A HIGH CAPACITY REVERSIBLE MULTIPLE WATERMARKING SCHEME
- APPLICATIONS TO IMAGES, MEDICAL DATA, AND BIOMETRICS

by

Behrang Mehrbany Irany

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Graduate Department of Electrical and Computer Engineering
University of Toronto

Copyright © 2011 by Behrang Mehrbany Irany
Abstract
A High Capacity Reversible Multiple Watermarking Scheme - Applications to Images, Medical Data, and Biometrics
Behrang Mehrbany Irany
Master of Applied Science
Graduate Department of Electrical and Computer Engineering
University of Toronto
2011

Modern technologies have eased the way for intruders and adversaries to bypass the conventional identity authentication and identification processes; hence security systems have been developed to a great extent for protection of privacy and security of identities in different applications. The focus of this thesis is digital watermarking as a part of Digital Rights Management (DRM), security and privacy, as well as the ability to employ electrocardiogram (ECG) as a method to enhance the security and privacy level.

The contribution of this work consists of two main parts: An application-specific high-capacity reversible multiple watermarking scheme is introduced in the first part to mainly target the medical images. The proposed data hiding method is designed such that the embedding of sensitive personal information in a generic image without any loss of either the embedded or the host information is possible. Furthermore, in the second part, the use of ECG biometric signals in the form of the embedded watermark is studied. Proposed framework allows embedding of ECG features into the host image while retaining the quality of the image, the performance of the security system and the privacy of the identity. Experimental results indicate that the reversible data hiding scheme outperforms other approaches in the literature in terms of payload capacity and marked image quality. Results from the ECG mark embedding also show that no major degradation in performance is noticeable compared to the case where no watermarking is needed.
Acknowledgements

First, I would like to sincerely thank my thesis advisor Prof. Dimitris Hatzinakos, as it is impossible to overstate my gratitude towards him. This work would not have been achievable without his help, support, guidance and encouragement. I would have been lost without his help. I would also like to thank my defense committee members for taking the time to review my work and offer their insightful comments and suggestions. I wish to thank the Communication Group faculty members and staff for their assistance during my study at University of Toronto. Financial support provided for this research from the Department of Electrical and Computer Engineering and Prof. Hatzinakos is truthfully appreciated. Last but not the least, I wish to thank my family for their constant support and love without whom I would not be able to carry on and offer my regards and blessings to all of those who supported me in any respect during the completion of this project.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>viii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>xii</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Digital Watermarking and Steganography</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Motivation and Problem Description</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Thesis Outline and Contributions</td>
<td>4</td>
</tr>
<tr>
<td>2. History and Background</td>
<td>6</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>6</td>
</tr>
<tr>
<td>2.2 Digital Watermarking History</td>
<td>7</td>
</tr>
<tr>
<td>2.3 Digital Watermarking Systems and Models</td>
<td>8</td>
</tr>
<tr>
<td>2.3.1 Communications Channel Based Model of Watermarking</td>
<td>9</td>
</tr>
<tr>
<td>2.3.2 Geometric Model of Watermarking</td>
<td>9</td>
</tr>
<tr>
<td>2.4 Digital Watermarking Properties</td>
<td>11</td>
</tr>
<tr>
<td>2.4.1 Effectiveness</td>
<td>12</td>
</tr>
<tr>
<td>2.4.2 Fidelity</td>
<td>12</td>
</tr>
</tbody>
</table>
# A. Reversible Watermarking Background

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1 Fragile Watermarking Algorithms</td>
<td>120</td>
</tr>
<tr>
<td>A.1.1 Fragile Algorithms Operating in Spatial Domain</td>
<td>121</td>
</tr>
<tr>
<td>A.1.2 Fragile Algorithms Operating in Transformed Domain</td>
<td>139</td>
</tr>
<tr>
<td>A.2 Semi-Fragile and Robust Algorithms</td>
<td>148</td>
</tr>
<tr>
<td>A.2.1 Semi-Fragile Algorithms Operating in Spatial Domain</td>
<td>148</td>
</tr>
<tr>
<td>A.2.2 Semi-Fragile Algorithms Operating in Transformed Domain</td>
<td>152</td>
</tr>
<tr>
<td>A.2.3 Robust Algorithms Operating in Spatial Domain</td>
<td>155</td>
</tr>
<tr>
<td>A.2.4 Robust Algorithms Operating in Transformed Domain</td>
<td>159</td>
</tr>
</tbody>
</table>

# B. Integer to Integer Wavelet Transform

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.1 The Basics</td>
<td>161</td>
</tr>
<tr>
<td>B.2 Implementation of the Integer Wavelet Transform</td>
<td>163</td>
</tr>
</tbody>
</table>

# Bibliography

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliography</td>
<td>166</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Summary of the significant reversible watermarking schemes . . . . . . . 35

4.1 Error Equal Rate and the corresponding verification rates of the verifica-
        tion process . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 112

4.2 Error Equal Rate and the corresponding verification rates in case of no 
        data quantization . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 112

4.3 Identification rate of a single ECG feature vector . . . . . . . . . . . . . 114
# List of Figures

1.1 Block diagram of a generic watermarking (steganography) system . . . . 3

2.1 Block diagram of a communications channel . . . . . . . . . . . . . . . . 10

2.2 Block diagram of a watermarking system based on communications model 10

2.3 Geometric representation of a watermarking system . . . . . . . . . . . . . 11

2.4 Categorization of reversible watermarking algorithms based on robustness 23

2.5 Categorization of the significant reversible watermarking schemes . . . . . . 32

3.1 Illustration of histogram shifting . . . . . . . . . . . . . . . . . . . . . . . 43

3.2 Examples of the proposed coefficient maps . . . . . . . . . . . . . . . . 47

3.3 Algorithm to derive the stopping parameters . . . . . . . . . . . . . . . . 54

3.4 Watermark encryption using AES . . . . . . . . . . . . . . . . . . . . . . . 56

3.5 Encryption method used in the proposed algorithms . . . . . . . . . . . . . 56

3.6 Watermark encryption test . . . . . . . . . . . . . . . . . . . . . . . . . . . 57

3.7 Multi-pass single watermark data embedding algorithm . . . . . . . . . . . 58

3.8 Multi-pass single watermark data extracting algorithm . . . . . . . . . . . 59

3.9 Coefficient masks . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 60

3.10 Test images used in the experiment . . . . . . . . . . . . . . . . . . . . . . 68

3.11 Watermarked natural images at different rates . . . . . . . . . . . . . . . . 69

3.12 Watermarked natural image with 1bb rate . . . . . . . . . . . . . . . . . . 69

3.13 Performance of the proposed watermarking scheme on natural images . . 70
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>Autocorrelation</td>
</tr>
<tr>
<td>AES</td>
<td>Advanced Encryption Standard</td>
</tr>
<tr>
<td>AWGN</td>
<td>Additive White Gaussian Noise</td>
</tr>
<tr>
<td>bpp</td>
<td>bits per pixel</td>
</tr>
<tr>
<td>CA</td>
<td>Confusion Avoidance</td>
</tr>
<tr>
<td>CC</td>
<td>Correlation Coefficient</td>
</tr>
<tr>
<td>CER</td>
<td>Crossover Error Rate</td>
</tr>
<tr>
<td>CPTWG</td>
<td>Copy Protection Technical Working Group</td>
</tr>
<tr>
<td>CRIMS</td>
<td>Criminal Records Information Management Services</td>
</tr>
<tr>
<td>CS</td>
<td>Candidate Set</td>
</tr>
<tr>
<td>DCT</td>
<td>Discrete Cosine Transform</td>
</tr>
<tr>
<td>DE</td>
<td>Difference Expansion</td>
</tr>
<tr>
<td>DES</td>
<td>Data Encryption Standard</td>
</tr>
<tr>
<td>DRM</td>
<td>digital rights management</td>
</tr>
<tr>
<td>ECC</td>
<td>Error Correction Codes</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>ECG</td>
<td>Electrocardiogram</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalogram</td>
</tr>
<tr>
<td>EER</td>
<td>Equal Error Rate</td>
</tr>
<tr>
<td>EHR</td>
<td>Electronic Health Record</td>
</tr>
<tr>
<td>FAR</td>
<td>False Acceptance Rates</td>
</tr>
<tr>
<td>FRR</td>
<td>False Rejection Rate</td>
</tr>
<tr>
<td>IDCT</td>
<td>Inverse Discrete Cosine Transform</td>
</tr>
<tr>
<td>IHW</td>
<td>Information Hiding Workshop</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>IWT</td>
<td>Integer Wavelet Transform</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear Discriminant Analysis</td>
</tr>
<tr>
<td>LSB</td>
<td>Least Significant Bit</td>
</tr>
<tr>
<td>MR</td>
<td>Magnetic Resonance</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Square Error</td>
</tr>
<tr>
<td>OTP</td>
<td>One-Time Pad</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PGP</td>
<td>Point to Point Graph</td>
</tr>
<tr>
<td>PSNR</td>
<td>Peak Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>RCM</td>
<td>Reversible Contrast Mapping</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of Interest</td>
</tr>
<tr>
<td>RONI</td>
<td>Region of Non Interest</td>
</tr>
<tr>
<td>SDMI</td>
<td>Secure Digital Music Initiative</td>
</tr>
<tr>
<td>WQM</td>
<td>Weighted Quantization Method</td>
</tr>
<tr>
<td>ZRE</td>
<td>Zero-Replacement Embedding</td>
</tr>
<tr>
<td>ZRX</td>
<td>Zero-Replacement Extraction</td>
</tr>
</tbody>
</table>
Introduction

Protection of digital multimedia contents, as well as security and privacy have become a major challenge due to the recent advances in the fields of networking, and digital multimedia, as well as the readily availability of copying and manipulative devices and software programs to the public. In order to prevent fraud and counterfeit in the past decade, Digital Rights Management (DRM) has been the focus of many researches in both academia and industry. According to the MIT Technology Review, DRM was one of the top ten emerging technologies that would "change the world." DRM technologies are engaged to control the use of digital media by preventing end user's access, copying, distribution, manipulation or conversion to other formats.

Digital watermarking and biometric signals are two components of DRM, serving the above purpose as well as controlling the rights and privacy of the owners. Due to the limitations in designs and performance rates, digital watermarking has not been fully successful and widely used, although it has already been implemented in many applications.
1.1 Digital Watermarking and Steganography

In the past decade, application of digital multimedia contents has grown rapidly because of their advantages over analog contents. Ease of transferring and broadcasting over networks, higher quality and durability, online/offline easy editing, copying, and simplicity of archiving or storing are just a few advantages of digital multimedia over analog contents. Ironically, all the above advantageous properties have raised the main concerns in copyright management and privacy protection of such contents. Encryption methods such as conventional connection-based security systems cannot carry out the required proper protection level as it is impossible to monitor how a legitimate user handles the content after decryption, which makes it possible for hackers and adversaries to illegally redistribute or manipulate the content.

*Digital watermarking* and *Steganography* are methods engaged to address such problems. Watermarking is defined as the practice of imperceptibly altering a work to embed a message about that work. On the other hand, steganography is the practice of undetectably altering a work to embed a secret message [3]. These alterations are called as the mark or watermark, which carry informative data for authentication, identification, privacy protection and controlled access purposes. Even though the aims of watermarking and steganography are quite different, both applications share certain high-level elements. Both systems consist of an embedder and a detector, as shown in Figure 1.1. The embedder takes three inputs, the to-be-embedded payload (watermark), the cover work and the secret key for the protection of the payload. Embedder’s output is typically transmitted or recorded. At the detector side, the marked work is presented as an input to the detector. Most detectors try to determine whether a payload is present, and if so, extract the detected payload using the secret key. Studies in this field have been mainly focused on marking methods for still images, digital audio and video contents.
Figure 1.1: Block diagram of a generic watermarking (steganography) system.

1.2 Motivation and Problem Description

Cryptography is the principal technology used to protect the rights of the content owners and probably the most widespread developed scientific method of protecting digital multimedia content. Unfortunately, as already explained, encryption cannot help the content owners or the distributors monitor how the content is handled after decryption which may lead to illegal copying and distribution or misuse of the private information. Consequently, it is not an overstatement to say that cryptography can protect content in transit, but once decrypted, the content has no further protection. Hence, there is a strong need for an alternative or compliment technology to cryptography which can protect the content even after it is decrypted.

Watermarking technology seems to have the potential of fulfilling such a need as it embeds imperceptible information into the content which is never removed during normal usage or causes inconvenience to the users. A watermark can be designed to survive different processes such as decryption, re-encryption, compression and geometrical manipulations. There are a number of other applications for which watermarking methods may be developed, used, or suggested, although major driving forces behind the watermarking technology have been copyright protection and copy prevention. Examples of
such applications are privacy protection, identification, media file archiving, broadcast monitoring, and device control. Chapter 2 of this research will introduce the available and potential digital watermarking applications in more details.

In general, applications of digital watermarking can be divided into three categories of copy protection, content authentication and information hiding. Different applications use and emphasize different combinations of these properties to address the needs.

To better serve the goals of digital watermarking technology and strengthen the privacy and security aspects of these applications, human biometrics can be employed. Biometric traits are highly correlated to an individual, so instead of utilizing secret messages, passwords, or personal data to verify an identity, these inherent characteristics can be employed for recognition as they offer superb security.

1.3 Thesis Outline and Contributions

The objective of this research is to design and develop a reversible data hiding scheme for specific applications in security and privacy field and to introduce a framework to engage electrocardiogram (ECG) biometric signals as the embedding mark. Designed system in this research is aimed towards the medical field, military and satellite imaging, and deep space photography as the primary targets, since in these fields high perceptual quality of the marked images is a must; but it can also be applied to other environments under different conditions. This being said, it is impossible to design a watermarking scheme that works for all environments and resists all kinds of manipulations either intentional or unintentional.

The outline of the thesis is organized as follows in the subsequent chapters:

Chapter 2 converses the history and background of digital watermarking technology in general; identifies the structure and parts of current systems while it defines the properties of a generic watermarking scheme. The current applications of digital watermarking
technology and the applications which this thesis research focuses on are then introduced. Furthermore, reversible digital watermarking techniques within different literature are categorized and discussed in this chapter.

In chapter 3, a high capacity reversible data hiding scheme is proposed. This process is capable of embedding multiple watermarks at different stages by using different secret keys into the cover image. The scheme employs integer-to-integer wavelet transform to decompose the original image; subsequently the necessary space to embed the watermark is created in high frequency wavelet subbands by using the histogram shifting technique. The introduced scheme exploits a scalable coefficient map and different embedding conditions, which allow high data hiding capacity while retaining the acceptable visual quality for the use in sensitive applications such as medical imaging, military and satellite imaging, and deep space photography. The proposed method is applicable to the Electronic Health Record (EHR) programs such as Ontario government supported "Smart Systems for Health Agency" [4] to provide privacy and security of the medical files.

In chapter 4, a framework to embed electrocardiogram (ECG) biometric signals as a watermark using the proposed scheme in chapter 3 is presented. The algorithm uses the autocorrelation feature extraction method along with Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA) dimensionality reduction methods to extract the biometric traits. The framework proposes suitable quantization and bit allocation methods to convert the extracted components into binary format for the embedding process. This scheme has the potential to be used in applications related to EHR, personal file archiving, copyright protection and distribution monitoring.

Chapter 5 concludes the thesis and summarizes the research presented in this work along with advantages of each proposed scheme. Future directions of the presented thesis work are suggested in this final chapter along with other possible improvements in the applications and frameworks discussed throughout the thesis.
Chapter 2

History and Background

This chapter presents a general overview on digital watermarking technology and illustrates the requirements and characteristics of the watermarking systems. Several applications employing digital watermarking are discussed. The targeted applications of the current work are then introduced and discussed. Finally, reversible watermarking schemes within different literature are presented and categorized by the degree of robustness against modifications.

2.1 Introduction

Information hiding, Watermarking, and steganography are three closely related areas which have a lot of common characteristics and share many technical approaches. Information hiding, also known as data hiding, is a general term containing a broad range of problems beyond embedding messages in content. The term hiding can refer to either making the information imperceptible or keeping the very existence of the information secret, as in watermarking and steganography respectively. Steganography is the art of concealed communication where the existence of the message is secret, simple example of such a method is invisible ink. On the other hand, in watermarking the embedded message is directly related to the cover work or host signal. Using these definitions,
information hiding systems can be divided into four categories:

1. **Covert watermarking**: the message is related to the cover work and the existence is hidden.

2. **Overt watermarking**: the message is related to the cover work and its existence is known.

3. **Steganography (covert communication)**: the message is independent of the cover work and the existence is hidden.

4. **Overt embedded communications**: the message is unrelated to the cover work and its existence is known.

Watermarking as it is used today may refer to all four categories mentioned above. To address the intended applications properly, in this thesis watermarking refers to the overt watermarking and steganography categories. Furthermore, this thesis focuses on the still images as a part of the multimedia contents; all the reviews, discussions, and the proposed reversible watermarking scheme and biometric watermark framework principally target digital still images. Although, the suggested algorithms can be modified and adjusted to extend to the video contents.

### 2.2 Digital Watermarking History

The earliest examples of general notion of watermarks probably go back to the earliest cavitations. There are samples of watermarks left from 1499. In 1954, a patent was filed by Emil Hembrooke for "watermarking" musical works, where a simple identification code was inserted in music by intermittently applying a narrow notch filter centered at 1 kHz to represent the Morse code. Later in 1979, a machine detectable pattern was described by Szepanski; the aforementioned pattern could be placed on documents for anti-counterfeiting purposes. However, the term *digital* watermark was first used...
Chapter 2. History and Background

by Komatsu and Tominaga in 1988. Interest in digital watermarking was increased throughout 1995 and in 1996 it was included in Information Hiding Workshop (IHW) as a primary topic. Later, in 1999 the SPIE started devoting a conference specifically to Security and Watermarking of Multimedia Contents. Several other organizations began including watermarking technology in different standards. The Secure Digital Music Initiative (SDMI) adopted watermarking as a central component to their music protection system. The Copy Protection Technical Working Group (CPTWG) considered watermarking technology for video content protection on DVDs. The International Organization for Standardization (ISO) showed an interest in watermarking for designing MPEG standards. VIVA and Talisman, both sponsored by European Union, employed the technology for broadcast monitoring. It was in late 1990s that companies using digital watermarking were established. An example of such companies in the area of image watermarking is Digimarc, which bundled its watermark embedders and detectors with Adobe’s Photoshop. Since after, there has been a huge interest in the field of digital watermarking and several different techniques are proposed through all these years. There are more than 900 annual publications in the area of watermarking and steganography by IEEE since 2004. All the emerged approaches are aiming towards the same target but in different areas, ”marking digital multimedia content with the lowest quality degradation possible.”

2.3 Digital Watermarking Systems and Models

There are two main conceptual models of watermarking systems; these models help clarify the actual watermarking systems and the ways they operate. The first model is based on a view of watermarking as a method of communications like in communications channels, and the other model is based on the geometric views of watermarking methodologies.
2.3.1 Communications Channel Based Model of Watermarking

To define the structure of the watermarking systems, communications channel model can be considered due to the fact that watermarking is, in essence, a form of communication. Figure 2.1 illustrates the basic elements of a communications channel model. The encoder creates the symbols from the input message and transmits them across a noisy channel, then at the receiver side the encoder reconstructs the original message from the received noisy transmitted symbols. To ensure the security of the model and the transmission secret keys can be employed at the encoder and decoder. The same generic model can be adopted to illustrate a digital watermarking system. In a digital watermarking model the to-be-embedded watermark can be considered as the input message, the cover media plays the same role as the noisy channel and the detector has the same function as the decoder [3]. The challenge is to design and develop encoders and decoders that lead to correct detection and extraction of the watermark at the receiver, this requirement implies that the watermark and the cover signal are independent of each other. However, the dependence of the embedding algorithms in some methodologies on the cover signal suggests that the encoder should employ side information of the cover signal in the embedding process. Examples of methodologies using side information in the encoder, are Cox’s spread spectrum watermarking [14] and Kundur’s multi-resolution watermarking [15]. Figure 2.1 shows the adopted communications channel model used for generic digital watermarking systems.

2.3.2 Geometric Model of Watermarking

Watermarking algorithms can be conceptualized in geometric terms besides the communications channel model mentioned earlier. To present a watermarking system geometrically, the cover image is considered to be a point in a high dimensional space, the media space. Within this space different probability distributions and regions of interest can be
defined as follows [3]:

- The *distribution of un-watermarked works* shows how possible each work is.

- The *region of acceptable fidelity* is a region where all works seem basically the same to a given work. All the signals in this region are considered to be identical to the original signal.

- The *detection region* describes the behaviour of the detection algorithm.

- *Successfully watermarked versions of the work* are basically the intersection of the acceptable fidelity region and the detection region.

- The *embedding distribution* or the *embedding region* describes the effects of an
Figure 2.3: Geometric representation of a watermarking system.

The challenge in the design of such systems is to develop the detection region to yield robust and effective watermarking algorithm. Watermarking schemes aim to survive specific modifications or signal processes are the ones fitting in this category.

2.4 Digital Watermarking Properties

Watermarking systems can be characterized by a number of properties. The importance of each property is relevant to the requirements of the application and the service the watermarking method offers. In this section the most common properties of a digital watermarking scheme are highlighted. There are properties associated with embedding process such as effectiveness, fidelity, and payload, and there are those associated with
detection process such as blind or informed detection, false positive behavior, and robustness. Security and watermark keys are integrated parts that ensure the protection of the watermark and the content.

2.4.1 Effectiveness

The effectiveness of a watermarking system is defined as the probability that the output of the embedder is watermarked, or in other words, the effectiveness is the probability that the embedded mark is detectable immediately after embedding process [3]. Although 100% effectiveness is always desirable, the definition indicates that a watermarking system might have effectiveness less than 100%. Getting full effectiveness often imposes very high cost with respect to other properties, so in some cases watermarking schemes targeting specific applications might sacrifice some effectiveness to achieve better performance in other characteristics, such as fidelity, security or robustness. For example the watermarking scheme proposed in chapter 3 is effective and appropriate for grayscale images.

In some cases, the effectiveness can be determined analytically, but in most of the schemes this property is estimated empirically by embedding a watermark in a large test set of images. Given sufficiently large test set of images a watermarking scheme targets, such as grayscale or color, leads to a good estimation of effectiveness characteristic of the scheme.

2.4.2 Fidelity

The perceptual similarity between the original signal and the watermarked version of it defines the fidelity of a watermarking system. The fidelity measure depends on the embedding process and the transmission of the marked signal. In the case of a watermarked video content transmitted using NTSC standard, due to relatively low quality of the broadcast technology, channel degradations may let the difference between the original
and watermarked signals become imperceptible. But in case of high quality signals such as HDTV and DVD video, it is undesirable to have perceptual distortions; hence much higher fidelity watermark systems are required. There are cases that mildly perceptible watermarks are accepted in exchange for higher robustness or lower cost.

The best method to evaluate fidelity of a watermarking system is based on examining both original and watermarked version of the work by human subjects [16], but due to high volume of test sets such an evaluation method is impractical. The most common evaluation method used in all the literature is the peak signal-to-noise ratio (PSNR) defined between the host and watermarked signals.

2.4.3 Data Payload (Embedding Capacity)

Data payload refers to the number of bits a watermark system embeds within a unit of time or within a unit of cover signal. In photographs, the number of bits embedded into the image is referred to as the data payload and it is usually expressed in bits of information embedded per host image pixel, bits per pixel (bpp). In audio and video, the measures are number of embedded bits per second and the number of bits per frame or second, respectively. A watermarking scheme that embeds N bits into the cover signal is referred to as an N-bit watermarking system.

The required data payload may differ greatly for each application. Copy protection or copy control applications may require only a few bits of information received over a period of few seconds or minutes in case of audio or video respectively, and for images maybe only a secret message or a name string does the whole job. However, in other applications such as broadcast monitoring this rate might be three times larger than the previous case, or in case of forensic applications the necessary to-be-embedded information should be complete enough to prevent any modification of the content. Practically a watermark system can embed as much information as possible into the cover signal, up to the saturation point, in exchange of fidelity loss. A common method to evaluate
the performance of the watermarking systems is to observe the fidelity characteristic at different data payload sizes, in case of images this evaluation method is the plot of fidelity (PSNR) versus data payload (bpp).

2.4.4 Blind or Informed Detection

Applications where the original signal or a part of it is available during watermark detection are referred to informed detection methods (private watermarking systems). This method sometimes substantially improves detector performance as the original version can be subtracted from the marked copy to extract the watermark pattern alone. Blind detection (public watermarking system) refers to applications in which detection must be performed without any access to the original signal, as in copy control application. This property of a watermarking scheme is critical in determining if the method is suitable for a given application.

2.4.5 False Positive Rate

False positive rate refers to the detection probability of a watermark in a signal in which no mark is present, or in other words, this rate is the probability that given a specific watermark and randomly selected host signals, the detector reports the presence of the watermark. As with other properties, the required false positive rate depends on the application the scheme is intended for. In copy control applications, if an un-watermarked content consistently generates false positives, it could cause serious trouble, so in such a case the rate is expected to be infinitesimal.

2.4.6 Robustness

Robustness is defined as the ability to detect the watermark after common signal processing operations. Examples of such operations are lossy compression, spatial filtering,
additive noise, and geometric distortions. It is impossible that a watermarking scheme can survive all the signal processing operations, and yet not all applications require robustness against all such operations. Therefore, robustness requirements are application dependent. For example in the case of video broadcasting, the watermarking system used should show good robustness against lossy compression, digital-to-analog conversion, additive noise, and small horizontal and vertical changes, but on the other hand the robustness against scaling or any other degradations that occur prior to the embedding or after the detection of the watermark is not required. In some applications, it is irrelevant or even undesirable to have robustness in any sense. This need is the base of an important branch of watermarking research that focuses on fragile watermarking schemes. Fragile schemes are those that do not tolerate any kind of signal processing operations, as such operations cause the watermark to be lost. Usually fragile methods gain other properties in exchange, such as higher data payload, better security or better fidelity.

The reversible watermarking scheme introduced in chapter 3 is an example of fragile systems. The proposed scheme benefits from high fidelity and data payload along with desirable security in exchange of being fragile. This scheme can be employed in sensitive or confidential applications where even small degradations in the marked images are not acceptable or allowed and might lead to serious issues; consequently, the operating environments are either secure or private. Examples of such applications are the ones dealing with medical, deep space and military images within institutes or organizations.

2.4.7 Security

The security in a digital watermark system is defined as the ability to resist any intentional process or attack intended to destroy the watermark’s purpose. These attacks can be divided into three categories, unauthorized removal, unauthorized embedding and unauthorized detection of the watermark. Unauthorized removal and embedding which
modify the host signal are called active attacks in contrary with unauthorized detection which is referred to as a passive attack. Similar to other properties, the relative importance of these attacks depends on the application the watermark system is designed for. There are even instances in which the watermark does not need to be secure against any type of attack.

Unauthorized removal attacks aim to destroy or remove the embedded watermark so it cannot be detected at the receiver. There are two different kinds of such attacks, elimination attacks and masking attacks. Elimination of a watermark refers to a condition in which the detector cannot locate the watermark at all. The intention here is to make a new signal that is perceptually similar to the original one but cannot be considered to contain a watermark. This does not imply that elimination attacks reconstruct the original signal. Masking of a watermark is the condition in which the attack runs the watermark in a way it is not detectable by the existing detectors, but more sophisticated detectors might be able to detect the watermark.

Unauthorized embedding or forgery is the attack where the adversary inserts a legitimate watermark in an illegal media or an illegitimate watermark into a media that should not contain it. These types of attacks require the adversary to have enough knowledge about the watermark, watermarking scheme and also the security keys associated with the system.

Unauthorized detection or passive attacks can be categorized into three different levels of severity. The case that an adversary detects and deciphers an embedded watermark is the most severe case of all. The case in which the hacker detects and distinguishes the watermark but cannot decipher it is the less severe level. Finally, the least sever form is when the adversary is only able to determine that the mark is present, but can neither decipher the message nor distinguish the actual mark.
2.4.8 Watermark Keys

In most of the applications the watermarking algorithms introduced, evolved and used cannot be kept secret, so there should be other methods to secure the watermarking scheme against forgery and counterfeit. Cryptography provides the necessary security level to the watermarking systems; this security is derived from securing only the key used in the encryption algorithm, and not the entire algorithm. Secret keys are arbitrary sequences of bits that determine how the messages are encrypted. In most of the watermarking systems the watermark or the cover signal is encrypted using a given key and can only be decrypted by the same key. In such systems the embedding method depends on the secret key employed and in the detection step, a matching key must be used to detect the marks. In this case if the secret key used in the process is compromised, only a new secret key should be selected and there is no need to change the entire algorithm. As cryptography only deals with the prevention of unauthorized reading and writing of the message, it can only be used in watermarking systems to secure them against passive attacks and forgery and it cannot be used in case of watermark removal attacks. There are various methods similar to spread spectrum communications cryptography which can be adopted to address this issue [3]. The very presence of the watermark message without any knowledge of the key is impossible in such algorithms, even if the watermarking scheme is known.

2.4.9 Multiple Watermarks

In some applications it is desirable and necessary to embed multiple watermarks at different instances. Multiple watermarking systems allow several independent watermark embedding and extraction instances in different periods of time. This property is crucial in applications where there is a need to embed several watermarks used for different purposes. In copy control applications, multiple watermarks are used to represent the
number of times a work can be copied or modified\cite{17}. In transactional watermarking services each intermediary might require to embed a specific watermark before handing it over to the end user, such as music distribution networks. Another viable example would be in medical and health care services where multiple information including the personal and medical data of the patients are required to be embedded into the medical images by different personnel using different secret keys.

2.5 Watermarking Applications

Watermarking can be employed in a variety of applications where there is a need to associate certain information with a multimedia content or signal. Watermarking is different from other techniques used to serve the same purpose in three important ways: first, watermark is imperceptible, second, it is inseparable from the host content, and finally it undergoes the same transformations and translations as the cover content. These attributes are the main reasons employing watermarking is invaluable in some applications. In recent years watermarking is engaged in many applications such as owner identification, proof of ownership, broadcast monitoring, transaction tracking, privacy protection, copy control, authentication, and legacy enhancement.

This section only discusses the applications the proposed schemes in the following chapters are intended to aim: privacy protection, owner identification and authentication.

2.5.1 Privacy Protection

Recently various governments have been promoting the *Electronic Health Record (EHR)* which is collecting the health information of individual patients in the electronic form and sharing it across different health care settings. This approach allows computerizing and simplifying workflow in health care system, increases safety through evidence based decision support, eases quality management, and outcomes reporting; which in
turn facilitates diagnosis of patients and education of medical care personnel. All the aforementioned records contain sensitive personal information which should be protected properly. In majority of medical records, personal information is stored in separate files which may be accessed when needed. The current system enables adversaries or intruders to obtain access to patients’ personal information which increases the probability of the data being misused. A case of personal identification theft at the University Health Network in Toronto\cite{18}, drew attention to privacy protection and access control concerns.

Exploiting the properties of watermarking mentioned earlier can profoundly address such issues as the private personal information can be embedded into the health record files (images or scanned records). Due to low visual degradation, health care personnel and physicians, who are not required to have the authority to access the detailed personal information, can still use the marked files. In health care system it is crucial to have the original copy of the medical images and records with no degradation in quality, hence non reversible digital watermarking systems may not be appropriate. Coatrieux\ et al.\cite{19} examined the relevance of watermarking for medical images. They concluded that digital watermarks could be used in addition to the current security tools, such as firewalls and encryption, in order to better protect medical records. Giakoumaki et al.\cite{20,21} proposed a wavelet transform based watermarking technique for medical data. They were able to address medical data authentication, archiving, and retrieval along with source and data authentication. They could embed multiple watermarks containing patient’s personal and examination data, keywords for information retrieval, physician’s digital signature for authentication, and a reference message for data integrity control in ultrasound images. The main drawback of this method is that the original medical records or images are overwritten and the original images cannot be retrieved from the watermarked ones. Although the physicians participated in their tests did not notice any distortions in the images, however, further information embedding can cause distortion that negatively impacts the resolution of the medical images which may be
unacceptable in diagnosis. Chapter 3 proposes a reversible watermarking scheme which can be employed in applications where the original image should be available such as military, satellite, deep space, and medical imaging.

2.5.2 Owner Identification

Textual owner tags have several limitations and disadvantages. It is easy to remove these signs from a content even in cases that no intention is involved, textual tags degrade the cover image aesthetically, and may cover some parts of the image. However, they can be less obtrusive if positioned properly on the image. Due to imperceptibility and inseparability characteristics of watermarks, they are more suitable for owner identification or identification in general than textual tags.

The two most important factors in medical care systems are privacy protection and accuracy of data. As medical records storage and archiving takes place in all health centers around the world, medical cases may become lethal due to errors caused by misplacement of test results or records. In such environments digital watermarking can optimize issues of identification in medical file archiving. Incorporating patient’s or individuals’s identification features such as biometrics can help automate, secure, and unify the process.

Other Identification applications may also employ watermarking techniques along with individual biometrics to address issues in the current systems. Military organizations and criminal records centers, such as Criminal Records Information Management Services (CRIMS) in Canada, may benefit from the advantages of such systems. The reversible watermarking scheme proposed in chapter 3 and the framework introduced in chapter 4, address the application of Owner identification.
2.5.3 Owner Authentication

Owner authentication application is similar to owner identification with an exception that its objective is to verify the identity of the owner in order to validate or reject a claim. Identity theft is a serious concern and is of great importance especially in handling sensitive, secure and private information. Other examples of owner authentication applications are copyright claims and proof of ownership. The proposed watermarking scheme in chapter 3 combined with framework explained in chapter 4 can serve cases of great need for identity verification. It is imperative to note that the reversible watermarking method proposed can be used in environments that network is not open to public, hence there is no chance of content modification post embedding process and prior to watermark extraction.

2.6 Reversible Digital Watermarking

Reversible watermarking techniques are also referred to as invertible or lossless data hiding schemes and were originally born to be engaged mainly in situations where the authenticity of a digital image has to be granted and the original content is definitively needed at the decoding side. It is important to point out that, originally, a high perceptual quality of the watermarked image was not a requirement due to the fact that the original copy was retrievable; hence, other problems such as overflow and underflow caused by the watermarking process were not taken into account either. In addition, by employing reversible watermarking the access to the original content can be controlled, and only the authorized person can access the original content by erasing the watermark while the watermarked content is available to everyone. Successively, this aspect of these schemes are considered the basis to permit the end user to operate on the watermarked image and to possibly decide to recover the original version at a later time if needed. This flexibility in operation is important within different applications in which reversible watermarking
is essential. Examples of such application are military and satellite imaging, deep space photography, and medical imaging. For instance, in cases where watermarking is deployed on a medical image to secure the privacy of the patients, the physicians or related entities should have access to high quality watermarked medical images to avoid any false or wrong diagnosis that may endanger the health of the patients. Also in most applications, such as data storage and transmission in sensitive fields, it is of great interest to keep the watermark embedded as long as possible in order to continuously protect the information; this means that if the watermark is removed, or at least the part which secures the integrity of the information is extracted, the image is no more protected just like the case of data encryption \[22\]. Hence, having high perceptual quality in the marked image while retaining the properties such as reversibility and high capacity is extremely important in reversible watermarking schemes.

Reversible watermarking is the method to provide the three mandatory security characteristics in different applications \[23\]. These characteristics are: \textit{Confidentiality}, which ensures that only the entitled and eligible users have access to the information; \textit{Availability}, that is the ability of an information system to be accessible; and \textit{Reliability}, which is based on the \textit{integrity} and \textit{authenticity} of the information. Integrity of the information ensures that the data have not been modified by unauthorized people; authenticity, on the other hand, provides the proof that the information relates to the correct person and issued from the reliable source.

Reversible watermarking algorithms can be subdivided into two main categories \[24\], \textit{Fragile} and \textit{Semi-Fragile} as illustrated in Figure \[2.4\]. Majority of the developed techniques belong to the family of fragile watermarking schemes, which implies that the embedded watermark disappears or is corrupted if a slight modification occurs on the watermarked image thus, revealing that data integrity has been compromised. In semi-fragile watermarking schemes, it is essential that the embedded watermark be able to tolerate possible unintentional processes which may be applied to the host image, such
2.7 Reversible Watermarking Background

In this section, different methods and algorithms developed throughout the years to tackle the issues in the area of the reversible watermarking are presented. A complete review of the following reversible watermarking methods are presented in Appendix A where each method is explored and discussed in detail.

Reversible watermarking was first introduced by Mintzer et al. \cite{25} in 1997. They embedded a visible watermark which could have been removed from the original media.
Other early methods were mainly based on 256 modulo addition which introduces "salt-and-pepper" effect in the cover images [26].

2.7.1 Fragile Watermarking Algorithms

The bulk of the literature and published works in the field of reversible watermarking is on fragile watermarking algorithms. Being a fragile watermark implies that the information embedded in the original media is not recoverable or readable as soon as the watermarked signal is modified or altered. Consequently, once the embedded watermark information is lost, the original data is not recoverable. Furthermore, fragile watermarking techniques can be divided into two subdivisions, spatial domain and transformed domain techniques.

2.7.1.1 Fragile Algorithms Operating in Spatial Domain

This section presents the main and significant fragile reversible watermarking techniques operating in the spatial domain.

One of the most significant works in this area was done and presented by Tian [27]. The technique can achieve high payload capacity, high visual quality, and reversible data embedding properties in digital grayscale images. The methodology is based on the calculation of differences of neighboring pixel values and upon selection of certain differences, the difference expansion (DE) is executed. There are a number of drawbacks with this method. In Tian’s algorithm it is impossible to evaluate whether embedding of a certain payload is feasible prior to the actual embedding process starts, as the location map itself depends on the expansion coefficients used. Moreover, the lossless compression of the location map and LSB plane imposes a great cost, which leads to a significant larger payload. As a result of such an increase, the watermarked image has the effect similar to mild sharpening in the mid tone regions even in the case of relatively small watermark data.

Later, Tian’s method is generalized and extended by Alattar [28]. In his scheme,
instead of using the Haar transform difference expansion applied to pairs of pixels to embed the watermark bits, Alattar employed difference expansion method on spatial and cross-spectral triplets of pixels in order to increase the capacity used for embedding; the proposed algorithm embeds two bits in each triplet. An advantage of Alattar’s method over Tian’s is the possibility to control the size of the generated payload by adjusting the threshold values.

Furthermore, Alattar proposed an extension of the previously mentioned algorithm in order to embed triplets of bits in the difference expansion of quads of adjacent pixels; where quads are defined as $1 \times 4$ vectors containing the pixel values from different locations within the same color component of the host image. Even though this method shows better results comparing to those of earlier works, it has a higher computational complexity.

Finally, Alattar proposed a further generalization of his algorithm in which the difference expansion of vectors composed by adjacent pixels are used. This method increases the overall embedding capacity and the computational efficiency. This technique is capable of achieving the embedding rate of 1 bpp and can be employed recursively to increase the embedding capacity.

Ni et al. presented a reversible data hiding algorithm which utilizes the zero or the minimum points in the histogram of an image, in the spatial domain, to embed data by slightly modifying pixel values. The lower bound of the PSNR achieved employing this scheme is larger than 48 dB. An advantage of this scheme is the low computational complexity.

Furthermore, Thodi and Rodriquez proposed different methods based on the difference expansion technique. These schemes use the histogram shifting methodology to embed the marks. The proposed technique improves the distortion performance at low embedding capacities and mitigates the capacity control problem. This scheme benefits from a highly compressible overflow map.
Thodi and Rodriguez [32, 33] proposed a histogram shifting method in order to embed data in prediction errors. The location map used in this scheme covers all cells that cannot be decoded without a location map. The combination of prediction error expansion and histogram shifting technique is the novelty of this scheme. The maximum embedding capacity of this scheme in a single pass is 1 bpp.

Weng et al. [34] proposed a high capacity reversible watermarking scheme and tackled the preexisting problems of predefined thresholds on differences to allow expansions, and large location maps recording all the expanded positions. This would consume most of the available capacity especially when the threshold is small.

Coltuc and Chassery [35] proposed a high capacity low cost reversible watermarking algorithm. They suggested a generalized integer transform on pairs of pixels obeying some simple constraints to embed the watermark and the correction data which are required to recover the original image. The novelty of this scheme is that no particular location map is needed to identify the transformed pairs of pixels. This scheme can provide the capacity of 1 bpp.

Furthermore, Coltuc proposes an improvement of his previous scheme [36]. In this method a revised integer transform is presented which enables the algorithm to embed a watermark codeword into a single transformed pixel instead of inserting a codeword into a pair of transformed pixels [35]. This method is capable of reaching capacities more than 1 bpp, if certain conditions are met.

Chang et al. [37] introduced two spatial quad-based schemes based on Tian’s difference expansion method [27]. They exploit the fact that the differences between the adjacent pixel values in the local region of an image are small. The difference expansion technique is applied to the image in row-wise and column-wise simultaneously, promising good use of both row-wise and column-wise pixel pairs with small differences. It is shown [37] that the proposed schemes have higher embedding capacity than Tian’s [27] and Thodi’s [32, 33] methods, and quite competitive with Alattar’s [30] scheme.
In [38], Weng et al. proposed a reversible data hiding scheme based on an invertible integer transform which exploits the correlations among four pixels in a quad. In this scheme data embedding is accomplished by expanding the differences between a pixel and each of its tree neighboring pixels. As high embedding capacity cannot be achieved only by the means of difference expansion, the companding technique is employed into the process. The comparison of this scheme with Tian’s [27] and Alattar’s [30] algorithms shows its higher embedding capacity at almost all PSNR values [38].

2.7.1.2 Fragile Algorithms Operating in Transformed Domain

In this section, the significant schemes in the area of fragile reversible watermarking, which operate in a transformed domain, are presented.

Chen and Kao [39], proposed a simple watermarking approach operating in Discrete Cosine Transform (DCT) domain that uses quantized DCT coefficients of the host image. The embedding and extraction algorithms in this scheme are based on parameters adjustment rules.

Yang et al. [40], proposed another method based on integer DCT coefficients modification with peak amplitudes in each coefficient histogram. Lossless integer DCT transform, which guarantees reversibility, is applied on $8 \times 8$ blocks of host image, this method employs the histogram modification principal proposed by Ni et al. [31].

Xuan et al. [41] presented a reversible data embedding method using integer wavelet transform and companding technique. The presented method exploits the Laplacian distribution of integer wavelet coefficients in high frequency subbands ($LH, HL, \text{ and } HH$), which facilitates the selection of the compression and expansion functions and keeps the distortion low in the watermarked image. The results [41] show better visual quality in term of PSNR comparing to Tian’s difference expansion method [27].

Weng et al. [42], proposed a reversible watermarking scheme based on the companding technique and an improved difference expansion (DE) method. The watermark is
embedded into high frequency subbands of the integer wavelet transform (IWT), using the companding technique. The results show high visual quality in moderate capacities.

In [43], Lee et al. proposed a high capacity reversible image watermarking scheme based on integer-to-integer wavelet transform for both grayscale and color images. The proposed technique divides an input image into non-overlapping blocks and embeds a watermark into the high frequency wavelet coefficients of each block. To avoid any loss of information in the forward and inverse transforms, integer-to-integer wavelet is used, by applying the lazy wavelet and the lifting construction. Comparing the experimental results [43] with other preexisting reversible watermarking techniques reveals that the proposed scheme has higher embedding capacity with better visual quality.

2.7.2 Semi-Fragile and Robust Algorithms

This section introduces the significant semi-fragile and robust reversible watermarking schemes in the literature. These schemes show a certain degree of robustness when the watermarked image undergoes specific alterations or processes. In the case of semi-fragile methods, the tolerable process is usually confined to a slight compression process or other mild intentional or unintentional changes. On the other hand, robust schemes often present good tolerance against specific intentional attacks or unintentional severe modifications, depending on the purpose they are designed for. Hence, a watermarking scheme is called semi-fragile or robust if the extracted watermark from the modified/processed marked image stays ascertainable and valid.

2.7.2.1 Semi-Fragile Algorithms Operating in Spatial Domain

De Vleeschouwer et al. [44] proposed two semi-fragile reversible data hiding algorithms based on patchwork theory [45], which show certain robustness against JPEG lossy compression. These methods operate on image tiles by identifying a robust feature in the luminance histogram of each one of such tiles. Both approaches tolerate slight lossy
attacks such as minor cropping and light JPEG compression.

Ni et al. [46], presented a lossless watermarking scheme based on De Vleeschouwer work [44]. By then, the only existing semi-fragile scheme which could tolerate JPEG compression process was based on 256 modulo addition to achieve losslessness and robustness, but this technique suffered from the annoying salt-and-pepper noise caused by using 256 modulo addition to prevent overflow and underflow. Ni et al.’s proposed scheme does not generate salt-and-pepper noise in the marked image. The scheme operates based on the patchwork theory by identifying a robust statistical quantity. The differences between couples of pixels in an image tile are analyzed employing Error Correction Codes (ECC) and permutation techniques. The comparison of the experimental results with those of De Vleeshouwer [44] shows that not only a significant improvement in both data hiding capacity and perceptual quality of marked image is achieved, but also robustness is enhanced in the case of a lossy process such as JPEG compression with higher compression rates.

### 2.7.2.2 Semi-Fragile Algorithms Operating in Transformed Domain

Zou et al. [47] proposed a semi-fragile lossless digital watermarking scheme based on integer wavelet transform. The wavelet family adopted is the LeGalle 5/3 filter bank which is the default transformation technique in JPEG2000 for lossless compression. This characteristic makes it possible for this scheme to be integrated into the JPEG2000 standard. Experimental results show that the salt-and-pepper noise is not present in the marked images and their visual quality is much higher compared with De Vleeshouwer’s work [44].

Wu [48] presented a semi-fragile reversible watermarking scheme for image authentication. In this method the watermark is embedded into $LL_4$ subband of the integer wavelet domain. In addition to the reversibility attribute, this scheme has the property of tamper localization. The provided experimental results indicate low embedding
distortion and good visual quality of the watermarked image.

2.7.2.3 Robust Algorithms Operating in Spatial Domain

The algorithm presented by Chrysochos et al. [49] is a reversible watermarking scheme based on histogram modification which is resistant to geometrical attacks. The hiding capacity of this method is its major downside, the maximum capacity is 128 bits for an 8-bit grayscale image. On the other hand, perceptual quality of the marked image is high, and this algorithm presents a good robustness in case of geometrical modifications, such as rotation, upsizing, and cropping (up to 80%). On the contrary, this scheme does not tolerate JPEG compression, low pass filtering and noise addition.

Coltuc and Chassery [50] proposed a technique based on Reversible Contrast Mapping (RCM) which is a simple integer transform applied to pair of pixels. The maximum data hiding capacity achieved in the experiments is very close to this scheme’s theoretical upper bound of 0.5 bpp. The proposed technique outperforms all other compression based schemes except Tian’s difference expansion method [27].

Coatrieux et al. [51] proposed an image reliability protection technique in which the robustness is achieved by combining two different approaches, one approach is based on a reversible technique and the other is based on a robust watermarking scheme. Due to the second protection level, the reversible watermarking, the global robustness of the scheme is limited. It is asserted that a JPEG compression with quality factor greater than 70 does not generate any bit error.

2.7.2.4 Robust Algorithms Operating in Transformed Domain

Saberian et al. [52] presented a reversible watermarking algorithm based on a quantization approach, called Weighted Quantization Method (WQM). The robustness of this scheme is limited to Additive White Gaussian Noise (AWGN).

Gao and Gu [53] proposed a reversible watermarking scheme employing wavelet lift-
ing algorithm based on Alattar’s difference expansion \[30\]. In the experimental results provided, image reversibility is granted when there is no attack performed on the marked image and watermark robustness is partially provided against cropping, salt-and-pepper noise, and some other image modifications localized in restricted zones.

2.7.3 Summary of Reversible Watermarking Literature Review

This section summarizes the aforementioned schemes in the form of a chart and a table. The reviewed schemes are classified under fragile, semi-fragile, and robust categories in Figure 2.5. Furthermore, the methodology and properties of each method is presented in Table 2.1.
Figure 2.5: categorization of the significant reversible watermarking schemes.
<table>
<thead>
<tr>
<th>Scheme</th>
<th>Category</th>
<th>Methodology</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tian [27]</td>
<td>Fragile-Spatial domain</td>
<td>Difference expansion (DE) technique in high frequency sub-bands of a one level Haar transform</td>
<td>High capacity, Grayscale images, Uncompressed location map</td>
</tr>
<tr>
<td>Alattar [28]</td>
<td>Fragile-Spatial domain</td>
<td>Difference expansion method on spatial and cross-spectral triplets of pixels</td>
<td>Controllable payload size</td>
</tr>
<tr>
<td>Alattar [29]</td>
<td>Fragile-Spatial domain</td>
<td>Difference expansion method on quads of adjacent pixels</td>
<td>High PSNR, High computational complexity</td>
</tr>
<tr>
<td>Alattar [30]</td>
<td>Fragile-Spatial domain</td>
<td>Difference expansion method on vectors composed by adjacent pixels</td>
<td>Embedding rate of 1 bpp, Multi-pass, High capacity, Low distortion</td>
</tr>
<tr>
<td>Ni et al. [31]</td>
<td>Fragile-Spatial domain</td>
<td>Histogram modification technique in spatial domain</td>
<td>PSNR lower bound &gt; 48db, Low computational complexity</td>
</tr>
<tr>
<td>Thodi and Rodriguez [32]</td>
<td>Fragile-Spatial domain</td>
<td>Histogram shift method to embed data in prediction errors</td>
<td>Highly compressible map</td>
</tr>
<tr>
<td>Thodi and Rodriguez [33]</td>
<td>Fragile-Spatial domain</td>
<td>Histogram shift method based on difference expansion</td>
<td>Uncompressed location map, 1 bpp in single-pass</td>
</tr>
<tr>
<td>Weng et al. [34]</td>
<td>Fragile-Spatial domain</td>
<td>Prediction of adjacent pixels followed by Companding technique</td>
<td>High capacity, Compressed location map</td>
</tr>
<tr>
<td>Coltuc and Chassery [35]</td>
<td>Fragile-Spatial domain</td>
<td>A generalized integer transform on pairs of pixels obeying some constraints</td>
<td>Bit rate &gt; 1bpp, Low PSNR, No location map is needed</td>
</tr>
<tr>
<td>Coltuc [36]</td>
<td>Fragile-Spatial domain</td>
<td>A generalized integer transform to embed a codeword into a single transformed pixel</td>
<td>Bit rate &gt; 1bpp, Low PSNR</td>
</tr>
<tr>
<td>Chang et al. [37]</td>
<td>Fragile-Spatial domain</td>
<td>Two spatial quad-based schemes based on Tian’s difference expansion method</td>
<td>Compressed location map, High capacity</td>
</tr>
<tr>
<td>Weng et al. [38]</td>
<td>Fragile-Spatial domain</td>
<td>Use an invertible integer transform which exploits the correlations among four pixels in a quad</td>
<td>High PSNR</td>
</tr>
</tbody>
</table>

Continued on Next Page...
<table>
<thead>
<tr>
<th>Scheme</th>
<th>Category</th>
<th>Methodology</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen and Kao [39]</td>
<td>Fragile-Transformed domain</td>
<td>Discrete Cosine Transform (DCT) domain that uses quantized DCT coefficients of the host image</td>
<td>Low computational complexity</td>
</tr>
<tr>
<td>Yang et al. [40]</td>
<td>Fragile-Transformed domain</td>
<td>Integer DCT coefficients modification with peak amplitudes in each coefficient histogram</td>
<td>Uncompressed location map</td>
</tr>
<tr>
<td>Xuan et al. [41]</td>
<td>Fragile-Transformed domain</td>
<td>Integer wavelet transform and Companding technique in high frequency subbands</td>
<td>High PSNR</td>
</tr>
<tr>
<td>Weng et al. [42]</td>
<td>Fragile-Transformed domain</td>
<td>Companding and difference expansion (DE) methods employed in high frequency subbands of integer wavelet transform</td>
<td>High PSNR, Moderate capacity</td>
</tr>
<tr>
<td>Lee et al. [43]</td>
<td>Fragile-Transformed domain</td>
<td>LSB-substitution or the bit-shifting techniques in high frequency subbands of Integer wavelet transform</td>
<td>Grayscale/color images, High PSNR, High capacity</td>
</tr>
<tr>
<td>De Vleeschouwer et al. [44]</td>
<td>Semifragile-Spatial domain</td>
<td>Based on Patch work and robust feature in the luminance histogram</td>
<td>Certain robustness against cropping and lossy JPEG, Salt and pepper noise</td>
</tr>
<tr>
<td>Ni et al. [46]</td>
<td>Semifragile-Spatial domain</td>
<td>Based on Patch work in [44] and the use of ECC and permutation techniques</td>
<td>Certain robustness against lossy JPEG, High PSNR</td>
</tr>
<tr>
<td>Zou et al. [47]</td>
<td>Semifragile-Transformed domain</td>
<td>Embeds bits into high frequency subbands of integer wavelet transform employing threshold levels</td>
<td>Certain robustness against lossy JPEG, High PSNR</td>
</tr>
<tr>
<td>Wu [48]</td>
<td>Semifragile-Transformed domain</td>
<td>Histogram shifting method employed in LL4 subband of the integer wavelet domain</td>
<td>Tamper localization, Good visual quality</td>
</tr>
</tbody>
</table>

Continued on Next Page...
<table>
<thead>
<tr>
<th>Scheme</th>
<th>Category</th>
<th>Methodology</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chrysochos et al. [49]</td>
<td>Robust-Spatial</td>
<td>Based on modification of the selected couples in histogram bins</td>
<td>Resistant to geometrical attacks, Fragile to lossy JPEG, High PSNR</td>
</tr>
<tr>
<td>Coltuc and Chassery [50]</td>
<td>Robust-Spatial</td>
<td>An integer transform applied to pair of pixels called Reversible Contrast Mapping (RCM)</td>
<td>Bit rate &lt; 0.5bpp, Low complexity, Robustness against cropping</td>
</tr>
<tr>
<td>Coatrieux et al. [51]</td>
<td>Robust-Spatial</td>
<td>Combine two different approaches and the other based on a robust watermarking scheme, one based on a reversible technique</td>
<td>Robustness against lossy JPEG</td>
</tr>
<tr>
<td>Saberian et al. [52]</td>
<td>Robust-Transformed</td>
<td>Based on a quantization approach Weighted Quantization Method (WQM),</td>
<td>Robustness against AWGN</td>
</tr>
<tr>
<td>Gao and Gu [53]</td>
<td>Robust-Transformed</td>
<td>A wavelet lifting algorithm based on difference expansion, $LL_1$ subband is used to embed data</td>
<td>Partial robustness against image modifications localized in restricted zones</td>
</tr>
</tbody>
</table>

Table 2.1: Summary of the significant reversible watermarking schemes.
Chapter 3

Reversible Multiple Watermarking

Various techniques have emerged to verify integrity and prevent forgery. Most methods rely on embedding a perceptually invisible mark called a digital watermark or signature into the multimedia file \[16\]. Digital watermarking techniques are so far a conceivable solution whenever there is a need to embed, in an imperceptible way, informative mark or message in a digital document with applications in copyright protection, authentication, verification, authorization, and biometrics. This goal is essentially attained by performing a slight modification in the original media (also known as host or cover media) considering the limitations and essentials within different applications to satisfy targets such as payload capacity, original data degradation level, security and robustness. An imperative point to highlight is that any ”slight modification” is irreversible meaning the watermarked media is different from the original one. Although the goal of watermark insertion has been fulfilled but, the effect causes that any successive assertion, usage, and assessment to occur on an altered, tainted version of the host if the original host data have not been stored and/or are not readily available. In some applications, depending on the purpose of watermarking, it is acceptable to have a corrupted version of the original data and deal with it in different situations but in most cases having the original copy of the host media is essential and crucial. When dealing with sensitive applications and imagery
such as deep space exploration, identity identification/verification, military investigation, recognition, and medical diagnosis, the entity using the watermarked data cannot tolerate the risk of getting inaccurate and distorted information instead of the original copy. An example that shows the sensitivity of the issue in medical imaging applications would be the case of radiology images. A radiologist looking at a watermarked radiographic image in order to verify whether certain pathology is present or not, cannot depend on a visually degraded image quality. A visually degraded image can cause false diagnosis which primarily puts the patient’s health and also the radiologist’s credibility at risk. In other words, the constraint is that the presence of the inserted watermark does not have any influence on the diagnosis stage. Hence, in such applications, irreversible watermarking methods undoubtedly are not practical. Due to such applications and this critical and strict requirement, another set of watermarking techniques has been introduced in the literature classified as reversible watermarking, also called lossless data embedding. Using the reversible term for this category of watermarking techniques clearly shows, it is to be intended, that the original host data is fully recoverable beside the watermark signal from the watermarked document in a way that any evaluation can be performed on the recovered version of the original media. Having reversible watermarking techniques employed allows retrieving and keeping the original media, which is a zero impact process along with embedding and conveying an informative message. However, perceptual quality of the marked media is also of great importance in variety of applications where there may be a need to examine the marked signal without extracting the embedded watermark.

3.1 Preliminaries

Reversible watermarking is a feasible concept due to the fact that the original media usually has a strong spatial or temporal redundancy. Reversibility is guaranteed if enough
free space can be found or created to embed the watermark within the host signal while retaining the characteristic of the host untainted. This task is possible in the appropriate transform domain by employing specific suitable techniques and methods. Similarly to each generic watermarking technique, there are a number of properties that a reversible watermarking algorithm should entail. Here are the list of the most important properties and their measure:

**Fidelity:** A watermark is said to have high fidelity if the degradation it causes is very difficult for a viewer to perceive, this property is measured with PSNR of the original image against the watermarked version.

**Computational cost:** Different applications require the watermark embedders and detectors to work at different speeds. For example, in broadcast monitoring, both embedders and detectors must work in, at least, real time. The embedders must not slow down the media production schedule, and the detectors must keep up with real time broadcasts. Hence, reliable fast watermarking schemes are an essential aspect of such systems.

**Efficiency:** A watermarking scheme’s efficiency is measured by the amount of payload (watermark) that can be embedded in the host media versus the degradation this process causes in the media. The evaluation of this property is based on PSNR (the most common measure of watermarked media degradation in the literature [22, 43, 54–57]) versus bits of information embedded per pixel (a measure of embedded payload).

**Security:** An important property of each and every watermarking scheme is the security. It is measured by how readily and easily the embedded information can be extracted and decoded.

As stated earlier, the majority of the existing reversible watermarking techniques are categorized as fragile schemes. Since the main goal of a reversible watermarking algorithm
is to make it possible to retrieve the original version of the host media without any losses from the watermarked copy, these techniques are not required to show the robustness property. To compensate this factor, in applications where robustness and reversibility properties are required simultaneously, a reversible watermarking technique can be used in combination with a robust one, like the proposed method by Coatrieux et al. [51].

3.2 Proposed Methodologies

This chapter introduces a high capacity reversible watermarking scheme, operating in wavelet transformed domain, which has significant payload capacity and improved visual quality comparing to most of the existing algorithms in the area of lossless data hiding. This scheme is an improvement and extension of an earlier work introduced in [58].

At first integer-to-integer wavelet transform is applied to the host image to decompose the image into the wavelet domain, integer-to-integer wavelet guarantees the reversibility. In this scheme, the high frequency subbands of the wavelet coefficients are used for data embedding, due to the fact that alterations in these coefficients create less distortion in the marked image than those of the low frequency subband. To embed bits into the chosen subbands, histogram shifting technique is employed in high frequency subbands of different wavelet levels. A scalable coefficient map is formed and used to prevent the probable overflow and underflow in the marked image, the size of this map is insignificant comparing with the size of the payload this scheme is capable of embedding. To increase the data hiding capacity and yet improve the resultant perceptual quality and PSNR of the watermarked image, novel stopping conditions are applied on different subbands of each level through a systematic algorithm. This algorithm helps to find the best stopping condition on each step which yields the best visual quality and least distortion in the created marked media.
3.2.1 Integer-to-Integer Wavelet Transform

Conventional wavelet transform cannot be used in the reversible watermarking methods as it doesn’t support full reversibility. In all images, such as an 8-bit grayscale image, the pixel values are integer numbers; transformation into wavelet domain by the mean of a floating point wavelet transform implies the risk of non-integer pixel values of the watermarked image due to the changes of the wavelet coefficients. Furthermore, as in conventional floating point wavelet transform rounding or truncation is applied to the wavelet coefficients the overall forward and inverse transform is not lossless; hence, original image cannot be reconstructed from the watermarked image if such a method is engaged during the embedding procedure. To avoid these problems and to preserve the reversibility during image transformation, invertible integer-to-integer wavelet transforms based on the lifting scheme proposed by Calderbank et al. [59, 60] are utilized. This transform maps integers to integers and does not cause any loss of information through forward and inverse transforms. A thorough introduction of integer wavelet transform, the lifting scheme and its properties is given in Appendix B.

The interpolating transform used in the proposed scheme is an instance of a family of symmetric, biorthogonal wavelet transforms built from interpolating Deslariers-Dubuc scaling functions defined by Calerbank et al. [59]. According to [59], every wavelet transform associated with finite length filters can be obtained as the Lazy wavelet followed by a finite number of primal and dual lifting steps and scaling. Among the instances defined in [59], (2, 2) integer wavelet transform, also known as LeGall 5/3 wavelet, which shows one of the best coding performances comparing to others [59], is the chosen transform to be employed in the proposed watermarking technique. In a one-dimensional (1-D) signal \( s = [s_{0,1}, s_{0,2}, \ldots, s_{0,N}] \), where \( s_{0,i} \)'s are integers, the forward transform of the one level
decomposition is defined by the low and high frequency coefficients \( s_{1,n} \) and \( d_{1,n} \) as

\[
\begin{align*}
  d_{1,n} &= s_{0,2n+1} - \lfloor 1/2(s_{0,2n} + s_{0,2n+2}) + 1/2 \rfloor, \\
  s_{1,n} &= s_{0,2n} + \lfloor 1/4(d_{1,n-1} + d_{1,n}) + 1/2 \rfloor.
\end{align*}
\] (3.1)

where \( s_{j,n} \) and \( d_{j,n} \) are, respectively, the \( n \)th low frequency and high frequency wavelet coefficients at the \( j \)th wavelet transform level \( [59] \). The function \( \lfloor \ast \rfloor \) rounds \( \ast \) to the nearest integer towards minus infinity. Similar to floating point wavelet and wavelet packet transforms, the low and high frequency coefficients, generated in the first decomposition level, can be further decomposed into more levels. To perform the one level decomposition on a two-dimensional (2-D) signal, such as an image, the aforementioned Equations (3.1) are applied to both the horizontal and vertical directions. The reverse transform of the decomposition is easily calculated as

\[
\begin{align*}
  s_{0,2n} &= s_{1,n} - \lfloor 1/4(d_{1,n-1} + d_{1,n}) + 1/2 \rfloor, \\
  s_{0,2n+1} &= d_{1,n} + \lfloor 1/2(s_{0,2n} + s_{0,2n+2}) + 1/2 \rfloor.
\end{align*}
\] (3.2)

Integer-to-integer wavelet transforms, like floating point wavelet transform and other wavelet transforms, are not showing normalized behavior throughout different frequency subbands of the decomposed signal. The (2, 2) interpolating transform used in this scheme follows the same pattern and modifications in different frequency subbands lead to diverse effects with different intensities in the spatial domain. For example, increasing \( s_{1,n} \) by 1 results the original data \( s_{0,2n} \) to be increased by 1. However, the same effect can only be achieved if \( d_{1,n} \) is increased by 2, given that all other coefficients stay the same. As a result of such characteristic pattern through experiments and analysis, it is shown that modifications in low frequency coefficients have greater impact in the reconstructed signal. In the other words and from the energy point of view, this behavior shows that wavelet transform does not conserve the total energy and its homogeneity in the transform process. In a one-dimensional signal, to retrieve the total energy in a one-level transform
it is required to scale the low frequency coefficients by $K$ and high frequency coefficients by $1/K$. In the case of an image, the scaling factors to obtain the total energy for the $LL$, $LH$, $HL$, and $HH$ subbands are $K$, $1$, $1$, and $1/K$ respectively. In integer wavelet transforms, such a scaling is not exact; however, the scaling factors provide a good estimate of the impact of the examined subband coefficients on the reconstructed signal. The scaling factor, $K$, for the $(2,2)$ interpolating transform, which approximates LeGall $5/3$ filters, is shown to be $2$. 

3.2.2 Histogram Shifting and Data Embedding Techniques

As mentioned in section 2.7, histogram shifting technique is employed for reversible watermark embedding in a number of previous algorithms. Histogram shifting method is based on creating a gap, used as the free space during data embedding process in the histogram of either the original or transformed data. An example of the procedure is explained here to fully demonstrate the technique. Figure 3.1(a) shows the histogram of the $HH$ subband coefficients of a grayscale image. In order to create a gap to embed bits at a coefficient with value equals $n$ ($n > 0$), one is added to all coefficients with values greater than $n$ as shown in Figure 3.1(b). After the gap is created for embedding at the desired coefficient value $n$, $n = 0$ in this example, the binary watermark message bits are embedded into the free space. The embedding process scans the subband in which the gap is created to locate the coefficients with value $n$ in order to use them in the process. In this example process searches for coefficients with value 0. If the to-be-embedded watermark bit is 0, the coefficient is not modified and left intact. On the other hand, if the to-be-embedded bit is 1, the value of that coefficient is increased by one; in this example 0 is changed to 1. The histogram after embedding is shown in Figure 3.1(c). If the number of the bits to be embedded is greater than the number of coefficients with value $n$, embedding procedure goes on using other coefficient values to create more free space.
To extract the embedded bits, the detection procedure follows the reverse pattern exactly. At first, all coefficients which values are used in the embedding process (the ones the shifting technique is performed on), 0 and 1 in the above example, are selected. Then the subband is scanned; receiving a coefficient with value $n$ means that bit 0 has been embedded and getting $n+1$ translates as bit 1. The original data can be completely recovered by subtracting one from all coefficients greater than the used coefficient value $n$, 0 in this example. The same procedure can be applied on all the coefficients used in the embedding process.

To shift the coefficients with negative values, the explained procedure is equally valid. For instance, starting from -1, all the coefficients smaller than -1 can be shifted towards negative infinity to create a gap to embed bits at -1.

Figure 3.1: Illustration of histogram shifting (a) original histogram, (b) shifted histogram, (c) histogram after watermark embedding.
3.2.3 Proposed Solution for Underflow and Overflow

Applying histogram shifts on the wavelet coefficients of an image usually results in an expansion of the histogram in the spatial domain which might lead to inevitable underflow and overflow problem. Such a problem refers to the condition in which the reconstructed marked image pixels have values outside the acceptable range corresponding to the original image. For an 8-bit grayscale image, it means that pixels of the watermarked image have values smaller than 0 or greater than 255. Such a phenomenon is defined as underflow and overflow in the previous watermarking works [27]. In order to prevent the aforementioned problem and minimize the number of affected pixels in spatial domain, most proposed schemes employ an underflow/overflow map of all coefficients to identify and locate those which cannot be altered. To convey this map to the decoder side, it is compressed and embedded into the host media. As the compression is lossless, it does not reduce the size of the map to a great extent which means such a map takes a large portion of the embedding capacity and leaves less for the to-be-embedded watermark message. This problem has a great impact as can be seen in a lot of the early works in the reversible watermarking area. [27-30, 64].

The proposed scheme introduces a novel location map exploiting the characteristics of the wavelet transform in order to tackle the large size coefficient map problem. A useful characteristic of the wavelet transform, which comes handy here is the fact that it shows good correlation and resolution in both spatial and frequency domains, due to the fact that high frequency subbands carry spatial information of the image. As discussed earlier, underflow and overflow issue may affect the pixels that have values close to the bounds of the acceptable range i.e. 0 and 255 in an 8-bit grayscale image, so it is possible to locate the corresponding coefficients in the frequency domain. Later during the embedding process the location map is employed to exclude coefficients corresponding to the pixels close to the boundaries of the acceptable range from being modified.

To clarify the technique, an example is given below:
• Let the original signal be a $4 \times 4$ two dimensional array of pixels, and the condition set on the assumption is to exclude pixel values less than 10 or greater than 250 from the embedding process.

\[
\begin{bmatrix}
157 & 255 & 160 & 2 \\
15 & 26 & 253 & 4 \\
120 & 11 & 9 & 245 \\
200 & 3 & 254 & 2
\end{bmatrix}
\]

• Following the rule the corresponding binary coefficient map is a $4 \times 4$ array of 1’s and 0’s indicating the pixels that are allowed to be modified and those which are not, respectively.

\[
\begin{bmatrix}
1 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 \\
1 & 0 & 0 & 0
\end{bmatrix}
\]

• To find the corresponding wavelet coefficients, integer wavelet is performed on the 4x4 array. Now, with no loss of generality, lets assume that the embedding process is performed in the $HH$ subband by embedding a single bit 1.

\[
\begin{bmatrix}
142 & 185 & -156 & 86 \\
91 & 112 & 91 & 139 \\
65 & -128 & -130 & -203 \\
-135 & 129 & -198 & -245 \\
\end{bmatrix}
\Rightarrow HH = \begin{bmatrix}
-130 & -203 \\
-198 & -245
\end{bmatrix}
\]

• To embed this bit in the coefficient that has the least chance of overflow or underflow, the coefficient map is considered. To cope with the size of $HH$ subband and also to reduce the size of the coefficient map, this map is downsampled in this example by taking every other sample.

\[
\begin{bmatrix}
1 & 1 \\
1 & 0
\end{bmatrix}
\]
• Even though the dimensions are not the same but since $HH$ subband contains a great deal of spatial information of the original image the method is effective. The process considers one of the coefficients $\{-130, -203, -198\}$ for embedding, and it proceeds with histogram shifting and bit embedding.

As discussed, the original coefficient map is the same size as the original image. However, it shows a considerable amount of redundancy as in most images the tone varies smoothly except on the edges and abrupt changes are rare. In general to reduce the size of the used coefficient map, the original map can easily be downsampled. As an experimental fact, it is possible to generate the map for the $HH_1$ from the one for the $HH_2$ subband by upsampling. This means that the size of the general coefficient map can be confined to the map of one subband, and in order to make the approach more efficient the size of the map can be reduced further. In the proposed scheme, the coefficient map employed has the size of the $HH_3$ which is $1/64^{th}$ of the original image. This map is embedded during embedding process as the side information along with the watermark message and used at the decoder for data extraction and image recovery. This approach prevents the pixels of the marked image to suffer from most of the probable underflow and overflow incidents, and leaves no noticeable visual artifacts such as salt-and-pepper noise.

As an alternative approach, integer wavelet transform can be applied to the coefficient map and the $HH, HL$ or $LH$ sections of the transformed map can be chosen to be used. The advantage of this approach is the better identification and isolation of those coefficients causing underflow and overflow in the transformed domain, but on the other hand this approach may reduce the effective hiding capacity to a great extent in the images having lots of pixel values close to the bounds of the acceptable range. Therefore, in applications where no underflow and overflow is tolerated the second approach is favorable in the cost of losing some hiding capacity.

Figure 3.2 shows examples of the coefficient maps generated using both approaches described above. The coefficients corresponding to pixel values between 10 and 245 are
Figure 3.2: Examples of coefficient maps (a) original image, (b) corresponding spatial domain coefficient map, (c) corresponding transformed domain coefficient map.

used for data embedding in these cases. As mentioned earlier, in the proposed scheme only high frequency subbands are used in data embedding process and as the result the coefficient maps are only applied to these subbands in the figure.

3.2.4 Derivation of the Embedding Conditions

To increase the data hiding capacity of the proposed scheme, and also to improve the visual quality and subsequent PSNR of the marked image, systematic stopping conditions are employed. This section discusses the proposed techniques used to derive the efficient stopping conditions affecting each subband in the embedding process.

3.2.4.1 Definition and Determination of Shifting Parameters

In the aforementioned example in section 3.2.4.1, the maximum number of embedded bits is the number of coefficients with values equal 0. In general, the number of coefficients to be modified has to be at least the size of the embedded watermark message. Therefore, more than one histogram shifts are usually needed to embed a large amount of data information.

In the proposed scheme, the watermark message is embedded in the high frequency
coefficients of the transformed image. As mentioned in section 2.7, it is shown that coefficients in high frequency subbands, also known as detailed subbands, follow an approximately Laplacian like distribution centered at zero [65], as illustrated in Figure 3.1(a). Due to this fact, to minimize the resultant distortion while maximizing the capacity, histogram shifts are started from zero and follow the technique described in section; bits are embedded into the first created gap. If there are un-embedded bits left after all coefficients with value 0 are used up, coefficients that equal to -1 are then shifted. By continuing this pattern, coefficients with values equal to 2 (which are originally coefficients with values equal to 1 but due to the first shift they are relocated to 2), -3, 4, -5, and so on are shifted respectively through the same method until either all the to-be-embedded data are inserted or the stopping conditions are encountered in the examined subband. If the latter happens, the scheme moves to another subband or deeper wavelet levels and starts the aforementioned procedure.

The advantage of choosing coefficients with value 0 to start in each subband during histogram shifting technique can be shown to be fewer shifts to embed the same number of bits. The number of shifts in the histogram directly affects the changes in coefficients with high magnitude, and leads to less distortion in the subsequent reconstructed watermarked image. On the other hand, it guarantees higher embedding rates and hiding capacity keeping the same level of distortion.

As elucidated and discussed in section 3.2.1, the wavelet coefficients in different subbands have different energies and so are of different importance concerning the resultant marked image and the level of distortion imposed. The proposed algorithm exploits this property to achieve higher visual quality and PSNR at different bit rates in the watermarked image. To accomplish this goal, the embedding process explained in section always starts in the $HH_1$ subband of the first level integer wavelet transform to embed the data. To achieve the best visual quality and the least distortion in term of PSNR, a stopping parameter $\gamma_1$ is defined through a well-defined systematic procedure considering
the performance and efficiency (measured in PSNR) of the process. The stopping parameters are the order of the chosen coefficients in the set \( \{0, -1, 2, -3, \ldots, m\} \), where \( m \) is the last available coefficient in the considered subband. As an example, \( \gamma_1 = 4 \) means that the shift is performed at coefficients with value equal to the fourth element in the above mentioned set which is -3. This procedure is introduced and thoroughly discussed in section 3.2.4.2. If the required space to embed all watermark bits is not acquired when \( \gamma_1 \) is reached, the process moves to either \( HL_1 \) or \( LH_1 \) subbands considering which subband results in a better performance and less distortion. Then, the most efficient stopping parameter \( \gamma_2 \) for these subbands to achieve the least possible distortion in the marked image is found. The process proceeds to the subbands of deeper wavelet transform levels (\( HH_2, HL_2/LH_2 \), and so on) if more space is needed. It should be noted that \( \gamma_1 \) is set to be engaged in all \( HH \) subbands involved in the embedding process and \( \gamma_2 \) is the chosen parameter for all \( HL \) and \( LH \) subbands of different wavelet levels. Although it is possible to define more consecutive stopping parameters for each subband of each wavelet level but due to high computational complexity such an approach imposes and the trivial performance improvement it yields, only the aforementioned parameters, \( \gamma_1 \) and \( \gamma_2 \), are introduced in this scheme. The stopping parameters are chosen to ensure that the total number of coefficients with absolute values less than or equal to the coefficient values corresponding to \( \gamma_1 \) and \( \gamma_2 \) in all engaged \( HH \) and \( HL/LH \) subbands, are greater than the size of the to-be-embedded watermark message and the necessary side information. In cases that the to-be-embedded bits are more than the total number of coefficients in all the wavelet levels, it is granted that the system fails to embed the data in a single-pass embedding process.

It is essential to note that the \( LL \) subband is never used for data embedding hence never modified in any of the levels for two main reasons. First, the coefficient distribution does not necessarily follow the Laplacian shape as in the high frequency subbands; therefore the start or end shifting coefficients cannot be easily defined. Moreover, and most
importantly, changes in the $LL$ subband have the greatest impact in the resulting quality and PSNR of the reconstructed image, as discussed in section 3.2.1, so if it is tampered with, it causes the highest degradation in the visual quality of the watermarked image. Although there are cases where this subband is used for data embedding in deep levels of integer wavelet transform to achieve a degree of robustness.

### 3.2.4.2 Derivation of Stopping Parameters

The proposed scheme benefits from a well-defined procedure to determine and derive the most suited stopping conditions and factors to maximize the hiding capacity while retaining the distortion level to minimum. These stopping conditions are in the forms of the stopping parameters introduced in section 3.2.4.1 briefly. A stopping parameter determines the last shifted coefficient in a subband during the embedding process. It means if there are to-be-embedded bits left when the coefficient corresponding to the stopping parameter is reached in a particular subband, the embedding process switches to the next subband to embed the rest of the bits. This section introduces the technique to find the most optimum and efficient stopping parameters $\gamma_1$ and $\gamma_2$, which play a crucial role in gaining the best visual quality and the least degradation and distortion in the marked images in this scheme. These parameters guide the encoder and decoder through embedding and data extraction processes, respectively. As mentioned before, $\gamma_1$ is the stopping parameter used in $HH$ subbands and $\gamma_2$ determines the stopping coefficients of both $HL$ and $LH$ subbands. As indicated in section 3.2.4.1, stopping parameters engage and incorporate the importance of different subbands in the wavelet transform by treating the $HH$ subbands differently from the $HL$ and $LH$ subbands. Furthermore, these parameters optimize the use of wavelet coefficients in different subbands by considering the resultant PSNR of the marked image within each stage.

The following is the procedure through which the proposed scheme defines the stopping parameters $\gamma_1$ and $\gamma_2$: 
1. The procedure starts by setting the initial values of $\gamma_1$ and $\gamma_2$ 0, then the 1st level integer-to-integer wavelet transform is performed on the host image.

2. An iteration process is engaged to evaluate the performance of the marked image at different stages. The process starts by shifting all the coefficients in the $HH_1$ subband according to the value of $\gamma_1$ (this value can be positive or negative) in each stage. The value of the stopping parameter $\gamma_1$ equals 1 in the first iteration, so the first gap is created by shifting all coefficients with values greater or equal to 1. Then the process checks if the created free space, considering the coefficient map, is enough to embed the total payload. If there is more space required, the resultant PSNR of the reconstructed image from the shifted $HH_1$ subband for an increased value of $\gamma_1$ by 1 is computed and compared to the resultant PSNR of the reconstructed image from each one of the $HL_1$ and $LH_1$ subbands as the first shift at 0 is applied to them. For instance, in the first iteration and as $\gamma_1=1$, the PSNR of the reconstructed image using the $HH_1$ subband with two gaps at 0 and -1 ($\gamma_1+1$) is computed and compared to the PSNRs of the reconstructed images obtained from each of the following cases: $HH_1$ and $HL_1$ with gaps created at coefficient value 0, $HH_1$ and $LH_1$ with gaps created at coefficient value 0. If the PSNR obtained from $\gamma_1$ is still greater than PSNRs in both $HL_1$ and $LH_1$ subbands this step is repeated and process proceeds with the iteration, else the procedure sets $\gamma_1$. This step guarantees that the final value of $\gamma_1$ obtained through this method, results in the highest PSNR in the marked image using histogram shifting technique to embed the watermark bits.

3. PSNR of the reconstructed image is calculated in both $HL_1$ and $LH_1$ subbands, which histograms are modified with the first shift at 0, the one imposing less distortion, measured in PSNR, is taken as the succeeding subband used for defining $\gamma_2$ and data insertion, following $HH_1$. The chosen subband’s name is recorded, de-
noted by $sub_{idx}$, as the side information or the overhead which specifies the subband amongst $HL_1$ and $LH_1$ being used following $HH_1$ during data embedding process. This information is used during data extraction and original image recovery phase.

4. The process shifts the histogram of $sub_{idx}$ subband up to and including the coefficient corresponds to $\gamma_2$. Then it checks, considering the location map, if the total created hiding space is enough for all the bits to be embedded, if not it starts the shifts in the $sub'_{idx}$ (e.g., if $sub_{idx} = HL_1$ then $sub'_{idx} = LH_1$) coefficients to the limit set by $\gamma_2$ and once more checks the availability of enough space to embed the total payload. In case the payload is not enough, PSNR of the reconstructed image resulted from shifts in $HL_1$ and $LH_1$ subbands is computed and compared against the resultant PSNR of the reconstructed image from the first shift in the $HH_2$ subband. If PSNR from $HL_1$ and $LH_1$ is found to be greater then step is repeated with an increase in $\gamma_2$ value, else the system sets $\gamma_2$.

5. Having $\gamma_1$ and $\gamma_2$ set through the previous steps, these stopping parameters are used to check the availability of enough pure free space to embed all the payload bits. If the payload fits in the host image, using the obtained stopping parameters in different wavelet levels, system sets the stopping parameters which ensure the least possible degradation of the marked image. In the case the to-be-embedded payload bits do not fit in the created gaps, the scheme sacrifices the resultant degradation level with an increase in capacity and repeats the above steps using the obtained values of $\gamma_1$ and $\gamma_2$ at the start. Although this final procedure causes the stopping parameters to be moved from the optimum point, highest resultant PSNR in the marked image using this method, but it increases the data capacity while retaining the PSNR at a very acceptable value.

The procedure mentioned above indicates that the stopping parameters $\gamma_1$ and $\gamma_2$ are pre-computed in this scheme, or in other words, these values are computed and set prior
to the actual embedding process.

In cases where the size of payload is greater than the total number of available coefficients in the original image; the system fails to obtain any stopping parameters for a single-pass embedding procedure. In this case, multi-pass embedding scheme can be adopted, which means to perform multiple iterations of the watermarking process using the previously marked image as the input. In multi-pass embedding, the previously watermarked image is used as the host image to embed more data. Multi-pass technique is explained in section 3.2.6.1 in detail.

Figure 3.3 illustrates the flowchart of the proposed algorithm, itemized above, to obtain the stopping parameters $\gamma_1$ and $\gamma_2$. This algorithm can easily be modified to cope with multi-pass watermark embedding by simply replacing the host image with the previously embedded image in Figure 3.3.

### 3.2.5 Security Techniques

This section discusses one of the most important issues in watermarking and its applications in the privacy area, the security and protection of the embedded mark. As for most of the watermarking techniques, the embedding methods employed are revealed through several links. So in order to secure the privacy, there should be a second layer of security which protects the content of the embedded marks. To provide the security of the watermarked image in the proposed scheme, either both the location and number of the watermarked coefficients or the watermark signal itself has to be encrypted. Therefore, adversaries or attackers cannot obtain the embedded information even if they are aware of the watermarking scheme.

The proposed security algorithm exploits two different encryption schemes in series to provide a reliable security level for the sensitive watermarked data. The first suggested encryption scheme is the XOR cipher, also known as *One-Time Pad* (OTP), this technique employs a keystream generated by a pseudo random number generator. Despite
Figure 3.3: Algorithm to derive the stopping parameters.
the proven perfect secrecy of this scheme \cite{66}, and the fact that if the key is truly random and as large as or greater than the length of the to-be-encrypted signal, the encrypted data is impossible to be decrypted without knowing the used key; but as true randomness is required which is practically impossible, it might not provide the best reliable secrecy required in a scheme used in applications dealing with important and sensitive data. The second encryption technique employed is the *Advanced Encryption Standard* (AES) method \cite{67} which is based on the block cipher Rijndael \cite{68} and can be named as the designated successor of the *Data Encryption Standard* (DES). AES uses a fixed block size of 128 bits and the key sizes of 128, 196 and 256 bits. In AES algorithm the strength of all key lengths is enough to protect secret information up to the secret level, for applications require *top secret* protection either 192 or 256 key lengths are suggested \cite{67}. Although AES is providing good security but as it is operating on a $4 \times 4$ array of bytes, in the case of images the trace of the encryption may be apparent on the encrypted image. Figure 3.4 shows the watermark logo used in this work along with the encrypted version of it using AES. It is apparent that the pattern of the encrypted logo is traceable and in some cases recognizable. This pitfall might make it possible for adversaries to decode the encrypted watermark in cases where alphabets, famous patterns or other recognizable forms are used. To provide the best secrecy, our proposed scheme benefits from both of these algorithms in series, as shown in Figure 3.5. At first the to-be-embedded data is encrypted using One-Time Pad technique and then the resultant encrypted mark is ciphered using AES method. This process ensures that the encrypted watermark is completely random and bears no correlation with the original mark.

The result of a systematic experiment using 100 different keys on the encrypted watermark and an example of the extracted mark using the wrong key are presented in Figure 3.6. For all the keys used in this experiment, the correlations between the extracted and the original watermark are computed. As expected, only the correct key can lead to correct decryption of the encrypted watermark, which shows the highest
Figure 3.4: Watermark encryption using AES (a) original watermark, and (b) encrypted watermark using AES algorithm on $4 \times 4$ array of bytes

Figure 3.5: Encryption method used in the proposed algorithms benefiting from both one-time pad and AES encryption schemes

correlation, 1, among the rest of the results. All other keys fail to decrypt the encrypted mark and generate images that have near zero correlation with the original watermark. As depicted in Figure 3.6(b), the extracted watermark using any wrong key is a random image and carries no information relevant to the original watermark image. The security algorithm proposed in this section is not only confined to image watermarks, it can be adopted in cases where other kinds of media are used as watermark messages during the embedding process.
3.2.6 Proposed Encoding and Decoding Algorithms

This section describes the encoding and decoding processes thoroughly. The proposed scheme can be adopted to be used for embedding a single watermark or in applications where insertion of multiple marks into the original image is required. An example for the latter case is in medical applications where different informational pieces of data have to be embedded into medical images, data such as patients’ personal information, medical history, and doctors’ reports. In this section both cases and their proposed algorithms are explained in steps.

3.2.6.1 Single Watermark - Embedding and Extracting Processes

Figures 3.7 and 3.8 illustrate the flowcharts of the single watermark embedding and decoding processes respectively. In this case, the goal is to embed a single watermark into the original image achieving the best PSNR by defining the most suited stopping parameters. Here we propose to use an n-level \((2,2)\) interpolating wavelet transform, where \(n\) is defined by the size of the to-be-embedded payload and the values of stopping parameters. As the general case, we suppose that the watermark message is larger than
the capacity of the host image and the embedding cannot be done in a single-pass process. Hence, the watermark message is divided into parts and each part is embedded through a single-pass process using the previously marked image as the host image. This method is called multi-pass watermarking.

The single watermark embedding process can be summarized in the following steps:

1. Coefficient map is formed, as explained in section 3.2.3, using the original image to secure the pixel values close to the boundaries of the acceptable range from falling off the range.

2. Integer wavelet transform is applied to the host image. Process determines stopping parameters $\gamma_1$ and $\gamma_2$ considering the number of embedding passes in case of multi-pass embedding and the coefficient map as the guide, as depicted in section 3.2.4.2. Two coefficient masks are proposed for employing the stopping parameters in different wavelet transform levels by imposing different conditions on each parameter.
Figure 3.8: Multi-pass single watermark data extracting algorithm.

in each step. Algorithm 1 uses the same $\gamma_1$ and $\gamma_2$ for all the wavelet levels involved in embedding process as shown in Figure 3.9(a). Algorithm 2 incorporates the importance of different wavelet levels by assigning different weights to the stopping parameters as shown in Figure 3.9(b).

3. The watermark data is divided into parts considering the capacity of each watermarking step calculated in the previous step. As mentioned earlier, $\gamma_1$ and $\gamma_2$ are computed prior to the actual embedding process, hence the hiding capacity in each stage is defined. Later, the part that is considered to be embedded in this stage is converted into bits, and the encryption algorithm introduced in section 3.2.5 is applied to the binary sequence.
Chapter 3. Reversible Multiple Watermarking

4. Histogram shifts, according to the stopping parameters, are performed in different levels, depending on the size of the payload and the number of watermarking passes defined.

5. Data including payload and overhead information is embedded using the technique described in section 3.2.4.1. The overhead information includes the coefficient map, stopping parameters, the number of wavelet levels involved, and the number of watermarking passes. After embedding both the overhead and the watermark, the modified signal is converted back to spatial domain.

In cases where the size of the watermark is greater than the total number of available coefficients in the host image, the scheme either fails if single-pass watermarking is intended or proceeds with multi-pass process and embeds the remaining watermark bits into the previously marked image following the steps mentioned above.

In order to present the general single watermark data extraction process, multi-pass watermarking is considered to be engaged in the embedding process. It should be noted that the extracting process starts from the last pass and proceeds to the first, it means that the multi-pass watermark embedding and extracting processes operate based on
"first embedded last extracted", or in other words this process is a "last in, first out" (LIFO) process.

The steps to extract the embedded watermark is listed as follows:

1. First, after applying integer wavelet transform to the marked image the overhead information is extracted.

2. The total payload is extracted using the retrieved overhead information; the changes done to the histogram during the embedding process are reversed through shifting the histogram back in different subbands of different wavelet levels, wherever applicable.

3. If the index extracted from the overhead shows there are still more parts of the original watermark message left embedded in the image, the scheme proceeds with another extraction process till the last part is extracted. In other words, the process proceeds to reverse all the passes used in the multi-pass embedding step.

4. After all parts of the embedded watermark are extracted, or in other words, the multi-pass embedding process is reversed completely, the inverse integer wavelet transform is applied to obtain the original image. The extracted parts are merged to form the original watermark in the same order they have been parted. Then the recovered watermark is decrypted using the decryption key and the required information is retrieved.

In all the steps above, the overhead information includes the coefficient map, the stopping parameters $\gamma_1$ and $\gamma_2$, $\text{sub}_{idx}$ parameter which indicates the subband (HL or LH) engaged in the embedding process after HH subband in each level, the stop subband (since not all subbands might be needed in the last used level), the last shifting coefficient in the stop subband, and the number of passes in case of multi-pass embedding process.
3.2.6.2 Multiple Watermarks - Embedding and Extracting Processes

The proposed scheme has the capability to be used in multiple watermarking applications using both single-pass and multi-pass embedding processes. As discussed in the previous section the multi-pass process imposes the constraint that the embedded watermarks during the last pass of the insertion process (the ones embedded in the last layer or into the last marked image) should be extracted first in order to ensure availability of the information inserted in the lower layers. Multi-pass process follows the same rule as mentioned before, the embedded watermarks in the first pass are extracted last.

The proposed scheme can be adopted to guarantee multiple watermark embedding in a single-pass process; this means that the process ensures the embedding of several independent watermarks into the host image so each can be retrieved in any order at any time, in a single-pass run. The same applies in case of the multi-pass embedding in which independent marks can be embedded into the image in each pass. So in situations where both the incredible hiding capacity of multi-pass process and the multiple watermark insertion properties are needed, the marks can be categorized in order of importance and availability urgency and then inserted into the original image employing multi-pass embedding process. Considering a patient’s medical file as an example for which the general case of multi-pass multiple watermark embedding may be required; patient’s identification, medical history and allergies to various medications are essential information that need to be readily accessed so will be embedded independently through the last pass in order to be retrieved first. On the contrary, the thorough report of the medical image can be the mark that is embedded during the first embedding pass which will be extracted last.

The $n$-level $(2,2)$ interpolating wavelet transform is employed the same as in single watermark embedding process. It is worth noting that all the steps of multi-pass watermark embedding explained in the previous section can be adopted here as well, but for sake of simplicity of the algorithm only single-pass multiple watermark embedding and
Single-pass multiple watermark embedding process can be summarized in the following steps:

1. The acceptable distortion level, measured by PSNR, is set and given as an input to the system. If the multi-pass embedding is in use, the satisfactory PSNR for each pass should be provided.

2. The watermark message is converted into bits, and then encrypted employing the method explained in section 3.2.5.

3. The process starts by performing integer wavelet transform on the original image, then scans the image to locate any watermarking overhead embedded in the ”host image” from a prior embedding. If any overhead is available, then it is extracted and the existing information is used for data embedding of the new watermark. In this case the next two steps are passed and process proceeds from step 6.

4. If there is no other watermark present, the coefficient map is formed, as explained in section 3.2.3 using the original image to secure the pixel values close to the boundaries of the acceptable range from falling off the range.

5. The stopping parameters $\gamma_1$ and $\gamma_2$ are defined considering the acceptable PSNR and the coefficient map. In this algorithm the stopping parameters are chosen to satisfy the PSNR set as the input, this means the process derives the parameters resulting the given PSNR as the condition. Either one of the two proposed coefficient masks introduced in section 3.2.6.1 (Figure 3.9) can be engaged.

6. Histogram shifts are performed in different levels according to the stopping parameters.

7. Data including payload and overhead information is embedded in the created gaps.
After embedding both overhead and watermark, the modified signal is converted back to spatial domain.

As mentioned before, in case the watermark size is bigger than the total number of coefficients available in the original image, the scheme either fails if set to single-pass watermarking or proceeds with multi-pass method and embeds the rest of the bits on top of the previously marked image following the same steps as above.

In this algorithm the overhead information consists of the coefficient map, the stopping parameters $\gamma_1$ and $\gamma_2$, $sub_{idx}$ parameter which indicates the subband (HL or LH) engaged in the embedding process after HH subband in each level, indices of multiple watermarks which indicate the size of each embedded watermark and the coefficient number its embedding begins at, the stop subband (since not all subbands might be needed in the last used level), the last shifting coefficient at the stop subband, and the number of passes in case of multi-pass embedding process.

In order to present the multiple watermark data extraction process, without loss of generality single-pass watermarking is considered to be engaged in the embedding process. In this process to extract a specific embedded watermark, the identification of the to-be-extracted mark should be provided as an input to the system. If multi-pass scheme is employed in the embedding process, the watermark extracting process starts with marks embedded in the last pass and proceeds to the first.

Extraction of an embedded mark can be done in the following steps:

1. First, after applying integer wavelet transform the overhead information is extracted.

2. Using the provided identification as the input, the wanted mark index and size are extracted from the overhead. The total payload of the called mark is extracted using the retrieved overhead information; and the changes done to the histogram during the embedding process due to this single mark are reversed through shifting
the histogram back in different subbands of different wavelet levels. The overhead information is updated showing the empty space size and its starting coefficient for the future use.

3. The inverse integer wavelet transform is applied to obtain the host image which contains the rest of the embedded marks.

4. The extracted watermark is decrypted using the decryption key and at the end the required information is retrieved.

Using the mentioned methods, the proposed scheme can be used in variety of different occasions with good flexibility. Depending on the application the proposed reversible watermarking technique can be hired to serve to the best, from high capacity single watermarking needs to applications requiring multiple pieces of information embedded into the host image.

3.3 Experimental Results

The proposed scheme can be adopted to serve a large variety of applications in which restoring the original image is a must. Such applications can include military imaging and remote sensing, satellite imaging, high-energy physics imaging, deep space high-resolution photography, medical images and files archiving and privacy protection, legal evidence, digital product promotion, E-business, advertisement business and many more.

This section focuses on the performance of the proposed scheme on both general natural images with broad histogram, which covers most of the grayscale values, and grayscale medical images. Military, satellite and deep space imaging are examples of applications in which natural images are involved. In these applications, both the perceptual quality of the marked image and the ability to retain the original image are of great importance. On the other side, the application of reversible watermarking in medical images is of
great attention due to the privacy and security issues. In other words, in consequence of the recent advances of Internet technology in healthcare systems, medical images can be exchanged to allow new medical practices such as telediagnosis and teleconsultation, therefore providing the security and privacy of the exchanged medical information is crucial. Hence, the proposed reversible watermarking scheme can be employed in the field of medical to address these concerns, while preserving the perceptual quality of the marked medical images to a great extent; i.e the constraint is that the presence of the inserted watermark does not have any influence on the diagnosis stage. In addition, in both fields of natural and medical images, the access to the original content can be controlled by employing the proposed reversible watermarking scheme such that only the authorized person can access the original content by removing the watermark while the watermarked content is available to everyone.

Reversible watermarking, as discussed in section 2.6 is the method to provide confidentiality, availability, and reliability of the contents in both medical and natural fields. In this section, the experiments are setup to demonstrate the ability of the proposed reversible watermarking scheme to fulfill the aforementioned goals in sensitive applications.

The proposed schemes are tested with images from variety of sources including USC-SIPI database [69]. In all tests 512 × 512 grayscale images are used to compare the results obtained from the proposed algorithms against other elected well performed methods. A 1024 × 1024 image is chosen to test the performance of the proposed scheme on larger image sizes. The scheme is tested on a series of natural and medical images. Natural images include the commonly used Lena, Mandrill, Airplane(F-16), Peppers, Barbara, and Man(1024 × 1024). The medical images used are the MRI images of brain and breasts, CT images of brain and liver and Ultrasound image of a fetus. Figure 3.10 shows all the images used in this experiment respectively.

Integer-to-integer (2, 2) interpolating wavelet transform is used [59]. The embedding capacity is represented by bit per pixel (bpp) and the watermark is generated using the
MATLAB \texttt{rand()} function. The proposed methods for embedding single or multiple watermarks are tested using single-pass and multi-pass approaches in order to get the highest capacity available. The image quality is evaluated by the peak signal-to-noise ratio (PSNR) in dB as the most common measure used in the literature \cite{20, 22, 27, 43, 51, 54, 57, 62, 63} for both non-reversible and reversible schemes which are applied to natural or medical images. PSNR is given by

\[
\text{PSNR} = 10\log_{10}\frac{255^2}{\text{MSE}},
\]

and

\[
\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (I_m(i) - I_w(i))^2,
\]

where in Equation (3.4), $I_m$ and $I_w$ are the original and watermarked images respectively, and $n$ is the total number of pixels in the image.

Although PSNR is the common measure to evaluate the image quality in the literature, but sometimes marked images with acceptable PSNR can present low visual quality and vice versa. Examples of such cases are included in the results.

### 3.3.1 Performance of the Scheme on Natural Images

Figure 3.11 shows Lena’s original image along with different watermarked images at various single-pass embedding capacity rates. The figures are zoomed in so that the difference in visual quality and the distortion level is more obvious. Since the embedding is done in high frequency wavelet subbands, each of the watermarked images have high frequency noise and appear sharper than the original image due to presence of larger high frequency coefficients; although this sharpening filtering effect does not affect the visual quality to a great extent even at low PSNRs.

It is noticeable that proposed scheme achieves good visual quality at low and moderate embedding capacities and even at high embedding capacity rates the distortion and
degradation in visual quality is acceptable. This scheme is capable of achieving high embedding rate of 1bpp in a single-pass embedding process. As an example, Mandrill’s watermarked image, which has the lowest PSNR at any capacity rate among other marked images in this experiment, is shown in Figure 3.12 at 1bpp capacity rate; the rate most algorithms fail to achieve for this particular image in single-pass embedding process. As mentioned earlier, comparing the original image with the high capacity watermarked image suggests that having substantially low PSNR does not necessary mean poor visual quality. Although the sharpening effect is intensified and quite noticeable, but the image perceptual quality is acceptable to human’s eyes. In general the distortion performance in each payload capacity depends on the characteristic of each image as it can be seen in Figure 3.13 which shows the resultant PSNR at different payload rates for three different natural images. This is due to the difference in the amount of high and low frequency components of each image and the shape of the grayscale histogram that affects the coefficient map engaged to prevent underflow and overflow.

To achieve higher capacity rates, multi-pass embedding process can be used. In multi-pass embedding the watermarked image is used to undergo the embedding steps again
Figure 3.11: Lena’s image: (a) original images, (b) bpp=0.3 PSNR=46.07dB, (c) bpp=0.6 PSNR=41.83dB, and (d) bpp=0.9 PSNR=36.58dB

Figure 3.12: Mandrill’s image: (a) original images, (b) bpp=1 PSNR=18.03dB

for insertion of more bits. Figure 3.14 shows the Lena’s marked image when the proposed scheme is used in 2-pass and 3-pass runs to achieve bit rates as high as 2.7 bpp for the latter case. As can be seen in Figure 3.14(b), although the resultant PSNR is above 23dB
Figure 3.13: Comparison of embedding capacity in bpp versus distortion in PSNR for different natural images

![Graph showing PSNR vs Payload](image)

Figure 3.14: Multi-pass watermarked images of Lena (a) 2-pass watermarked image (bpp=1.8 PSNR=27.48dB), (b) 3-pass watermarked image (bpp=2.7 PSNR=23.31dB)

for the 3-pass watermark embedding trial, but the perceptual quality is lower than those of single-pass marked images with lower PSNRs such as the Mandrill’s marked image in Figure 3.15(b). To further investigate the technique, the same multi-pass method is applied to Mandrill’s image and the marked images with corresponding PSNRs are shown in Figure 3.15.

Four seminal algorithms proposed in the literature, are implemented for comparison purpose. Lee’s algorithm [43], using block size $16 \times 16$, is chosen as it shows the best
Figure 3.15: Multi-pass watermarked images of Mandrill (a) 2-pass watermarked image (bpp=1.8 PSNR=18.11dB), (b) 3-pass watermarked image (bpp=2.7 PSNR=14.25dB)

performance in the expansion based scheme proposed with low complexity comparing to others in this field. Thodi and Rodriguez [32, 33] proposed different versions of the reversible data hiding methods. The two methods that achieve better performances are chosen in these tests. The first incorporates histogram shifting and the prediction errors (method P3). The second is a combination of histogram shifting and difference expansion (method D3). Xuan et al. scheme [62], exploiting integer wavelet and companding technique, is the fourth selected method among all those employing companding technique in their schemes.

In all the performed tests the proposed algorithm 2, which employs stopping parameters with different weights in each level, outperforms other methods in the high capacity embedding rates (i.e. bpp ≥ 0.4). At low capacities, Lee’s method has higher PSNR in some of the test images (e.g. Airplane(F-16)) because the blocks it chooses for embedding are optimized according to PSNR in comparison to the proposed methods in which all the coefficients having the same value are shifted at once. However, it is possible for the proposed algorithms to achieve higher PSNR at low capacities by carefully choosing and adjusting the start of the histogram shifting coefficients and limiting the shift of certain values to the number of to-be-embedded bits. By doing so, it adds computational
complexity in return of very limited perceptual and visual enhancement. In some images, such as Mandrill, that most of the methods fail to achieve embedding rates higher than 0.7 bpp, the proposed scheme can embed up to 1bpp in a single-pass run. The results of the experiment is shown in Figure 3.16.

Based on the illustrated results, it is apparent that generally the algorithms which operate in frequency domain achieve better results than those operating in spatial domain. This is due to the fact that data is more de-correlated in the frequency domain, which makes it possible to generate more space for information hiding while keeping the distortion level in the acceptable range. It should be noted that this effect is more significant at higher capacities.

To compare the performance of the proposed algorithms on images of larger sizes, both algorithms are applied on a 1024 × 1024 test image. Based on the results depicted in Figure 3.17, algorithm 2 performs slightly better than algorithm 1, regardless of the size of the test image.

The proposed scheme shows zero tolerance in case of any modifications or alterations on the marked image. This scheme, like majority of reversible watermarking schemes, is fragile to geometrical modifications and lossy compressions. However, the proposed technique can be employed along with a robust watermarking technique to compensate this imperfection.

### 3.3.2 Performance of the Scheme on Medical Images

Five medical images of different variety shown in Figure 3.10 are selected to test the proposed algorithms. All the images have the same size of 512 × 512 pixels.

The medical images are different from the natural images to a great extent as in medical images the pixels cover the entire range from 0 to 255 in most cases. On the contrary, most of the natural images such as those introduced in the previous section usually have pixel values ranging from 25 to 230. In other words, in case of strict
Figure 3.16: Distortion versus capacity graphs of the tested methods for different natural images (a) Lena, (b) Mandrill, (c) Airplane(F-16), (d) Peppers, and (e) Barbara
underflow and overflow protection by introducing larger safe regions on the boundaries of the acceptable grayscale values (here 0-255), the embedding capacity of the medical images will be smaller, as many wavelet coefficients cannot be modified. To investigate the issue further, Figure 3.18 compares the histogram of Lena and the MRI output of the brain. It is clearly noticeable that there are more pixels with values between 0 and 10 presented in the histogram of the medical image. However, having strict coefficient maps to protect the underflow and overflow to a large extent, does not always lead to the best performance. Choosing smaller safe regions despite the chance of underflow and overflow for some pixels, results in a better overall performance of the system. On the
Figure 3.19: Wavelet coefficient histogram of (a) MRI of the brain, and (b) Liver CT scan.

other hand, due to the characteristic of medical images, the histogram of the wavelet coefficients is more compactly centralized around 0, while still retaining the Laplacian shape distribution. Figure 3.19 shows the histogram of the brain MRI and Liver CT scan. Both images show that most of the coefficients have values close to 0. This suggests that few histogram shifts can yield high available watermark embedding space with less distortion for medical images than in case of natural images. Figure 3.20 shows the brain MRI image and the watermarked images at different rates, up to 0.9 bpp in a single-pass run. Keep in mind, figures are zoomed in so the difference can be magnified. Although the coefficient map used in this particular experiment is set to employ values between 5 and 250 in the embedding process, but no noticeable visible degradation due to underflow and overflow is present. The performance of the system in both cases, having a strict coefficient map and a loose one, is depicted in Figure 3.21 where the test image is the brain MRI. In case of a less-strict coefficient map not only the overall performance of the system is better, up to 3dB, but also the system provides a higher embedding capacity, which is not achievable when strict coefficient map is used due to the fact that fewer wavelet coefficients are permitted to be modified.
In case of medical images, the proposed scheme can be used in multi-pass runs to increase the overall capacity in exchange of higher degradation in visual quality. Figure 3.22 shows the MRI of the brain at different multi-pass capacity rates with the corresponding PSNRs.

Figure 3.23 shows the PSNR versus the embedded payload for the proposed algorithms and Lee’s algorithm [43]. The proposed algorithms outperforms Lee’s algorithm for capacities greater than 0.1 bpp. Comparing the results obtained for natural images (Figure 3.16), the improvement in hiding capacity while achieving higher PSNR is more significant in case of medical images.

The difference between the proposed algorithm 1 and 2 is more significant in high embedding capacity rates, where algorithm 2 performs slightly better than algorithm 1.
Figure 3.21: Comparison of performance effect caused by choosing either a strict or loose coefficient map.

Figure 3.22: Multi-pass watermarked images of the brain MRI (a) 2-pass watermarked image (bpp=1.8 PSNR=34.68dB), (b) 3-pass watermarked image (bpp=2.7 PSNR=29.12dB)

Utilization of weighted stopping parameters as in algorithm 2, can force the effect of the histogram shifts towards the subbands of the lower wavelet levels. This results in a slight improvement of the image quality, but since some of the wavelet coefficients in deeper levels are not used in the embedding process, the overall capacity is partly lower.

Furthermore, in medical applications in which the perceptual quality of the marked image is not important due to the fact that reversibility of the scheme ensures the preser-
Figure 3.23: Distortion versus capacity graphs of the tested methods for different medical images (a) MRI of the brain, (b) MRI of the breast, (c) CT scan of the brain, (d) CT scan of the liver, and (e) Ultrasound of the fetus.
vation of the original host image at any time, the proposed scheme can be applied to the
host image in multi-pass trials to embed enormous amount of watermark bits, multiples
of the highest single-pass capacity. Although as mentioned earlier, this method is appli-
cable where there is no concern about the visual and perceptual quality of the marked
image and the the amount of the to-be-embedded bits is of great interest.

3.3.3 Complexity Analysis

The computational complexity of the proposed scheme depends on different factors such
as the computational cost of the wavelet transform and histogram shifting. Discrete
wavelet transform imposes a standard complexity depending on its implementation \[70\].
The computational cost of histogram shifting is directly related to the number of input
pixels, \(N\), and the shifts, \(k_1\), as \(O(k_1N)\). In addition to these costs, the proposed scheme
requires to compute the resultant PSNR of the reconstructed image affected by each
shift in high frequency subbands of the first level wavelet transform in order to compute
the stopping parameters. The cost of such computations are proportional to the total
number of shifts in high frequency subbands of the first level wavelet transform, \(k_2\), and
the cost of discrete wavelet transform in each step.
3.4 Chapter Summary

In this chapter, a high capacity reversible watermarking scheme based on integer-to-integer wavelet transform and histogram shifting technique has been proposed. The technique can be employed in different scenarios to embed single watermark as well as multiple watermarks using single or multi-pass embedding procedures. The scheme is capable of achieving capacity rate of 1 bpp in a single-pass run for which most of the reversible methods fail to yield. Employing multi-pass runs can help the scheme to achieve even higher bit rates such as 2.7 bpp in a three-pass trial of Lena’s image. Multiple watermark embedding is an important advantage of this technique which makes it an applicable scheme in settings where multiple marks are needed to be inserted at different instants, using different keys, or for different means. This property can be of great significance in medical imaging applications, where diverse information with different importance and privacy index form the to-be-embedded data. In such a case multiple watermarking schemes are of great interest.

Through the experimental results demonstrated in sections 3.3.1 and 3.3.2 and also by comparing the achieved results from the proposed scheme with those claimed in several literature, one can observe that the proposed scheme performs better. This holds true for both natural and medical images at the embedding rate of 0.1 bpp or higher, in comparison to most of existing methods operating in either transformed or spatial domains.

The key points and advantages of the proposed system can be listed as follows:

- One-Time Pad (OTP) and AES algorithms are employed in concatenation to ensure the security and protection level of the marks. OTP, a strong encryption algorithm, is used to empower AES in case the encrypted marks’ patterns are traceable. The secrecy level of the proposed technique is guaranteed using keys with minimum length of 128 bits.
• Two methods to form scaleable coefficient maps are presented. These coefficient maps serve the scheme to prevent underflow and overflow in the marked images. These maps are generated by downsampling either the original image mask that singles out the pixels with values closed to the boundaries of the acceptable grayscale range or the transformed version of this mask in wavelet domain. Maps, generated through the proposed methods, benefit from very small sizes in comparison with coefficient maps in most of the existing techniques [28–30, 33, 62, 63] which require lossless compression.

• Integer-to-integer wavelet transform is used to decompose the original image rather than the floating point transforms in order to preserve the reversibility. To minimize the distortion level, high frequency subband coefficients are used for data insertion employing histogram shifting method.

• Stopping conditions are set through finding the most suited and efficient parameters yielding the best PSNR through a systematic algorithm. Two approaches concerning the ways to employ these parameters on different wavelet levels are proposed. The scheme is capable of performing single-pass and multi-pass runs, and most importantly, multiple watermarks can be embedded in each run using the proposed scheme.

• The proposed system outperforms most of the currently available reversible watermarking techniques in both hiding capacity and resultant perceptual quality (measured in PSNR) at rates higher than 0.1 bpp. The performance of the system on both natural and medical images for single-pass and multi-pass trials proves the usefulness of the proposed scheme.
Chapter 4

Electrocardiogram Biometric

Features Embedding

Accurate identification and validation of human identity is of great need in many recent applications. Access control, medical files privacy, copyright, and criminal investigation are some of the areas exhibiting the need for such reliable identity authentication. To achieve this goal, recent studies focus on characteristics that are physiognomy dependent for every individual, biometrics. Biometric characteristics show high correlation in a particular person and differ in traits from person to person. This makes it a unique self signature employed in different privacy secured applications for identification and verification.

Biometric features should comply with a few conditions and properties in order to be engaged widely in different applications. The first and most important criterion is the universality of the feature, which is connected to the natural aspect of the attribute. Furthermore, a biometric trait should be distinctive in a population, which guarantees the uniqueness of the formed signature. Another quality of biometrics is permanence which dictates no considerable change over a period of time. And finally, such a feature should satisfy measurability which stands for the quality that the feature is collectable
Biometric traits are categorized in behavioral and physiological groups. Behavioral characteristics are those revealing personalized patterns such as gait and keystroke, which are of great interest in surveillance applications. Physiological biometrics on the other hand, are the properties rooted in physical appearance and characteristics of an individual. This group includes the well known features such as fingerprint, iris, face, and Electrocardiogram (ECG).

In this chapter, the proposed reversible watermarking scheme introduced in chapter 3 is adopted to be used in the case of embedding biometric features into an image, as the mark or signature. This approach is applicable in various areas such as copyright, medical files security and privacy, automatic archiving, legal distribution, and other areas in which identification and/or verification of individuals are the main concern. To generalize the statement, it can be indicated that each and every application relates to security, privacy, and personalization of multimedia, specifically images, may benefit from employing the proposed scheme.

Electrocardiogram (ECG), the signals reflecting cardiac electrical activity over a period of time, are considered as biometric traits to be used in embedding process in this work. This chapter covers the techniques to extract the features, methods to yield the best quantized values (as the features should be in binary format to be embedded), and different approaches to allocate bits to the quantized features. In the end, the results from the proposed scheme are presented and discussed.

### 4.1 Electrocardiogram Preferences

Recently, there have been numerous researches in the concept of using medical attributes of the human body for privacy and security purposes. Examples of such features are Electrocardiogram (ECG), Electroencephalogram (EEG), heart rate, heart sound, and
blood pressure. In this work ECG is chosen as it satisfies the requirements for universality
and permanence, and also it has become a well defined mature subject in recent years.

The universality requirement is met given that ECG can be monitored from every
individual. Also the criterion of permanence is satisfied due to the fact that the main
structure of such signals is invariant over a large period of time except in rare cases
with abnormality. The diacritical waves in a subject’s heart beat can be monitored and
recorded through the life time. in addition, the heart, the structure creating the waves,
is very well protected inside body thus environmental factors cannot impact its activity
significantly. Another advantage of ECG as an effective biometric signal comes from the
structural functioning of each subject’s nervous system and heart. The electrocardiogram
waveform is managed by the autonomic nervous system, a combination of sympathetic
and parasympathetic factors. Therefore, reproduction of such a waveform pertaining to
an individual through systematic methods is almost impossible.

Besides these advantages there are some disadvantages to this biometric, such as the
noticeable impacts caused by mental stress and exercise or the presence of diseases such
as arrhythmia and their effects on the morphological properties of the waveform. On
top of these, the method used to collect ECG from each subject is not an easy private
procedure and can be categorized as an invasive acquisition procedure.

4.2 Ethical Issues in Biometrics

Biometric based systems and their applications are not a dissolved issue in the mind of
society even in recent days. Usually there is a public fear and mistrust formed in most of
the people regarding biometrics and their applications, mainly rooted in concerns about
violation of human rights and the possibility of relating the biometric features to personal
data [71]. Such issues are more intense in case of a non intrusive biometric system which
can capture the subject’s biometric signals without the individual’s consent. However,
even intrusive biometrics can lead to such concerns among the masses; as such information can be linked to some private data of each individual [71].

Therefore, the system employing the biometric features should be equipped with a mechanism to guarantee confidentiality. There are different methods allowing such confidentiality; the first option can be encryption which ensures the privacy of the biometric within a system. Another method is to store as little information about the raw biometric feature as possible, such that reconstruction of the original biological data is impossible.

To ensure the privacy and confidentiality of the proposed scheme, the encryption method described in section 3.2.5 of chapter 3 is used. In addition, other techniques are employed to reduce the size of the private biometric information in the to-be-embedded mark in order to prevent the reconstruction of the original biometric signal.

4.3 Fundamentals of ECG

ECG signals reflect the variations around the cardiac electrical potential in a period of time. The diversity in voltage is because of the action potentials of cardiac cells. The electrocardiogram is a non periodic but highly repetitive signal which means the signal has high correlation in periods of time. ECG is composed of three main waves created by the sequential depolarization and re-polarization of the heart. Figure 4.1 depicts the most significant components of an ECG signal, i.e. the P wave, QRS complex and the T wave.

The P wave is formed because of the depolarization of the right and left atria and has a duration of approximately 120 ms. The QRS complex which is the largest wave, reflects the depolarization of right and left ventricles, and its normal duration is between 70 to 110 ms. T wave represents the depolarization of the ventricles, it lasts for 300 ms following the QRS complex.

In spectral realization of an ECG wave, the P wave usually corresponds to the low
frequency components of about 10 to 15 Hz, the QRS complex contributes to higher frequencies which lie in the range of 10 to 40 Hz, due to its steep slopes.

### 4.4 Proposed Methodologies

This section introduces the techniques used in the proposed algorithm in order to embed ECG features into an image. To achieve this goal, the ECG signal is processed to a certain extent to carry enough distinctive information to allow reliable identification and verification and at the same time, make it impossible to reconstruct the original ECG signal due to privacy and security concerns.

All the identification systems can be modeled as a pattern recognition problem involving three main steps, original signals pre-processing, feature extraction, and classification. This section discusses the methodology adopted to fulfill the requirements in each step in order to extract the required reliable information from ECG waves to be used in data embedding.

The pre-processing step mainly deals with noise and artifact removals. Feature extraction and dimensionality reduction techniques are performed on windows of the ECG waves to form distinctive personalized signatures of each subject. Finally, classification
among a gallery set completes the process.

4.4.1 Preprocessing and Noise Cancelation

ECG waves suffer from a lot of noise in their raw shape just after collecting. The most common types of noise present in ECG waves are the baseline wander and the powerline interference. The baseline wander is the low frequency noise makes the signal to wander away from the isoelectric line; the source of such a noise is the possible respiration, body movements or inadequate electrode attachment. The powerline interference usually is the result of poor grounding or conflicts with other electrical devices. Noisy ECG signals cannot be used in a reliable feature extraction process; therefore, pre-processing in the shape of noise cancelation is inevitable. There are different choices of filtering techniques, both linear and non-linear, to be engaged in this step. In the proposed methodology a Butterworth band pass filter of order 4 is applied to the ECG raw signals. The cutoff frequencies of the used filter are set to 0.5 Hz and 40 Hz for low and high frequencies respectively. This filter shows effective results in cutting out the noise at low frequencies while retaining the rest of the significant information. Later, the filtered signal is divided into non-overlapping windows. It is preferred that the length of the windows to be chosen larger than the average heart rate such that each window contains multiple heart beats.

4.4.2 Autocorrelation Feature Extraction Technique

Feature extraction can be divided into two main steps. The first step mainly includes the employed method to extract the distinctive features. Subsequently in the second step, the dimensionality of the extracted vectors is reduced. The second step is discussed in the next section.

Autocorrelation (AC) feature extracting technique, a Non fiducial point method, is employed to extract the reliable features in the proposed scheme. Autocorrelation in general gets the repetitive property of the electrocardiogram, and its shape is highly
Figure 4.2: Autocorrelation feature waveforms of two different subjects dependent on the P, QRS and T waves. Distinctive patterns related to characteristic of each subject’s ECG can be obtained by studying its autocorrelation. The AC represents the similarity features over multiple heart beat cycles of an individual’s ECG. The AC equation which is used to obtain these features is presented as

$$\hat{R}_{xx}[m] = \frac{\sum_{i=0}^{N-|m|-1} x[i] x[i + m]}{R_{xx}[0]}, \quad (4.1)$$

where $x[i]$ is a window of the ECG signal segmented in the pre-processing step and $x[i + m]$ is the time shifted version of that window, and $m = 0, 1, ..., (M - 1)$ for which $M \ll N$. Although the AC signal is formed by contribution of all three major waves in an ECG, but because of large variations in amplitudes, normalization is necessary. Therefore, after applying AC to ECG windows of each subject, the value of each obtained AC is normalized by its maximum. Figure 4.2 illustrates the distinctive characteristic of AC features obtained from two different subjects. It is clear that AC features of each subject are all aligned and they almost follow the same pattern. On the other hand, these features differ from one subject to the other to a great extent.

Although for a particular subject, all the AC vectors are extracted from the individual’s ECG signal, some of the used ECG instances (windows) might have been recorded
in situations where the subject was not at rest or did not stay in the same mental or physical condition consistent to the rest of the recording. An outlier removal technique is employed to exclude the AC features generated from such instances in order to increase the efficiency and accuracy of the scheme.

The outlier removal method starts with computing the average AC vector of all the obtained AC vectors pertaining to a subject. Then, the correlation of the average AC vector with all the AC vectors are calculated. Each AC vector which has a correlation less than a predefined threshold is excluded from the procedure. This process ensures that the method proceeds with the most reliable steady features of a subject.

### 4.4.3 Dimensionality Reduction Techniques

The obtained AC features carry enough distinctive information so they can directly be used for classification or quantization in this work. However, the dimensionality of each of the AC features, depending on the sampling frequency of the ECG, can be significantly high and make it unsuitable as a watermark due to the limited space available in images. To reduce the dimensionality and consequently the payload size of the to-be-embedded mark, special techniques are acquired. To serve this purpose two techniques are chosen among many available, *Principal Component Analysis* (PCA) and *Linear Discriminant Analysis* (LDA)\[71\]. It has been shown through experiments\[71\] that PCA serves better in cases where the goal is to verify the identity of a subject. In the watermarking applications, verification may refer to instances where there is a need to compare the embedded ECG features with the features of a nominee to verify the authenticity of the identification. On the other hand, LDA is better suited in applications where identification is required. This requirement imposes the need of identifying the embedded ECG features among a set or a pool of ECG features; in other words, identification means to point out the subject, whose ECG is embedded, in a set including that subject and others.
### Principal Component Analysis (PCA)

PCA is a well known unsupervised learning technique to yield optimal projection in lower dimensions by transforming possibly correlated variables into a smaller number of uncorrelated variables. This process retains all the useful information, known as principal components, and eliminates redundancy.

PCA is proposed to be applied to the training set which contains all of the autocorrelated ECG windows corresponding to different subjects. Given a training set \(X = \{X_i\}_{i=1}^U\), where \(X_i\) is the class of each subject and \(U\) the total number of classes. It is assumed that each subject has \(U_i\) autocorrelated ECG windows, \(x_{ij}\). Then, the class of each subject is given by \(X_i = \{x_{ij}\}_{i=1}^{U_i}\) and the total number of AC features for all subjects is \(N = \sum_{i=1}^U U_i\). The PCA technique is applied to the training set \(X\) to find the \(L\) most significant eigenvectors of the covariance matrix corresponding to the training set \(X\),

\[
S_{\text{cov}} = \frac{1}{N} \sum_{i=1}^U \sum_{j=1}^{U_i} (x_{ij} - \bar{x}) (x_{ij} - \bar{x})^T,
\]

where \(\bar{x} = \frac{1}{N} \sum_{i=1}^U \sum_{j=1}^{U_i} x_{ij}\) is the average of the ensemble. After obtaining the \(L\) eigenvectors, denoted by \(\Psi\), the original AC vectors are transformed to the lower \(L\)-dimension subspace by a linear mapping

\[
y_{ij} = \Psi^T (x_{ij} - \bar{x}),
\]

in which the basis vectors \(\Psi\) are orthonormal \([71]\).

As stated earlier, PCA is usually employed in applications with the requirement of verification and authentication to verify the authenticity of an identity. In PCA, Cosine distance is used to compare the embedded feature vectors with feature vector of the authentication nominee. The output of this measure is compared against a predefined threshold to decide the final result. The Cosine distance can be computed using the
following relation

\[ D(x_i, x_j) = 1 - \frac{x_i \cdot x_j}{\|x_i\| \|x_j\|}. \] (4.4)

### 4.4.3.2 Linear Discriminate Analysis (LDA)

LDA is a well known method with applications in different areas such as statistics, pattern recognition and machine learning. The method is used to find a linear combination of features which characterize or separate two or more classes of objects in order to reduce the dimensionality of features besides having the classes better distinguished.

Given the definitions of sets and classes specified in PCA section, a set of \( K \) feature basis vectors \( \{\psi_m\}_{m=1}^K \) are estimated by maximizing Fisher’s ratio or equivalently solve the following equation

\[ \psi = \arg \max_{\psi} \frac{|\psi^T S_b \psi|}{|\psi^T S_w \psi|}, \] (4.5)

where \( \psi = [\psi_1, ..., \psi_K] \), and \( S_b \) and \( S_w \) are the between and within class scatter matrices respectively. These matrices can be computed as follows

\[ S_b = \frac{1}{N} \sum_{i=1}^{U} U_i (x_i - \bar{x}) (x_i - \bar{x})^T, \]

\[ S_w = \frac{1}{N} \sum_{i=1}^{U} \sum_{j=1}^{U_i} (x_{ij} - \bar{x}_{i}) (x_{ij} - \bar{x}_{i})^T, \] (4.6)

where \( \bar{x}_i = \frac{1}{U_i} \sum_{j=1}^{U_i} x_{ij} \) is the mean of class \( X_i \) and \( N = \sum_{i=1}^{U} U_i \) is the total number of ECG windows per subject.

This maximization forces the projected features of different subjects to separate further while keeping the variance between features of the same subject small. In LDA, \( \psi \) is defined as the \( K \) most significant eigenvectors of \( (S_w)^{-1} S_b \) which correspond to the first \( K \) largest eigenvalues. The input autocorrelated ECG window \( x \) is transformed by linear projection \( y = \psi^T x \).
LDA dimensionality reduction method is better suited to identification applications. In this method, a measure based on the *normalized Euclidean* distance and the nearest neighbor is employed to find the best match of the embedded ECG features. The normalized Euclidean distance is defined as

\[
D(x_1, x_2) = \frac{1}{V} \sqrt{(x_1 - x_2)^T (x_1 - x_2)},
\]

where \( V \) is the dimensionality of the feature vectors, \( x_i \).

### 4.4.4 Quantization

The necessary ECG feature vectors are obtained following the aforementioned processes. In order to embed a watermark into the host image, the mark must be converted into binary format. The output of the feature extraction methods are real value vectors, so the first step towards embedding is the quantization of these real values.

This section discusses the proposed methods to quantize the output of the feature extraction algorithm. In this study, both linear and non-linear quantization methods are adopted to cope with the requirements. In the proposed scheme, *uniform* quantization is the linear quantization method suggested and among non-linear methods, \( \mu \)-Law quantization is adopted.

#### 4.4.4.1 Uniform Quantization

Uniform quantization is the straightforward solution to the quantization problem. This technique is used whenever the signal is in a finite range \((f_{\text{min}}, f_{\text{max}})\). To start with quantization, the entire data range is divided into \( L \) equal intervals of length \( Q \), known as *quantization interval* or *quantization step size*. The step size can be computed as follow

\[
Q = \frac{(f_{\text{min}} - f_{\text{max}})}{L}.
\]
In the proposed scheme, the values within the range of each interval are mapped to the middle value of the range. This means that in the uniform quantization technique all the real values within the range of an interval are mapped to the middle point in that interval. This process can be shown as follows

\[
Q_i(f) = \left\lfloor \frac{f - f_{\text{min}}}{Q} \right\rfloor, \tag{4.9}
\]

\[
Q(f) = Q_i(f) Q + \frac{Q}{2} + f_{\text{min}},
\]

where \(Q_i\) is the index of quantized value.

To make the case simpler, the quantization range can be shifted up, so that \(f_{\text{min}} = 0\). This modification is necessary here, as it omits the need to save the negative values and to convert them into binary format.

In uniform quantization method the number of levels, the quantization step size, and the minimum value of the original signal prior to any shifts are the necessary information to be recorded. This information is used at the decoder to reconstruct the quantized values from the saved level indicator of each value.

Uniform quantization is optimal for signals with uniform distribution, but since extracted features through both PCA and LDA methods are not uniformly distributed, non-uniform quantization may serve the purpose better.

### 4.4.4.2 Non-Uniform Quantization

Using non-uniform quantization allows of having smaller quantization intervals near zero value. Since the feature vectors have a monotonic push and many values concentrated near zero, non-uniform quantization is adopted in the proposed scheme.

There are two ways to perform non-uniform quantization on a set of data. The first method is to directly design the intervals and reconstruction levels, which seems practical in situations where the input data has similar waveform and range. This method is practically impossible in this work as neither the waveform nor the range of data are
identical in all the extracted features. The other method is to use a non-linear mapping scheme to impose the required non-linearity level on the input data followed by uniform quantization of the mapped data. A well known non-linear method which adopts the latter technique is the $\mu$-Law quantizer.

The implementation of the $\mu$-Law quantization method can be summarized in the following steps:

- First, the input signal $x$ is transformed using the $\mu$-Law to form signal $y$ through the following relation

$$y = F(x) = X_{\text{max}} \frac{\log \left( 1 + \mu \frac{|x|}{X_{\text{max}}} \right)}{\log (1 + \mu)} \text{sgn}(x). \quad (4.10)$$

- The transformed value, $y$, is quantized using the uniform quantizer introduced in section 4.4.4.1.

- Finally, the quantized value, $\tilde{y}$, is transformed back using the inverse $\mu$-Law

$$\tilde{x} = F^{-1}(\tilde{y}) = \frac{X_{\text{max}}}{\mu} \left( 10^{\frac{\log(1+\mu)}{X_{\text{max}}} |\tilde{y}|} - 1 \right) \text{sgn}(\tilde{y}). \quad (4.11)$$

It should be noted that the above algorithm holds in case the input signal is in the form of a vector or a matrix. The equations in the above algorithm can be simply adjusted to let them fit in vector or matrix algebra rules.

The proposed $\mu$-Law quantization method is employed in two different ways to obtain the most acceptable results. In one approach, the extracted features are all shifted towards positive infinity, so that $f_{\text{min}} = 0$ and then the $\mu$-Law is performed on the shifted features following with the uniform quantization operation. In the other approach, the $\mu$-Law is applied to the features, then the transformed results are shifted up towards positive infinity, and finally the uniform quantization is performed.

In non-uniform quantization method, there are necessary information required at the receiver side to allow full reconstruction of the quantized values from the embedded indicators. This information and the adopted method are discussed in section 4.5.
4.4.5 Binarization and Bit Allocation

To convert the quantized values into binary format, there are different techniques employed in the proposed scheme. It should be stated here that due to the shifts done in the quantization algorithms to level the minimum value of the input signal with zero, all the quantized values resulting from either methods are positive values; hence, there is no need to consider negative values in binarization algorithms.

The main concern in this stage is the total number of allocated bits. This factor relates directly to the overall to-be-embedded payload size in the embedding process. Two approaches are considered in this work, the first approach is to assign a predetermined number of bits to each quantized value, or in other words, to treat each quantized value independently and identically. The other approach is to determine the total quantity of available bits and then distribute them among the quantized values through a sensible practical scheme while considering statistical factors. This approach is a bit allocation process which assigns different number of bits to different quantized values considering the reliability of each value and the amount of information it carries. It is stated that optimized bit allocation is possible if there is enough side information and statistical knowledge.

This section introduces three different approaches to fulfill binarization.

4.4.5.1 Predefined Bit Allocation

In this approach predefined number of bits is allocated to all feature values regardless of their characteristics. The number of allocated bits to each feature is directly related to the number of levels in the quantization step, the more bits allocated to a single value the more quantization levels are possible and better resolution is achieved. The relation states that if \( n \) bits are allocated to a quantized value there are \( 2^n \) levels in the designed quantizer.
4.4.5.2 Greedy Bit Allocation

The Greedy bit allocation approach is used in cases where the total number of bits to be assigned is defined. In such a case the resource is limited so the bit allocation should be done considering the importance and informative property of each value. Greedy bit allocation follows a two-pass uniform bit allocation algorithm. Given $n_b$ to be the total number of bits as the resource and $n_c$ the number of elements in a feature vector, in the first pass $\left\lfloor \frac{n_b}{n_c} \right\rfloor$ bits are allocated to each element of the feature vector. The remaining bits, $n_r = n_b - n_c \times \left\lfloor \frac{n_b}{n_c} \right\rfloor$, are then allocated to the first $n_r$ feature elements in the second pass [72].

This approach allocates an extra bit to the most significant components of each feature vector, if the total number of bits is not a multiple of the number of feature components.

If Greedy bit allocation is employed, quantization step can be done in two different approaches. In the first method, the quantization can be done by treating all the values equally and later in the bit allocation step more bits can be assigned to the most significant values. The alternative approach divides the elements into two groups and quantization is done separately considering the number of bits allocated to each group.

4.4.5.3 Component Reliability Based Bit Allocation

In cases where the feature elements are not in a decreasing order of reliability, bit allocation is done through computing and considering a reliability factor for each element [72]. In the proposed ECG feature extraction method, several ECG windows from each subject are taken to calculate the AC vectors. Therefore, each subject $i$ is defined with a number of different feature vectors. The proposed reliability factor for a particular subject $i$, and feature element $t$, is denoted as $R_{i,t}$ and defined as

$$ R_{i,t} = \frac{1}{2} \left( 1 + \text{erf} \left( \frac{|\mu_{i,t} - \mu_i|}{\sqrt{2\sigma_i^2}} \right) \right), $$ (4.12)
where $\mu_{i,t}$ is the subject mean of element $t$, $\mu_i$ the population mean of total elements belonging to the subject, and $\sigma^2_t$ the subject population variance.

The reliability factor defines the probability that a new measurement from the same subject results in the same feature element, in the case that a Gaussian distribution is assumed for the set of the feature elements. As stated earlier, computation of the reliability factor requires information about the examined subject mean and variance, which indicates multiple training (enrollment) instances should be used in order to estimate these values.

In order to adopt this method in the current work, the following steps are proposed:

1. The reliability factor of each feature element is computed.

2. In cases where the bit resource is so limited, a threshold can be considered to filter a Candidate Set (CS) among all components to be used in the bit allocation process. Those elements falling out of this set are discarded and no bit is allocated to them.

3. To allocate bits to each element considering the reliability factor, a linear rescaling should be performed at this point. The scaling factor represents the number of bits allocated to a feature component per one unit of reliability factor. This parameter can be computed as

$$\rho_t = \frac{n_b}{\sum_{t \in CS} R_t},$$

where $n_b$ is the total number of bits available, $R_t$ the reliability of element $t$, and $CS$ the candidate set defined earlier.

4. To allocated the bits to each element, the following steps are taken:

   - Scaling factor is multiplied to each element reliability factor, those belong to CS, and rounded towards minus infinity to form $n_{c,t} = \lfloor \rho_t \times R_t \rfloor$.

   - $RS$ the sum of $n_{c,t}$’s are computed, and is compared to the total number of bits, $n_b$: 
- If $RS = n_b$ then the allocation is complete and $n_{c,t}$ defines the number of bits allocated to each component.

- If $RS < n_b$ then there are bits left to be allocated, so the $n_{c,t}$ of the $(n_b - RS)$ feature components with highest reliability is incremented by 1.

- If $RS > n_b$ is the case, then some allocated bits should be revoked from those elements having lowest reliability. The $n_{c,t}$ of the $(RS - n_b)$ with lowest reliability is decremented by 1.

It should be noted that employing component reliability method in bit allocation increases the complexity of the system in total.

### 4.4.6 Verification (Authentication)

Verification is a one to one process in which the validation of an identity claim is examined. In a verification scenario, one subject claims an identity for which the system determines the degree of similarity between the subject’s biometric features with the claimed identity registered features, and then based on the obtained similarity the access to that identity is granted or rejected. The measure used in verification is usually based on a predefined threshold applying to the distance between the feature vectors of both parties, the subject who claims the identity and the claimed identity. For verification applications both PCA and LDA methods can be employed in feature extraction process. The terms corresponding to different aspects of the verification (authentication) process are listed as follow:

- Verification accuracy or authentication of legitimate subjects, measured in authentication or verification rates.

- Deny identity authentication to legitimate subjects, measured in False Rejection Rate (FRR).
• Authenticating intruders or illegitimate subjects, measured in False Acceptance Rate (FAR).

Each of these measurements is computed as fractions of the desired set divided by the probe set. The relations for false acceptance and rejection statistics are computed as

\[
\begin{align*}
\text{FAR} &= \frac{\text{Number of falsely authenticated subjects}}{\text{Total number of intruders}}, \\
\text{FRR} &= \frac{\text{Number of rejected legitimate subjects}}{\text{Total number of subjects}}. \\
\end{align*}
\]  

(4.14)

The Equal Error Rate or Crossover Error Rate (EER or CER) is the case that both FAR and FRR are equal, i.e. EER = FAR = FRR. This measure shows the authentication performance of the system, the lower the EER is the more reliably authentication is done in the biometric system.

The proposed verification method is summarized in the following steps:

1. The Cosine distance in case of PCA and Euclidean Distance for LDA, between the feature vector of the claimer and the registered feature vector of the claimed identity are computed.

2. The calculated distance in each case is compared against a predefined threshold:

   • If the distance is less than the threshold the identity claimed is granted and validated.

   • If the distance is greater than the threshold the identity claimed is rejected.

4.4.7 Identification

Identification as a one-to-many process is the procedure of distinguishing an identity among others, where no prior information about the identity of the subject is available. In this work, the group in which the individual has to be identified considered to be a closed set of subjects whose ECG features are known to the system, known as registered subjects.
Although the proposed reversible watermarking scheme can be used in applications where blind identification process is required by adopting methods which perform better in such environments, this concern is not the main goal of this research. In identification applications, LDA is employed in the feature extraction process due to its properties. In a closed set, applying LDA moves the projected features belonging to each subject to a more centered cluster while keeping them further separated from features of other subjects, such that the classes or subjects are better distinguished. In order to perform identification, all the extracted biometric features registered to the system have to be compared to the set of features which identifies the wanted individual, to find the best possible match.

The main terms which are frequently used in the identification process are elaborated as follows [71]:

- Identify an individual correctly, which is measured in *identification* rates.
- Misidentify an enrolled individual, which is measured in *miss-identification* rates.

The proposed method to serve the purpose consists of following steps:

1. The Euclidean distances between all the compact ECG features from the testing set and those extracted from the image are computed.

2. To narrow down the list of nominated identities, a predefined threshold is used to discard those features which distances with the target are larger than the set number. This helps to filter those features outside the reasonable distance and decrease the comparison time and complexity in the next step.

3. The remaining features are indexed and sorted. The voting process decides the chosen identity. The subject which has the most number of the remaining features is selected as the determined identity. In case of a draw, the one with the smallest distance is chosen.
4.5 Proposed Biometric Embedding Algorithm

This section illustrates the proposed algorithm to embed a biometric signal, specifically electrocardiogram (ECG), into a cover image. This framework can be employed in applications where identification or verification processes are required. Examples of such applications are patients’ personal information, medical or personal files archiving, copy right issues dealing with images, ID cards and many more.

Figure 4.3 shows the general components of the proposed algorithm. The procedure can be divided into four major entities with defined tasks. The first entity deals with obtaining ECG, its feature extraction, quantization and binarization, or in other words, processing the obtained ECG to be embeddable considering criteria of watermarking and biometrics. The second unit embeds the binary ECG features as watermarks into the cover image using the scheme described in Chapter 3. The third part is responsible for the embedded data extraction and original image recovery at the detector or receiver. Finally the last entity compares the extracted features employing methods described in 4.4.6 and 4.4.7 to verify or identify an identity.

The aforementioned entities and the steps to complete the overall process are ex-
plained in the following. The first section of the procedure explains the ECG feature extraction and mark embedding and the second part discusses the watermark extraction and profile match up.

1. The ECG of the subject whose identity needs to be used as the mark is acquired during enrollment. It is preferred that the collected ECG duration be a few minutes long to ensure the precision of the extracted features.

2. The collected ECG is preprocessed to yield a clean and noise free signal employing the method described in 4.4.1. Then windowing is applied to the clean signal in such a way that each window contains at least a few heartbeats or equivalently a few QRS complex waves.

3. To extract the features of the registered ECG, autocorrelation method is applied to each window followed by AC normalization. Each of normalized autocorrelation waveforms are then cut to a point to preserve only the magnified effect of a QRS complex which is its duration just before the T wave starts to pick up.

4. To reduce the dimensionality of the extracted features either PCA or LDA transformation methods are employed. It is essential to note that if the objective of the application is verification (authentication), then either one of the methods can be engaged. However, if the purpose is identification, then the LDA method is suggested.

5. The bit allocation technique determines how many bits are assigned to each of the feature components. Depending on the chosen method for bit allocation, the side information also known as overhead in the embedding process may differ. In case of the predefined bit allocation, the only information needed to be carried on from this step is the number of bits per component. If Greedy bit allocation is chosen, the number of minimum bits assigned to all components along with indices of those
components earned an extra bit are saved as overhead information. In component reliability method, a vector with the same size of the feature vector is engaged to record the necessary information; each element of this vector represents the number of allocated bits to the corresponding feature component. It can be deduced that although the purpose of the component reliability method is to make the number of assigned bits as efficient as possible in order to reduce the payload size, in cases that the number of all feature vectors belonging to a subject is relatively small in comparison to the total number of elements in each feature vector, i.e. a subject with two PCA feature vectors each having 10 components, this method can result in an increase in total number of assigned bits and consequently the size of the payload.

6. The compact extracted features are quantized by the means of either uniform or non-uniform quantization methods explained in sections 4.4.4.1 and 4.4.4.2. The quantization method in case of predefined bit allocation is straightforward, as the quantity of bits assigned to each value (feature component) determines the number of quantization levels. In case of Greedy or Component reliability methods, two approaches are proposed. In the first approach, the quantization levels are chosen according to the highest number of bits allocated to a value, afterwards, the values of those components with fewer assigned bits are rounded to the closest level the bits can represent. The advantage of this method is that a global quantization condition is applied to all the components; therefore, the side information which is required at the receiver is limited to the number of levels, maximum value of the range, and the quantization step size of a single quantizer. The other proposed approach which provides better accuracy in the expense of higher computational complexity and more side information, suggests of employing different quantization structures for each group of components assigned the same number of bits. In the Greedy method, there are two designs (two different quantizers) in total, hence two
sets of quantization side information needed to be embedded as the overhead. In
the case of Component Reliability approach, the number of designs relates directly
to the number of bit assignment groups. The more diverse the numbers of assigned
bits are, the more groups are shaped and more information is needed at the decoder
side. The latter approach increases the total payload but in cases that the extracted
ECG features are obtained from several instances, the side information payload can
be negligible in comparison to the size of the total feature components, therefore,
this approach can be efficient while serving a better accuracy. Although this is not
a major concern as the proposed reversible watermarking scheme in chapter 3 is
a high capacity method, in cases where other schemes are of interest or employed
on top of the proposed one in order to serve other properties such as robustness
against geometrical modifications [51, 53], this may be a concern.

7. After binary form of all feature vectors are obtained, the embedding process starts
as described in chapter 3, till all necessary bits of the payload, the mark (ECG fea-
tures) and overhead, are embedded. The overhead embedded in this step contains
the same parameters defined in chapter 3 with an addition of the side information
that is required to retrieve the quantized value at the receiver and the necessary
bit allocation vectors in case of Greedy or Component Reliability methods. After
all bits are embedded the marked image is created.

The marked image produced through the aforementioned steps can be stored and used
in different occasions. Furthermore, the marked image can be used for multi-pass and
multiple watermark embedding processes, as explained in chapter 3.

The following steps explain ECG feature mark extraction and identification/verification
algorithms:

1. The procedure stars with watermark extraction and original image restoration. The
necessary side information is extracted and kept to be used in the upcoming steps.
2. Using the quantization and bit allocation extracted information, the quantized values of feature components are recovered.

3. If verification is the objective of the application, the ECG signal of the subject who claims the identity is collected to be compared to the extracted ECG mark. On the other hand, if identification is the main objective, the ECG of all the registered subjects are used to identify the identity extracted from the marked image. The same feature extraction method as in embedding process (PCA or LDA) is applied to the collected ECG signals and the feature component vectors are obtained.

4. To verify or identify the identity using comparison methods explained in 4.4.6 and 4.4.7, two different approaches are proposed. One approach suggests the comparison of the extracted quantized ECG features to the collected real value features with no further processing done, i.e. quantization. In this method the distance between the recovered features from the image and collected real value features from the subject(s) during the test are computed using Cosine or Euclidean measures; this information can be used to verify or identify an identity. The other proposed method imposes quantization on the collected features using the quantizer information extracted from the overhead, so that the basis of the comparison using distance measurements can be the same. In other words the collected features are quantized and then used in comparison techniques to determine the validity of the claim or the identity of the subject.
4.6 Experimental Results and Applications of the Proposed Scheme

Due to privacy and security needs, there is a growing interest in ECG biometric signal and its applications in different fields. The proposed scheme can be used in environments where there is a need for authentication or identification. Prime examples of such applications are Electronic Health Record (EHR), automatic personal file archiving, copyright ownership, and media distribution. The performance of the proposed algorithm within different scenarios is evaluated and results are compared to the case where no embedding is required hence, no quantization is performed. This comparison allows to evaluate the effect of the quantization and binarization of ECG features on the accuracy and performance of the system. The scenarios are divided into two groups of applicability, verification experimental results and identification experimental results.

4.6.1 Verification Experimental Results

The proposed scheme is tested on The Biometric Security Laboratory (BioSec.Lab) database [73]. This database consists of ECG recordings of 52 healthy subjects in 2 instances of 3 minutes, collected in two occasions over a month at university of Toronto. The collected ECGs are then sampled with the frequency rate of 200 Hz. The first 5 seconds of all the recordings are cut out due to the presence of strong noise and relatively low SNR. The remaining signals are filtered using a Butterworth filter of order 4 with cut off frequencies of 0.5-40 Hz, these cut off frequencies are empirically defined. Figure 4.4 shows an ECG sample used in this experiment before and after filtering. Later, each cleaned ECG signal is divided into windows of 5 seconds, then the autocorrelation method is applied to each window. The autocorrelation waveforms are single sided and then chopped to 400 ms (keeping 80 samples) to retain the characteristic of the main QRS complex, as this wave is less affected by emotional conditions [74]. The ACs are
divided into two sets, training (gallery) and testing. The training set features are used in the processes prior to watermark embedding; and the testing set features are employed to be compared to the embedded marks. To evaluate the performance of the proposed algorithm, both PCA and LDA methods are applied to the training set. To reduce the size of the feature vectors in each of these cases, the most significant feature components are retained which possess the majority of the signal energy. In PCA method the criterion is to keep the components holding 98% of the total energy of the signal, which results in feature vectors with the length of 10. In LDA the condition is on the eigenvalues of \( S_b \) which are chosen to be greater than 0.0001; this results in feature vectors which each consist of 32 elements. Using the bit allocation methods described in section 4.4.5, the training set features are then quantized. At this stage, only the method using multiple quantizers as Greedy and Component Reliability bit allocation techniques is considered due to its better performance. Quantized values are then converted to binary format according to the number of allocated bits to each of them. Throughout the experiments, and for all the bit allocation methods, the average number of bits per element in each feature vector is set to be 6 bits. Therefore, if PCA is employed the to-be-embedded watermark for each feature vector has a size of 60 bits; on the other hand, if LDA is used, the watermark size is 192 bits.

Lena’s image is chosen to test the performance of the system throughout the experiment. To start with the embedding process, the binary ECG features are encrypted. Then the encrypted watermark is embedded into the cover image along with all the necessary side information. The watermark size depends on the chosen ECG feature extraction method and the criterion applied to it. The side information to be embedded in this scheme includes all the data detailed in chapter 3 along with the necessary details used in quantization part, such as number of bits allocated to each component, maximum of the quantization range and quantization step. The size of quantization and bit allocation side information depends greatly on the size of the feature vector and the
Figure 4.4: Collected ECG signal from BioSec.Lab \[73\] database (a) ECG signal before filtering, (b) filtered ECG signal

bit allocation method engaged. In case of Component Reliability technique which is the most excessive technique among the others in terms of side information volume, if PCA is used the required data at the receiver side is a bit allocation vector with the same size of a PCA feature vector. This vector keeps the information of the original maximum or minimum values for all the used quantization designs (the number of different designs, at greatest, can be the same as the number of elements of a feature vector in the case no two components are assigned the same number of bits), and the quantization step value of each quantizer employed.

To verify the identity, the embedded ECG mark is extracted and the original image is restored. The extracted mark is used to authenticate the identity claim exploiting methods explained in 4.4.6. To test the performance, all the subjects in the training set are used to be validated with all the subjects within the testing set. The threshold values used on the Cosine and Euclidean distances are defined empirically through inspection of the experimental results to cover the most meaningful range. To evaluate the verification performance of the framework under both PCA and LDA techniques, Predefined and Component Reliability bit allocation methods are employed. $\mu$-Law quantizer is the
chosen quantization method during the tests because of its better performance compared to uniform quantizer. Figures 4.5 and 4.6 depict the verification rates of the system at different distance threshold values in case PCA and LDA are employed respectively. The authentication method employed suffers from the undesired false rejection and acceptance effects, where the system does not validate the identity of a legal user or authenticates the wrong identity or an intruder. Figures 4.7 and 4.8 illustrate the false acceptance and false rejection rates (FAR and FRR) at different threshold values for PCA and LDA features respectively, employing Component Reliability and Fixed bit allocation techniques. The Equal Error Rates (EER) and the corresponding verification (authentication) rates of PCA and LDA features are shown in Table 4.1. For comparison purposes and to evaluate the effect of quantization and binarization on the extracted ECG features, the case that no embedding is present, i.e. no quantization and binarization, is experimented. The authentication rates and the FAR and FRR diagrams of this case are depicted in Figure 4.9 and the ERRs and the related validation rates are presented in Table 4.2. It can be seen that the quantization and binarization processes have substantially minor effects on the whole performance. At ERR the verification rate is dropped from 92.99% to 92.22% for PCA features and in case LDA is used the verification rate is 91.21%.
Chapter 4. Electrocardiogram Biometric Features Embedding

Figure 4.5: Verification performance of the framework employing PCA (a) Predefined bit allocation method, (b) Component Reliability bit allocation method

Figure 4.6: Verification performance of the framework employing LDA (a) Predefined bit allocation method, (b) Component Reliability bit allocation method

4.6.2 Identification Experimental Results

Linear Discriminant Analysis (LDA) is the method used in applications with identification objectives. All the processes prior to embedding step are done as explained in the verification performance section, except that in this case, LDA is the only method employed in the feature extraction phase. All subjects in the training set are processed
and converted into binary format to be used in the comparison to those from the testing set; the data in the latter set is employed in "identification matching" procedure in two different ways, real value feature vectors and quantized feature vectors of the testing set. The performance is computed for different bit allocation techniques to present the effect of binarization methods proposed.
Table 4.1: Error Equal Rate (EER) and the corresponding verification rates based on PCA and LDA feature extraction.

<table>
<thead>
<tr>
<th></th>
<th>Equal Error Rates (EER) (%)</th>
<th>Verification Rate (%)</th>
<th>corresponds EER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PCA - Predefined bit allocation</strong></td>
<td>6.21</td>
<td>92.16</td>
<td></td>
</tr>
<tr>
<td><strong>PCA - CR bit allocation</strong></td>
<td>6.19</td>
<td>92.22</td>
<td></td>
</tr>
<tr>
<td><strong>LDA - Predefined bit allocation</strong></td>
<td>7.32</td>
<td>90.86</td>
<td></td>
</tr>
<tr>
<td><strong>LDA - CR bit allocation</strong></td>
<td>7.14</td>
<td>91.21</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2: Error Equal Rate (EER) and the corresponding verification rates in case of no data quantization for evaluation of binarization effect on the verification process.

<table>
<thead>
<tr>
<th></th>
<th>Equal Error Rates (EER) (%)</th>
<th>Verification Rate (%)</th>
<th>corresponds EER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PCA</strong></td>
<td>6.19</td>
<td>92.99</td>
<td></td>
</tr>
<tr>
<td><strong>LDA</strong></td>
<td>6.55</td>
<td>91.98</td>
<td></td>
</tr>
</tbody>
</table>

The identification process starts with extraction of the embedded ECG feature vectors and restoration of the original cover image, Lena’s image is the host image used in the identification experiments in this work. The extracted feature vectors are used to be compared with all of the subjects’ features from the testing set. The testing set is used for performance evaluation. In this experiment, the results obtained from both real value and quantized versions of the testing set are almost identical therefore, only the results from the real value version of the testing set are demonstrated. The Euclidean distance between the feature vectors of all subjects in the testing set and the to-be-identified
Chapter 4. Electrocardiogram Biometric Features Embedding

Figure 4.9: FAR, FRR, verification rates in case of no data quantization (a) PCA verification rate, (b) PCA false acceptance and rejection rates, (c) LDA verification rate, (d) LDA false acceptance and rejection rates.

feature vectors are computed. Then, using the method explained in 4.4.7, the subject is identified. In all tested scenarios all subjects are fully identified (identification rate of 100%) considering multiple compact ECG feature vectors embedded for each subject. In this experiment 30 feature vectors are embedded into the image and the final decision is made upon the voting method described in 4.4.7. The identification rate in case of embedding a single ECG feature vector into the host image is lower in all the cases. Table 4.3 presents the subject identification rates of a single ECG feature vector for each of the
aforementioned scenarios, also for the comparison purposes the identification rate of the case with no quantization and binarization is given. As within each case, the obtained results are different for each subject, the average value over 52 subjects are stated in the table. It can be seen that the identification rates are almost identical in all the cases, but the use of Component Reliability bit allocation results in the best performance. It can be noticed that introduction of quantization and binarization to the system slightly affects the system’s performance. This shows that quantization, binarization, and watermark embedding can be employed in ECG biometric systems without affecting the verification and identification performances of the system to a great extent.

Table 4.3: Identification rate of a single ECG feature vector.

<table>
<thead>
<tr>
<th>Identification Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predefined bit allocation - real value testing set</td>
</tr>
<tr>
<td>Greedy bit allocation - real value testing set</td>
</tr>
<tr>
<td>CR bit allocation - real value testing set</td>
</tr>
<tr>
<td>Case of no quantization and binarization</td>
</tr>
</tbody>
</table>

4.7 Chapter Summary

This chapter proposed a framework to embed a specific biometric signal, ECG, into a cover image for the verification and identification purposes to accommodate a large variety of applications. The scheme methodology was explained thoroughly and all the components were listed. The method employs autocorrelation feature extraction technique along with PCA or LDA dimensionality reduction methods depending on the serving purpose, i.e. identification or verification. Uniform and non-uniform quantization methods are engaged according to Predefined, Greedy and Component Reliability bit
allocation techniques to produce the embeddable binary sequence. The novel Reversible Watermarking scheme proposed in chapter 3 is employed to embed the ECG marks into the cover image, which allows the ECG feature embedding framework to benefit from all the advantages and features the proposed reversible watermarking scheme offers. The extracted marks are used to identify or verify a subject’s identity through well defined algorithms using different measures depending on the dimensionality reduction technique applied. Finally performance of the proposed scheme is tested in different scenarios on The Biometric Security Laboratory (BioSec.Lab) database [73]. The obtained results show that the degradation in performance due to the presence of quantization and binarization noise is minor, therefore, the ECG watermarking scheme is highly reliable to be employed in sensitive applications where data hiding and privacy protection are needed.
Chapter 5

Conclusions

5.1 Research Summary

This thesis introduces a novel high capacity reversible multiple watermarking scheme applicable to all environments where security and privacy are the main concerns, such as medical record protection and archiving. The motivation of this work is to improve digital watermarking performance in different aspects to better serve the multimedia communication and its related concerns. The designed and developed scheme benefits from three important and crucial properties in the watermarking area: high capacity, reversibility, and multiple watermark embedding capability. It uses the integer wavelet transform to create the necessary embedding space in high frequency subbands of the host image. The proposed scheme is tested on natural and medical images in instances where single or multiple watermarks are embedded showing significant improvement in data hiding capacity and visual quality of the marked image. The proposed scheme is then compared to four seminal methods, each exploiting different techniques; the obtained results show significant enhancement in performance that satisfies the criteria of high fidelity and data payload for digital watermarks.

In order to expand and utilize the advantages of the proposed reversible watermarking
technique in applications where verification and identification are the needs, a framework is introduced to embed ECG biometric signal as the watermark. Having ECG as the mark embedded in an image allows the watermarking scheme to be engaged in several different applications in which the goal is to control and validate the identities while keeping them private. This framework employs the autocorrelation technique for the feature extraction step and different quantization and bit allocation techniques to fulfill the goal. The suggested algorithm is tested on BioSec.Lab database [73] and the obtained results show comparable performance of the system with the case that has no quantization or binarization, i.e. no ECG watermarking is used.

5.2 Future Research Directions

The research basis of this thesis can be extended into two major directions. First, each of the proposed methodologies can be improved and enhanced in order to better serve the applications and purposes they are developed for, also the work can be modified to address other applications where privacy and security properties are needed. Furthermore, the proposed frameworks can be developed to offer added properties and features so they can be engaged in applications for which the proposed schemes with current properties are not suitable for.

5.2.1 Future Directions of the Proposed Algorithms

Reversible Multiple Watermarking

The proposed watermarking schemes can be improved in different components. A more sophisticated coefficient map can prevent all underflow and overflow incidents. Although the coefficient map methods proposed in the proposed scheme satisfy the underflow and overflow criterion for most of the applications, it might be of interest to have an underflow and overflow free marked image in some specific applications which
the proposed scheme in chapter 3 targets. Examples of such applications are deep space photography, military, and satellite imaging where in an image there are a lot of small particles present and a minor pixel size change can yield wrong perceptions and have great impacts.

The Structural Similarity (SSIM) index can be employed as a replacement of PSNR as the ”measure” to evaluate the performance of the proposed reversible watermarking scheme in chapter 3. Furthermore, this measure can be engaged to derive the stopping parameters introduced in section 3.2.4.1.

The security layer (encryption) used in the scheme can be modified such that embedding conditions and parameters are merged with encryption algorithm.

Also the proposed method can be adjusted to suit other purposes in applications where fragile reversible data hiding is required. The modifications can even be taken to the point of applying the scheme to video content.

**Embedding ECG Biometric**

The proposed framework can be improved and optimized by introducing other statistical parameters in bit allocation phase. The Component Reliability method described can be modified to address PCA and LDA dimensionality reduction in a better way so the number of bits used in total can be lower. Although uniform and mu-law quantization techniques show satisfying performances, but other quantization procedures may even lead to better results.

Furthermore, the proposed framework can be modified to be used for embedding of other biometrics such as finger print, iris, electroencephalogram (EEG), and heart sound. In all these cases the proposed reversible watermarking technique may be the selected embedding system, but the biometric feature extraction, dimensionality reduction, bit allocation and quantization methods may need to be adjusted and adopted to result in an acceptable performance.
5.2.2 Robust High Capacity Reversible watermarking Scheme

In some applications it is desirable and necessary to employ a watermarking scheme which is robust against some signal processing operations. Applications dealing with open network copy protection, content authentication, and owner identification are examples where a robust watermark is a must.

The possibility of adding the robustness feature to the proposed scheme can be explored, by either engaging other properties of wavelet transform as in the method proposed by Wu [48] or the help of combination of different techniques and frameworks such as the work by Coatrieux et al. [51]. In the first approach, the mark which requires the robustness property is usually for the authentication purpose and can be formed from a few bits or strings. Therefore, this mark can be embedded into a deep low frequency subband such as $LL_4$. The watermark embedded with this method has been shown to be robust against lossy JPEG compression at low rates [48]; the rest of the marks can be inserted using the proposed reversible multiple watermarking technique into the high frequency subbands. The other approach suggests to employ a robust watermarking scheme on top of a fragile reversible watermarking technique in order to secure the robustness of specific marks in certain regions of the host image.
Appendix A

Reversible Watermarking

Background

Reversible watermarking has gained a lot of attentions since applications of watermarking and data hiding expand into areas where original data preservation is a must. Several methods and techniques have been introduced throughout the years which have their own advantages and disadvantages. As mentioned in Chapter 3, reversible data hiding schemes can be categorized into **fragile** and **semi-fragile** groups. This appendix presents and discusses, from each category, several different techniques developed to tackle the reversible watermarking issues.

Reversible watermarking was first introduced by Mintzer et al. [25] in 1997. They embedded a visible watermark which could have been removed from the original media. Other early methods were mainly based on 256 modulo addition which introduces "salt-and-pepper" effect in the cover images [26].

A.1 Fragile Watermarking Algorithms

The bulk of the literature and published works in the field of reversible watermarking is on **fragile** watermarking algorithms. Being a fragile watermark implies that the information
embedded in the original media is not recoverable or readable as soon as the watermarked
signal is modified or altered. Consequently, once the embedded watermark information is
lost, the original data is not recoverable. Furthermore, fragile watermarking techniques
can be divided into two subdivisions, *spatial domain* and *transformed domain* techniques.

### A.1.1 Fragile Algorithms Operating in Spatial Domain

This section presents the main and significant fragile reversible watermarking techniques
operating in the spatial domain.

One of the most significant works in this area was done and presented by Tian [27].
The technique can achieve high payload capacity, high visual quality, and reversible
data embedding properties in digital grayscale images. The methodology is based on
the calculation of differences of neighboring pixel values and upon selection of certain
differences, the difference expansion (DE) is executed. The payload $B$ to be embedded
in Tian’s method includes a compressed location map, the original least significant bit
(LSB) values, and the watermark payload which contains an image hash.

Tian’s method modifies the high frequency coefficients of a one level Haar transform.
If $x$ and $y$ are the intensity values of a pair of pixels in a grayscale image, the procedure
starts to define two amounts, the low-pass and high-pass Haar transform coefficients $l$
(the average) and $h$ (the difference)

$$ l = \left\lfloor \frac{x + y}{2} \right\rfloor, \quad h = x - y, \quad (A.1) $$

for $x, y \in \mathbb{Z}, \ 0 \leq x, y \leq 255$, and given $l$ and $h$, the inverse transform can be
respectively computed as

$$ x = l + \left\lfloor \frac{h + 1}{2} \right\rfloor, \quad y = l - \left\lfloor \frac{h}{2} \right\rfloor, \quad (A.2) $$

where $\lfloor x \rfloor$ is the greatest integer less than or equal to $x$.

The technique separates the pixel sets into different categories according to the char-
acteristics of the corresponding $h$ and behaves slightly different on each of these groups
during the embedding process.

**Changeable difference**: For a gray scale valued pair \((x, y)\) a difference number \(h\) is changeable if \(|2 \times \left\lfloor \frac{h}{2} \right\rfloor + b| \leq \min(2(255 - l), 2l + 1)\).

**Expandable difference**: For a gray scale valued pair \((x, y)\) a difference number \(h\) is expandable if \(|2 \times h + b| \leq \min(2(255 - l), 2l + 1)\).

This is applied to prevent underflow and overflow for the watermarked pixels \((x', y)\).

In the watermark embedding procedure to embed a message bit, \(b \in \{0, 1\}\) of the payload, the amount \(h\) should be modified to obtain \(h'\) which is called Difference Expansion (DE) as follow for the expandable differences

\[
h' = 2 \times h + b, \quad b = \text{LSB}(h'),
\]

(A.3)

and for the changeable ones as

\[
h' = 2 \times \left\lfloor \frac{h}{2} \right\rfloor + b, \quad b = \text{LSB}(h').
\]

(A.4)

By replacing \(h\) with \(h'\) in Equation (A.2), the watermarked pixel values \(x'\) and \(y'\) are obtained. The basic feature which distinguishes expandable differences from changeable ones is the possibility to embed watermark bits into pairs of the first set without the need of saving the original LSB values of them which reduced the total amount of extra information required on top of the watermark itself which is known as the header information.

A location map is considered to record the diverse categories of differences. In order to embed the location map of the expandable coefficients, the method proposes to substitute the least significant bit plane of the high-pass coefficients with the header information. To ensure reconstruction, the original LSBs are compressed and embedded together with the location map. Hence, the final total payload consists of the watermark message bits, \(B\), location map, \(L\) and compressed LSBs, \(C\).
To extract the embedded watermark and retrieve the original values, the decoder follows the same pattern adopted during embedding procedure and applies the same routines to each pair. It starts with applying Equation (A.1) to each set of pairs, creating the two sets of differences as $C$ for changeable $h$ and $NC$ for not changeable ones, and then taking all LSBs of differences belonging to the changeable set (set $C$) and forming the bit stream $B$. The first step is to recover the location map and use it along with $B$ to restore the original $h$ values, then by employing Equation (A.2) the original image is obtained, and finally the embedded payload, the remaining part of $B$, is extracted and used for authentication and validity check. There are a number of shortcomings and drawbacks for this method. In Tian’s algorithm it is impossible to evaluate whether embedding of a certain payload is feasible before the actual embedding process starts, which is due to the fact that the location map itself depends on the expansion coefficients used. Moreover, the lossless compression of the location map and LSB plane imposes a great cost, the size of the overhead, which leads to a significant larger payload than of what it should be, the watermark. As a result of such an increase, the watermarked image has the effect similar to mild sharpening in the mid tone regions even in case of relatively small watermark data.

Tian’s method is generalized and extended by Alattar [28]. In his scheme, instead of using the Haar transform difference expansion applied to pairs of pixels to embed the watermark bits, Alattar employed difference expansion method on spatial and cross-spectral triplets of pixels in order to increase the capacity used for embedding; the proposed algorithm embeds two bits in each triplet. The term triplet refers to a $1 \times 3$ vector containing the pixel values of a colored image. The triplets are divided in two categories:

**Spatial triplet:** Three pixel values of the image are chosen from the same color component within the image according to a predetermined order.

**Cross-spectral triplet:** Three pixel values of the image are chosen from different color components (RGB).
Appendix A. Reversible Watermarking Background

For the triplet vector \( t = (u_0, u_1, u_2) \), the forward difference expansion is defined as

\[
\begin{align*}
v_0 &= \left\lfloor \frac{u_0 + wu_1 + u_2}{N} \right\rfloor, \\
v_1 &= u_2 - u_1, \\
v_2 &= u_0 - u_1,
\end{align*}
\]

(A.5)

where \( N \) and \( w \) are constants with values of \( N = 3 \) and \( w = 1 \) for spatial triplets, and \( N = 4 \) and \( w = 2 \) for cross-spectral triplets. In this method, the inverse transform to be applied on the transformed triplets \( t' = (v_0, v_1, v_2) \) is defined as

\[
\begin{align*}
u_1 &= v_0 - \left\lfloor \frac{v_1 + v_2}{N} \right\rfloor, \\
u_0 &= v_2 + u_1, \\
u_2 &= v_1 + u_1.
\end{align*}
\]

(A.6)

To embed the watermark bits into the expandable triplets, which are those satisfying a condition to avoid overflow/underflow, the values of \( v_1 \) and \( v_2 \) are considered as below

\[
\begin{align*}
v'_1 &= 2 \times v_1 + b_1, \\
v'_2 &= 2 \times v_2 + b_2.
\end{align*}
\]

(A.7)

According to the definition stated, the triplets are classified in the following categories:

- \( S_1 \) contains all the expandable triplets in which \( v_1 \leq T_1 \) and \( v_2 \leq T_2 \), where \( T_1 \) and \( T_2 \) are predefined threshold.
- \( S_2 \) contains all the changeable triplets that are not in \( S_1 \).
- \( S_3 \) contains all the not changeable triplets.
- \( S_4 \) contains all the changeable triplets in \( S_1 \) and \( S_2 \).

The embedding process starts with transforming the triplets using Equation (A.6) and categorizing them into \( S_1, S_2, \) and \( S_3 \) divisions. The triplets in \( S_1 \) and \( S_2 \) groups are used
for the watermarking process and transformed into $S_1^w$ and $S_1^w$, then the pixel values of the
original image $I(i, j, k)$ are replaced with the corresponding watermarked triplets in $S_1^w$
and $S_1^w$ to create the watermarked image $I_w(i, j, k)$. The overall overhead of the method
includes the binary compressed location map which identifies the location of triplets in $S_1$, $S_2$, and $S_3$ together with the LSB of the changeable triplets. The watermark extraction
and original image retrieving procedure simply follows inverse steps of the watermark
embedding phase.

An advantage of Alattar’s method over Tian’s is the possibility to control the size
of the generated payload by adjusting the threshold values. Though Alattar’s algorithm
undergo less distortion given the same payload and outperforms Tian’s technique at lower
PSNRs, but at higher PSNRs Tian’s method outperforms the latest [27, 28].

Furthermore, Alattar proposed [29] an extension of the previously mentioned algo-
rithm [28] in order to embed triplets of bits in the difference expansion of quads of
adjacent pixels; where quads are defined as $1 \times 4$ vectors containing the pixel values from
different locations within the same color component of the host image as shown in Figure A.1 The difference expansion and its inverse transforms are the generalized forms
of the previous methods applied on quad vectors $q = (u_0, ..., u_3)$ and $q' = (v_0, ..., v_3)$,
respectively. The difference expansion transform is

\[
\begin{align*}
  v_0 &= \left\lfloor \frac{a_0 u_0 + a_1 u_1 + a_2 u_2 + a_3 u_3}{a_0 + a_1 + a_2 + a_3} \right\rfloor, \\
  v_1 &= u_1 - u_0, \\
  v_2 &= u_2 - u_1, \\
  v_3 &= u_3 - u_2,
\end{align*}
\]

and the inverse transform is

\[
\begin{align*}
  u_0 &= v_0 - \left\lfloor \frac{(a_1 + a_2 + a_3)v_1 + (a_2 + a_3)v_2 + a_3 v_3}{a_0 + a_1 + a_2 + a_3} \right\rfloor, \\
  u_1 &= v_1 + u_0, \\
  u_2 &= v_2 + u_1, \\
  u_3 &= v_3 + u_2.
\end{align*}
\]

As discussed in the previous algorithm, similarly quads are categorized in expandable or changeable sets and treated differently during watermark embedding process. Adjustable threshold values are considered to control the payload size.

In the embedding procedure the quads are transformed using Equation (A.9) and then divided into the categories. The expandable and changeable quads are modified to form the watermarked sets, later the pixel values of the original image \(I(i, j, k)\) are replaced with the corresponding watermarked quads in watermarked sets to form the watermarked image \(I_w(i, j, k)\). Watermark extraction and original image restoring process follows the inverse steps.

In the experiments done employing this method, images with a lot of low frequency contents and high correlation produce more expandable triplets with lower distortion than high frequency images. This method outperforms Tian’s algorithm, using gray scale images, at PSNR higher than 35 dB, it also outperforms the previous work of Alattar described in [28] and shows a huge increment in the payload size retaining the same PSNR. Even though it shows better results comparing to those of earlier works,
Appendix A. Reversible Watermarking Background

has a higher computational complexity comparing to those.

Finally, Alattar proposed [30] a further generalization of his algorithm in which the difference expansion of vectors composed by adjacent pixels are used. This method increases the overall embedding capacity and the computation efficiency. A vector is defined as $u = (u_0, ..., u_{N-1})$, where $N$ is the number of pixel values chosen from $N$ different locations within the same color component according to a secret key, from a pixel set of $a \times b$ size. The forward difference expansion transform for the vector $u$ is defined as

$$
v_0 = \begin{bmatrix}
\sum_{i=0}^{N-1} a_i u_i \\
\sum_{i=0}^{N-1} a_i
\end{bmatrix},
$$

$$
v_1 = u_1 - u_0,
$$

$$
\vdots
$$

$$
v_{N-1} = u_{N-1} - u_0,
$$

where $a_i$ is an integer and for the pixel set size the conditions are: $1 \leq a \leq h$ (image height), $1 \leq b \leq w$ (image width) and $a + b \neq 2$. The inverse transform for the transformed vector $v$ is then defined as

$$
u_0 = v_0 - \begin{bmatrix}
\sum_{i=0}^{N-1} a_i v_i \\
\sum_{i=0}^{N-1} a_i
\end{bmatrix},
$$

$$
u_1 = v_1 + u_0,
$$

$$
\vdots
$$

$$
u_{N-1} = v_{N-1} + u_0.
$$

Expandable vectors are those that can be modified through the embedding process $(b_1, b_1, ..., b_1) \in \{0, 1\}$ to create $ar{v} = (v_0, \bar{v}_1, ..., \bar{v}_{N-1})$ yet the reverse transform applied
on them do not cause the underflow/overflow problem in the retrieved image.

\[
v_0 = \begin{bmatrix} \sum_{i=0}^{N-1} a_i u_i \\ \sum_{i=0}^{N-1} a_i \end{bmatrix},
\]

\[
\tilde{v}_1 = 2 \times v_1 + b_1,
\]

\[
\vdots
\]

\[
\tilde{v}_{N-1} = 2 \times v_{N-1} + b_{N-1}.
\]

To prevent underflow/overflow the following conditions have to be satisfied

\[
0 \leq \tilde{u}_0 \leq 255,
\]

\[
0 \leq \tilde{v}_1 + \tilde{u}_0 \leq 255,
\]

\[
\vdots
\]

\[
0 \leq \tilde{v}_{N-1} + \tilde{u}_0 \leq 255.
\]

On the other hand, vector \( u \) is considered \textit{changeable} if Equation (A.13) holds when \( v_i \) is substituted by \( \lfloor v_i / 2 \rfloor \).

Similar to the previously discussed Alattar algorithms, the defined sets of vectors can be classified in the following groups, \( S_1 \), \( S_2 \), \( S_3 \), and \( S_4 \) containing all the expandable vectors satisfying the threshold conditions; all the changeable vectors that are not in \( S_1 \), all the vectors that are not changeable and the union of \( S_1 \) and \( S_2 \), respectively.

Embedding process starts by transforming the vectors and then dividing them into the aforementioned groups. Later \( S_1 \) and \( S_2 \) are modified according to watermark bits to form \( S_1^w \) and \( S_2^w \), following with replacing the pixel values of the original image with the corresponding watermarked versions kept in \( S_1^w \) and \( S_2^w \) to create the marked image. As in all the previous methods, a location map that locates \( S_1 \), \( S_2 \), and \( S_3 \) is created and embedded as the overhead to make the reverse process possible. Watermark extraction and original image restoring processes are simply the inverts of the above procedure.
This technique is capable of achieving the embedding rate of 1 bpp and can be employed recursively to increase the embedding capacity. Considering the results from this algorithm [30] and comparing them with Tian’s [27] and even other Alattar’s [28, 29] algorithms, spatial quad based technique provides high capacity and low distortion in most of the images and is one of the best algorithms available.

Ni et al. [31], presented a reversible data hiding algorithm which utilizes the zero or the minimum points in the histogram of an image, in the spatial domain, to embed data by slightly modifying pixel values. In this method, the algorithm first looks into the histogram of the host image to locate a zero point, which corresponds to the grayscale value which no pixel in the given image assumes, or in case that zero point does not exist, it looks for a minimum point. Afterwards the algorithm searches for a peak point, which corresponds to the grayscale value which the maximum number of pixels in the host image assumes. In Figure A.2, histogram of Lena’s image, \( h(236) \) is the zero point and \( h(155) \) represents the peak point. The objective of looking for a peak point is to increase the hiding capacity as in this method, the number of bits that can be embedded into a host image equals to the number of pixels related to the peak point. To explain the method,
Lena histogram can be taken as an example. The first step in the embedding process is to scan the whole image in a sequential order, row-by-row or column-by-column, and increase the value of pixels between 155 and 236 (including 155 and 236) by 1. After this step, the histogram is shifted to the right hand side from pixel value 155 by 1, which creates an empty gap at value 155 which can be used for watermark insertion. To embed the watermark bits, the image is scanned again, in the same sequential order as used for locating values between the zero (minimum)-peak points range, once a pixel with grayscale value of 154 is located, it is incremented by 1, if the to-be-embedded bit value is 1; otherwise, if the to-be-embedded value is 0, the pixel value remains untouched. As can been seen, the data embedding capacity in this method, in case only one pair of zero and peak points is used, equals to the frequency of the peak point.

To extract the watermark and recover the original host image, having a as the peak point and b as the zero (minimum) point, supposing $a < b$, the algorithm scans the whole watermarked image in the same sequential order used in the embedding process. Whenever a pixel with grayscale value of $a + 1$ is located, a bit '1' is extracted. If a pixel with value of a is encountered, a bit '0' is extracted. Following this pattern, by scanning the whole watermarked image the embedded watermark can be extracted, later all the pixels with values in between the zero-peak points in the histogram are decremented by 1 so the original histogram, hence host image, retained. The example described is the simple case of employing one pair of zero and peak points. An extension of the proposed algorithm exploits the usage of multiple pairs of maximum and minimum points. This extension can be treated as the multiple repetition of the technique for one pair case described earlier. The lower bound of the PSNR of a watermarked image created by this algorithm versus the original image is larger than 48 dB. As discussed, in the embedding process all the pixels with grayscale values between the minimum and the maximum points are either incremented or decremented by 1. Therefore, in the worst case, the grayscale values of all pixels are modified by 1, which implies that the resultant mean
square error (MSE) is at most equal to one, which yields to

$$\text{PSNR} = 10 \log_{10} \left( \frac{255^2}{\text{MSE}} \right) = 10 \log_{10} (255^2) = 48.13 \text{ dB}$$ (A.14)

Another advantage of this scheme is the low computational complexity. The results demonstrated in [31], show that the overall performance of the proposed technique is occasionally better than many other reversible watermarking algorithms.

Furthermore, Thodi and Rodriguez [32] proposed different methods based on the difference expansion technique [27]. These schemes use the histogram shifting methodology [31] to embed the marks. The proposed technique improves the distortion performance at low embedding capacities and mitigates the capacity control problem. This scheme benefits from a highly compressible overflow map.

Thodi and Rodriguez [32, 33] proposed a histogram shifting method in order to embed data in prediction errors. The location map used in this scheme covers all cells that cannot be decoded without a location map. The combination of prediction error expansion and histogram shifting technique is the novelty of this scheme. The maximum embedding capacity of this scheme in a single pass is 1 bpp.

Weng et al. [34] proposed a high capacity reversible watermarking scheme and tackled the preexisting problems of predefined thresholds on differences to allow expansions, and large location maps recording all the expanded positions. This would consume most of the available capacity especially when the threshold is small.

In Weng’s method, after all the pixel are ordered into an one dimensional list \{S₁, S₂, ..., Sₘ×ₙ\}, each pixel \(S_i\) is predicted by its adjacent pixel on the right side (\(s_i + 1\)) to form the predicted value, \(\bar{S}_i\), for which the prediction error is \(P_{e,i} = S_i - \bar{S}_i\).

After prediction, the companding technique is applied, which is consisted of a compression \(C\) function and an expansion function satisfying, for an arbitrary signal \(x\), the \(E(C(x)) = x\) relationship. \(C_Q\) and \(E_Q\) respectively represent the quantized version of \(C\) and \(E\).
Using the aforementioned defined concepts, $P_{e,i}$ is companded to $P_{Q,i}$ using the quantized compression function $C_Q$ and then expanded by applying the quantized expansion function according to the following expressions

\[
P_Q = C_Q(P_e) = \begin{cases} 
P_e & |P_e| < T_h \\
\text{sgn}(P_e) \times \left(\left\lfloor \frac{|P_e| - T_h}{2} \right\rfloor \right) & |P_e| \geq T_h
\end{cases}
\]

\[
E_Q(P_Q) = \begin{cases} 
P_Q & |P_Q| < T_h \\
\text{sgn}(P_Q) \times (2|P_Q| - T_h) & |P_Q| \geq T_h
\end{cases}
\]

where $T_h$ is a predefined threshold. The companding error, $r$, is given as $r = |P_e| - |E_Q(P_Q)|$ which is 0 if $|P_e| < T_h$.

Embedding procedure is performed according to Equation (A.16) in which $S_i^w$ is the watermarked pixel and $w$ is the watermark. The pixels are categorized into two different sets, $C_1$ if $S_i^w$ does not cause any underflow/overflow and $C_2$ otherwise.

\[
S_i^w = \tilde{S}_i + 2P_Q + w.
\]

$C_1$, the pixels considered for watermarking purpose, further are classified into two subsets $C_{<T_h}$ and $C_{\geq T_h}$ in dependence of $P_{e,i}$ value compared to $T_h$. In Weng’s method the to-be-embedded bits consists of a losslessly compressed location map created by assigning 1 to all the pixels in $C_1$ and 0 to all the pixels in $C_2$, the bitstream $R$ containing the companding error $r$ for each pixel in $C_{\geq T_h}$ and the watermark $w$. So the maximum available hiding capacity at a given threshold $T_h$ is given as

\[
D = \|C_1\| - L_S - \|R\| = \|C_{<T_h}\| - L_S.
\]

where the operator $\|\cdot\|$ represents the length of a sequence or the cardinality of a set.

The extraction process follows the same steps applied in the embedding process backwards. All LSBs are gathered; the location map is identified, recovered and decompressed. Later the classification is obtained, and then watermark extraction and host
image restoring is performed through prediction using the following equations

\[ P_{Q,i} = \left\lfloor \frac{S^w_i - \tilde{S}_i}{2} \right\rfloor, \quad (A.18) \]

\[ w = \text{Mod} \left( \left( S^w_i - \tilde{S}_i \right), 2 \right), \]

where the predicted value \( \tilde{S}_i \) is equal to \( S_{i+1} \). Based on the presented experimental results \cite{34}, Weng’s scheme outperforms Tian’s \cite{27} and Thodi’s \cite{33} methods globally.

Coltuc and Chassery \cite{35} proposed a high capacity low cost reversible watermarking algorithm. They suggested a generalized integer transform on pairs of pixels obeying some simple constraints to embed the watermark and the correction data which are required to recover the original image. The proposed scheme can yield capacities more than 1 bpp in a single pass.

To study the algorithm it is essential to know the basics of the integer transform used. Let \( x_i \) be the image pixels in an 8 bit gray level image with \( L = 255 \) and let \( n \geq 1 \) be a fixed integer. The forward transform which defines the relationship between a pair of the host image pixels \( x = (x_1, x_2) \) and the transformed pixels \( y = (y_1, y_2) \) is given as below

\[ y_1 = (n + 1) x_1 - nx_2, \quad (A.19) \]

\[ y_2 = -nx_1 + (n + 1) x_2. \]

In the above equation to avoid the problem of under/overflow for the transformed pixels, \( y_i \), a pixel \( x_i \) is transformed if and only if the result lies within the Cartesian product \([0, L] \times [0, L]\). The transform is invertible and is given as

\[ x_1 = \frac{(n + 1) y_1 + ny_2}{2n + 1}, \quad (A.20) \]

\[ x_2 = \frac{ny_1 + (n + 1) y_2}{2n + 1}. \]

Since \( x_i \)'s, \( y_i \)'s, and \( n \) are integers, the above equation surely holds the divisibility criterion, which is written as a congruence equation.

\[ (n + 1) y_1 + ny_2 \equiv 0 \text{mod} (2n + 1) \quad (A.21) \]

\[ ny_1 + (n + 1) y_2 \equiv 0 \text{mod} (2n + 1) \]
In other words, if a pixel $x_i$, is transformed by the forward transform the pixel $y_i$ obeys the congruence equation. As can be easily deducted, modifying the transformed pixels through an additive insertion of a value $w \in (0, 2n]$ (i.e. watermark) as in Equation (A.22) causes Equation (A.22) not to be satisfied by the modified pair of pixels.

\[
(y_1, y_2) \xrightarrow{\text{watermark insertion}} (y_1 + w, y_2)
\]  

Although a non-transformed pair does not necessarily satisfy Equation (A.22), but it can be shown that there always exists a $w \in (0, 2n]$ which can modify the pair so that the equation holds. On this basis, prior to watermark embedding stage phase, all the pairs of pixels are modified to satisfy Equation (A.22), later using Equation (A.22) the watermark is embedded into the transformed couples. We can assume the watermark data is confined to the range $[1, 2n]$ without loss of generality. This method guarantees that for the watermarked pairs the congruence Equation (A.22) no longer stands, which makes the detection of the watermarked couples possible. To prevent pixel overflow, a supplementary constraint must be imposed

\[
x_1 + 2n \leq L,
\]

\[
x_2 + 2n \leq L.
\]

Those pairs satisfying the above equations are transformed and used during the embedding phase. The pairs excluded, not transformed, are modified using Equation (A.22) to satisfy the congruence equation, and the corresponding correction information is appended to watermark payload as the overhead.

During the detection and extraction phase, the watermark is sequentially extracted and, simultaneously, the original image is recovered. In this procedure, the same pairs of pixels used in the embedding process are identified and using congruence equation, depending on the result whether it is 0 or 1, the pairs are categorized as not-transformed or transformed respectively; those transformed are the pairs carrying the total payload. The payload is recovered and then divided into the watermark and the correction data.
Using the extracted watermark and correction data all the changes on the transformed pairs are inverted to recover the original image.

Let \( p \) be the number of pixel pairs in the original image, and \( t \) be the number of transformed ones. As each transformed pixel allows the insertion of a codeword in the range \([1, 2n]\), and each one of the not-transformed pairs should be corrected by an integer ranging in \([0, 2n]\), the theoretical bit-rate (hiding capacity) of this scheme is equal to

\[
b(n) = \frac{t}{2p} \log_2 (2n) - \frac{p - t}{2p} \log_2 (2n + 1) \quad \text{bpp.}
\]  

(A.24)

If the number of the transformed pixel pairs is large enough, it can be shown that this method can provide capacities more than 1 bit per pixel as long as \( n \geq 2 \). Furthermore, Coltuc proposes an improvement of his previous scheme in [36]. In this method a revised transform is presented such that instead of inserting a single watermark codeword into a pair of transformed pixels, the algorithm embeds a codeword into a single transformed pixel. The direct transform on the sequence of pixels is defined as

\[
y_i = (n + 1) x_i - nx_{i+1}
\]

(A.25)

To avoid underflow/overflow the following constraint is imposed on pixels to be transformed

\[
0 \leq (n + 1) x_i - nx_{i+1} \leq L
\]

(A.26)

The inverse transform is defined as

\[
x_i = \frac{y_i + nx_{i+1}}{n + 1}
\]

(A.27)

Again the divisibility can be written in the form of a congruence relation

\[
y_i + nx_{i+1} \equiv 0 \text{ mod } (n + 1)
\]

(A.28)

Late on the watermark embedding, extraction and original image restoration procedures follow the same steps as the previous scheme. Let \( t \) be the number of transformed pixels
and $N$ the number of image pixels, the theoretical bit-rate of this scheme is shown to be

$$b(n) = \frac{t}{N} \log_2(n) - \frac{N - t}{N} \log_2(n + 1) \text{ bpp (A.29)}$$

It can be shown that if the number of the transformed pixel pairs is large enough, this method can provide capacities more than 1 bit per pixel as long as $n > 2$, and more than 2 bpp in case $n > 4$. Though according to the results [35, 36] the latter technique shows significant improvement in data hiding capacity, but it achieves low perceptual quality in terms of PSNR.

Chang et al. [37] introduced two spatial quad-based schemes based on Tian’s difference expansion method [27]. They exploit the fact that the differences between the adjacent pixel values in the local region of an image are small. The difference expansion technique is applied to the image in row-wise and column-wise simultaneously, promising good use of both row-wise and column-wise pixel pairs with small differences.

Let $(x_1, x_2)$ be a pixel pair, the integer Haar wavelet transform is given as

$$a = \left\lfloor \frac{x_1 + x_2}{2} \right\rfloor, \quad (A.30)$$
$$d = x_1 - x_2.$$

Let $m$ be 1-bit message, it can be inserted into the high frequency coefficient using $d' = 2 \times d + m$. The inverse transform is given as below

$$x_1 = a + \left\lfloor \frac{d + 1}{2} \right\rfloor, \quad (A.31)$$
$$x_2 = a - \left\lfloor \frac{d}{2} \right\rfloor.$$

Furthermore, $m$ and $d$ can be restored using the following relation

$$m = d' - 2 \times \left\lfloor \frac{d'}{2} \right\rfloor, \quad (A.32)$$
$$d = \left\lfloor \frac{d'}{2} \right\rfloor.$$

The embedding process starts by partitioning the host image of size $n \times n$ into $n^2/4 \times 2 \times 2$ blocks (spatial quad-base expansions) as shown in Figure A.3. To locate and choose
Figure A.3: A 2 × 2 block in the partitioned image.

the suitable blocks in the host image for watermarking, a simple measure function is considered as

\[
\rho(b, T) = (|a_{11} - a_{12}| \leq T) \wedge (|a_{21} - a_{22}| \leq T) \wedge (|a_{11} - a_{21}| \leq T) \wedge (|a_{12} - a_{22}| \leq T),
\]

where \( b \) is a 2 × 2 block in the host image, \( T \) is a predefined threshold, \( a_{11}, a_{12}, a_{21}, \) and \( a_{22} \) are pixel values in block \( b \). \( \wedge \) denotes the logical operator "AND". As \( \rho(b, T) = \) is a Boolean value, so if it is true, \( b \) is a candidate block for data embedding, and if not, \( b \) is discarded and not chosen for watermarking. In this work, two different spatial quad-based expansion approaches are proposed. In the first scheme, only blocks satisfying \( (a_{11} - a_{12}) \times (a_{21} - a_{22}) \geq 0 \) are taken for row-wise watermarking following with column-wise expansion. Other constraints are imposed to watermarked both for row-wise and column-wise watermarking to avoid under/overflow. In the second scheme initial relation is not a condition anymore. The requirements to avoid under/overflow are checked and later a 4-bit message is hidden in each block. In both approaches, a location map to identify and locate the expanded and watermarked blocks is employed, compressed and
embedded as a part of the payload. the algorithm is compared to Thodi’s, Alattar’s and Tian’s algorithms in [37]. It is claimed that the proposed schemes have higher embedding capacity than Tian’s [27] and Thodi’s [32, 33] methods, and quite competitive with Alattar’s [30] scheme.

In [38], Weng et al. proposed a reversible data hiding scheme based on an invertible integer transform which exploits the correlations among four pixels in a quad. In this scheme data embedding is accomplished by expanding the differences between a pixel and each of its tree neighboring pixels. As high embedding capacity cannot be achieved only by the means of difference expansion, the companding technique is employed into the process.

Given a grayscale image $I$, each $2 \times 2$ adjacent pixels are grouped into non-overlapping quad denoted by $q$,

$$\mathbf{q} = \begin{bmatrix} u_0 & u_1 \\ u_2 & u_3 \end{bmatrix}, \quad u_0, u_1, u_2, u_3 \in \mathbb{N}. \tag{A.35}$$

The forward integer transform is defined as

$$v_0 = \left\lfloor \frac{u_0 + u_1 + u_2 + u_3}{4} \right\rfloor,$$

$$v_1 = u_0 - u_1,$$

$$v_2 = u_0 - u_2,$$

$$v_3 = u_0 - u_3. \tag{A.36}$$

The inverse integer transform is given by:

$$u_0 = v_0 + \left\lceil \frac{v_1 + v_2 + v_3}{4} \right\rceil,$$

$$u_1 = u_0 - v_1,$$

$$u_2 = u_0 - v_2,$$

$$u_3 = u_0 - v_3. \tag{A.37}$$
The watermarking procedure starts by applying the forward integer transform on each quad and then proceeds with the application of the companding technique, which is covered previously in Weng’s other method [34] (please refer to [34, 38] for further details). The output from the companding is categorized in three different sets, \( C_1 \), \( C_2 \), and \( C_3 \) according to specific characteristics. All the quads in \( C_1 \) and \( C_2 \) are used in the watermarking process, and the rest left untouched. Finally the inverse transform is performed on all the quads to yield the watermarked image. The total payload to-be-embedded consists of the watermark, location map and the original LSBs modified during embedding process. The watermark extraction and image restoration procedure begins with forming the quads and applying the forward integer transform on them, following with classifying the quads using the restored location map. And finally the watermark is extracted and the host image is restored by employing the inverse transform. The proposed scheme is tested and compared with Tian’s [27] and Alattar’s [30] algorithms, from results [38] the suggested technique presents better PSNR with a payload of the same size and in general outperforms these methods at almost all PSNR values.

A.1.2 Fragile Algorithms Operating in Transformed Domain

In this section, the significant schemes in the area of fragile reversible watermarking, which operate in a transformed domain, are presented.

Chen and Kao [39], proposed a simple watermarking approach operating in Discrete Cosine Transform (DCT) domain that uses quantized DCT coefficients of the host image. Watermark embedding and extracting algorithms are based on three parameters adjustment rules: Zero-Replacement Embedding (ZRE), Zero-Replacement Extraction (ZRX), and Confusion Avoidance (CA). ZRE and ZRX are the rules used to embed and extract one bit, respectively, and CA is to prevent confusion during these processes. Below the aforementioned rules are presented.

**Zero-Replacement Embedding (ZRE):** Embeds one bit into three successive num-
bers of \((a, 0, 0)\) satisfying \(a \neq 0\) as follows

1. Change \((a, 0, 0)\) to \((a, 1, 0)\) if the embedding bit is 1.
2. Change \((a, 0, 0)\) to \((a, -1, 0)\) if the embedding bit is 0.

**Zero-Replacement Extracting (ZRX):** Extracts one bit from \((a, b, 0)\) when \(b = 1\) or \(-1\) as follows

1. Extract bit 1 from \((a, 1, 0)\) and modify them to \((a, 0, 0)\).
2. Extract bit 0 from \((a, -1, 0)\) and modify them to \((a, 0, 0)\).

**Confusion-Avoidance (CA):** When applying ZRE two possible patterns are generated according to each embedding bit. As these patterns are valid in other coefficients sets, so to avoid watermark embedding or extracting errors, CA is proposed.

1. In embedding, each \((a, k, 0)\) set is modified to form \((a, k + 1, 0)\) when \(a \neq 0, k > 0\) or changed to \((a, k - 1, 0)\) when \(a \neq 0, k < 0\).
2. In extracting, each \((a, k, 0)\) set is changed to \((a, k - 1, 0)\) when \(a \neq 0, k > 0\) or changed to \((a, k + 1, 0)\) when \(a \neq 0, k < 0\).

In the embedding process, the first step is to partition the host image into sets of \(8 \times 8\) blocks and apply DCT transform on each of these blocks, later the result is quantized according to a predetermined quantization table. Then, all patterns of three successive coefficients are chosen according to a pre-determined selection sequence and preprocessed by applying CA rule. Later, the watermark bits are embedded into valid corresponding patterns, where \(a \neq 0\) in \((a, 0, 0)\), using ZRE rule. Finally, the Inverse Discrete Cosine Transform (IDCT) is applied to the watermarked DCT coefficients and all blocks are combined to build up the final watermarked image. To recover the original image and extract the embedded watermark, all the initial steps in the embedding process are followed to construct the coefficient triplets. Later, ZRX rule is applied on all the valid patterns,
the watermark bits are extracted from each block and compared to the corresponding original binary watermark bits to determine whether the block has been attacked or not. By extracting the watermark bits the original coefficients used for watermarking are recovered, and applying the extracting step of CA rule ensures to convert back all the other coefficients to their original values. Finally, IDCT is applied on each block and using the blocks the original image is recovered.

Yang et al. [40], proposed another method based on integer DCT coefficients modification with peak amplitudes in each coefficient histogram. Lossless integer DCT transform, which guarantees reversibility, is applied on $8 \times 8$ blocks of host image, this method employs the histogram modification principal proposed by Ni et al. [31] to create the free space used for embedding the watermark. Similar to float-point DCT, an integer DCT transform has the energy concentration property, which can be used to enhance the capacity of histogram modification scheme, but unlike the float-point version, it is lossless hence suitable for reversible watermarking purposes. The first step in watermark embedding process is to generate the coefficient histograms in the integer DCT domain to do histogram modification. The original image is divided into $M$ image blocks with size $8 \times 8$, then integer DCT of each block is computed. Inside each transformed block, the coefficient in the position $(p, q)$ ($1 \leq p, q \leq 8$), where $p$ and $q$ are row and column indices respectively, is chosen. Therefore, for each coefficient position $(p, q)$, $M$ coefficients can be gathered from all the transformed blocks to form a coefficient group $G(p, q)$. Having blocks of $8 \times 8$, there are totally 64 coefficient groups, and as only one histogram is created for each group, 64 histograms are formed. To embed the watermark, histogram modification technique is applied only to the AC coefficient groups ($p + q > 2$), which each can be modeled with a General Gaussian distribution concentrated around the 0 scale value, a desirable property for histogram modification. If security is needed in some cases, a secret key $K_c$ can be used to select $N$ ($N \leq 63$) coefficient groups for watermarking. Watermark embedding process is exactly the same as histogram modification method used
in [31]. To ensure reversibility, the positions of the original peak point $P$ and zero point $Z$ in each histogram of all the $N$ coefficient groups used in the embedding process, should be kept as the overhead information, needed during watermark extraction. Watermark extraction and original image recovery process is simply the reversed steps taken in the embedding process.

Xuan et al. [41] presented a reversible data embedding method using integer wavelet transform and companding technique. The presented method exploits the Laplacian distribution of integer wavelet coefficients in high frequency subbands ($LH$, $HL$, and $HH$), which facilitates the selection of the compression and expansion functions and keeps the distortion low in the watermarked image. The embedding process starts with histogram modification, as a preprocessing step, which narrows the host image histogram from both sides to prevent over/underflow problem. After histogram adjustment, Integer Wavelet Transform (IWT) is performed on the image. Data embedding starts using companding technique, explained in the previous section, more details in [41], in the high frequency subbands. After embedding the total payload consisting of the watermark payload and some overhead data used in the extraction process, inverse IWT is performed to obtain the marked image. The extraction process is the reversed order of the embedding process. After the watermark is extracted, the original histogram is recovered using overhead data. The results [41] showed better visual quality in term of PSNR comparing to Tian’s difference expansion method [27].

Weng et al. [42], proposed a reversible watermarking scheme based on the companding technique and an improved difference expansion (DE) method. The watermark is embedded into high frequency subbands of the integer wavelet transform (IWT), using the companding technique. To avoid the overflow/underflow in the watermarked image, a method based on histogram modification is adopted. Though this method is based on the technique developed by Xuan [41], but by changing the order the method is applied, Weng could avoid the overflow/underflow issue Xuan’s method suffers from, as in the
latter method all pixel values of the predefined range are modified without considering whether they really suffer overflow/underflow or not in the embedding process. The proposed data embedding method offers some more advantages over Xuan’s technique in the forms of an increment in the hiding capacity with the PSNR value slightly increased, and the overall PSNR enhancement. The reason lies as the histogram modification is taken place after inverse IWT, hence all the high frequency coefficients, less than the predefined threshold, are entirely utilized for carrying the watermark. The watermark embedding process has two main steps, the first step consists of applying the IWT to the original image $I$ and embedding the to-be-embedded data bits, including the watermark bits and the companding errors, into the LSBs of one bit left shifted version of selected coefficients, later for the final part of this step, inverse IWT is performed and the image $I'$ is constructed. As some $I'$ pixel values would fall out of the grayscale level range (i.e. $[0, 255]$) and to cancel out this problem, in the second step of embedding process, a histogram modification method is used to make these values fall into the acceptable range. Later an improved DE technique is engaged to embed the information related to such modifications into the modified image $I'_H$, which results the final watermarked image. The improved DE method is based on classification of differences into three separate categories of, expandable, changeable and non-changeable. Data extraction process is also composed of two stages corresponding to the embedding process. In the first stage, difference classification is performed again and DE technique inverted to retrieve the histogram modification information along with $I'_H$. In the second stage by using the information retrieved in the previous stage, histogram modification is inversely applied, and the result is transformed by IWT. Embedded watermark is extracted from the high frequency subbands used during embedding process. As the final step, inverse IWT is performed to retrieve the original image. Experimental results stated in [42] show an embedding rate of 0.6 bpp with a correspondent PSNR of 40 dB for Lena image.

In [43], Lee et al. proposed a high capacity reversible image watermarking scheme
based on integer-to-integer wavelet transforms for both grayscale and color images. The proposed technique divides an input image into non-overlapping blocks and embeds a watermark into the high frequency wavelet coefficients of each block. To avoid any loss of information in the forward and inverse transforms, integer-to-integer wavelet is used, by applying the lazy wavelet and the lifting construction. As this method works on blocks of an image, the conditions to avoid under/overflow are derived for an arbitrary wavelet and block size. To increase the final PSNR of the watermarked image, the proposed method uses an adaptive technique to embed the watermark bits into the original image. In the embedding process, the watermark is embedded into the wavelet coefficients of the host image using either the LSB-substitution or the bit-shifting, specifically, $p$-bit-shifting technique. In the LSB-substitution method, the watermark is embedded by replacing the LSB of the selected wavelet coefficient with the to-be-embedded bit.

$$c^w = 2^0 \cdot \left\lfloor \frac{c}{2} \right\rfloor + w$$  \hspace{1cm} (A.38)

where $c$ is the original wavelet coefficient, $c^w$ is the watermarked coefficient and $w$ is the watermark bit. In the $p$-bit-shifting technique, the original wavelet coefficient $c$ is multiplied by $2^p$ where $p$ is a positive integer and a watermark $w$ is embedded into its $p$ LSBs as follows

$$c^w = 2^p \cdot c + w,$$ \hspace{1cm} (A.39)

$$(A.40)$$

where $w = 2^0 \cdot w_0 + 2^1 \cdot w_1 + \cdots + 2^{p-1} \cdot w_{p-1}$ and $\{w_0, w_0, \ldots, w_{p-1}\}$ are a set of $p$ watermark bits, the value of $p$ is adaptively determined for each block to minimize the perceptual distortion of the watermarked image. It’s during the embedding stage that a probable underflow/overflow can occur in the spatial domain. To guarantee reversibility, since it is lost if underflow/overflow occurs, such an issue must be predicted prior to watermark embedding by identifying and locating the LSB-changeable and bit-shiftable blocks of the host image. An image block is called LSB-changeable, if a watermark bitstream
can be embedded into the LSBs of its high frequency wavelet coefficients avoiding any underflow/overflow in the spatial domain. As another definition, a block is called to be bit-shiftable or, specifically, $p$-bit-shiftable, when a watermark bitstream can be embedded into its high frequency wavelet coefficients using bit-shifting by $p$ bits without any underflow/overflow in the spatial domain. The scheme of forward and inverse wavelet transform and watermark embedding is depicted in Figure A.4. Here the conditions to avoid underflow/overflow are discussed. As shown in Figure A.4 first, an $M \times N$ pixel block $S$ is transformed into a block of $M \times N$ wavelet coefficients $C$ using the $2-D$ non-expansive integer-to-integer transform $\text{IntDWT}_2(.)$. Next, a block $C_M$ is obtained either by setting the LSBs of the chosen coefficients to zero or by applying bit-shifting to the chosen coefficients in $C$. The modified pixel block $S_M$ is constructed by applying the $2-D$ inverse floating-point wavelet transform, $\text{fDWT}_2^{-1}(.)$, to $C_M$. By adding a watermark bit block $W$ to $C_M$, a block of watermarked wavelet coefficients $C_W$ is obtained. Then, $S_{WF}$ and $S_{WI}$ are obtained by applying $\text{fDWT}_2^{-1}(.)$ and $\text{IntDWT}_2^{-1}(.)$ to $C_W$, respectively. The embedding error $E_W$ is computed by applying $\text{fDWT}_2^{-1}(.)$ to $W$. Using a floating-point wavelet transform, overflow and underflow, caused by watermarking in the wavelet domain, can be predicted exploiting the linearity of the transform. The
watermarked block $S_{WF}$ is given by

$$
S_{WF} = fDWT^{-1}(C_W)
$$

$$
= fDWT^{-1}(C_M + W)
$$

$$
= fDWT^{-1}(C_M) + fDWT^{-1}(W)
$$

$$
= S_M + E_W
$$  \hspace{2cm} (A.41)

As $S_M$ can be determined easily by having an image block $S$, underflow and overflow is solely dependent on the error $E_W$ caused by the embedded watermark $W$. In this way, two matrices $E_{WP}$ and $E_{WN}$, which elements represent limits of max positive and negative errors caused by the embedded watermark can be obtained as follow

$$
E_{WP} = \sum_{i,j \in (HL_1 \cup LH_1 \cup HH_1)} 1/2 \{Q_{ij} + |Q_{ij}|\}
$$  \hspace{2cm} (A.42)

$$
E_{WN} = \sum_{i,j \in (HL_1 \cup LH_1 \cup HH_1)} 1/2 \{Q_{ij} - |Q_{ij}|\}
$$

where $Q_{ij} = fDWT^{-1}(O_{ij})$, and $O_{ij}$ is the matrix with only one nonzero element of value 1 in the $i^{th}$ row and $j^{th}$ column. Since the elements of $E_W$ satisfy the inequality $E_{WN} (m,n) \leq E_W (m,n) \leq E_{WP} (m,n)$, neither underflow nor overflow occur in $S$ for any watermark $W$ satisfying the following inequality

$$
s_{\text{min}} - E_{WN} (m,n) \leq S_W (m,n) \leq s_{\text{max}} - E_{WP} (m,n),
$$  \hspace{2cm} (A.43)

for $0 \leq m < M, 0 \leq n \leq N$. Since in the proposed method, integer-to-integer wavelet transforms are used, the watermarked image block obtained is not $S_{WF}$ but $S_{WI} = \text{IntDWT}^{-1}(C_W)$. The roundoff error matrix $E_R$, the matrix formed of errors introduced by integer-to-integer wavelet transforms due to the truncations of coefficients during the lifting steps, can be defined by two matrices, $E_{RP}$ and $E_{RN}$. The elements of these matrices represent limits of maximum positive and negative roundoff errors. Introducing
such error, the watermarked image block $S_{WI}$ is given by

$$S_{WI} = \text{IntDWT}_2^{-1}(C_W)$$

$$= \text{IntDWT}_2^{-1}(C_M + W) = fDWT_2^{-1}(C_M + W) + E_R$$

$$= S_M + E_W + E_R$$

An image block $S$ can be said to be LSB-changeable or bit-shiftable for any watermark block $W$ if the following inequality is satisfied,

$$s_{\text{min}} - E_{WN}(m,n) - E_{RN}(m,n) \leq S_W(m,n) \leq s_{\text{max}}$$

$$- E_{WP}(m,n) - E_{RP}(m,n),$$

for $0 \leq m < M, 0 \leq n \leq N$. A binary location map matrix, $L$, that indicates which blocks are watermarked, is embedded as the overhead information, used by decoder to retrieve the message bits and to reconstruct the original image, along with the watermark in the embedding process. Since a bit-shiftable block can change into a non-bit-shiftable block after embedding, location map is stored and embedded in a number of LSB-changeable blocks. The watermark extraction and original image recovery process starts with dividing the marked image into non-overlapping blocks of $M \times N$. Later, the same wavelet transformation engaged in the embedding process is applied to each block. While the LSB-changeable blocks are searched and located, the location map matrix is recovered from the LSBs of the high frequency wavelet coefficients, and by searching through the watermarked blocks; the original LSBs and the watermark bits are extracted. Using the retrieved location map and extracted original LSBs, the original host image blocks can be reconstructed. Applying the inverse integer-to-integer wavelet on each retrieved block finally ends the process. Comparing the experimental results [43] with other existing reversible watermarking techniques reveals that the proposed scheme has higher embedding capacity with better visual quality. The size of the non-overlapping blocks is an
important factor which determines the performance and efficiency of this method; too small (i.e. $4 \times 4$) or too large (i.e. $32 \times 32$) sizes degrade the algorithm performance.

A.2 Semi-Fragile and Robust Algorithms

This section introduces the significant semi-fragile and robust reversible watermarking schemes in the literature. These schemes show a certain degree of robustness when the watermarked image undergoes specific alterations or processes. In the case of semi-fragile methods, the tolerable process is usually confined to a slight compression process or other mild intentional or unintentional changes. On the other hand, robust schemes often present good tolerance against specific intentional attacks or unintentional severe modifications, depending on the purpose they are designed for. Hence, a watermarking scheme is called semi-fragile or robust if the extracted watermark from the modified/processed marked image stays ascertainable and valid.

A.2.1 Semi-Fragile Algorithms Operating in Spatial Domain

De Vleeschouwer et al. [44] proposed two semi-fragile reversible data hiding algorithms based on patchwork theory [45], which show certain robustness against JPEG lossy compression. These methods operates on image tiles by identifying a robust feature in the luminance histogram of each one of such tiles. Each bit of the watermark message is associated with a group of pixels in this method. To start with the embedding process, the host media is tiled in non-overlapping pixel blocks, each tile is reserved for a bit of the to-be-embedded message. Each tile is equally divided into two pseudo-random sets of pixels, i.e. zones A and B. The luminance histogram of each zone is computed and mapped around a circular support (positions on the circle are indexed by the corresponding luminance). A weight, proportional to the occurrence frequency of each luminance value within the group, is assigned to the corresponding position of that luminance on the
circle. Then the center of the total mass is calculated and localized respect to the center of the circle. Since zones A and B are pseudo-randomly determined, it is highly probable that the localized mass centers are very close to each other before any alteration due to watermark embedding, as they are representing average values in the patchwork algorithm. Hence slight rotations of these centers of mass, relating to zones A and B, in two opposite directions allow for embedding a bit of the message. A clockwise rotation of the zone A center of mass can be related to the embedding of a bit 1, and a counterclockwise rotation can be linked to a bit 0. As mentioned earlier, the center of mass in the other zone is rotated in the opposite direction. Although there are some cases in which the centers of mass in two generated zones are not close enough or properly positioned and cannot be used for data embedding, but in general these cases are negligible comparing to the those of interest and do not affect the available hiding capacity significantly.

Watermark detection and extraction process is done easily following the reverse pattern. It is straightforward to detect if a bit 1 or 0 is embedded in a certain tile of the marked image, and to recover the original image, the mark is removed from each block by counter rotation of each luminance histogram along the circular support.

If the luminance histogram is mapped linearly into the circular support, because of the abrupt transition on the occurrences of the higher and lower bounds values of the grayscale acceptable range, salt and pepper noise can appear even for a small support rotation. To prevent such an issue, the luminance histogram can be mapped to the circular support in an alternative fashion. This rearrangement of the histogram causes zones A and B centers of mass appear very close to the center of the circle making the watermark detection less reliable. In this case, the center of mass computation is substituted with the minimal inertia axis computation that can be detected more easily. This alternative technique makes the salt and pepper noise disappear. Both these approaches can deal with slight lossy attacks such as cropping and JPEG compression. The proposed methods show good robustness; though the latter, while more favorable
from a perceptual point of view, is more fragile to JPEG compression.

Ni et al. [46], presented a lossless watermarking scheme based on De Vleeschouwer work [44]. By then, the only existing semi-fragile scheme which could tolerate JPEG compression process was based on 256 modulo addition to achieve losslessness and robustness, but this technique suffered from the annoying salt-and-pepper noise caused by using 256 modulo addition to prevent overflow and underflow. Ni et al.’s proposed scheme does not generate salt-and-pepper noise in the marked image. The scheme operates based on the patchwork theory by identifying a robust statistical quantity. The differences between couples of pixels in an image tile are analyzed employing Error Correction Codes (ECC) and permutation techniques.

The method starts by dividing the host image into non-overlapping pixel blocks, then each block is divided into pixel couples. The arithmetic average, $\alpha$, of differences of grayscale values of each of these couples is computed, the average is expected to be close to zero for most of the blocks due to the fact that these values are highly correlated and have spatial redundancy. The main idea for data embedding is that the difference value $\alpha$ is kept within specified limit $K$ and $-K$, which is found through numerous experiments to be less than 5, to embed bit 0 and move outside this range in case of a bit 1. Using this pattern, a categorization of $\alpha$ with respect to $K$ is shaped. Later, the shift quantity, which is also called the embedding level, adds another parameter, $\beta$ usually twice of the threshold $K$, to define the categories. This technique is engaged to avoid the underflow/overflow errors causing significant visual degradation in the marked image. Following the patterns mentioned earlier, four categories are formed:

**Category 1:** The pixel grayscale values of a block under consideration are far enough away from the two bounds of the histogram, let’s say 0 and 255 in case of an 8-bit gray scale image. In this category, according to the value of $\alpha$, the following two cases are considered:

- **Case 1:** $\alpha$ value is within the threshold range, between $K$ and $-K$. 
• Case 2: The absolute value of $\alpha$ exceeds the threshold $K$.

Category 2: Some pixel grayscale values of the block under consideration are very close to the lower bound of the histogram, 0 in case of an 8-bit grayscale image, and at the same time there is no pixel value close to the upper bound of the histogram. According to the value of $\alpha$, three different cases are considered here:

• Case 1: The value of $\alpha$ is between $K$ and $-K$, the threshold range.

• Case 2: The value of $\alpha$ is beyond the threshold $K$ located on the right hand side.

• Case 3: The value of $\alpha$ is beyond the threshold $-K$ located on the left hand side.

Category 3: Some pixel grayscale values in the block under consideration are very close to the upper bound of the histogram, 255 in an 8-bit grayscale image, and no pixel values are close to the lower bound. This category is similar to Category 2 except that the distribution of grayscale values of the block is close to the upper bound instead of the lower one in the histogram.

Category 4: Some pixel grayscale values in the block under consideration are close to the upper bound, while some other pixel grayscale values are close to the lower bound of the histogram. In this category two cases are further considered according to the value of $\alpha$:

• Case 1: The value of $\alpha$ is between $K$ and $-K$, the threshold range.

• Case 2: The absolute value of $\alpha$ is beyond the threshold $K$.

In the embedding process, the difference value $\alpha$ is increased or decreased by $\beta$, depending on the different categories and the cases defined above. In the cases that modifying $\alpha$ causes over/underflow and consequently salt-and-pepper noise, the value of $\alpha$ left intact
Appendix A. Reversible Watermarking Background

and regardless of the value of the to-be-embedded bit an error bit is inserted. These possible error bits introduced during embedding process, are corrected by employing Error Correction Code (ECC), which introduces sufficient data redundancy as a part the payload. The proposed ECC, claimed to correct most of the random errors generated during embedding process, is BCH (63, 7, 15). For some images, error bits may be concentrated in particular areas of the image and considered as bursts of errors, which leads to too many error bits in one codeword; in such cases, ECC is of no use to recover the data. To efficiently cope with this issue, the watermarking algorithm employs a message bits permutation scheme together with ECC to redistribute the errors along the entire image. Data extraction is a much simpler task and actually is the reverse process of data embedding.

Experimental results show that a significant improvement in both data hiding capacity and perceptual quality of marked image is achieved respect to the scheme proposed by De Vleeschouwer in [44]. In particular, robustness is enhanced in the case of a lossy process such as JPEG compression with higher compression rates.

A.2.2 Semi-Fragile Algorithms Operating in Transformed Domain

Zou et al. [47] proposed a semi-fragile lossless digital watermarking scheme based on integer wavelet transform. The wavelet family adopted is the LeGalle 5/3 filter bank which is the default transformation technique in JPEG2000 for lossless compression. This characteristic makes it possible for this scheme to be integrated into the JPEG2000 standard. Special measures are considered to prevent over/underflow issue and the resultant salt-and-pepper noise.

The algorithm embeds bits into the integer wavelet transform coefficients of a selected high frequency subband, $HL$, $LH$, and $HH$. The coefficients of the High frequency subband in a wavelet transform follow a zero mean Laplacian shape distribution. As a
deduced fact, if the considered high frequency subband is divided into non-overlapping blocks of size $n \times n$, the resulting mean values of these blocks have zero mean Laplacian like distribution too. The scheme starts by scanning all the blocks looking for the maximum absolute value of coefficients, $m_{\text{max}}$. A threshold $T$ is set to be the smallest integer number which is greater than $m_{\text{max}}$. During the embedding process, the mean value of the blocks is modified in order to embed bits. If a bit 1 is to be embedded, the mean value of the block in use is shifted away from zero by a quantity larger than $T$ denoted by $S$. If the to-be-embedded bit is 0, the block is left unchanged. In the data extraction process, anytime a mean value with absolute value larger than $T$ is located, a bit 1 is extracted. In blocks with mean values smaller than $T$, the recovered bit are 0’s. Since $S$ is a fixed value used for modifying all the blocks, the original coefficients can be recovered to reconstruct the original host image. The original mean values are obtained by deducting the value of $S$ from wavelet coefficients in the blocks from which bit 1’s are recovered. As stated, the embedding process is reversible, and original media is fully recoverable. On the other hand, since embedding the watermark is controlled by mean value of the IWT coefficients in one block, slight modifications on the image caused by unintentional processes such as JPEG/JPEG2000 compression do not impose a huge change on these values, hence the correct detection of the hidden data is expected even after the marked image has been undergone minor modifications. To prevent overflow and underflow, caused by a process the marked image may experience (i.e. format conversion from JPEG2000 to TIFF), the algorithms suggests a block classification method to locate those blocks which can be modified during embedding phase. This classification divides the blocks into four categories, and each category is represented by a spatial domain histogram corresponding pixel values of the examined block. Assuming the maximum absolute pixel grayscale value change is $S_{\text{max}}$, the underflow condition occurs when there are pixels with grayscale values less than $S_{\text{max}}$ and the grayscale values need to be decreased. Overflow may take place if there are pixels with their grayscale values greater than $(255 - S_{\text{max}})$ and the grayscale
values need to be increased in the embedding process. To avoid overflow and underflow, the above two scenarios should not occur. To better clarify the approach, the $[0, S_{\text{max}}]$ and $[255 - S_{\text{max}}, 255]$ are called 0-zone and 255-zone respectively. Depending on the presence of any of these two regions the blocks are classified. The worst case is where pixels in 0-zone and pixels in 255-zone are both presented in a block, such a block is called not-embeddable block and it is not used for data embedding. If the embedded bit in a block is 1 then its detection causes no problems, but in case a bit 0 is recovered during detection process, the fact that the examined block is used to embed a bit 0 or it is a not-embeddable one, cannot be judged. To overcome this issue and correct the probable errors during data extraction process, an ECC technique is employed. Experimental results in [47] show that the salt-and-pepper noise is not present in the marked images and their visual quality is much higher compared with De Vleeshouwer’s work [44].

Wu [48] presented a semi-fragile reversible watermarking scheme for image authentication. In this method the watermark is embedded into $LL_4$ subband of the integer wavelet domain. In addition to the reversibility attribute, this scheme has the property of tamper localization.

To embed data, the method employs histogram shifting of integer wavelet coefficients which results in higher visual quality of the marked image compared with other schemes presented. The method can also tolerate JPEG compression with a low compression rate. To reconstruct the original image, the scheme performs a four-level CDF 9/7 integer wavelet transform based on the lifting scheme. As can be concluded, the original image can be obtained, only if the marked image has not been altered or modified. As for most of the images, the integer wavelet coefficients histogram of the high frequency subbands are concentrated near zero and follow a Laplacian shape distribution, this property can be used to implement reversible data hiding. Prior to data embedding, the host image is pre-processed using histogram modification to avoid underflow or overflow during payload bits insertion phase. later, four level IWT is performed on the pre-processed
image, and the watermark is embedded in \( LL_4 \) subband by inserting specific five-bit codes identifying 0 and 1 into the 5 LSBs of selected wavelet coefficients. Information required to reconstruct the original image are later embedded reversibly as the overhead employing histogram shifting in high frequency subbands of the IWT domain. After the embedding process is done, the inverse IWT is performed to form the marked image. To extract the watermark and recover the original image, four level IWT is performed on the watermarked image, and embedded data is extracted from \( LL_4 \) subband. To authenticate and detect the tampering area, the difference between the extracted watermark and original one is computed. If a watermarked image suffers from unintentional incidental attacks most of the watermark error pixels are isolated points on the difference image, but in case of malicious attacks the watermark error pixels are grouped together with a high probability. If the watermark is authentic, as the next step the original image is recovered from the marked one. The experimental results provided in [48] show that the embedding distortion is small and the watermarked image has good visual quality.

### A.2.3 Robust Algorithms Operating in Spatial Domain

The algorithm presented by Chrysochos et al. in [49] is a reversible watermarking scheme resistant to geometrical attacks which is based on histogram modification. Embedding process is done by inserting the watermark bits 0 or 1 in the selected couples of histogram bins, \( hist(a) \) and \( hist(b) \), following the below relations

\[
\begin{align*}
  m = 0 & \rightarrow \text{hist}(a) < \text{hist}(b), \\
  m = 1 & \rightarrow \text{hist}(a) > \text{hist}(b).
\end{align*}
\]

(A.46)

If the relation does not hold between the selected histogram bins, pixels belonging to each bin are swapped with the other one. In case an equality happens, the selected bins are skipped. The most important parameter in this scheme is the public key, a real number which its integer part (start) indicates the point the embedding procedure starts choosing.
histogram bin couples and the decimal part (step) defines the minimum distance, two histogram bins of a couple may have. In order to embed all the watermark bits, couples are chosen sequentially over the histogram. In addition to the watermark payload, side information consisting of records of swapped bins or the ones not used, watermark length and the public key is embedded as overhead. The hiding capacity of this method is the downside of it, the maximum capacity is 128 bits for a 8-bit grayscale level image. On the other hand, perceptual quality of the marked image is high, and this algorithm presents a good robustness in case of geometrical modifications, such as rotation, upsizing, cropping up to 80% and so on. But on the contrary, JPEG compression, low pass filtering and noise addition are not tolerated by this method.

In [50] Coltuc and Chassery proposed a technique based on Reversible Contrast Mapping (RCM) which is a simple integer transform applied to pair of pixels. RCM can be invertible even in case the LSBs of the transformed pixels are lost. For an 8-bit grayscale image, the forward RCM transform for the pair \((x,y)\) is

\[
x' = 2x - y, \quad y' = 2y - x,
\]

in which \(x'\) and \(y'\) are limited to grayscale bounds (0 and 255 in this case) to avoid overflow and underflow. The inverse RCM transform is defined as

\[
x = \left\lceil \frac{2}{3} x' + \frac{1}{3} y' \right\rceil, \\
y = \left\lceil \frac{1}{3} x' + \frac{2}{3} y' \right\rceil.
\]

The proof that \(x\) and \(y\) can be retrieved even if the LSBs of the transformed pixels are lost is given in [50]. Due to this property, LSBs are used in the embedding process to be replaced by watermark bits. It can be said that ceiling operation is robust to the loss caused by watermarking only if both \(x'\) and \(y'\) are not odd numbers, which means both \(x\) and \(y\) should not be odd numbers. Although even such pairs can be used in data embedding by employing a technique which modifies the LSB of the first pixel, but in
order to avoid decoding ambiguities, the odd pixels located on the borders of grayscale level should be eliminated from the embedding procedure, for which the domain can be denoted by \( D_C \). After the image is partitioned into pairs, embedding goes on as follows:

1. If \((x, y) \in D_C\) and it is not composed of odd pixel values, the forward RCM transform is applied and the LSB of \(x'\) is set to 1, to indicate a transformed pair, and the LSB of \(y'\) is used for watermark bit insertion.

2. If \((x, y) \in D_C\) and it is composed of odd pixel values, LSB of \(x'\) is set to 0, to indicate an odd pair, and consider the LSB of \(y'\) available for data embedding.

3. If \((x, y) \notin D_C\) the LSB of \(x\) is set to 0 and the true value is saved as the payload.

The payload to-be-embedded is composed of the watermark message and the bits saved in the step 3. The watermark detection process starts by partitioning the marked image into pairs \((x', y')\) and then proceeds as follows:

1. If the LSB of \(x'\) is 1 then the LSB of \(y'\) is a watermark bit; after setting the LSBs of \(x'\) and \(y'\) to 0 the original pair \((x, y)\) is recovered by applying the inverse RCM transform.

2. If the LSB of \(x'\) is 0 and the pair \((x', y')\), after setting the LSBs to 1, belongs to \(D_C\), then the LSB of \(y'\) is a watermark bit; after setting the LSBs of \(x'\) and \(y'\) to 1 the original pair \((x, y)\) is simply recovered.

3. If the LSB of \(x'\) is 0 and the pair \((x', y')\) with the LSBs set to 1 does not belong to \(D_C\), there is no watermark bit embedded in this pair; after replacing the LSB of \(x'\) with the true LSB taken from the overhead of the payload sequence, the original pair \((x, y)\) is reconstructed.

It is important to note that the true LSB of a non-transformed pair is embedded in a spatially close couple to make the scheme more robust in case of cropping, though experimental results of the robustness against cropping are not documented in the published
work. Taking \( p \) as the global number of couples and \( T \) the number of pairs used for data embedding, then \( (P - T) \) is the additional overhead on top of the watermark message; the bit-rate provided by algorithm will be

\[
B = \frac{T - (P - T)}{2P} = \frac{2T - P}{2P} \text{ bit/pixel} \quad (A.49)
\]

In order to increase the overall capacity, further iterations can be applied which surely lead to higher perceptual distortion. The bit-rate achieved in the experiments is very close to the theoretical upper bound of 0.5 bpp. The proposed technique outperforms other compression based methods but it is performing worse than Tian’s difference expansion method [27] though it shows less complexity.

Coatrieux et al. [51], proposed a image reliability protection technique in which the robustness is achieved by mixing two different approaches, one based on a reversible technique and the other based on a robust watermarking scheme. The embedding process is summarized in Figure A.5. This scheme is devoted to deal with Magnetic Resonance (MR) images, in which separation of the Region of Interest (ROI), such as any anatomical part of the body, from the Region of Non Interest (RONI), the black background behind the desirable object, is quite easy. It is of great importance to note that the capacity to make such a distinction between ROI and RONI is essential for the system to operate, and also the fact that the watermarking process should not affect the image segmentation at a later time. As shown in Figure A.5, the embedding consists of two separate watermarking procedures. In the first step, after segmentation, the RONI is watermarked employing a lossy robust watermarking method to make the image robust against modifications such as JPEG compression. The watermark inserted in RONI in this step consists of authenticity and integrity factors derived from the ROI to establish a secure link between these two regions. In the second step, a reversible watermarking technique embeds the factors generated from the entire image into the ROI. Because of the second protection level, the reversible watermarking, the global robustness of the scheme is limited, though it is asserted that a JPEG compression not lower than a quality factor of 70 does not
Figure A.5: Watermark embedding method employed in Coatrieux’s robust reversible watermarking technique.

generate any bit error.

A.2.4 Robust Algorithms Operating in Transformed Domain

Saberian et al. [52], presented a reversible watermarking algorithm based on quantization approach, named Weighted Quantization Method (WQM). Let $S = (s_1, ..., s_n)$ be the input signal and $Q = (q_1, ..., q_m)$ the corresponding quantization levels, embedding process employs a couple of functions $(f, L)$. Function $L$ operates depending on the value of the to-be-embedded bit, $m$.

1. $L_0(s) = $ the least quantization level greater than $s$,

2. $L_0(s) = $ the biggest quantization level smaller than $s$. 

Function $f$ operates as

$$f_m(s_i, L_m(s_i)) = \frac{s_i + dL_m(s_i)}{d + 1},$$

(A.50)

where $m = 0, 1$ and $d \geq 1$ is the designing parameter to make sure the embedding process is done correctly. But as $d$ grows larger the marked image distortion is getting worse, usually $d$ is set to 1. According to the definition of functions $f$ and $L$, it can be shown that $L_m(s') = L_m(s_i)$ where $s'$ is the watermarked signal for which the closest quantization level is chosen for during the extracting process. Later by employing $L_m(s_i)$ watermarking process can be inverted and the original value of $s_i$ can be recovered. Although this approach can be applied both in spatial and transformed domain, but a Point to Point Graph (PGP) transformation is engaged in the paper and experimental results are achieved on such a basis. Unfortunately robustness of such an algorithm is very limited and only results for image suffered from Additive White Gaussian Noise (AWGN) are presented.

In [53], Gao and Gu proposed a reversible watermarking scheme employing wavelet lifting algorithm based on Alattar’s difference expansion [30]. At first the host image is divided into non overlapping blocks of $8 \times 8$, then 1 level integer wavelet transform is applied on each of these blocks and the $LL_1$ subband is used for data embedding. Image blocks are shuffled according to a secret key to achieve the required security level and robustness against some malicious attacks. In particular, the four coefficients on the diagonal of this subband are grouped into two couples and used for watermarking considering the expandability. Expandability is checked to prevent the algorithm from the possible underflow and overflow, and the necessary data is saved as the side information to be used in the extraction process. In the experimental results provided, image reversibility is granted when there is no attack performed on the marked image and watermark robustness is partially provided against cropping, salt and pepper noise, and other image modifications localized in restricted zones.
Appendix B

Integer to Integer Wavelet Transform

The Integer to Integer Wavelet Transform (IWT) is the basis of the reversible watermarking scheme developed in Chapter 3. Integer wavelet transform is useful in applications where losses are not desirable. The main advantage of the wavelet transform comparing to the Fourier or the discrete cosine transforms is the property of time-frequency localization. The wavelet transform is employed in various fields such as image and video compression and signal analysis. This appendix briefly discusses the properties and implementations of the IWT used in this research work. For a more comprehensive description of the IWT, please refer to [59]. To better understand the IWT the basics of Discrete Wavelet Transform (DWT) is mentioned.

B.1 The Basics

Similar to the Fourier transform, the wavelet transform represents a signal as the projections to a series of basis functions. All basis functions $\psi_{s,u}(t)$ are generated by scaling and translating a mother wavelet $\psi(t)$, defined by the low and high frequency coefficients
$s_{1,n}$ and $d_{1,n}$ as

$$
\psi_{s,u}(t) = \frac{1}{\sqrt{s}} \psi \left( \frac{t-u}{s} \right).
$$

(B.1)

The mother wavelet has zero mean and finite support in time. Furthermore, $\psi(t)$ has finite energy,

$$
\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty,
$$

(B.2)

providing good resolution in both time and frequency [75]. In the DWT, the scaling and shifting are performed in discrete steps,

$$
\psi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi \left( \frac{t-2^j n}{2^j} \right),
$$

(B.3)

where $j$ and $n$ are integers. In the DWT, we usually first decompose a signal into approximate and detail components. The approximate component is then decomposed to approximate and detail components at other resolutions recursively. The wavelets basis provide the detail information; in order to ensure that the approximate coefficients are bounded, a scaling function $\phi(t)$ is introduced. The scaling function must satisfy the dilation equation,

$$
\phi(t) = \sum_{n} \sqrt{2} h[n] \phi(2t - n),
$$

(B.4)

as well as the relationship between $\phi(t)$ and $\psi(t)$

$$
\psi(t) = \sum_{n} \sqrt{2} g[n] \phi(2t - n).
$$

(B.5)

$h[n]$ and $g[n]$ are called the low-pass and high-pass analysis filters. It can be shown that $h[n]$ and $g[n]$ are related via the following equation,

$$
g[n] = (-1)^n h[1 - n].
$$

(B.6)

Therefore, it is sufficient to only specify the low-pass filter coefficients. For a 1D signal $x[n]$, the DWT $y[n]$ is of the same length with the first half of its elements representing
approximate information and the second half representing detail information,

\[ x[n] \rightarrow y[n] : \left\{ \sum_j x[j]h[n-j] : \sum_j x[j]g[n-j] \right\} . \]  

(B.7)

For a 2D signal such as an image, the filters are applied to the horizontal and vertical directions separately. After one stage, the DWT of an image consists four subbands, LL, LH, HL, HH, respectively. Figure B.1 shows the location the subbands. In the next iteration, the LL subband is further decomposed.

### B.2 Implementation of the Integer Wavelet Transform

The integer wavelet transform can be implemented using the so-called lifting scheme [59]. Computing the wavelet transform using lifting steps consists of several stages. The idea is to first compute a trivial wavelet transform (the Lazy wavelet or polyphase transform) and then improve its properties using alternating lifting and dual lifting steps. Figure B.2 shows the block diagram of the forward and reverse wavelet transform using lifting.
The lazy wavelet first splits the signal into its odd indexed samples $s_{1,l}$ and the even indexed samples $d_{1,l}$, then improve the calculation using a series of dual lifting and lifting steps. A dual lifting step consists of applying a filter to the even samples and subtracting the result from the odd ones,

$$d^{(i)}_{1,l} = d^{(i-1)}_{1,l} - \sum_k p_k^{(i)} s^{(i-1)}_{1,l-k}. \quad \text{(B.8)}$$

On the other hand, a lifting step consists of applying a filter to the odd samples and subtracting the result from the even samples,

$$s^{(i)}_{1,l} = s^{(i-1)}_{1,l} - \sum_k u_k^{(i)} d^{(i)}_{1,l-k}. \quad \text{(B.9)}$$

Eventually, the even samples will become the low pass (detail) coefficients while the odd samples become the high pass (approximate) coefficients. The inverse transform can be
found by reversing the operations and flipping the signs,

\[ s^{(i-1)}_{1,l} = s^{(i)}_{1,l} + \sum_k u_k^{(i)} d^{(i)}_{1,l-k}, \quad (B.10) \]

\[ d^{(i-1)}_{1,l} = d^{(i)}_{1,l} + \sum_k p_k^{(i)} s^{(i-1)}_{1,l-k}. \]

An integer version of every wavelet transform can be produced since every wavelet transform can be written using lifting. In each lifting step, the result of the filter can be rounded off right before adding or subtracting. An integer dual lifting step, which results the detail coefficients, is given as

\[ d^{(i)}_{1,l} = d^{(i-1)}_{1,l} - \left\lfloor \sum_k p_k^{(i)} s^{(i-1)}_{1,l-k} + 1/2 \right\rfloor, \quad (B.11) \]

and an integer primal lifting step, which results the approximate coefficients, is given by

\[ s^{(i)}_{1,l} = s^{(i-1)}_{1,l} - \left\lfloor \sum_k u_k^{(i)} d^{(i)}_{1,l-k} + 1/2 \right\rfloor. \quad (B.12) \]

The steps introduced above result in an integer to integer transform and the inverse is given by flipping the signs and reversing the operations. The issue with the scaling factor \( K \) is more complicated in the integer to integer transform since simply dividing by \( K \) might not result in integer coefficients. One solution is to ignore \( K \) and keep in mind that the coefficients calculated are off by \( K \) times \([59]\). In this case, it is preferable that \( K \) is as close to 1 as possible. For a complete review on Integer to Integer Wavelet transform please refer to Calderbank el al. “Wavelet Transforms That Map Integers to Integers” \([59]\).
Bibliography


[41] G. Xuan, C. Yang, Y. Zhen, Y. Q. Shi, and Z. Ni, “Reversible data hiding using integer wavelet transform and companding technique,” in *the 3rd International
Workshop on Digital Watermarking (IWDW 04), vol. 3304, Seoul, South Korea, 2004, p. 115124.


