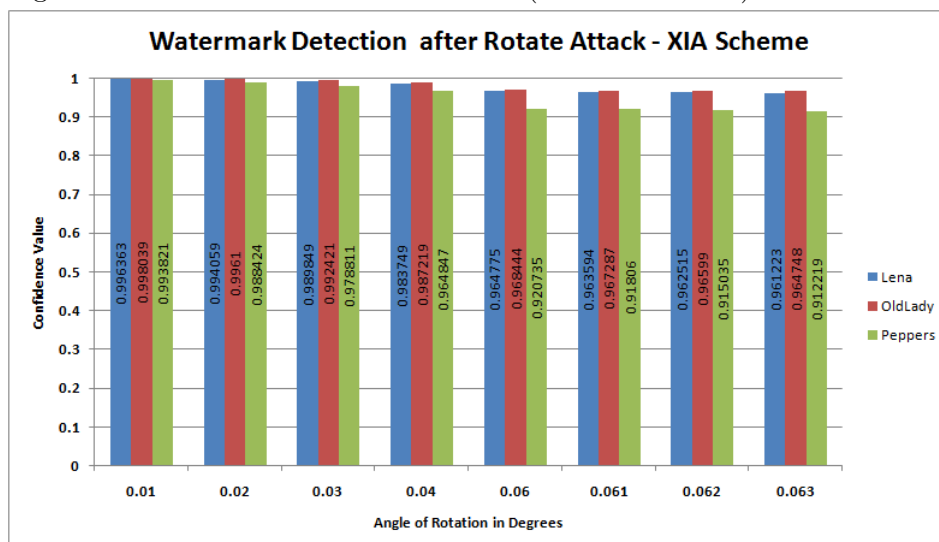


Figure 5.28: Rotation:XIA Scheme(Method:Bicubic)

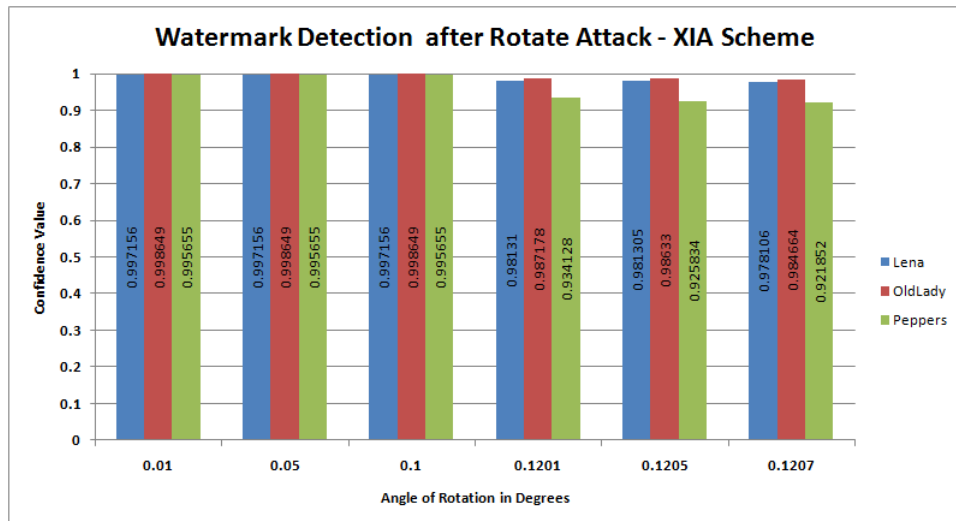


Figure 5.29: Rotation:XIA Scheme(Method:Nearest)

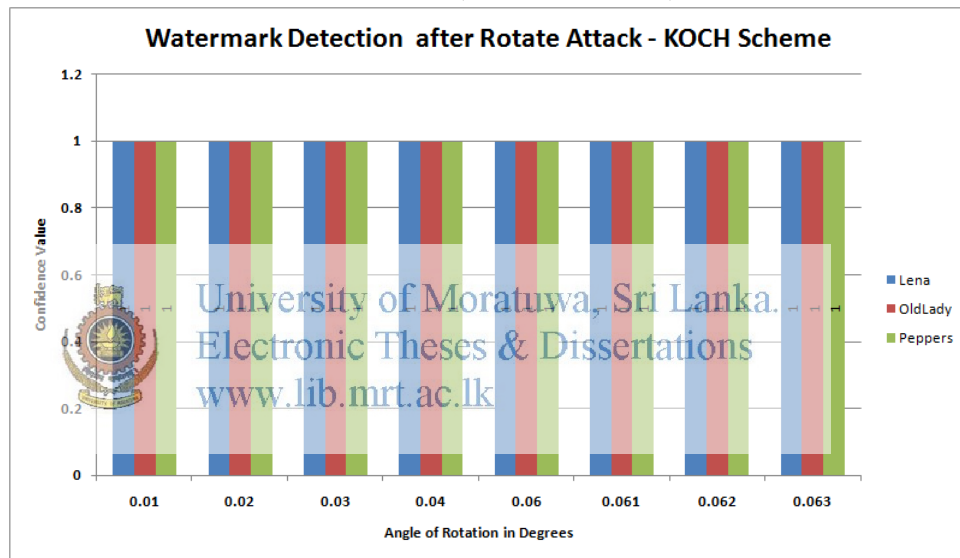


Figure 5.30: Rotation:KOCH Scheme(Method:Bicubic)

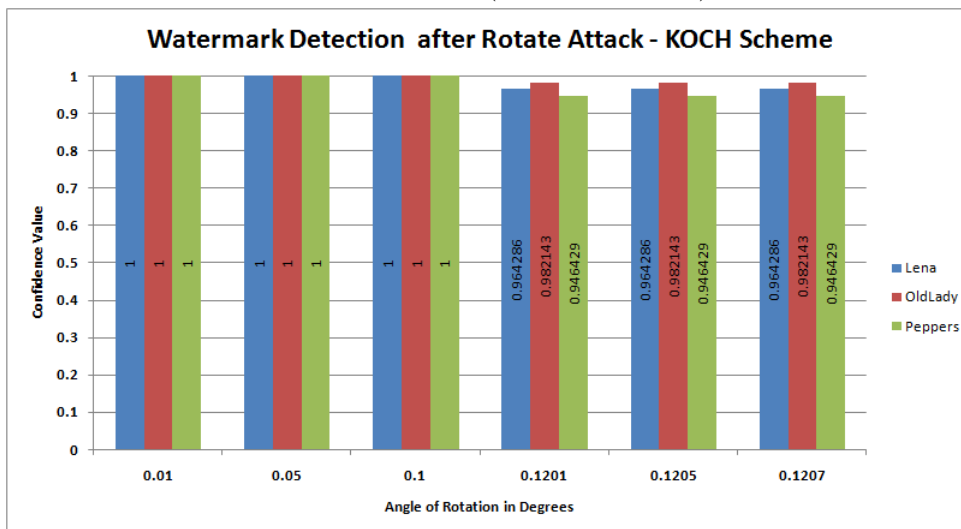


Figure 5.31: Rotation:KOCH Scheme(Method:Nearest)

Appendix B

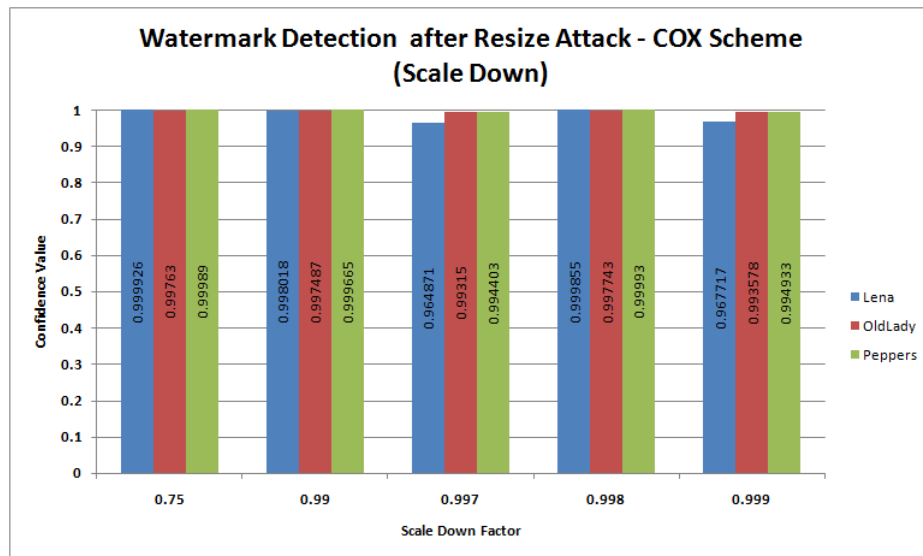


Figure 5.32: Resize:COX Scheme(Scale Down)

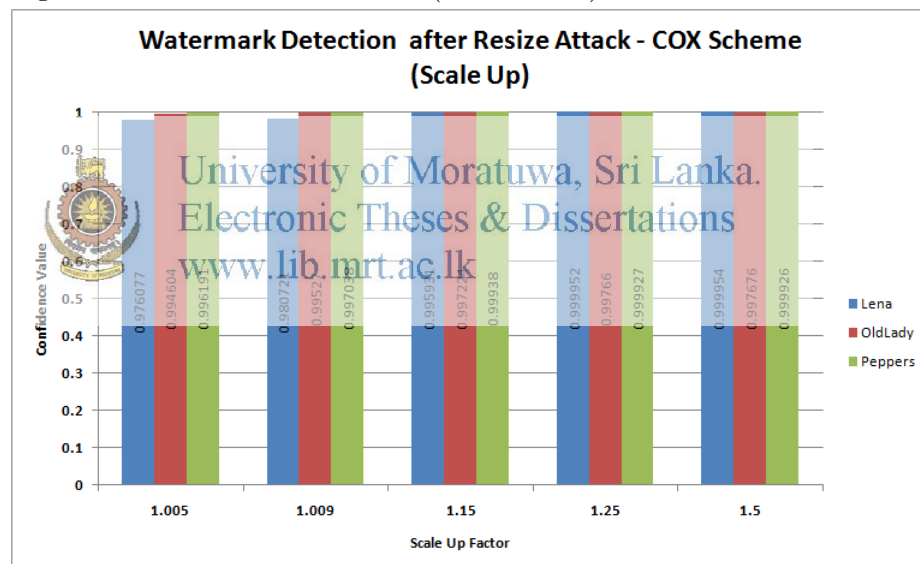


Figure 5.33: Resize:COX Scheme(Scale Up)

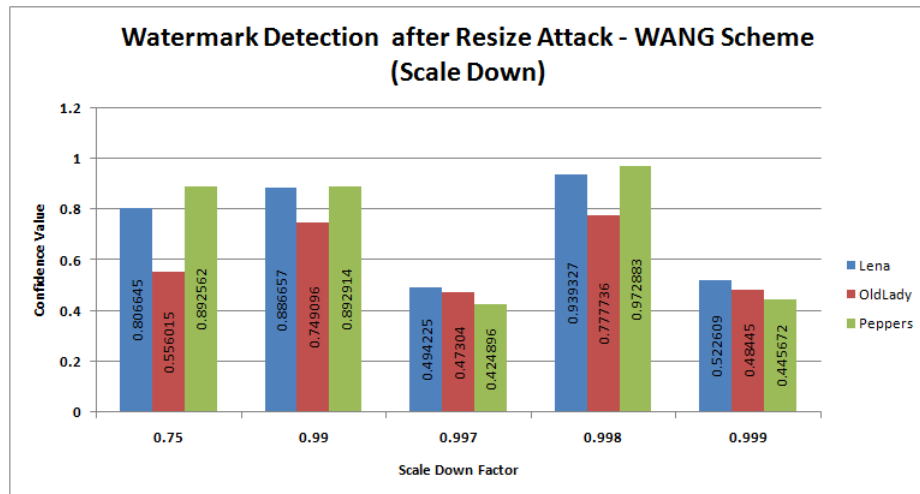


Figure 5.34: Resize:WANG Scheme(Scale Down)

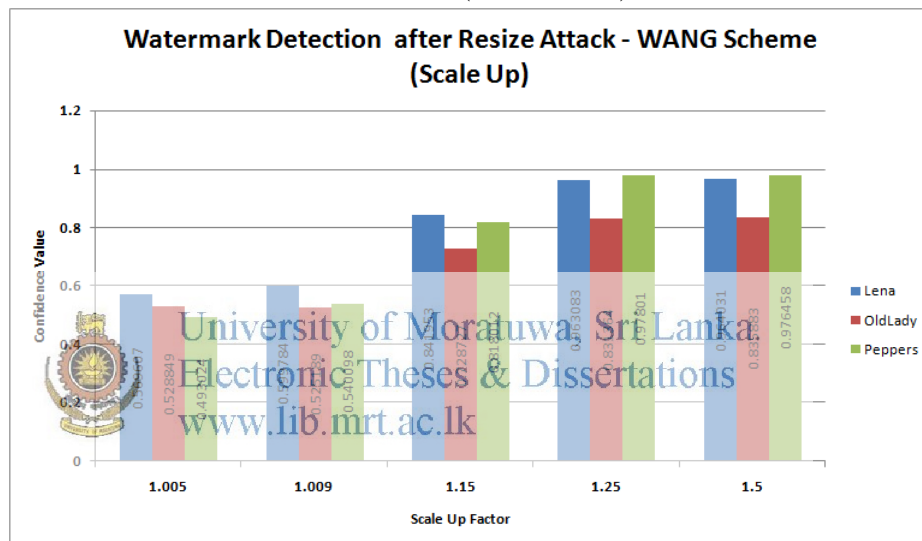


Figure 5.35: Resize:WANG Scheme(Scale Up)

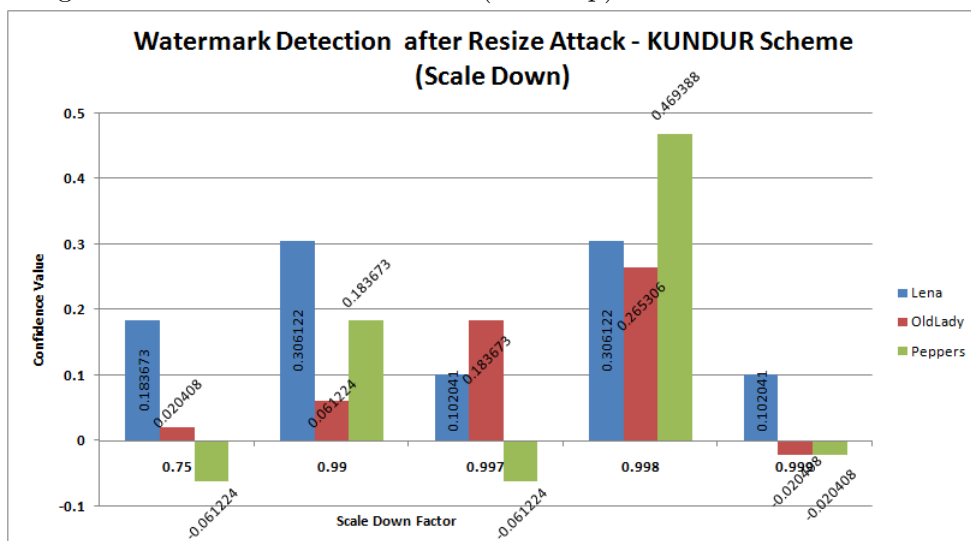


Figure 5.36: Resize:KUNDUR Scheme(Scale Down)

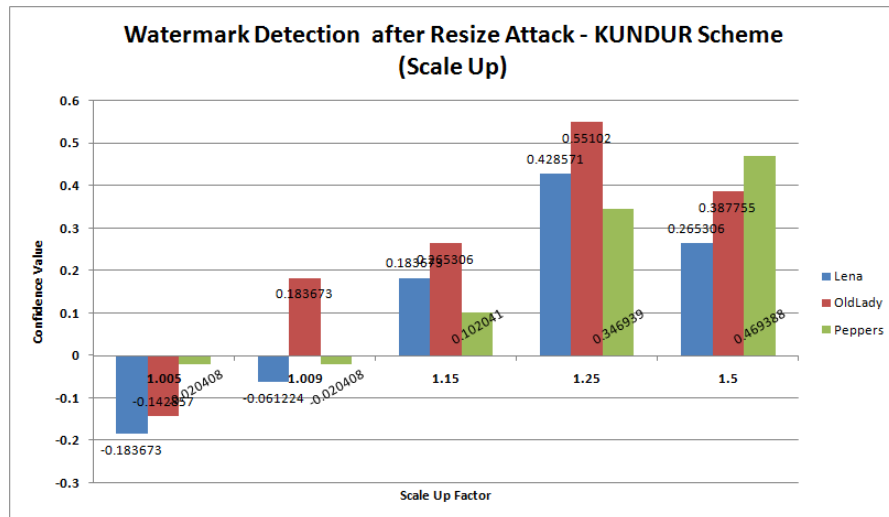


Figure 5.37: Resize:KUNDUR Scheme(Scale Up)

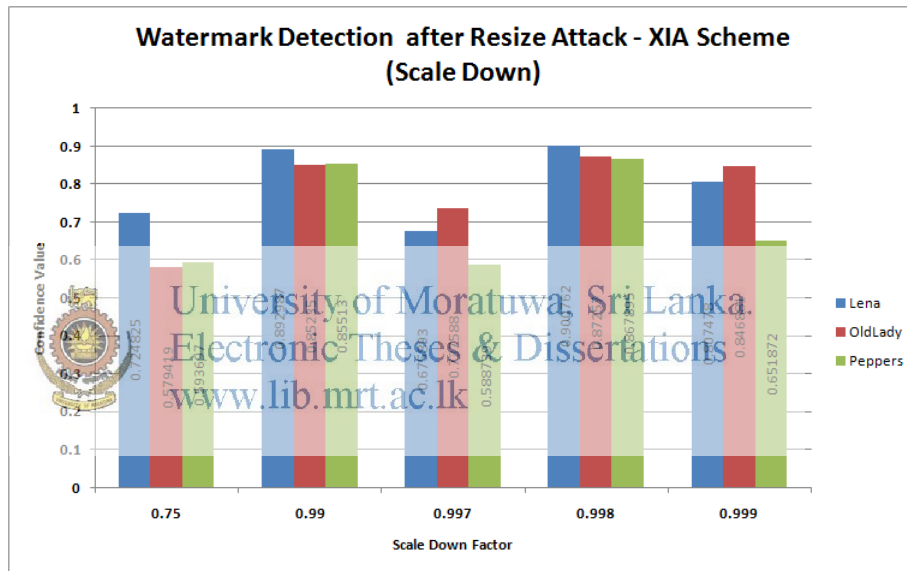


Figure 5.38: Resize:XIA Scheme(Scale Up)

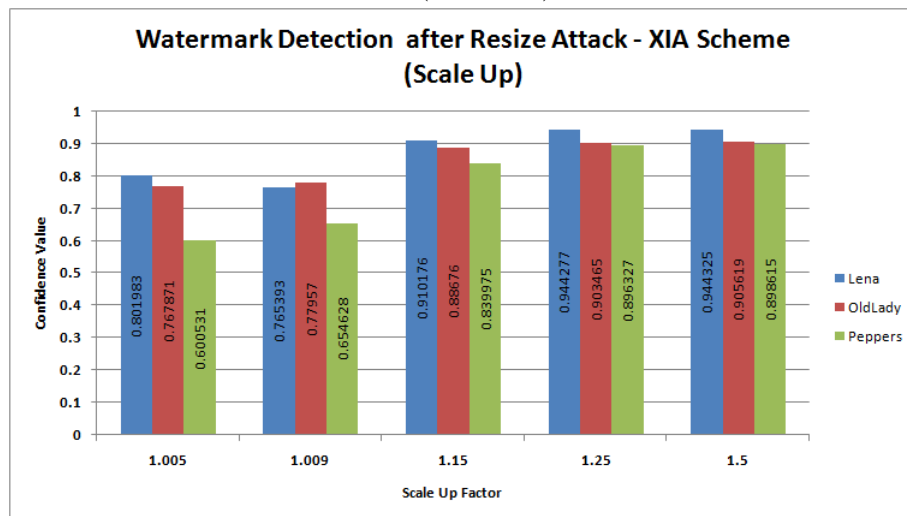


Figure 5.39: Resize:XIA Scheme(Scale Up)

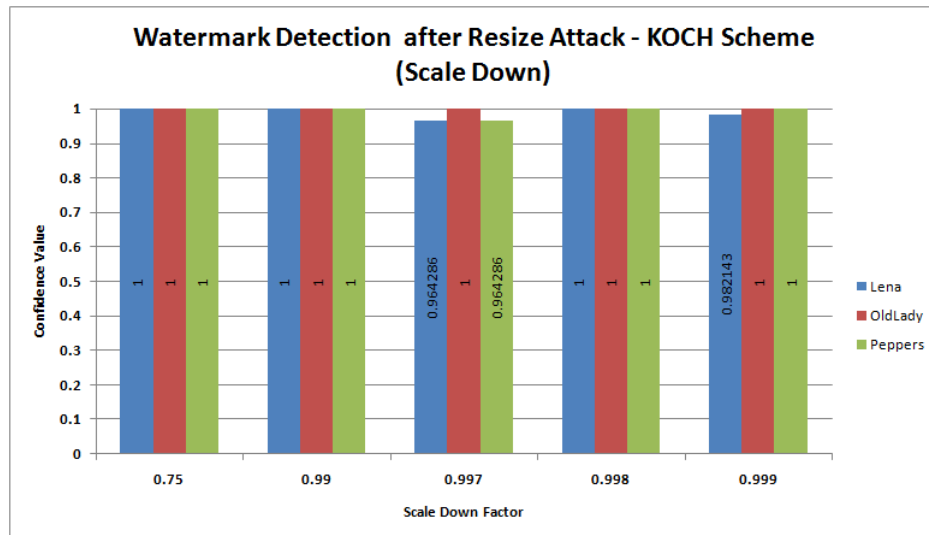


Figure 5.40: Resize:Koch Scheme(Scale Down)



Figure 5.41: Resize:Koch Scheme(Scale Up)

Appendix C

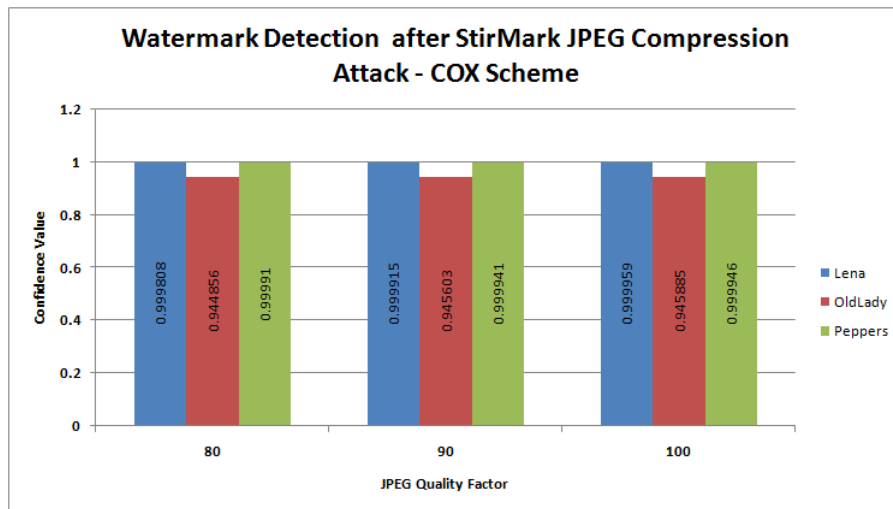


Figure 5.42: JPEG Compression:COX Scheme

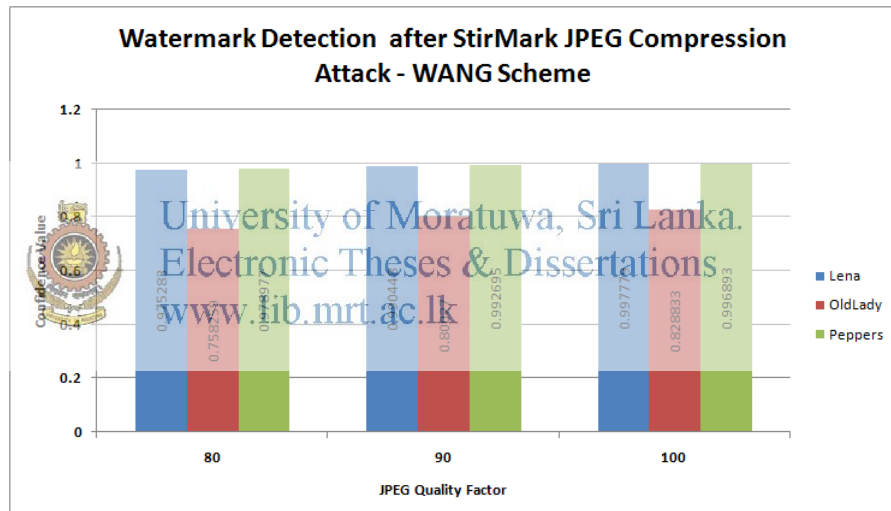


Figure 5.43: JPEG Compression:WANG Scheme

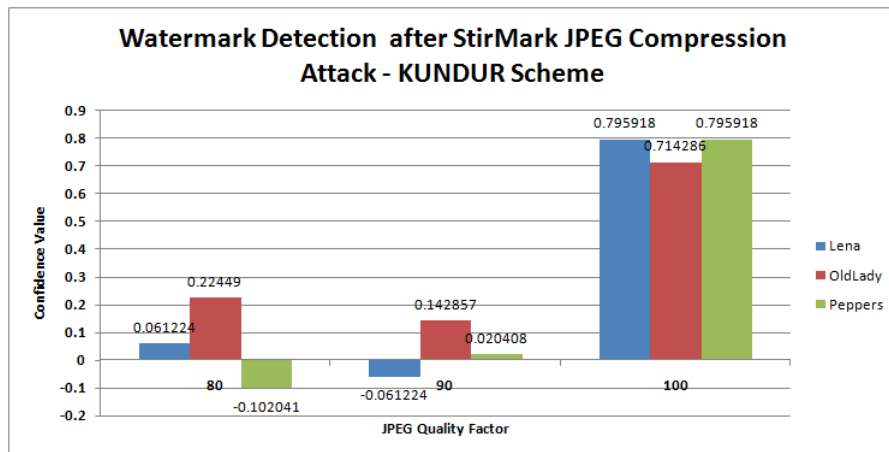


Figure 5.44: JPEG Compression:KUNDUR Scheme

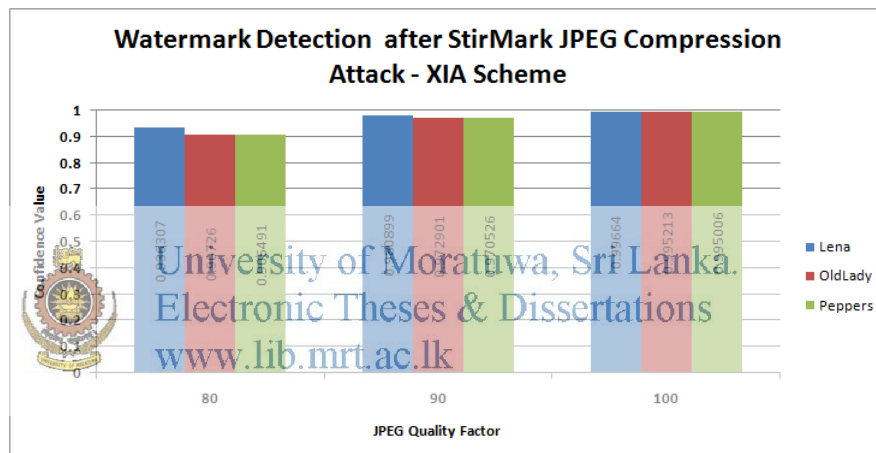


Figure 5.45: JPEG Compression:XIA Scheme

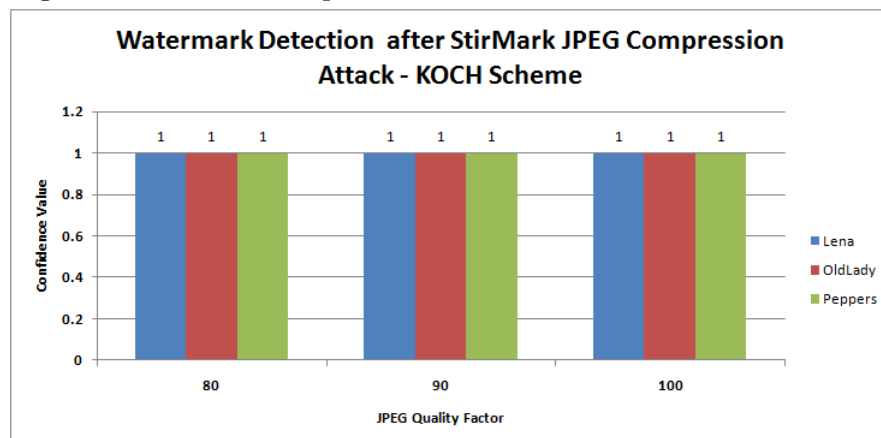


Figure 5.46: JPEG Compression:KOCH Scheme

Appendix D

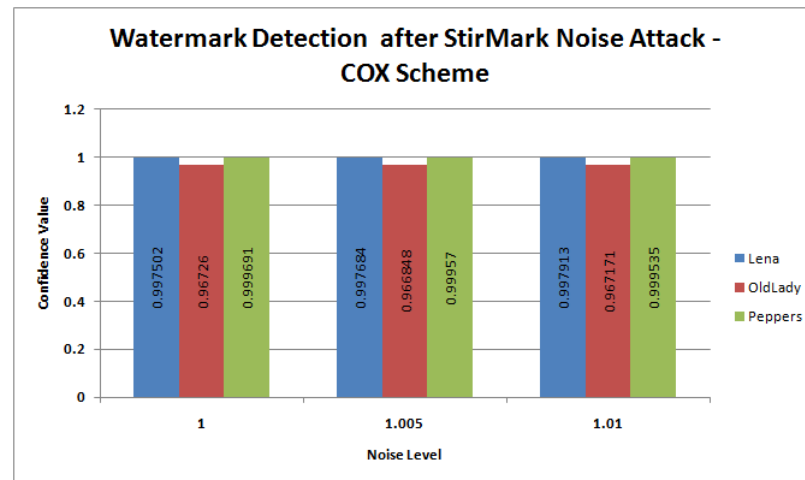


Figure 5.47: Noise Attack:COX Scheme

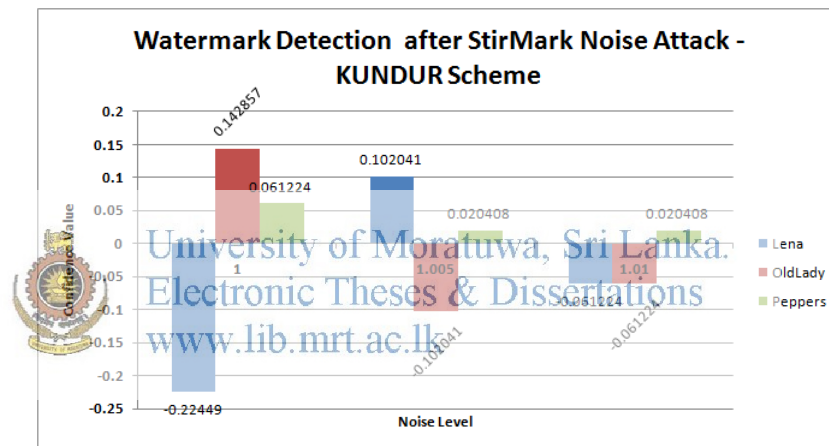


Figure 5.48: Noise Attack:KUNDUR Scheme

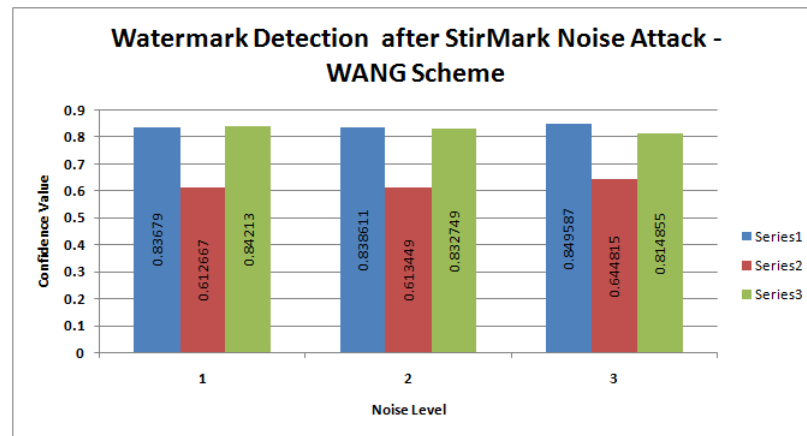


Figure 5.49: Noise Attack:WANG Scheme

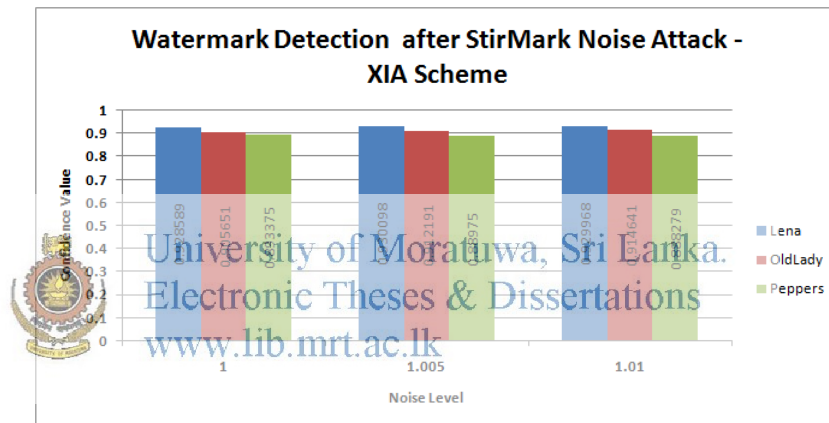


Figure 5.50: Noise Attack:XIA Scheme

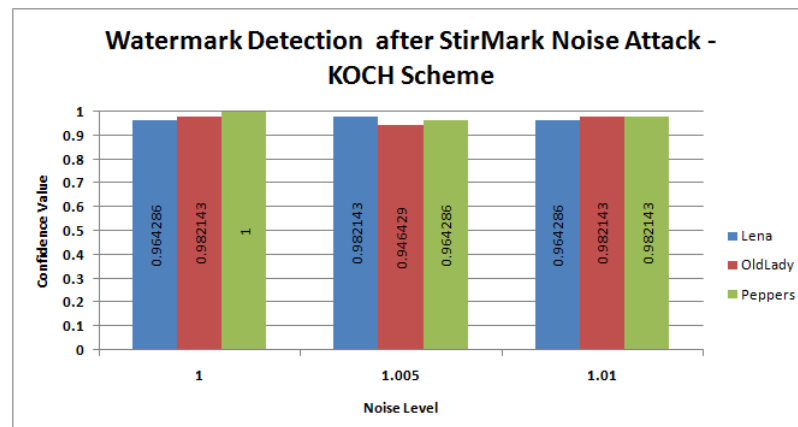


Figure 5.51: Noise Attack:KOCH Scheme

Appendix E

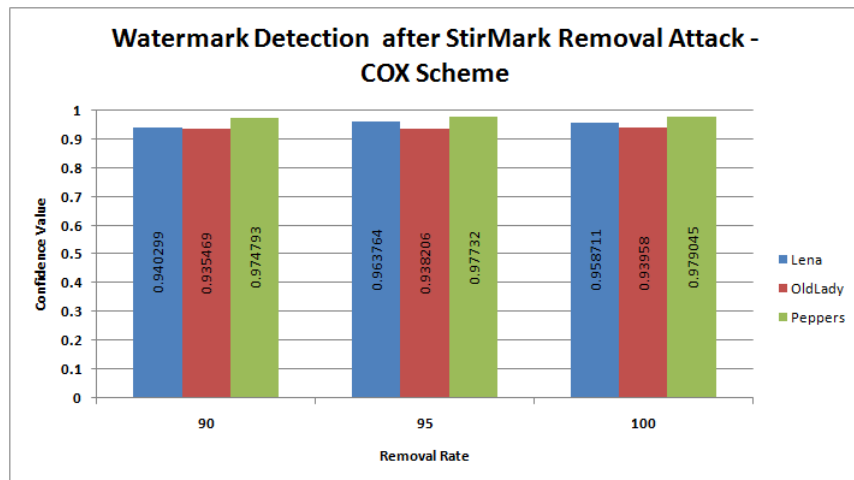


Figure 5.52: Removal Attack:COX Scheme

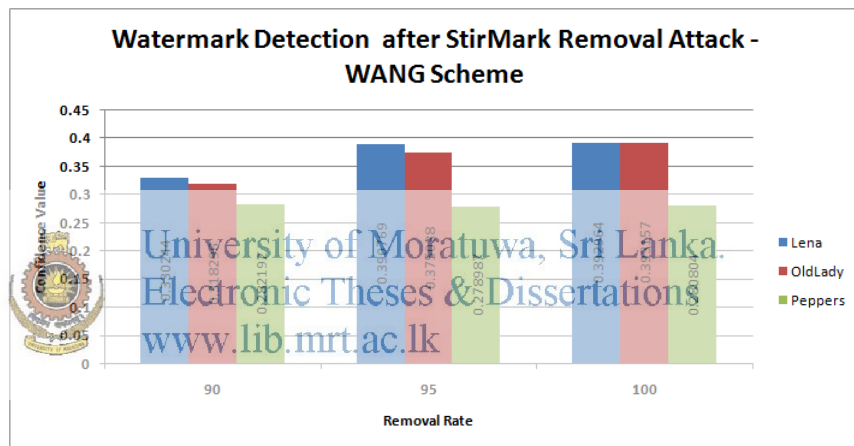


Figure 5.53: Removal Attack:WANG Scheme

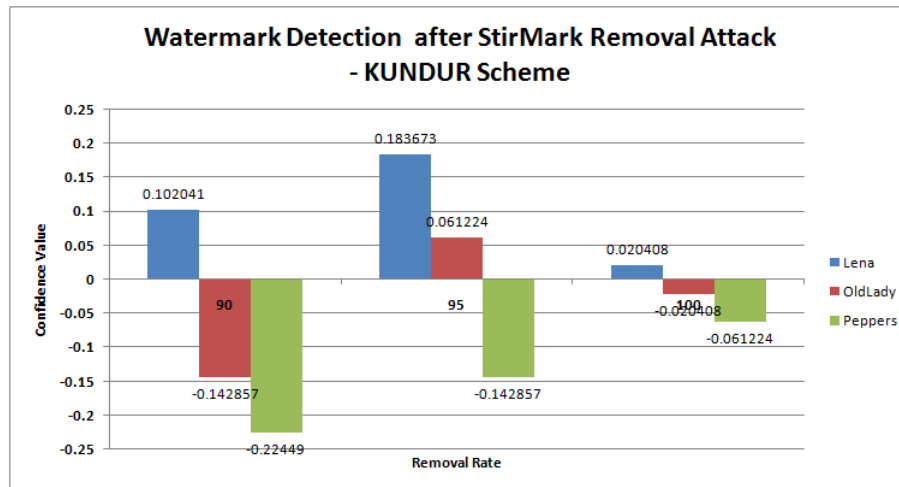


Figure 5.54: Removal Attack:KUNDUR Scheme

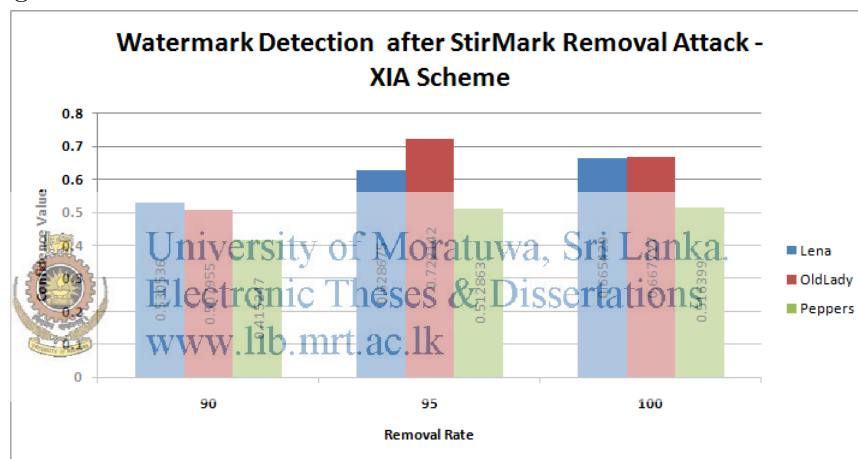


Figure 5.55: Removal Attack:XIA Scheme

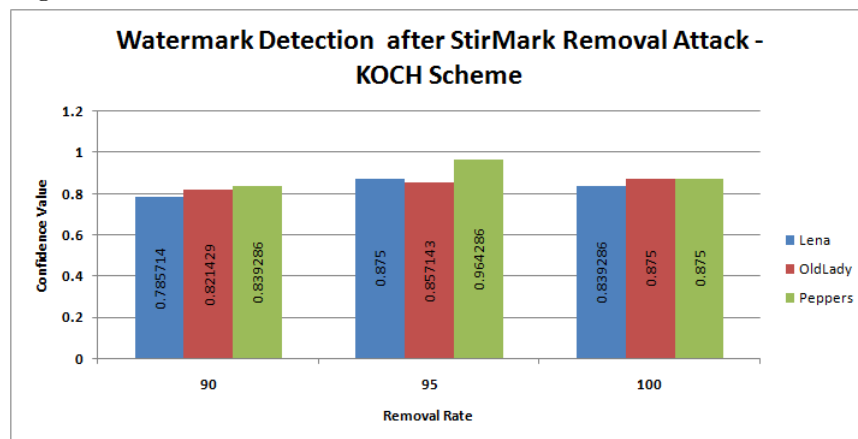


Figure 5.56: Removal Attack:KOCH Scheme

Appendix F

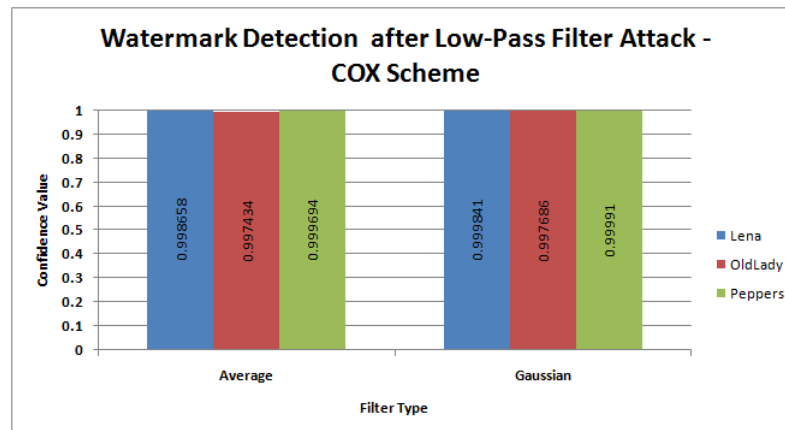


Figure 5.57: Filter Attack:COX Scheme

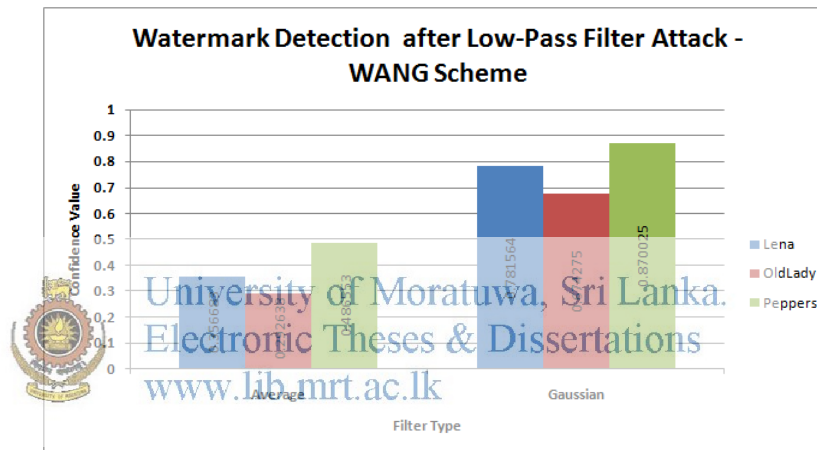


Figure 5.58: Filter Attack:WANG Scheme

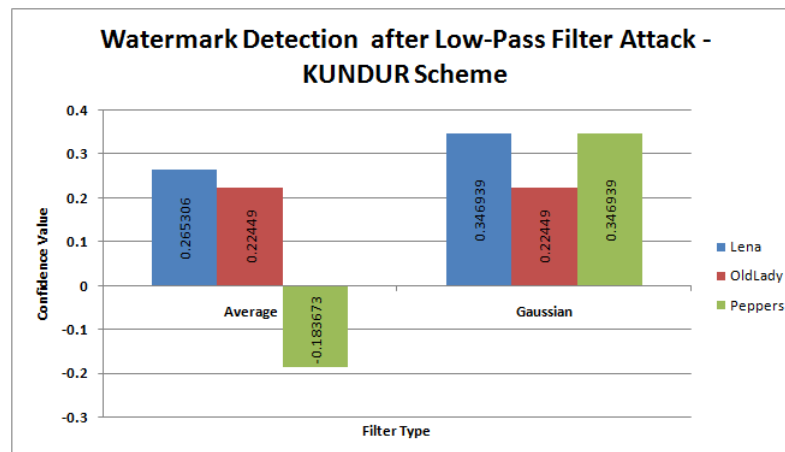


Figure 5.59: Filter Attack:KUNDUR Scheme

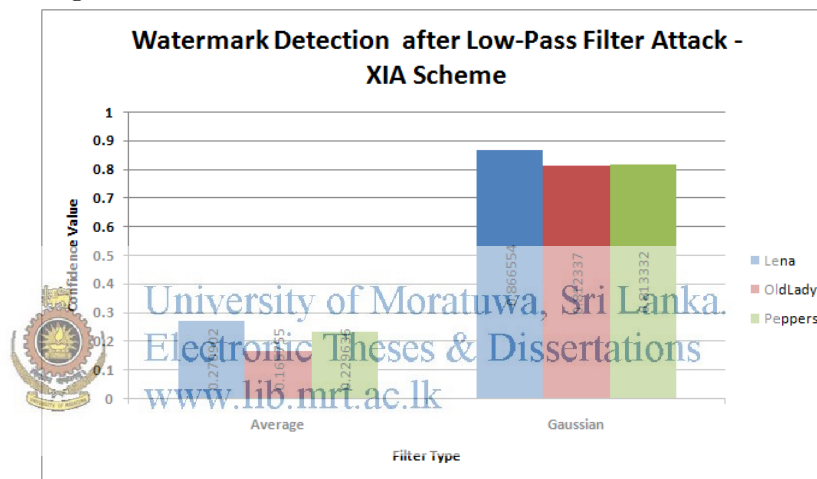


Figure 5.60: Filter Attack:XIA Scheme

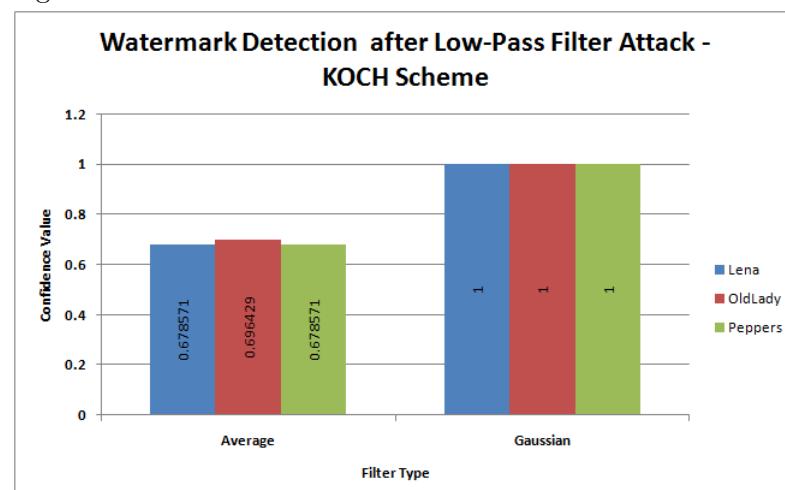


Figure 5.61: Filter Attack:KOCH Scheme

Appendix G

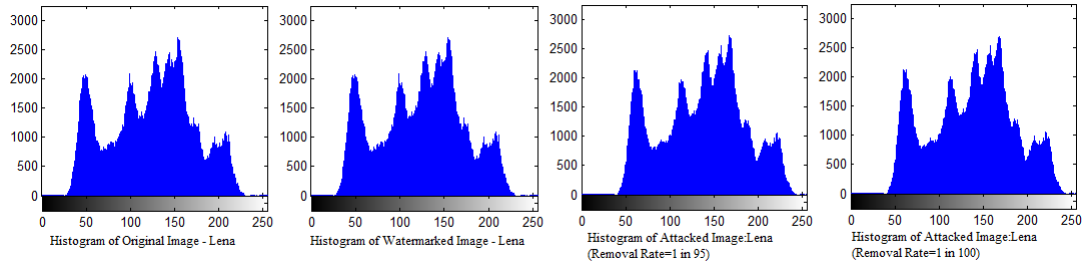


Figure 5.62: Histogram Comparison of Lena:Kundur(StirMark Removal Attack)

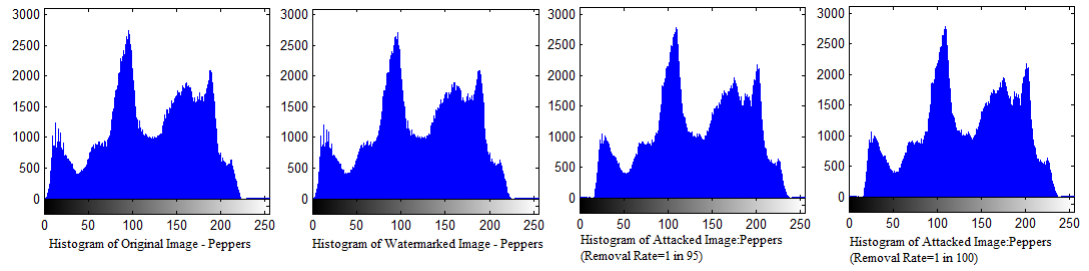


Figure 5.63: Histogram Comparison of Peppers:Kundur(StirMark Removal Attack)

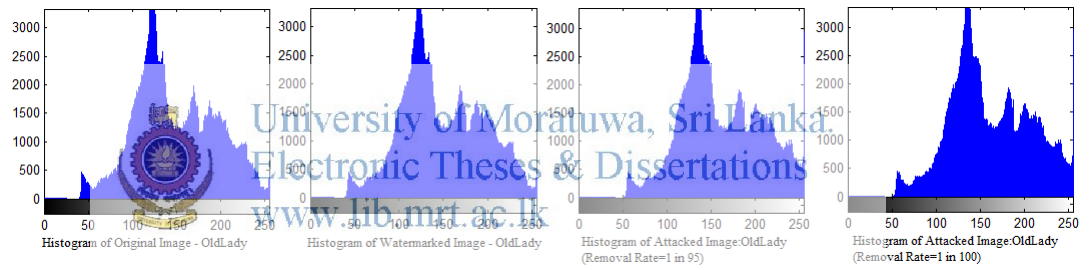


Figure 5.64: Histogram Comparison of OldLady:Kundur(StirMark Removal Attack)

Appendix H

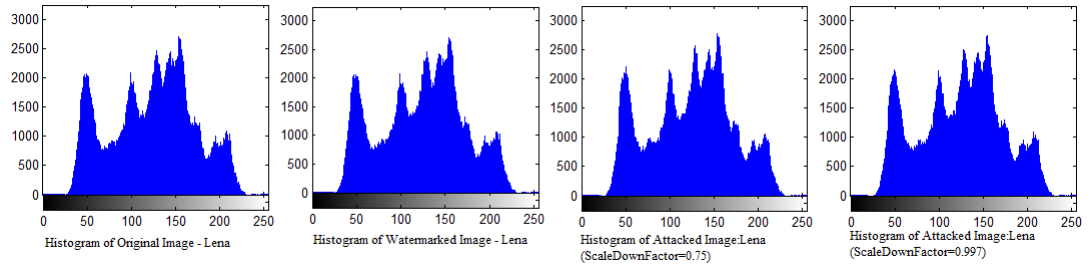


Figure 5.65: Histogram Comparison of Lena:Kundur(Scale Down Attack)

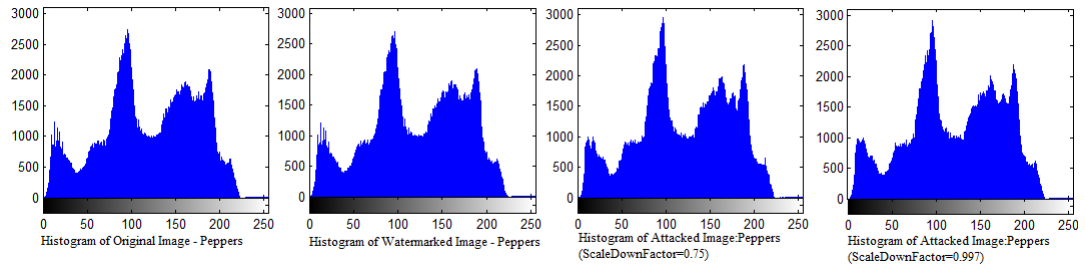


Figure 5.66: Histogram Comparison of Peppers:Kundur(Scale Down Attack)

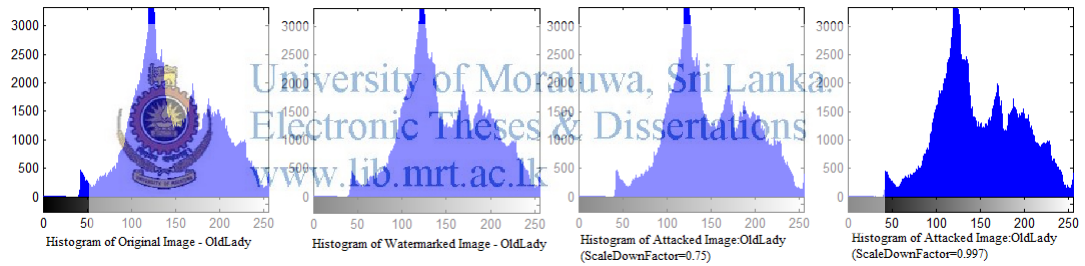


Figure 5.67: Histogram Comparison of OldLady:Kundur(Scale Down Attack)

Appendix I

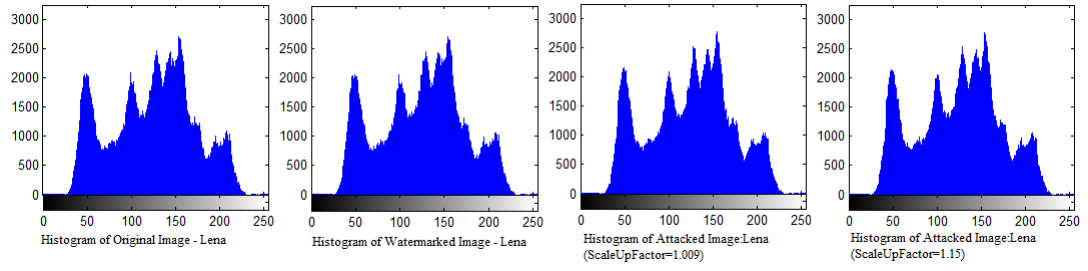


Figure 5.68: Histogram Comparison of Lena:Kundur(Scale Up Attack)

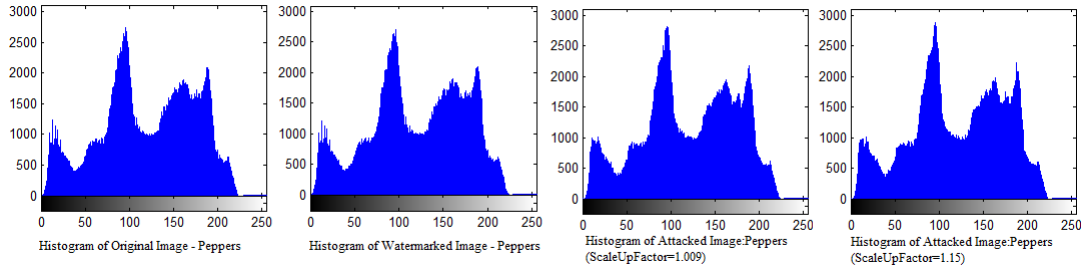


Figure 5.69: Histogram Comparison of Peppers:Kundur(Scale Up Attack)

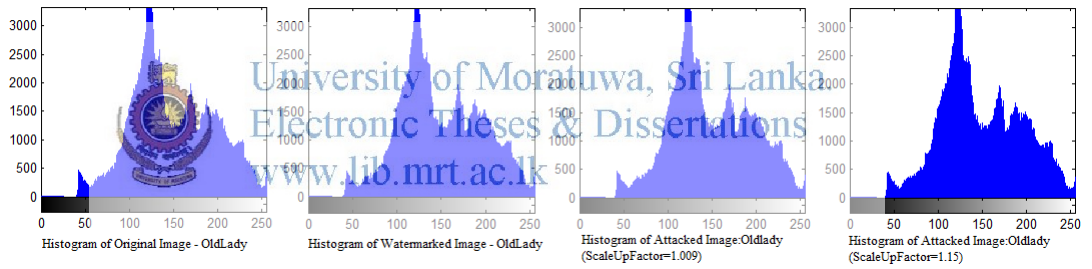


Figure 5.70: Histogram Comparison of OldLady:Kundur(Scale Up Attack)

Appendix J

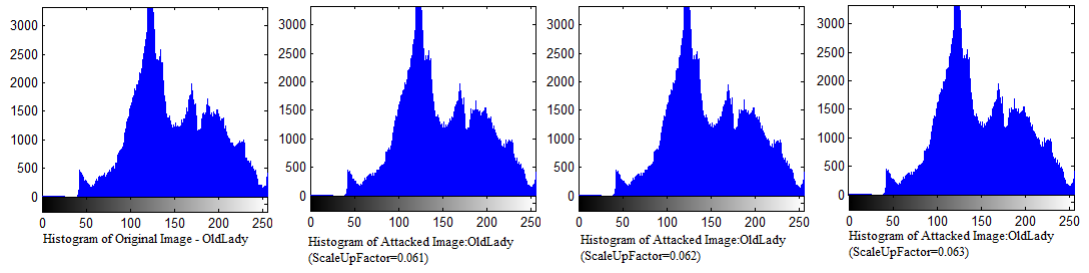


Figure 5.71: Histogram Comparison of Lena:Kundur(Rotate Attack)

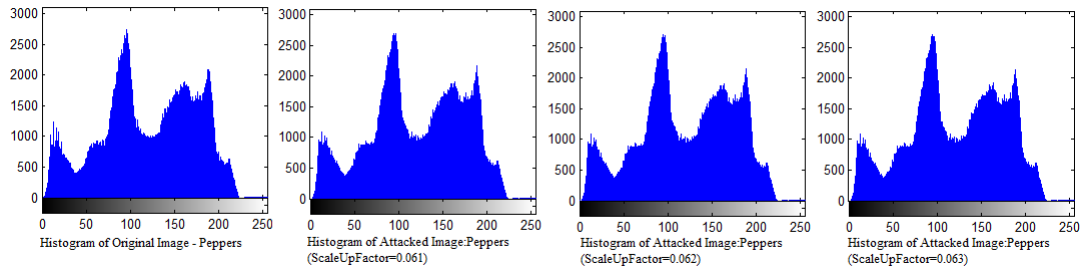


Figure 5.72: Histogram Comparison of Peppers:Kundur(Rotate Attack)

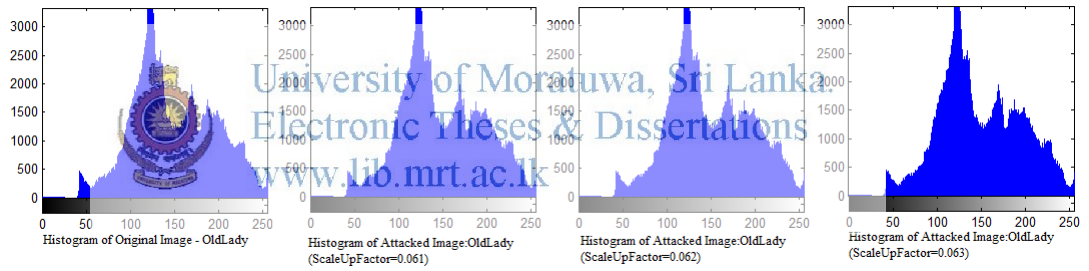


Figure 5.73: Histogram Comparison of OldLady:Kundur(Rotate Attack)

Appendix K

- Analysis of Flip Attack Results: All watermarking schemes were failed to detect the watermark

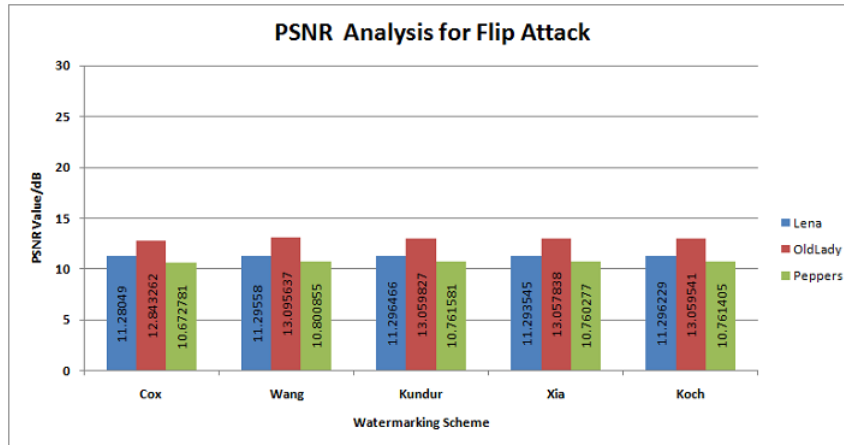


Figure 5.74: PSNR Comparison(Flip Attack)

- Analysis of JPEGCompression Attack:KUNDUR Scheme is failed to detect the watermark (JPEG Qulaity=80,90)

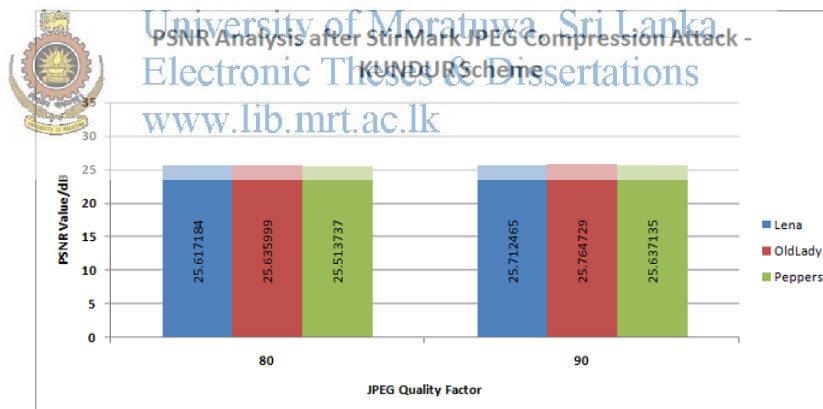


Figure 5.75: PSNR Comparison:Kundur(JPEG Attack)

- Analysis of FilterAttack Results: XIA and KUNDUR Schemes were failed to detect the watermark (Filter Type=Average)

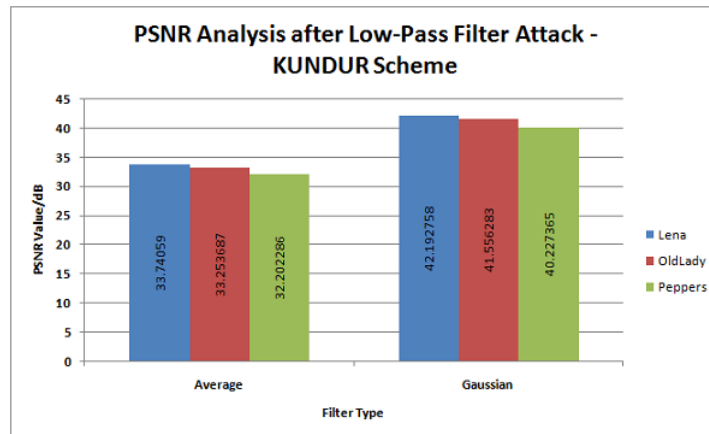


Figure 5.76: PSNR Comparison:Kundur(Filter Attack)

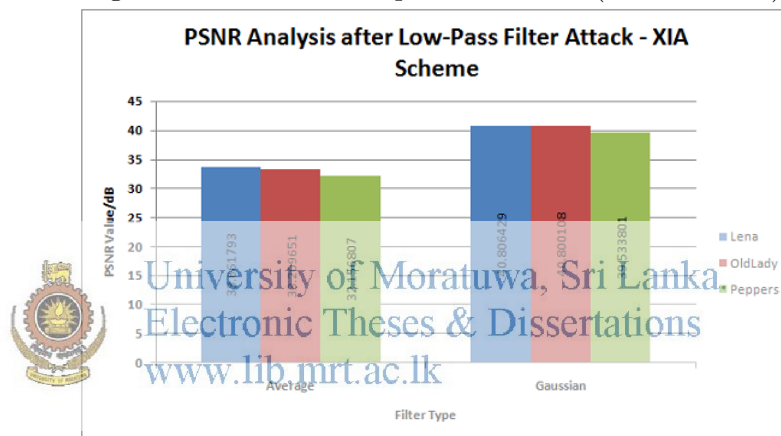


Figure 5.77: PSNR Comparison:Xia(Filter Attack)

- Analysis of Noise Attack-KUNDUR Scheme is failed to detect the watermark

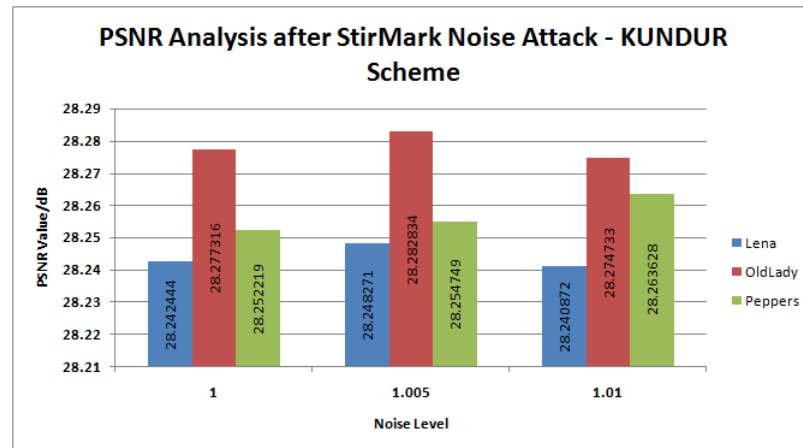


Figure 5.78: PSNR Comparison:Kundur(Noise Attack)

- Analysis of Removal Attack-KUNDUR Scheme is failed to detect the watermark

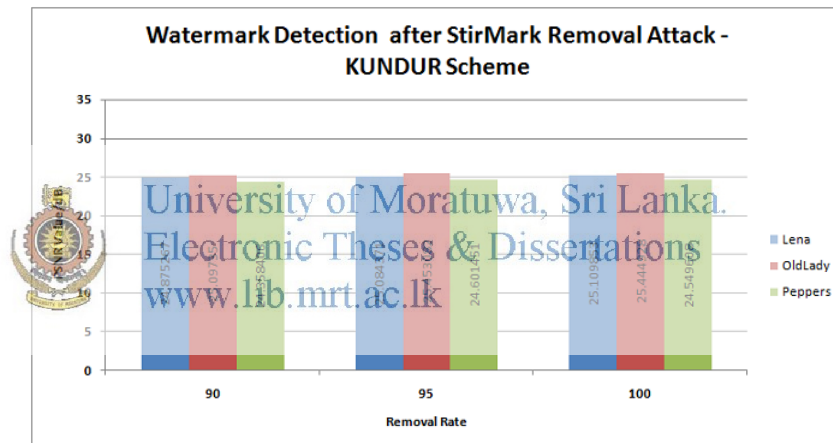


Figure 5.79: PSNR Comparison:Kundur(Removal Attack)

- Analysis of Resize ScaleDown/ScaleUp Attack-KUNDUR Scheme is failed to detect the watermark on certain scale factors

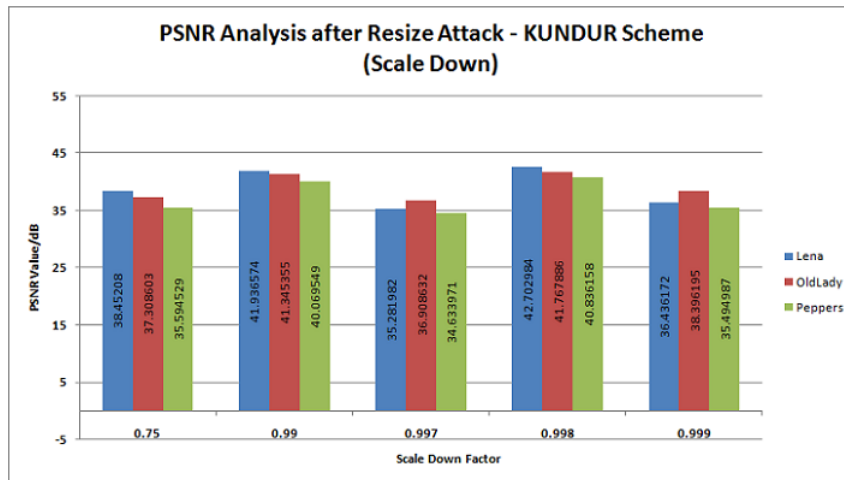


Figure 5.80: PSNR Comparison:Kundur(Scale Down Attack)

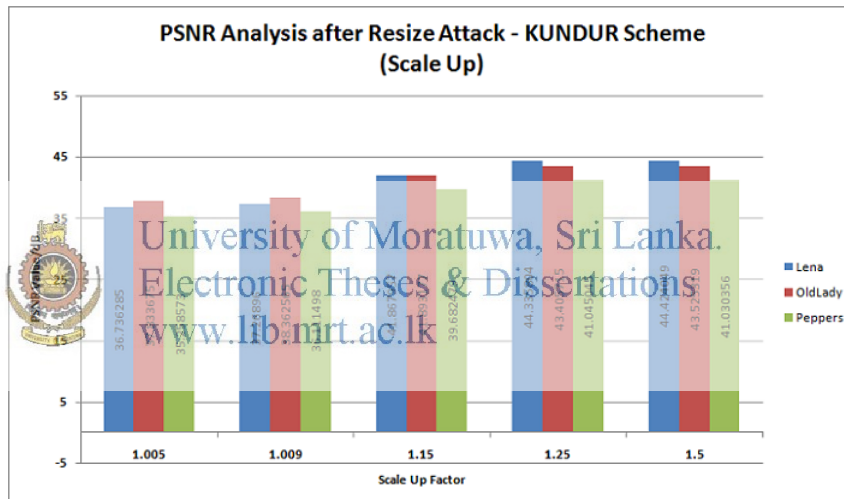


Figure 5.81: PSNR Comparison:Kundur(Scale Up Attack)

- Analysis of Rotate Attack: KUNDUR Scheme is failed to detect the watermark on certain rotation angles (Method:Bicubic)

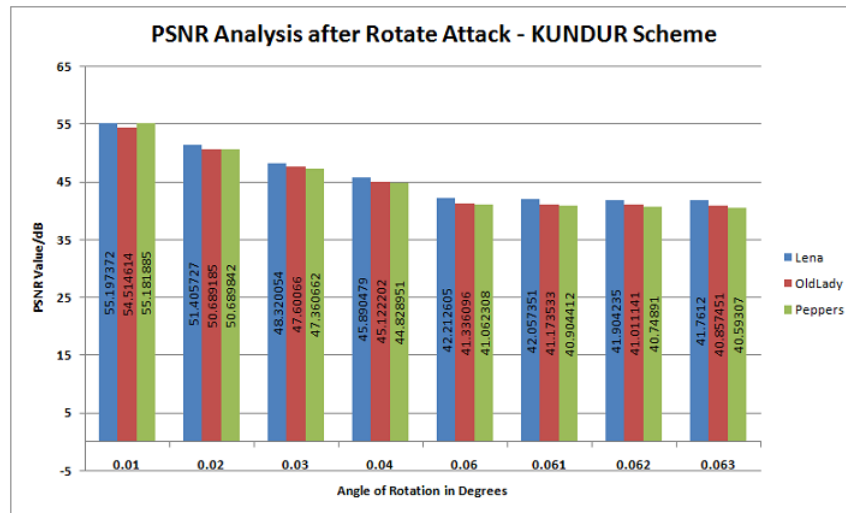


Figure 5.82: PSNR Comparison:Kundur(Rotate Attack)



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk