People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
University of Batna 2
Faculty of Mathematics and Computer Science
Department of Computer Science

Thesis

submitted in fulfillment of the requirements
for the degree of Doctor in Computer Science

A Robust Watermarking Approach for Images Authentication and Traceability

---

Author:
Wassila BELFERDI

Committee members:

President :	Kamal Eddine MELKEMI	Professor	University of Batna 2
Supervisor :	Ali BEHLOUL	Doctor	University of Batna 2
Examiner :	Mohamed BENMOHAMMED	Professor	University of Constantine 2
Examiner :	Lemnouar NOUI	Professor	University of Batna 2
Examiner :	Larbi GUEZOULI	Doctor	University of Batna 2
Examiner :	Rachid SEGHIR	Doctor	University of Batna 2

Abstract

A Robust Watermarking Approach for Images Authentication and Traceability

Digital watermarking is a promising technology that has shown the ability to achieve the security and protection of image data.The enormous growth, use, and distribution of these images reveal security threats with legal and ethical complexities. Though, despite the wide interest that has, digital watermarking still not been widely adopted in some applications like image authentication with restoration ability. Existing watermarking schemes often suffer from technical and security flaws. Validation of the suitability of those schemes for an application becomes more challenging. One main reason for these problems is the compromise between the watermarked image quality, tamper localization and recovered image quality, which is a serious problem to the majority of current watermarking schemes with restoration ability.

Addressing these gaps, in this thesis, a number of original contributions have been made. Starting with a comprehensive literature review on digital watermarking schemes and their applications and determine the requirements for watermarking, which has led to the following main contributions:

A novel self-embedding watermarking scheme aims to authenticate the content of a watermarked image and to detect any possible alterations and recover damaged area is developed, the proposed scheme improves the performance in terms of imperceptibility and robustness, focusing on three major considerations: the invisibility of the embedded watermark, the accuracy of detection and the high quality of the recovered color image.

Moreover, in support of developing a strict-authentication scheme for medical image applications, a novel digital signature-based scheme that uses the Cholesky decomposition is developed, which overcomes the limitations of existing schemes. Experimental results prove that the proposed scheme has higher capacity and more efficient detection ability

In addition, aiming at developing a blind dual-color image watermarking scheme, a new watermark embedding scheme is addressed, which is different from some existing schemes that use the binary or gray-level image as watermark. Experimental results demonstrate the stronger robustness of the proposed scheme against most common attacks including image compression, cropping, noising and scaling, etc, that allows this scheme to be applicable for traceability applications.

The presented new self-authentication model would help develop more secure self-authentication scheme with restoration ability. Additionally, the presented Cholesky digital signature-based scheme for medical images and its validation has created a new efficient approach, which can be used for different applications including content authentication and tamper detection. Moreover, the proposed blind dual-color image watermarking scheme offers an efficient robust tool for images traceability.

Key words: Digital watermarking; Strict authentication; Digital signature; Robust watermarking; Fragile watermarking; Traceability.

Résumé

Le tatouage numérique est une technologie prometteuse qui a la capacité d'assurer la sécurité des images. L'énorme utilisation et distribution de ces images provoque des menaces de sécurité avec des complexités juridiques et éthiques.

D'autre part, malgré, le grand intérêt que le domaine du tatouage numérique a connu, il n'est pas encore largement utilisé dans certaines applications surtout dans l'authentification d'image avec la capacité de restauration. Les schémas existants souffrent souvent de limitations techniques et de sécurité. La validation de la pertinence de ces schémas pour une application devient plus difficile. Un des raisons de ces problèmes est le compromis entre la qualité de l'image tatouée, la localisation d'altération et la qualité d'image récupérée, ce qui constitue un grand problème pour la majorité des schémas de tatouage actuels avec capacité de restauration.

Pour résoudre ces problèmes, dans cette thèse, un ensemble de contributions originales ont été faites. Après un état de l'art sur les algorithmes de tatouage numérique, leurs applications et les exigences de tatouage,

un nouveau schéma de tatouage basé sur l'auto-insertion pour authentifier le contenu de l'image tatouée est proposé. Ce schéma permet de détecter les altérations possibles et de récupérer l'information perdue, ainsi que l'amélioration des performances de notre méthode de tatouage en termes d'imperceptibilité et de robustesse. Dans notre proposition on se focalise sur trois considérations majeures: l'invisibilité du tatouage, la précision de la détection et la haute qualité de l'image couleur récupérée.

En outre, pour développer un schéma assurant une authentification stricte pour des applications d'images médicales, un nouveau schéma basé sur la décomposition de *Cholesky* est développé, ce qui permet de surmonter les limitations des schémas existants. Les résultats expérimentaux démontrent que le schéma proposé a une capacité plus élevée et une détection plus précise.

De plus, dans le but de développer une méthode efficace de tatouage aveugle d'image couleur, un nouveau schéma est proposé, qui est différent de certains schémas existants qui utilisent une image binaire ou en niveau de gris comme marque. Les résultats expérimentaux démontrent que la méthode proposée a une robustesse élevée face à la plupart des attaques courantes telles que la compression d'image, le recadrage, le bruitage et la mise à l'échelle, etc.

Le nouveau modèle basée sur l'auto-authentification présenté aiderait à développer un schéma plus sécurisé avec une grande capacité de restauration. De plus, le schéma basé sur la decomposition de *Cholesky* pour les images médicales et sa validation a créé une nouvelle approche efficace, qui peut être utilisée pour différentes applications telles que l'authentification du contenu et la détection des modifications.

Mots clés: Tatouage numérique; Authentification stricte; Signature digital; Tatouage robust; Tatouage fragile; Traçabilité.

ملخص

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

كلمات مفتاحية:

[illegible]

To my parents and all my family members who have given me the support and comfort that accompanied me during this path.

Acknowledgements

First and foremost, I would like to thank Allah almighty for giving me the strength, knowledge, ability and opportunity to undertake this research study. Praise be to Allaah.

I would like to express my deeply-felt thanks to my thesis supervisor, Dr. Ali BEHLOUL, for his encouragement and thoughtful guidance I am grateful for the time he gave me, his educational and scientific qualities, his frankness and sympathy, his ideas and advices. I learned a lot with him and I should send him my gratitude for all that. It has been an honor and pleasure working with him during the last years.

I am happy to acknowledge my deepest sincere gratitude to Prof. Lemnouar NOUI, for the truly inspiring teaching which turned me towards searching. I am happy to acknowledge my deepest sincere gratitude to Prof. Kamal Eddine MELKEMI for the honor that he makes to preside this jury.

I also thank the members of thesis committee: Prof. Mohamed BENMOHAMMED, Prof. Lemnouar NOUI, Dr. Larbi GUEZOULI, and Dr. Rachid SEGHIR for having accepted to assess my thesis.

An extra special recognition to my family whose love and aid have made this thesis possible, especially my parents, I love you both very much and am immensely grateful for all that you do. Thank you for everything.

I am indebted to my friends: Wafa A, Hadjer B.

This work could not achieve its goals without the contribution of many people whom I extend my deepest thanks.

Contents

List of Figures

List of Tables

List of Algorithms

List of Abbreviations

BCH	**B**ose **C**haudhuri **H**ocquenghem
BER	**B**it **E**rror **R**ate
BMP	**B**it**M**a**P**
CFA	**C**olor **F**ilter **A**rray
Corr	**Corr**elation
CQ	**C**orrelation**Q**uality
CRC	**C**orrelation **C**oefficient
DCT	**D**iscrete **C**osine **T**ransform
DFT	**D**iscrete **F**ourier **T**ransform
DIVX	**D**igital **V**ideo **Ex**press
DVD	**D**igital **V**ersatile **D**isk
DWT	**D**iscrete **W**avelet **T**ransform
FFT	**F**ast **F**ourier **T**ransform
HVS	**H**uman **V**isual **S**ystem
IBM	**I**nternational **B**usiness **M**achines Corporation
IF	**I**mage **F**idelity
IWT	**I**nteger **W**avelet **T**ransform
JPEG	**J**oint **P**hotographic **E**xperts **G**roup
LSB	**L**east **S**ignificant **B**it
MP3	**M**otion **P**icture **3**Layer-3
MSE	**M**ean **S**quare **E**rror
MSB	**M**ost **S**ignificant **B**it
NC	**N**ormalized **C**orrelation
NCC	**N**ormalized **C**ross**C**orrelation
NVF	**N**oise **V**isibility **F**unction
PDF	**P**ortable **D**ocument **F**ormat
PN	**P**seudo-random **N**oise
PSNR	**S**ignal to**N**oise**R**atio
ROI	**R**egion **O**f **I**nterest
RONI	**R**egion **O**f **N**on **I**nterest
RSA	**R**ivest **S**hamir **A**dleman
RST	**R**otation **S**caling **T**ranslation
SARI	**S**elf **A**uthentication **R**ecovery **I**mages
SNR	**S**ignal to**N**oise**R**atio
SS	**S**pread **S**pectrum
SSIM	**S**tructural **SIM**ilarity
SVD	**S**ingular **V**alue **D**ecomposition
TA	**T**orus **A**utomorphism
URL	**U**niform **R**esource **L**ocator
VEIL	**V**ideo **E**ncoded **I**nvisible **L**ight
VQ	**V**ector **Q**uantization
WPSNR	**W**eighted **P**eak **S**ignal to **N**oise**R**atio
3D	**3**three **D**imensional

Chapter 1

Introduction

1.1 Motivation and Overview

Over the last years, the digital multimedia data and digital networks have seen a high-speed evolution. Hence, this growth has made duplication of multimedia data much easier and faster, the idiom "seeing is believing" can no longer hold true. Digital images are perfectly reproduced and undetectably manipulated. However, in situations, where the reality is unknown, it can not be said, for sure, that an image has been modified or not, the protection of these images represents a big challenge for researchers.

One solution would be to restrict access to the data using some encryption techniques. However, encryption does not provide perfect security. Once the encrypted data is decrypted, they can be unconditionally distributed or manipulated [35]. Hence, to ensure the security new methods are needed.

In the late 1990s, there was a great interest in digital systems for the watermarking of various content, more papers are published in which the foundations of digital watermarking are shown. The main focus has not been only on images, videos, and audios, but other content such as binary images [173], text [17, 18, 97], line drawings [143], 3D models [13, 117, 177], executable code [146], and integrated circuits [63, 76] have also been watermarked. Proposed methods were in direct response to the need to design a system that is both efficient and reliable to be used in several applications including copyright identification, verification of content integrity since the watermark was embedded, and broadcast monitoring.

Today, with the unsatisfactory of conventional security tools, the digital watermarking has been considered as a promising solution for reinforcing the trustworthiness of digital images [30]; by embedding a secret imperceptible watermark directly into the original data in such a way that it always remains present, once created the watermark can be detected or extracted for the purpose of owner identification or/and integrity verification of questioned data [30, 134].

The construction of a watermarking scheme significantly relies on and varies with its objectives (e.g., content authentication, copyright protection ...) and application scenarios (e.g., video, audio, image).

The research presented in this thesis aimed to investigate digital watermarking to provide a systematic way for designing, analyzing, and applying it for color image authentication purpose, with a particular focus on three major considerations: the invisibility of the embedded watermark, the accuracy of detection and the high quality of the recovered color image.

Due to the unlimited access and distribution of digital multimedia over the internet, authentications today are more threatened than ever. The easy transmission and manipulation of digital data create a real threat for information creators. Furthermore, they want to be sure that their work is not used in an improper way or modified without their permission. Thus, unauthorized use of data creates several problems and three main issues may arise:

1. How will the owner know that the original content is modified?
2. If the owner knows about this fact, how he can identify modified parts exactly?
3. The last but very important issue is that even if the first two problems are resolved, how the owner will restore damaged parts?

The above problems can be solved by hiding some information into the multimedia data, which can be extracted later to prove the authenticity of the original content and restore the modified areas.

When dealing with digital watermarking for images authentication. Even with its great usefulness a compromise is raising between the watermarked image quality, tamper localization and recovered image quality, which is a serious problem to the majority of current watermarking schemes with restoration ability.

The first drawback is the compromise between the quality of the watermarked image and the embedded data amount, means that the watermarked image quality decreases when embedded data amount increases [127].

The second drawback is the accuracy of tamper localization, increasing watermarking payload to further improve tamper detection, decreases the quality of the watermarked image.

The third drawback is the ability of tamper recovery, to restore the tampered regions with a good visual quality more bits to represent the effective information are used, subsequently, the visual quality of the watermarked image is affected [179].

Despite its promising future, digital watermarking has not been widely exploited in all potential applications. The majority of literature methods for image authentication either lacking the ability of tamper recovery or uses only gray-scale images[99, 22, 44, 93, 171, 85, 82] or recover tampered areas in low tampering rates [164, 182, 81, 23] or even fail in tamper detection [21].

Therefore it is required and makes sense to investigate the systematic development and evaluation of digital watermarking schemes in general for digital image applications, addressing the identified research gaps.

1.2 Research Objectives

The basic idea is to decrease the amount of the authentication and recovery data that modifies the host color image, to enlarge the watermarking capacity, then re-use the same amount of data (watermark) in the authentication and restoration processes. To address these challenging gaps, the research presented in this thesis has the following objectives:

1. To conduct a thorough review of requirements for digital watermarking schemes and their suitability criteria for images authentication.
2. To investigate and develop a formal generic watermarking model, and define its properties.
3. To propose a new digital watermarking scheme for color images.
4. To exploit the color images characteristics, to reduce the watermark's size to perform tamper detection with high detection accuracy.
5. To conceive a method of recovering the tampered image devised for the quality improvement of the restored color image.

6. To develop the expected attack models for watermarking schemes using the developed watermarking model, that will help analyze the potential threats in an application scenario.

7. To develop a computationally efficient watermarking scheme and evaluate its performance systematically for image authentication using the above framework, i.e., using the watermarking model and its defining properties, and design and evaluation criteria.

1.3 Research Questions

In this thesis we address several research questions. In what's follow, the major research questions are formulated.

1. What is an appropriate model for describing and analyzing a digital watermarking scheme?

   Watermarking has already been explored for digital image applications. All the schemes, however, may not be equally suitable and thus, may not be directly applicable even to similar problems. The answer of the above question requires an investigation to identify the main properties of a watermarking scheme. This means identifying the respective inputs, outputs, and various design requirements of each building block.

   Multimedia data are generally compressed and quality enhanced. Thus, accepting lossy compressed multimedia and some content-preserving filtering is an essential requirement in many applications. Thus, developing a security framework for image authentication rises a security concerns through a suitable watermarking scheme. Therefore, another key research question can be posed as follows:

2. What are the security attacks against image authentication schemes?

   Identifying security attacks is another essential part in assessing how the security requirements of image authentication schemes are achieved.

1.4 Research Outcomes

For answering the research questions and achieving the desired objectives, this thesis presents several contributions in the field of digital watermarking and its applications. These contributions and findings that will be discussed in this thesis have been presented in reputable conferences, and published in leading journal (See Appendix A). The main contributions are summarized below.

1. A new blind watermarking scheme for embedding color image watermark into the color host image has been introduced, which offers three main features: The first one is to use the integer wavelet transform (IWT). The second one is to embed color image watermarking into the color host image. This is motivated by the fact that compared with gray level watermarking; digital color image has more amounts of data. The last but not the least is that the proposed detector is blind (Chapter 6 & Publications in Appendices A). Specifically:

   - A blind dual-color images watermarking scheme based on IWT and sub-sampling is developed. IWT can map integer to integer without the rounding error and can further concentrate the energy of each sub-image in few coefficients that are used

to embed the watermark. After sub-sampling the color host image, obtained sub-bands are used to embed the watermark, exploiting the advantage of the high correlation between LL sub-band coefficients for better imperceptibility and stronger robustness.

- A new blind watermarking model is proposed to address the lack of computational efficiency in existing watermarking schemes. In fact, there is a strong correlation between the LL sub-bands coefficients, this property can be used to embed and extract the watermark in a blind manner.
- A set of expected attacks on the proposed scheme have been investigated. These attacks are defined considering different inputs (e.g., watermarked image).
- Experimental results demonstrate that the proposed scheme not only guarantees the invisibility of watermarking but also has strong robustness against the operations of common image processing (image compression, cropping, noising and scaling...), and its performance outperforms other methods considered in this paper.
- The novelty of this scheme is the use of image sub-sampling that is used to ensure the imperceptibility of the watermark, the blindness of the detector, and the robustness against several attacks. This approach makes the solution more suitable for several applications including images traceability.

2. A new basic watermarking model that is based on Cholesky decomposition is developed for ensuring a strict authentication service of an important case of watermarking that aim detect any possible tampered region of an image with a special application in medical and military images (Chapter 5 & Publications in Appendices A). Specifically:

   - A Novel Cholesky decomposition-based scheme for Strict Image Authentication. To ensure the requirement of strict authentication where no changes are tolerated to the color host image, the Cholesky decomposition properties are used beside the use of cryptographic hash function and public key crypto-system.
   - For the purpose of reducing the authentication data size, diagonal coefficients of the Cholesky decomposition matrix are used instead of using the entire matrix.
   - The experimental results demonstrate that our technique is able to identify image tampering despite the minor modifications, even if only one bit is modified. In addition, the proposed scheme demonstrates its effectiveness in sensitive information systems such as medical images.
   - This scheme can be used in healthcare systems as an efficient tool that secures the sharing and control of medical images. A review of an application example in medical information security is shown to ensure both integrity and authenticity.

3. A new watermark embedding scheme is developed, in light of the offer mentioned research, this new strict authentication model devised for color image tamper detection and restoration offers three major considerations: the invisibility of the embedded watermark, the accuracy of detection and the high quality of the recovered color image. (Chapter 4 & Publications in Appendices A). Specifically:

   - An effective fragile watermarking scheme based on the Bayer pattern is proposed for color image tamper detection, and restoration, this scheme satisfies the requirements of self-embedding, tamper detection and restoration abilities.
   - The basic idea of this scheme is to decrease the amount of the authentication and recovery data that modifies the host color image, which allows embedding

several copies of a reduced watermark into the LSB bits of the host color image, then use it later in the authentication and restoration processes.

- Experimental results are concluded to prove the effectiveness of the proposed method and show the major contributions of this paper includes:
  - Firstly, the simple structure of the watermark self-embedding into color images that is visually indistinguishable.
  - Secondly, the exploitation of the color images characteristics, to reduce the watermark to a gray-scale image using the Bayer pattern and Torus Automorphism to perform tamper detection with high detection accuracy reaches up 100% and low false alarm ratio.
  - Finally, a method of recovering the tampered image by interpolating the extracted gray-level watermark has also been devised for the quality improvement of the restored color image which can achieve 34 (dB) in tampering rates greater than 25%.
- This scheme can find a niche area of application where the security becomes a serious problem such as in medical images and in law enforcement.

1.5 Thesis Organization

The remainder of this thesis is organized as follows.

Chapter 2, captures the background of the proposed research of this thesis, with an overview of the key research themes, namely digital watermarking as an effective security tool, their necessary and relevant components, properties, and digital watermarking applications. Moreover, watermarking schemes are classified, evaluation criteria, objectives, and applicability are also presented. As well as digital image watermarking techniques, advantages and inconveniences are investigated.

Chapter 3, captures the different watermarking theoretical aspects and their practical interpretations for several applications including image authentication application, the formal generic watermarking model is also investigated, main watermarking requirements, possible attacks, and an overview of the key research themes are also discussed. The review of image authentication schemes advantages, inconveniences, design and evaluation criteria, objectives, and applicability are presented. concluding with the future of watermarking as promising security technique.

Chapter 4, presents the development and validation of a new watermark embedding scheme. That can be classified as being a: spatial, fragile, additive, blind, invisible, non-reversible, and self-embedding scheme. Detailed steps of watermark embedding, extraction, localization, and restoration are discussed. To improve its performance in terms of imperceptibility and robustness, a self-embedding and a reduction watermark processes are presented. The proposed method focuses on three major considerations: the invisibility of the embedded watermark, the accuracy of detection and the high quality of the recovered color image. The justification for considering the task of developing the proposed embedding scheme and the challenges are identified. The new watermark embedding scheme, features, implementation, and applicability are discussed as well as findings and contributions presented in this chapter.

Chapter 5, describes the development and validation of a new watermark embedding scheme for strict image authentication that is based on Cholesky decomposition. A digital image watermarking based on formal generic model is used in defining the basic

properties. The proposed scheme can be classified as being a: frequency, fragile, non-reversible, and strict authentication scheme. Considering possible abilities of the adversary, a set of possible attacks are presented, research findings and contributions of this chapter are discussed.

Chapter 6, presents the development and validation of a new watermarking scheme, that is based on integer wavelet transform (IWT) and sub-sampling. The proposed scheme can be classified as being a: frequency, robust, additive, blind, invisible, non-reversible, and dual color image watermarking scheme. Justification for considering the task of developing embedding scheme is discussed and the challenges are identified, beside, the new watermark embedding scheme, and its features, implementation and applicability.

Chapter 7, concludes the thesis with a summary of the original contributions and future work.

Part I

State of the Art

Chapter 2

Watermarking Techniques, Applications and Classification

2.1 Introduction

The digital representations of multimedia material (movies, songs, photographs...) represent a main advance in the field of communication that offers many advantages. However, the unlimited reconstructions of illegal copies reveal a serious threat to the rights of content owners and a new challenges for researchers.

Until recently, encryption was the main tool that offers the security of content owners' rights during the transmission of the data, from the sender to the receiver. After encrypting it, the encrypted file is delivered via the Internet. Although, without the appropriate decryption key the delivered file would be useless to a pirate, only authorized peoples have the access to purchase legitimate copies of the content. Unfortunately, after decryption, an unprotected copy of the content can be reproduced and illegally distributed. In other words, after reception and decryption, the protected data is no longer protected [107, 30].

Thus, the urgent need for an efficient alternative or cryptography complement, to overcome the security gaps and provide better protection than what cryptography can offer, more sophisticated methods are designed to protect content even after decryption. This led to the discovery of the data hiding field.

Steganography and watermarking are two closely related concepts that belong to the data hiding field (See Fig.2.2). *The steganography is the practice of undetectably altering a cover object to embed a secret message*. The hidden message itself must be protected and can not be perceived by the user, while, the cover object is not valuable [30].

On the other hand, the term watermarking has slightly different meanings in the literature, one definition that seems to appropriate is the following, *the watermarking is the practice of imperceptibly altering a cover object to embed a message about that cover object, although in watermarking the hidden message is usually related to the cover object* [30]. Meaning that the coupling of the message to the cover object is of value and the protection of the content is critical [125, 154]

Encryption is considered a useful transmission tool. But, it doesn't provide original data manipulation in its protected form. Unlike encryption, the watermark persists in the content and perceive the same manipulation as the host content (listening to, viewing...). Also, unlike the steganography, where both message and hiding method are secret, in watermarking, the watermark embedding process and the message, except the secret key, do not have to be secret [126, 107, 30].

Even though the objectives of watermarking and steganography are quite different, steganography is used for secret communication, whereas the watermarking ensures that the content and the watermark are inseparable. This principal advantage makes watermarking suitable

for several applications including content protection, copyright management, content authentication, tamper detection, broadcast monitoring, copy and device control, transaction tracking... [126, 107, 30].

Digital watermarking is a relatively recent research area that has attracted the attention of researchers in academia and industry branches. It has become one of the hottest research topics in the multimedia signal processing community. The concept of digital watermarking arises as a powerful solution to several problems in digital media such as the copyright protection and content authentication as a means to identify the owner or distributor of digital data or to verify the content integrity [154].

This chapter captures the background of the proposed research of this thesis, with an overview of the key research themes, namely digital watermarking (Section 2.1) as an effective security tool and digital watermarking applications (Section 2.2). Classification of watermarking schemes (Section 2.3), where evaluation criteria, objectives, and applicability are presented. Digital image watermarking techniques (Section 2.4), are reviewed, advantages and inconveniences are investigated. Then a comparison between spatial and frequency techniques is reviewed in Section 2.5. Chapter summary is given in Section 2.6.

2.2 Digital Watermarking Applications

Watermarking is an interdisciplinary study that attracts the attention of researchers from communications, cryptography, audio and image processing. It can complement cryptography to provide a continues protection even after decryption and delivering the content over the Internet. Leading us to ask questions about whether watermarking can provide effective solutions to real problems, and if it can create new and interesting problems for basic and applied research.

New problems have been proposed. Thus, watermarking must meet difficult and often conflicting [1] constraints [32]. Low embedding distortion, imperceptibility, and security are the common requirements of all classes.

Generally, image watermarking application may have different objectives, that are classified into security and non-security objectives (See Fig. 2.1). Security objectives to achieve certain security properties such as integrity of the watermarked image and non-security objectives to annotate an image-database for efficient management [115].

Current interest is focused on a number of applications that broadly fall into the categories of security and device control [32], each different application imposes different requirements on the watermarking system. For example, robustness is a main requirement for copyright applications, while it is unneeded in most authentication applications [129, 154, 2].

Actual watermarking applications are broadcast monitoring, owner identification, ownership proof, transaction tracking, authentication, copy control, and legacy enhancements. Characteristics of each of these applications are identified to show the watermarking suitability as a solution. To do this we must carefully consider the application requirements and examine the limitations of alternative solutions [30].

2.2.1 Owner identification and proof of ownership

Textual copyright notices have several limitations as a technology for identifying the owner of a digital item. Even when it is copied, those documents are easily falsified. On the other

[1] Capacity, imperceptibility and robustness are three contradictory requirements, a higher capacity means the expense of either imperceptibility or robustness strength or both, thus, a good compromise between these requirements should be considered (See Section 3.3).

hand, using a central repository to register the copyright is too costly, especially with many documents to be registered [30].

Because watermarks can be made invisible and inseparable from the content, they are likely to be superior for owner identification and ownership proof [30].

This category of applications was the first concern of watermarking literature. The embedded data identifies the rights/content owner or distributor of a digital object and used for notifying and tracking illegal copies of the item. Also, it is used for ownership proving [154, 129].

2.2.2 Broadcast Monitoring

Over the last few years, the number of television and radio channels delivering content has seen an emerging growth. As well as the amount of content transmitted over these media. With this fast-growing, the broadcast reality has become an urgent need for content and copyright owners, distributors, and broadcasters.

Broadcast monitoring is a technique of cross-verifying if the supposed content to be broadcasted was actually broadcasted at the right time and duration [126, 129].

Before watermarking, a low technique of broadcast monitoring is used, a human has to observer the broadcasts then record what they see or hear. However, besides the high probability of error, this method is costly. Thus, automated monitoring is highly desirable. For doing this, two categories of techniques are proposed: Firstly, passive monitoring systems where a computer monitors broadcasts and compares the received signals with a database. Secondly, active monitoring systems rely on associated information that is broadcast along with the content. However, because of the database large size, comparing the received signals against a database is not trivial. As well, storing and managing the database can be expensive [30].

Watermarking is considered as an alternative solution to provide broadcast monitoring services. It has the advantage of existing within the content, rather than modifications, and is therefore completely compatible with the installed base of broadcast equipment, including both digital and analog transmission [30, 154].

2.2.3 Transaction Tracking

Transaction tracking[2] is another important application of watermarking to make a major contribution in electronic items traceability. In broadcast monitoring and owner identification applications, the same watermark is embedded in all copies of the same content. However, transaction tracking application provides a deterrent tool to illegal use, a unique watermark is embedded in each individual copy, thus, a content owner or distributor can identify the illegal copy source [33]. In addition, this watermark is embedded not only to provide information about the legal owner or distributor of the digital item but also to mark it whole past-history.

A copyright owner inserts different watermarks in each customer copy, which records one or more transactions to prevent illegal copy or distribution and for deterring such actions [154]. For example, for each document copy, the owner would place a different watermark, that might record the recipient in each legal sale or distribution of the document. Then, if the item is misused, the owner could find out who was responsible [30, 2, 129, 27, 126].

Transparency and robustness requirements must be satisfied for this application. Though, the main challenge that fingerprinting schemes face is the collusion attack (See Section 3.4.2)

[2]Transaction tracking is more often called traceability or fingerprinting, similar to a human fingerprint that uniquely identifies a person, the watermark can uniquely identify each legal copy of documents.

in which several legal copies of the same media are obtained to produce an approximation of the original unwatermarked version for illegal redistribution[3] [2, 129, 27].

Noting that transactional watermarks application has been implemented in the $DIVX$ digital video disk players and in the distribution of movie dailies. Each of which places a watermark that uniquely identifies the player in every movie that is played [107, 33].

2.2.4 Copy Control

Another common piracy scenario is illegal copying or recording. Watermarks for broadcast monitoring as well as for owner identification, proof of ownership and transactional watermarks do not prevent illegal copying. Rather, they serve as a deterrent against such wrongdoing [33, 30]. For example, broadcast monitoring and transaction tracking services help identify dishonest broadcasters and adversaries after they distribute illegal copies. Hence, preventing the illegal action is clearly better.

The first defensive line against illegal copying is encryption, where the decryption is not possible for anyone who does not have the encryption key. In fact, besides decryption, the user must also convert the content to an analog signal in order to perceive it. However, this conversion to the analog signal destructs all digital technologies for content protection. In other words, a high-quality illegal digital copy could be reconstructed by analog signal digitization, therefore, this copy can be illegally played and recorded on consumer devices. Thus, what we would like is to somehow allow media to be perceived, yet still, prevent it from being recorded. For doing this, existing technology are applicable only to analog television signals, rather than audio or to digital signals of any type [30].

Since watermarks can be embedded in the content itself, they are present in every representation of the content, could transmit a significant number of bits as well robust against removal and have the ability of blind detection, therefore, watermarking is one powerful tool to provide a better method of implementing copy control [30, 2].

Unlike the above-mentioned applications, where watermarking is used for identifying and deterring intellectual rights violation, in copy control watermarking applications, active protection is offered by controlling the terms of use of the digital content. The embedded watermark contains the rules of copying and use, then a detector installed in the recording device detects the watermark, disallows recording, and identify illegal copies and refuse to play them. This application is currently envisaged for digital video disks (DVD) [107, 154]. However, this mechanism requires every recording device contained a watermark detector which is capable, by law or patent license, of preventing copyrighted material copying. unfortunately, it is difficult to persuade the consumers to pay more for a device that restricts their "freedom" of making copies [129].

2.2.5 Legacy System Enhancement and Database Linking

To improve the functionality of legacy systems[4], a very large deployed system should be upgraded, although, this upgrade may be incompatible with the existing system. Ideally, the upgraded system should be backward compatibility with deployed devices, and it must continue to work with the existing legacy system. Therefore, data embedded through watermarking is one of the ways to achieve this end [30, 154, 32].

[3] To be vulnerable to collusion attacks, fewer than 20 copies are required. In fact, most existing watermarks can be removed using fewer than 20 copies [14, 147, 32].

[4] Legacy enhancement is sometimes called meta-data insertion which refers to the data that describes the content.

This application can be used in a wide range of area, for example, these techniques are used for generating robust watermarks that contain a digital image URL. Then the embedded URL can be used later for connecting her automatically to the corresponding web page, similarly, medical X-rays could store patient records [6, 142].

2.2.6 Authentication and Tamper-proofing

The availability of sophisticated processing tools allows the easy modification of digital items. Therefore, authentication and tamper detection are more needed. Thus, watermarking provides the authenticity and integrity of a digital item. Any changes made to the item will also be made to the watermark. As a result, modifications can be detected. Verification watermarks are required to be fragile so that any modification to the image will destroy the mark. Thus, degrade or destroy this fragile watermark indicates the presence of tampering and hence the content can not be trusted anymore [2, 154].

Tamper detection is very critical for some applications that use highly sensitive data such as medical or satellite images. It is also useful to prove whether the image is tampered or not in a law court. Some authentication watermarking methods can provide tampered localization [126].

These are the main watermarking applications that are currently being used, but, with the emerging development of this technology, several other promising applications are likely to appear.

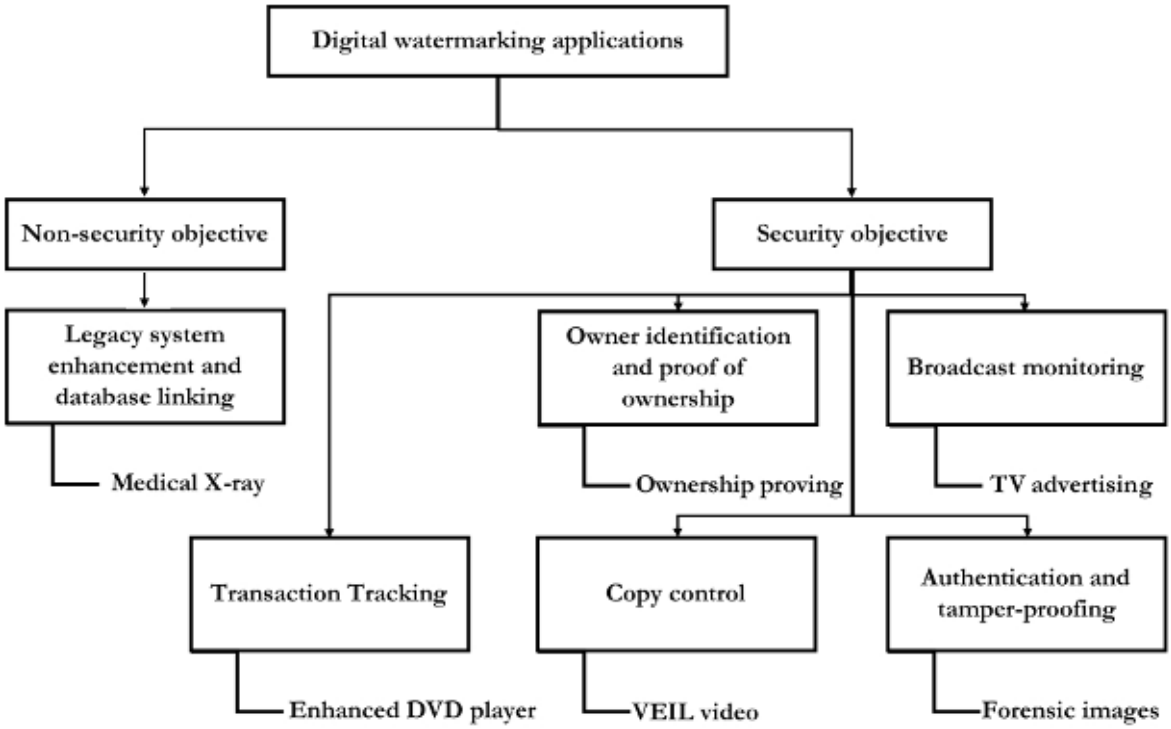


FIGURE 2.1: Classification of watermarking technology based on applications

2.3 Classification of Watermarking Schemes

In the literature, digital watermarks, their features, their techniques, and applications are classified into various categories each with their own distinct properties and characteristics.

However, there is no uniform criterion for the classification of image watermarking schemes [50].

Thus, with reference to various survey articles [30, 134, 142, 126], a watermarking system has certain generic requirements which must be met when implemented. According to these requirements, watermarking schemes can be categorized according to the type the host signal (still images, circuit design, video or audio signal), their resistance to host medium modifications (robust, semi-fragile, fragile), and the host signal availability during extraction (non-blind, semi-blind, blind). Also, they can be distinguished in terms of the watermarking embedding domain [142], as shown in Figure 2.2.

In the following, basic categories of watermarking schemes and descriptions for their properties are reviewed distinguishing each class from the rest.

A first classification can be organized according to the host signal. Thus, we can distinguish between the following categories:

2.3.1 According to the Host Signal

Watermarking technique can be classified depending on the media into which the watermark is embedded.

Text Watermarking

To check the alteration made to text files (e.g. Portable Document Format (PDF), DOC...), font shape and spaces between characters or inline are used to insert the watermark. Hence, in the case of fonts modulating, the watermark can't be detected [120, 9, 142].

Image Watermarking

This class of watermarking is the most widely used to protect the photos over the Internet, by profiting the image properties and human visual system characteristics, the secret data (Logo, Stamp, Label...) is hidden into an image then detect it later [120, 9, 142].

Audio Watermarking

Internet composition of tunes (e.g. $MP3$) attracts the interest to audio watermarking. Like other categories, this approach requires the robustness and inaudibility of watermarks [120, 142, 9].

Video Watermarking

Video refers to a three-dimensional $(3D)$ signal, with an image that preserves $2D$ in space and $1D$ in time. Thus, video watermarking is an extension of image watermarking, to provide video control, the watermark is embedded in the video stream. One of the main differences that distinguish video from image watermarking is time synchronization. This category requires the robustness and real-time extraction for compression [120, 9, 142]

Graphic Watermarking

In this case, the host signal is $2D$ or $3D$ computer generated graphics, where the watermark is embedded to provide a copyright service [9, 142].

Watermarking schemes can be also classified based on the embedded data type.

2.3.2 According to Watermark Type

Basically, there are two main types of watermarks that can be noise (pseudo-random noise PN, Gaussian random and chaotic sequences) or image (binary, stamp, logo, Grey-scale, and label).

Noise

Noise refers to a randomly generated sequence that can be a pseudo-random noise (PN), Gaussian, or Chaotic sequence.

PN sequence is the most widely used since it has the good feature of self-correlation, as well as the robustness against cryptographic attack. However, it is weak for image processing, noise and compression attacks [15]. On the other hand, Gaussian is a randomly generated sequence of numbers comprising 1 and -1, as a watermark, it is termed with zero and one variation $N(0,1)$. To detect the watermark, a correlation measure is used. Generally, such watermarks are used for extracting multiple inserted watermark [28]. Chaotic sequence is easier to generate, simply modifying initial values generates a completely different chaotic sequence [181, 120, 9, 126].

Image

Instead of embedding a noise as a watermark, meaningful data in form of a logo image is used. Such watermarks are termed as binary, gray-scale, or color image watermarks that are used for subjective detection [126].

2.3.3 According to Perceptibility

Based on the human visual perceptibility, watermarking schemes are generally classified into perceptible and imperceptible watermarking.

Perceptible Watermarking

As a perceptible watermark everyone can see, hear, or read what the sender wants others to know. It confirms the authentication and mainly used in logo or trademark label. The disadvantage of type watermarking is its fragility to attack. Also, in some cases, the watermark embedding process reduces the quality of the host signal [156, 120, 91].

Imperceptible Watermarking

In this category, the original host signal and the watermarked one are indistinguishable. Watermark can't be detected with a human eye or ear only authorized person can observe it to validate the owner authentication. The watermark is inserted by altering the host signal, in a perceptually unnoticeable way and it can be recovered only with an appropriate decoding mechanism. Compared to perceptible watermark, imperceptible one is very robust to proof the ownership, however, preventing modifications is not possible [120, 142, 138, 91, 9, 131].

2.3.4 According to the Embedding Domain:

In general, the embedding techniques can be classified into two categories: spatial domain approach or frequency domain approach.

Spatial Domain:

This domain focuses on modifying the pixels of one or two randomly selected subsets of images. It directly loads the data into the pixel values of an image. Some of its algorithms are LSB, spread spectrum (SS) Modulation based technique (See more details in Section 2.4.1) [114, 155, 149].

Modifications might include flipping the low-order bit of each pixel. The inserted information may be easily detected using computer analysis.

Frequency Domain

Also called transform domain. Values of certain frequencies are altered from their original. There are several common used transform domain methods, such as the discrete cosine transform (DCT) [34, 12, 100], discrete wavelet transform (DWT) [66, 11, 10], and discrete Fourier transform (DFT) [144, 87, 61].

2.3.5 According to Robustness

Modifications of the host signal can either be a result of common signal processing operations (e.g., lossy compression) or be specifically intended to destroy the watermark or affect the credibility of a watermarking system. Such modifications are usually referred to as attacks (More details will be discussed in Section 3.4) [154]. Depending on the resistance of watermarking schemes to attacks, one can distinguish between the following watermarking categories techniques:

Robust Watermarking Schemes

Robustness refers to the potential to detect the presence of the watermark after the host signal perceives common signal processing operations [30]. This class of schemes is mainly devised for copyrighted digital items, the embedded watermark designed so as to resist the common edit manipulations including lossy compression, filtering, printing, scanning, and geometric distortions (rotation, translation, scaling...). To provide certification, the watermark is not destroyed even after attacks. But, because of the severity of these attacks, no watermarking scheme can resist all kinds of modifications. Thus, robustness is considered as the subset of all possible manipulations that degrades the host signal up to a certain degree [154, 30, 131, 142].

This class of schemes has found its applications in many areas, including ownership proof, identification, transaction tracking, copy control, and broadcast monitoring. Thus, robustness to all possible processing operations is not required for all watermarking applications. Rather, a watermark need only survive the common signal processing operations and still being detectable between the embedding and detection processes [30].

In some applications, the robustness is considered as undesirable and completely irrelevant property. Actually, an important category of watermarking research focuses on fragile watermarks.

Fragile Watermarking Schemes

Unlike robust watermarking, fragile watermark is designed to not be robust. In this case, the watermarks can be easily destroyed by any attempt to tamper with them, in other words, fragile watermark is designed to consider all kind of malicious attacks, such as cut-and-paste and vector quantization and non-malicious attacks such as lossy compression, transcoding,

bit rate scaling, and frame rate conversion, though, even the slightest host data modification destroy the watermark.

Fragile watermarks are usually applied in authentication scenarios and content-integrity verification, they are designed to be vulnerable than robust ones, they must be very sensitive to the signal changes then the state of fragile watermark determines whether the data is tampered or not [30, 154, 131, 142].

Fragile watermarking is feasible only in sensitive applications, such as military intelligence, satellite image, and medical image archiving. Although, content-preserving operations (e.g., lossy compression) are sometimes necessary for many Internet and multimedia applications.

Semi-fragile Watermarking

In this class, a set of attacks are considered as legitimate and tolerable, providing selective robustness to a subset of attacks, while being fragile to others, such watermarks are proper for authentication applications instead of fragile ones [154, 126].

We should notice that in practice all robust watermarks are essentially semi-fragile, the selective robustness requirement is not imposed by the system designer but rather something that can not be avoided [154, 129].

Usually, the property of semi-fragility is achieved by exploiting the relationships among transformed coefficients of the host signal. These relationships are modified by malicious manipulations while invariant with content-preserving operations [129]. Although, distinguishing content-preserving operations from malicious attacks is a challenging issue for semi-fragile schemes. Therefore, this issue restricts the use of semi-fragile watermarking, where is not suitable for legal and national security issues applications [129].

In order to more enhance the robustness and increase the level of security, watermarking schemes are usually controlled by cryptographic keys.

2.3.6 According to Encryption Method

As in cryptographic systems, watermarking schemes can be classified into two categories based on encryption method used during embedding and detection processes:

Symmetric or Private Key

Such schemes are also called private key schemes, the same key is used to embed and detect the watermark, so, the sender and receiver share the same key. However, key management is a major disadvantage, where keys should be exchanged securely [144, 64].

Asymmetric or Public Key

In contrast to the symmetric encryption, two different but mathematically related keys are used, a public key and its associated private key. The watermark can be detected using a key that is different from the used one in the embedding process [52, 51]. A private key is used to embed the watermark and a public one for detection. For each private key, several public keys may be produced. Despite their advantages over their symmetric counterparts, asymmetric schemes are much more difficult to devise [154, 142].

2.3.7 According to the Embedding Process

Considering the information needed during the embedding process, two categories of watermarking schemes can be distinguished:

Blind Embedding Schemes

Schemes of this category see the host data as noise or interference. Therefore, these techniques consider watermarking as a classical communications problem, where a signal is transmitted over a noisy channel, in the case of watermarking, we should take into consideration the restrictions on the distortions amount imposed on the host signal by the watermark. Essentially all first watermarking schemes belong to this category and most of them use the spread spectrum principle, where the watermark signal is a pseudo-random sequence embedded, usually in an additive way, in the host signal [30, 154].

Informed Coding (Embedding) Schemes

Blind embedding scheme restricts the encoded watermark to be independent of the cover signal. The embedder already has knowledge about the original cover signal, thus there is no need to impose this restriction. This knowledge can be used to improve watermark detection performance. These methods are also called known host state methods, in which the watermarking, at the transmitter, is considered as communication with side information problem [102, 108]. The only difference from blind embedder is providing the host signal as an additional input to the watermark encoder [30, 154].

2.3.8 According to the watermarked Image Quality (Lossless)

According to the quality of a watermarked image, watermarking can be classified into reversible and irreversible watermarking.

Irreversible Watermarking Schemes

In such schemes, the watermark embedding process is not reversible; in such a way that the modifications to embed the watermark in the original host signal are permanent. Although these modifications are often insignificant. Though, one limitation of these schemes that they can't be used in applications of high significance, such as military, government, law, and medical images, where any tiny distortion, even if it were a result of the watermark embedding process itself is hard to be accepted [129, 91].

Reversible Watermarking Schemes

In some applications, perfectly recovering the original signal is required for watermarking schemes, by removing the digital watermark and replacing the data that had been overwritten. Reversible watermarking[5] is considered as a new authentication technique that is used for reconnaissance images and military applications [129, 91].

2.3.9 According to the Detection Process

This category defines the required resources to extract the watermark. As shown in Fig.2.2, When it comes to the detection process, watermarking schemes can be divided into three main categories:

Non-blind Watermarking

Non-blind are also known as visual, private, non-blind, or non-obvious schemes. These techniques require either the original signal or some information derived from it during the

[5] Also known as invertible or erasable watermarking.

detection phase. Those schemes can be considered as a more general of informed detection category. Despite their limited applications, these schemes are considered the most robust methods of watermarking [9, 142, 154, 91].

Semi-blind Watermarking

Such schemes do not require the original data for watermark detection. But, the original watermark and the key are required to extract the right watermark [113, 142, 9].

Blind Watermarking

This detection type requires neither the original image (or other information about it) nor watermark data. Due to their wider scope of application, blind (also known as public watermarking) techniques attract more researchers attention. Compared with non-blind detection schemes, unnecessary knowledge of the original host signal implies a higher watermark technology, which creates more challenges to design blind detection schemes. The main approach that belongs in this category is correlation-based detection, where the watermark presence is decided according to the correlation between the watermark and the signal [9, 142, 154, 91].

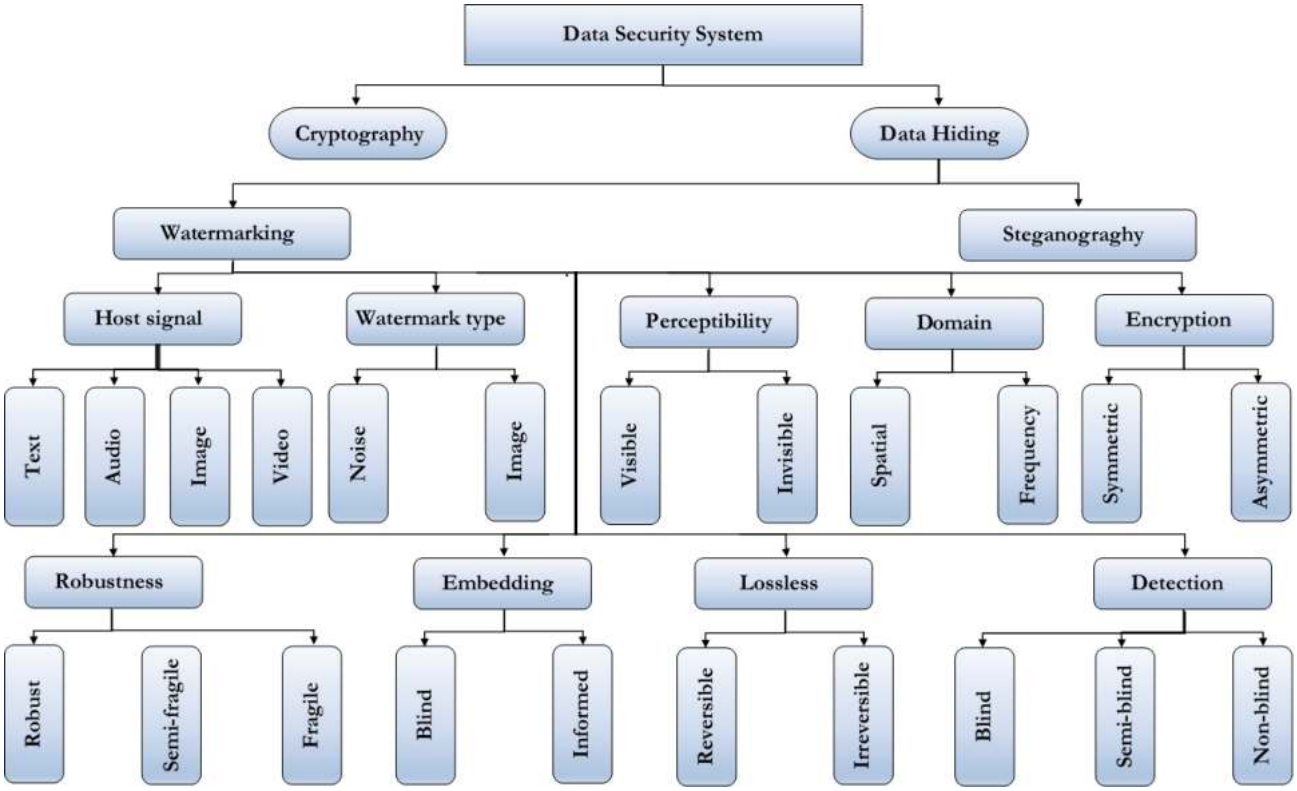


FIGURE 2.2: Classification of Watermarking Schemes

2.4 Digital Image Watermarking Techniques

Watermarking is considered as the process of combining two data parts in such a way that they can be independently detected at any moment [30]. One piece of information is the host data, that can be music, an image, or a movie ... The other piece of information is a watermark, that can be detected later. In fact, watermarking is possible because the human

visual system (HVS) discards significant amounts of data when processing it. This redundancy is, of course, important to the lossy compression field, which watermarking exploits to hide watermarks [107, 60].

The chosen method to embed the watermark has a direct influence on detection algorithm and the robustness against attacks, accordingly, designing an effective watermarking scheme implies finding a compromise amongst the basic features of robustness, fidelity, and payload (See Section 3.3) [91].

In general, there are two major approaches to represent digital images namely spatial domain and transform (frequency) domain.

2.4.1 Spatial Domain

Spatial domain-based approaches are designed to embed the watermark directly in the cover image, by slightly changing the value of randomly selected pixels or modifying host image characteristics. The number of changed bits in the pixels should carefully consider avoiding that the watermark becomes visible, to use it for document authentication and tamper detection. LSB and spread spectrum (SS) are the most used methods in spatial domain [16, 2].

Least Significant Bit (LSB) Method

This is the simplest and the most commonly used technique in spatial domain, for a given image each pixel is represented by an 8-bit sequence, the last bits (least significant bit LSB) of the pixel values of that image contain less relevant information while the most information is contained in the most significant bits $(MSBs)$. Thus, the modification of LSB bits does not cause perceptible changes. Before embedding the watermark, it is first encrypted then some random pixels of the cover image are selected using a key, which determines the pixels that will be modified by the embedding process. This technique of watermarking consists of replacing the LSB of the pixel values in the carrier image with the watermark's values. Thus, the watermark can be hidden partially or completely. To extract the watermark, embedding algorithm is reversed [91, 9, 62, 65, 120]. Fig.2.3 illustrates the LSB method sample of watermark embedding (to extract the watermark, same steps are applied in the inverse order.), moreover, embedding algorithm steps are shown in Algorithm 1.

Algorithm 1 Least Significant Bit (LSB) Algorithm

1: **Input: Cover image, Watermark**
2: **Output: Watermarked image**
3: Read both cover image and watermark.
4: Convert cover images into gray-scale.
5: Convert the watermark into bit-sequence.
6: Replace the low significant bits of the watermarked image with the watermark bit-sequence.
7: Make the least significant bits of host image zero.
8: Add shifted version (step 6) of watermarked image to modified (step 7) cover image.

Besides the easy implementation of this method (See Algorithm 1), it provides high perceptual transparency, and resistant against cropping attacks. Even though, LSB techniques are less robust to common signal processing operations and sensitive to noise and compression which can simply destroy the watermark [91, 9, 62, 65, 120].

Cover Image

Watermarked Image

111	113	108	110	106	115	104	108	111	103	97	114	100	101	105	99
111	113	106	108	107	105	107	104	101	108	97	112	106	119	141	118
106	107	104	103	99	101	96	99	97	96	99	125	111	141	168	163
104	102	106	102	95	96	100	96	98	97	99	120	132	169	202	190
107	97	106	109	107	98	100	105	131	115	119	143	171	191	206	217
98	96	119	112	97	85	76	77	131	142	120	137	168	205	228	227
109	116	128	126	99	88	79	83	102	127	118	99	114	199	232	232
116	112	144	146	122	97	94	87	110	156	128	99	111	191	230	229
133	108	137	149	140	127	126	115	139	162	119	98	127	202	230	227
149	115	108	136	152	136	135	138	133	124	110	105	159	218	226	227
169	135	107	114	135	141	144	137	127	107	116	145	200	226	224	227
188	164	138	111	111	108	115	100	112	122	137	186	220	229	230	235
180	177	163	147	134	122	123	121	128	162	186	208	220	222	235	238
192	186	196	183	162	150	144	153	169	192	206	220	219	224	231	234
199	204	201	199	187	182	184	186	196	214	226	231	230	234	232	235
183	194	209	203	205	205	213	213	212	218	224	224	228	219	226	240

110	112	108	110	106	114	104	108	110	102	96	114	100	100	104	98
110	112	106	108	106	104	106	104	100	108	96	112	106	118	140	118
106	106	104	102	98	100	96	98	96	96	98	124	110	140	168	162
104	102	106	102	94	96	100	96	98	96	98	120	132	168	202	190
106	96	106	108	106	98	100	104	130	114	118	142	170	190	206	216
98	96	118	112	96	84	76	76	130	142	120	136	168	204	228	226
108	116	128	126	98	88	78	82	102	126	118	98	114	198	232	232
116	112	144	146	122	96	94	86	110	156	128	98	110	190	230	228
132	108	136	148	140	126	126	114	138	162	118	98	126	202	230	226
148	114	108	136	152	136	134	138	132	124	110	104	158	218	226	226
168	134	106	114	134	140	144	136	126	106	116	144	200	226	224	226
188	164	138	110	110	108	114	100	112	122	136	186	220	228	230	234
180	176	162	146	134	122	122	120	128	162	186	208	220	222	234	238
192	186	196	182	162	150	144	152	168	192	206	220	218	224	230	234
198	204	200	198	186	182	184	186	196	214	226	230	230	234	232	234
182	194	208	202	204	204	212	212	212	218	224	224	228	218	226	240

FIGURE 2.3: LSB watermarking sample of embedding and extraction of the watermark [91]

Spread Spectrum (SS) Method

This technique consists of embedding the watermark by linearly combining the host image with a modulated pseudo noise signal, in other words, each bit of watermark is spread over the cover image, then, according to the visibility requirement, the spread bits are then modulated with a cryptographically secure pseudo noise signal and added to the image pixels. Then, using the means of high-pass filtering and a correlation-based method, the watermark can be retrieved. Fig.2.4 shows a simple example. This method can be used for establishing secure communications, increasing resistance to attacks and jamming, and to prevent detection [65, 16]. However, to achieve the security requirement, the watermark length needs to be quite large, though, detection complexity is increased. Moreover, one can extract the watermark and reconstruct the cover image [69].

Simplicity and the very low computational complexities are major strengths of spatial domain methods. However, these techniques are not efficient when exposing them to common signal processing operations such as filtering or lossy compression [139, 60]. In the other side, an attacker could effectively destroy the hidden information by simply randomize all LSB. Accordingly, more robust watermarks are needed, which is approved in the transform domain.

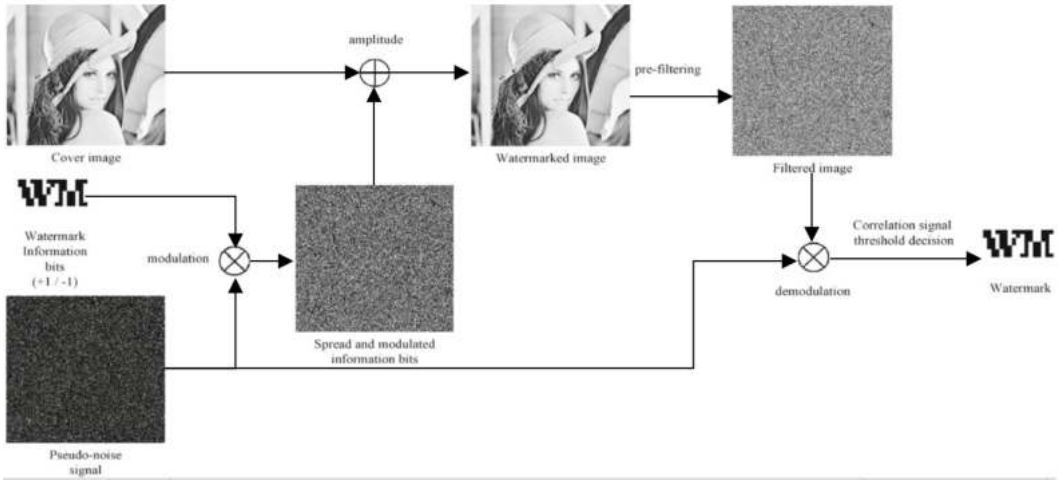


FIGURE 2.4: SS watermarking sample of embedding and extraction of the watermark

2.4.2 Frequency Domain

Frequency or transform domain approaches demonstrate its effectiveness by having some advantages since the frequency domain can well characterize most of the signal processing operations. These techniques take the advantage of the low sensitivity of the human visual system (HVS) to high and middle-frequency information to embed the watermark, these frequencies can be modified or even removed without significantly affecting the original image quality. In these techniques, the cover image is first transformed to the frequency domain by the use of any transformation methods, then, values of certain frequencies are altered to embed the watermark then, the inverse transform is applied (See Fig.2.5) [63, 153, 2].

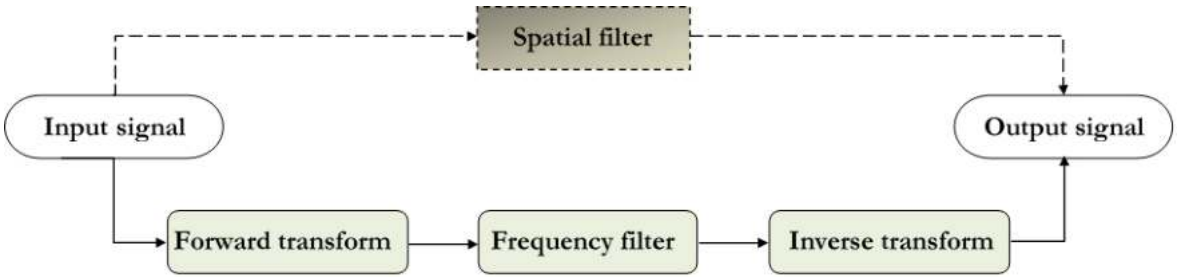


FIGURE 2.5: Basic concept of watermarking techniques: Dashed way refers to spatial domain, continues way refers to frequency domain

The fact of embedding data through the frequency coefficients, making it difficult to destroy and tolerant against a variety of signal processing manipulations types. To do that, several reversible transforms are used like Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Integer Wavelet Transform (IWT), Discrete Fourier Transform (DFT), or Singular Value Decomposition (SVD). Each of these transformations differs from the other with its characteristics that represent the image differently [91, 126, 62, 153].

Discrete Cosine Transform (DCT) Method

Compared to spatial domain techniques, DCT based watermarking techniques are robust against simple image processing operations like brightness, low pass filtering, blurring and contrast adjustment,etc. In fact, embedding a watermark into the cover image spectrum

does not directly influence the selected image quality. Characterized with its strong energy compaction property, *DCT* makes a spectral analysis of the signal and orders the spectral regions from high to low energy concentrating the high energy components into fewer transform coefficients as it is shown in Fig.2.6 [91, 68, 62].

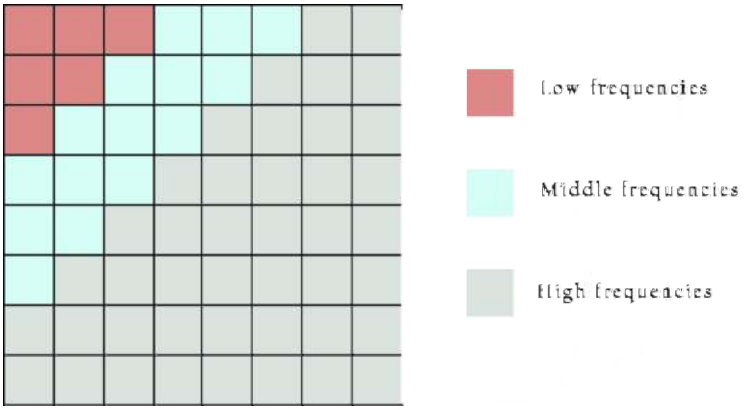


FIGURE 2.6: Spectral regions order from high to low energy concentration

The *DCT* formulates the selected set of data points as a sum of cosine functions oscillating at different frequencies (See Eq.2.1).

$$F(u,v) = \frac{2}{\sqrt{N \times M}} c(u)c(v) \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} f(i,j) \cos\left[\frac{\pi(2i+1)}{2N}u\right] \cos\left[\frac{\pi(2j+1)}{2M}v\right]$$
$$0 \leq u \leq N-1, 0 \leq v \leq M-1 \qquad (2.1)$$
$$c(u), c(v) = \begin{cases} \sqrt{\frac{1}{\sqrt{2}}}, & \text{if } u, v = 0. \\ 1, & \text{otherwise.} \end{cases}$$

$$f(i,j) = \frac{2}{\sqrt{N \times M}} \sum_{u=0}^{N-1} \sum_{v=0}^{M-1} F(u,v)c(u)c(v) \cos\left[\frac{\pi(2i+1)}{2N}u\right] \cos\left[\frac{\pi(2j+1)}{2M}v\right]$$
$$c(u), c(v) = \begin{cases} \sqrt{\frac{1}{\sqrt{2}}}, & \text{if } u, v = 0. \\ 1, & \text{otherwise.} \end{cases} \qquad (2.2)$$

Simulating human's way of perceiving the light to identify then modify parts that are not perceived, both *DFT* and *DCT* represent data in terms of frequency space rather than an amplitude space. Nevertheless, while *DFT* requires the use of complex numbers and both Sins and Cosines functions, *DCT* needs only real coefficients and Cosine functions. Thought, *DCT* is useful in a wide range of applications in science and engineering. However, *DCT*-based algorithms are difficult to implement that requires a higher computational cost. Moreover, they are vulnerable to geometric attacks like rotation, scaling, cropping ... [65, 120, 142, 126, 131].

Two approaches of *DCT*-based watermarking namely global and blocks watermarking. When *DCT* is applied to all parts of the image, this approach is the global one. While *DCT* block-based watermarking is applied to blocks, and the transform is applied to each block separately (See algorithm 2) [91, 62].

Algorithm 2 DCT Block Based Watermarking

1: **Input: Cover image, Watermark image**
2: **Output: Watermarked image**
3: Segment the cover image into non-overlapping blocks (e.g., each block of 8×8 pixels).
4: Calculate the forward DCT for each non-overlapping blocks.
5: Apply a block selection criterion (e.g. HVS).
6: Apply a coefficient selection criterion (e.g., the middle or high-frequency AC coefficients).
7: Embed the watermark by altering the selected coefficients.
8: Apply inverse DCT transform on each block (See Eq.2.2).

Discrete Fourier Transform (DFT) Method

DFT is considered as an important image processing tool in the field of watermarking because it controls the frequency of a host signal and decomposes it into its sine and cosine components which results generally complex values. The DFT can provide the selection of the adequate image parts for embedding the watermark with the highest invisibility and robustness, this feature makes DFT suitable for watermarking applications [62, 68]. Like DCT, DFT-based watermarking schemes use the magnitude of its coefficients to embed the watermark. Thus, middle frequencies are the best location to embed the watermark since modification to the low-frequency coefficients can damage the visibility while high frequencies coefficients are removable by $JPEG$ compression [122, 144]. For a square image of size $N \times N$, the DFT and $IDFT$ are given by equations 2.3 and 2.4 respectively. To do so, fast Fourier transform (FFT) is the algorithm used to compute the (DFT) and it's inverse (See Algorithm 3) [62].

$$F(u,v) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} f(i,j)e^{-i2\pi(\frac{ui}{N}+\frac{vj}{N})} \qquad (2.3)$$
$$0 \leq u \leq N-1, 0 \leq v \leq N-1$$

$$f(i,j) = \frac{1}{N^2}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} F(u,v)e^{-i2\pi(\frac{ui}{N}+\frac{vj}{N})} \qquad (2.4)$$

Algorithm 3 FFT Block Based Watermarking

1: **Input: Cover image, Watermark image**
2: **Output: Watermarked image**
3: Apply the FFT for host and watermark images to find out the FFT coefficients.
4: Define the value of the embedding factor to control the invisibility of watermarking.
5: Embed the watermark modifying the selected FFT coefficient of the host image.
6: Inverse FFT coefficients give the watermarked image.

Unlike the spatial domain, DWT, and DCT techniques, DFT is more robust against geometric manipulations like cropping. Furthermore, DFT is a rotation, scaling, and translation (RST) invariant, that can provide a recovery tool from geometric distortions. However, its implementation is complex and the computational cost is also higher [9, 131, 120, 142, 126].

Discrete Wavelet Transform (DWT) **Method**

Wavelet transform is a modern technique that is being widely studied due to their applications in the area of signal processing including the simulation of wireless antenna distribution, compression, removal of noise in audio, and watermarking... The interesting features of the DWT are the ability of image compression with resistant watermarks may be produced. Moreover, the possibility to select among different types of filter[6] banks, tuning for the desired bandwidth.

The basic idea of the DWT is to perform a more precise reflection of the HVS anisotropic properties using wavelet (i.e. small wave) of varying frequency and limited duration. DWT performs the (Eq.2.5) firstly for all image rows and then for all columns, decomposing the image into three spatial directions namely, LH (horizontal details), HL (vertical details) and HH (diagonal details) this latter represents the high frequencies with the finer scale wavelet coefficients, while the most energy in the image is placed in LL (low-frequency) components. For each decomposition level, the LL sub-band of the previous level is used as an input and the same process is repeated until completely decompose the image, as illustrated in Fig.2.7. Where, the coefficients magnitude is larger in the LL bands and smaller in HH, LH, and HL bands [142].

$$\begin{aligned} a_{j+1}(p) &= \sum_{u=-\infty}^{+\infty} l(n-2p)a_j(n) \\ d_{j+1}(p) &= \sum_{u=-\infty}^{+\infty} h(n-2p)a_j(n) \end{aligned} \tag{2.5}$$

Where elements a_j are used for next level of decomposition, d_j are wavelet coefficients, determine the output of the transform. $l(n)$ and $h(n)$ are coefficients of low and high-pass filters respectively.

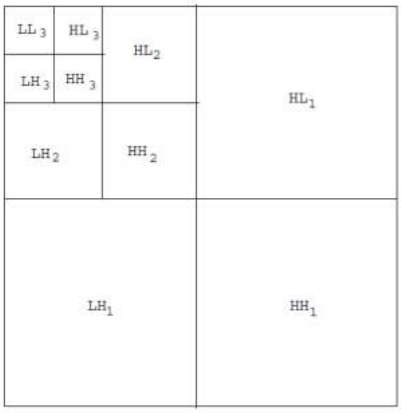


FIGURE 2.7: DWT 3-Level decomposition

Actually, most of the existing signals are time-varying, which makes DWT wavelets more appropriate for the analysis of transient and time-varying signals since DWT wavelets have their energy concentrated in time [65].

[6]For DWT image watermarking, there are many available filters, though, the frequently used filters are Haar Wavelet Filter, Daubechies Orthogonal Filters, and Daubechies Bi-Orthogonal Filters.

In a DWT-based watermarking selected coefficients of the cover image are substituted by the watermark's data (See algorithm 4) after decomposing the cover image into a high-frequency band that describes the edge components and a low-frequency band that is decomposed again into high and low-frequency bands (See Fig.2.7). Embedding the watermark into middle and high frequencies, since they are less sensitive to (HVS), is typically less robust against image distortions like lossy compression, low-pass filtering, and small geometric deformations of the image [91]. On the other hand, embedding the watermark into low frequencies is more robust to above-mentioned manipulations but less robust to histogram modifications such as gamma correction, contrast or brightness adjustment, and cropping [91]. Thus, watermark embedding depends on the compromise between robustness and invisibility. Changing the low-frequency (LL) image bands will damage the cover image and subsequently, the fidelity propriety is affected. However, when modifying spectrum region, robustness against compressions is achieved. While altering the middle and high-frequency (LH and HL) bands, increases the robustness against noise and several types of filtering. Therefore, an adaptation for the human visual system should be considered to avoid perceptibility of smaller cover image modifications [91, 54].

Algorithm 4 DWT Based Watermarking

1: **Input: Cover image, Watermark image**
2: **Output: Watermarked image**
3: Apply the two-dimensional DWT, to obtain the first level decomposition of the cover image.
4: Apply a coefficient selection criterion (e.g., the middle or high-frequency or HVS).
5: Embed the watermark modifying the selected DWT coefficients (e.g. LL band coefficients).
6: Apply inverse DWT transform on each block.

DWT and DCT-based watermarking schemes follow the same procedures (i.e. the same underlying concept) the only difference is the transformation process that resulting different coefficients which makes DWT much preferred, furthermore, it provides several features including the low linear complexity compared with the computational cost provided by Fourier and cosine transforms [68, 153, 9] moreover, the advantage of a wavelet transform is to allow a dual analysis taking into account both the frequency and spatial domains. Another feature of wavelet transform is that the HVS is more closely processed than the DCT [142, 65, 91, 9, 120, 153, 68, 62, 131].

Integer Wavelet Transform (IWT) Method

Compared to other signals, image characterized by the great data amount and the integer type of its pixel values. Thus, existing watermarking schemes that use the traditional wavelet transform assume that inputs are of a floating type. The conversion from integer to float and float to integer causes the loose of the perfect reconstruction property and lead to a rounding error. In fact, DCT, DFT, and DWT are simple but have the important drawback of using non-integer filter coefficients, which is why it produces non-integer transform coefficients [139]. Thus, these transforms are not more appropriate for image due to the disadvantages of computational complexity and rounding error [150, 7].

In order to avoid rounding errors, a lifting scheme[7] is used to implement the integer wavelet transform (IWT). This transform is reversible, i.e., the image can be fully reconstructed from the integer transform coefficients [139]. Integer wavelet transform (IWT) can

[7] A method for computing the integer wavelet transform in place, so no extra memory is required [139], it is designed based on matrix algebra theory and phase filter bank [174].

transform the pixel value to integer without any rounding error [167] and can quickly finish the watermarking algorithm because it uses only integers. Moreover, it can be implemented using three main steps, namely, split, predict and update that are explained as follows:

- Split: the original signal c_k^j is divided into even e_k^{j-1} and odd sequences o_k^{j-1} as follows [150, 7, 139]:

$$\begin{aligned} e_k^{j-1} &= c_{2k}^j \\ o_k^{j-1} &= c_{2k+1}^j \end{aligned} \tag{2.6}$$

- Predict: according to the data correlation, the even sequence e_k^{j-1} and a prediction operator P are used to predict the odd sequence, and the difference between the odd sequence o_k^{j-1} and the prediction results $P(e_k^{j-1})$ that is used as the high-frequency coefficient of next IWT-level [150, 7, 139]:

$$O_k^{j-1} = o_k^{j-1} - P(e_k^{j-1}) \tag{2.7}$$

  Where O_k^{j-1} is the predicted difference, P is the prediction operator.

- Update: to keep the same feature of even sequence e_k^{j-1}, O_k^{j-1} is calculated using the updating operator U_p, and then update the original even sequence.

$$E_k^{j-1} = e_k^{j-1} + U_p(O_k^{j-1}) \tag{2.8}$$

  Where E_k^{j-1} is the updated even sequence.

After the IWT, the even sequence is considered as the low-frequency component, while the odd sequence represents the high-frequency, then following the same steps, the low-frequency component can furthermore more transformed [167]. By replacing the split block by merge operation, the inverse operation can be performed [7].

Singular Value Decomposition (SVD) Method

From the viewpoint of linear algebra, a digital image is considered as a non-negative integer matrix. Which allows performing SVD on digital images directly [140].

Singular value decomposition (SVD) is seen as a data reduction method to transform correlated values into a set of non-correlated ones [157]. Thus, this numerical multi-variable technique can express the input data as three sub-matrices [120]. For a given matrix A, the singular value decomposition is represented by:

$$A = U \times S \times V^T \tag{2.9}$$

Where U and V are unitary matrices, and S is a diagonal matrix and the superscript T denotes matrix transposition. The diagonal elements of S are called the singular values of A known as eigen values and these are assumed to be arranged in decreasing order. The columns of U are called the left singular vectors while the columns of V are called the right singular vectors of A.

For image processing applications, SVD has two interesting properties: firstly, the efficient representation of the intrinsic algebraic properties of an image, where singular values represent the image luminance while singular vectors represent geometry image characteristics. Secondly, the good stability of singular values which does not significantly change when a small perturbation is added to the image [55, 140].

These properties prove that SVD can be used as a fundamental mathematics tool to develop watermarking schemes because of its translation and scaling properties, largest modifications of singular value, which represent the energy of the signal, does not meaningfully affect the image, which why they are used to embed the watermark. Moreover, SVD provides more accuracy and less memory space [157].

Despite, the robustness provided by the SVD properties in watermarking systems, they can not outperform robustness against different attacks provided by the frequency-based methods [160]. Thus, to improve the robustness of SVD-based schemes, another transform is used beside the SVD [111]. In addition, SVD-based watermarking schemes are generally vulnerable to false positive attack, an attacker can produce his fake watermark using the singular vectors of any watermark and he can claim the ownership of the watermarked image. Thus, a technique for correcting this weakness should be considered when using these methods[8] [153].

Cholesky Decomposition

The Cholesky decomposition is considered an important tool in matrix computation because it offers several properties that make it useful for image application. Those properties include the destruction of the relation between each input parameter and their proceeding ones and the creation of the new correlation between the output parameters; this means that the same information used to determine the value of one parameter is also used to partially conclude the value of the other. Moreover, the Cholesky decomposition can indicate the correlation degree of its variables. Furthermore, compared with the other entries in a row, the size of the diagonal entries indicates the independence degree in a correlated variable [41].

This sort of result and properties can be potentially useful when considering the value of future research since this may give an indication of the degree to which variables are related. The Cholesky decomposition of a positive definite matrix A is a decomposition of the form:

$$A = L \times L^T \tag{2.10}$$

In which L is upper triangular with positive diagonal elements and L^T denotes the conjugate transpose of L, if A is positive definite[9]; the decomposition is unique and there is only one lower triangular matrix L with strictly positive diagonal entries [58, 41].

2.5 Spatial Domain vs Frequency Domain

The need for protecting shared images is growing quickly. Several digital watermarking techniques are designed for different applications. Spatial and frequency domain are two main techniques[10] for protecting security and privacy. Owner identification, proof of ownership and several other watermarking applications require the robustness grants by the use of transforms. In contrast, authentication and tamper-proofing application require the fragility as a main property for such use of watermarking. Each one of these two techniques has its own advantages and inconveniences (See Table.2.1), thus favoring one domain is not

[8] In [165] and [153] several algorithms were proposed to overcome this limitation.

[9] In linear algebra, a positive definite $N \times N$ matrix M is a symmetric real matrix, where the scalar $z^T Mz$ is strictly positive for every non-zero column vector z of n real numbers. In which z^T denotes the transpose of z [3].

[10] Noticing that recent researches suggest the combination of spatial and frequency domains in order to embed more watermark data while reducing the distortion effects on the watermarked image.

possible, a compromise between several requirements that are defined by the watermark application should be considered.

Spatial domain watermarking is interesting because it offers a better intuition on how to attain an optimal tradeoff between robustness, capacity, and imperceptibility, furthermore, spatial domain algorithms can resist a wide range of manipulations (See Table.2.1). The main advantage of such methods is the easy use, less complexity and also the less time. As a result, most of these methods are applicable in real-time applications such as electronics, control systems engineering, and statistics [16, 153, 9]. A detailed investigation of watermarking techniques in the spatial domain is given in Table.2.2, where *LSB* and *SS* methods are compared.

However, the most serious problem of spatial domains is the weakness of robustness, they can hardly survive under attacks such as lossy compression and low-pass filter. On the other hand, compared with spatial domain methods, frequency domain watermarking methods are relatively robust to noise, image processing, and compression. Unfortunately, too much data can not be embedded in the frequency domain because the quality of the host image will be distorted significantly [16, 153, 2]. A complete investigation for various transforms is discussed in Table.2.3. The basic idea of frequency domain techniques is explained, furthermore, the principle of the frequently used transformation in watermarking is illustrated.

This investigation also captures limitations and advantages of each transform. For example, *DCT*-based techniques are vulnerable to geometric attacks like rotation, scaling, cropping while *DFT*-based techniques are not.

TABLE 2.1: Comparison between spatial and frequency watermarking techniques according to several criteria

Criteria	Spatial domain	Frequency domain
Advantage	Simplicity	Robust against general image processing
Disadvantage	Fail in blind watermarking	Low resistance against some process as rotation, cropping and changing size
Computational complexity	**Low**	High
Computation time	**Low**	High
Robustness	Fragile	**More robust**
Capacity	**High**	Low
Perceptual quality	**High control**	Low control
Example of application	Authentication	Copy control

TABLE 2.2: Investigation of watermarking techniques in spatial domain

Basic idea	Method	Principe	Advantages	Disadvantages
Spatial domain-based approaches are designed to modify randomly selected pixels subsets of images. They directly embed the watermark in the cover image pixels.	LSB	This technique consists of replacing the LSB of the pixel values in the carrier image with the MSB of the pixel values of the watermark.	• Easy to implement. • Low computational complexity. • Does not generate serious distortion to cover image. • Resistance to geometric attacks such as removal of inner distance, scaling, rotation.	• Robustness restriction. • Vulnerable to noise, compression, low-pass filter, and cropping attack.
	SS	This technique consists of embed the watermark by linearly combining the cover image with a modulated pseudo noise signal.	• Establishment of secure communications. • Resistance to natural interference. • Relatively high security. • Prevent detection.	• Increased complexity. • Needs synchronization between transmitter and receiver. • Large watermark length.

Basic idea	Method	Principe	Advantages	Disadvantages
In these techniques, the cover image is first transformed to the frequency domain by the use of any transformation methods, then, values of certain frequencies are altered to embed the watermark.	*DCT*	*DCT* makes a spectral analysis of the signal and orders the spectral regions from high to low energy concentrating the high energy components into fewer transform coefficients.	• Low sensitivity of the human visual system. • Difficult to destroy. • Tolerant against signal processing manipulations like brightness, low pass filtering, blurring and contrast adjustment, etc.	• Difficult to implement. • Computationally more expensive. • They are vulnerable to geometric attacks like rotation, scaling, cropping, etc.
	DFT	The *DFT* can provide the selection of the adequate image parts for embedding the watermark into the coefficients of the middle frequency, with the highest invisibility and robustness.	• More robust against geometric manipulations like cropping. • *DFT* is *RST* invariant. • Useful in recovering tool from geometric distortions.	• Complex implementation. • Higher computational cost. • Certain higher frequency components tend to be suppressed during the quantization step.

DWT	*DWT* decomposes an image into low and high frequency components, where the most energy in the image is placed in low-frequency components, in which the watermark is embedded to increases the robustness.	• *HVS* is more closely processed. • Higher compression ratio. • Good localization both in time and spatial frequency domain.	• Higher computational complexity. • Longer compression time. • Noise/blur near edges of images or video frames.
IWT	The lifting scheme is used to provide the basic structure for the adaptive wavelets in order to remove rounding errors.	• Map the pixel value to integer without any rounding error. • Image can be fully reconstructed from the integer transform coefficients. • Quick and efficient implementation. • Low computational complexity. • Strong robustness and good invisibility.	• Longer compression time. • Noise/blur near edges of images or video frames.

SVD	The main idea is to embed the watermark into the singular values by applying the SVD onto the cover image.	• Efficient representation of the intrinsic algebraic properties of an image. • Good stability of singular values. • More accuracy and less memory space.	• Less robustness against different attacks. • Vulnerable to false positive attack.
Cholesky	The Cholesky decomposition destroys the relation between each input parameter and their proceeding ones and creates of the new correlation between the output parameters.	• The same information used to determine the value of one parameter is also used to partially conclude the value of the other. • Indication of the correlation degree of its variables. • The size of the diagonal entries indicates the independence degree in a correlated variable.	• Complex implementation. • Higher computational cost.

TABLE 2.3: Investigation of watermarking techniques in frequency domain

2.6 Conclusion

This chapter has presented a comprehensive literature review and the findings of the initial investigation in three parts; namely, digital watermarking and its application, digital watermarking classification, and digital watermarking techniques. Specifically:

- The suitability of digital watermarking has been justified for security applications. Cryptography and steganography have been studied and their general limitations are identified. None of those disciplines are found capable to individually address the rising security concerns, where digital watermarking is found promising to address the limitations and thereby to complement the security protection.
- Applications of digital watermarking are studied, further, security requirements for each application (i.e., robustness, visibility...) have been discussed.
- Classification of watermarking schemes depending on several criteria has been studied and an example of an application has been given.
- Techniques of watermarking are investigated in spatial and frequency domains, this would naturally make it difficult to choose a suitable watermarking technique for an application thus, a concise comparison between those techniques is illustrated.

Work in the area of watermarking is still in its beginning. However, it is expected that security will not become an issue. Rather, robustness, fidelity, and payload requirements are the key issues. But, an incomplete consideration of these requirement and evaluation criteria when designing a watermarking scheme can affect its efficiency. To address this problem, the need for developing a watermarking model, defining fundamental watermarking properties, defining attack models and evaluation measures, etc. have been pointed out in the next chapter.

Chapter 3

Image Authentication Watermarking Models

3.1 Introduction

Since digital media processing tools have seen a wide availability, their manipulations and reuse become easier and faster, as well as making unauthorized copies that make it difficult to establish the multimedia content authenticity. Those illegal modifications affect the multimedia value and cause economic loss especially with sensitive digital multimedia content like judicial evidence, confidential government documents, or other significant information.

Over multimedia types, digital images are considered as the most important type, as well it is the basis of videos. Digital images are used in wide range of applications including medical image archiving, pharmaceutical research, military target images, media recording of criminal events, accident scene capturing for insurance and forensic purposes, digital notaries documents, and quality control images, etc. Therefore, in order to avoid false judgments, the content protection of these images is becoming a problem in the study of information security [129, 91]. A simple example can better clarify this problem. Based on medical diagnostic images, a patient may eventually get better due to the adequate medical treatments. A possible false diagnosis can expose the patient life to danger, if the stored image is illegally manipulated, stored or compressed, in such a way a doctor can not detect these distortions. Thus, in such case, even slight modifications are not tolerated. On the other hand, many other applications tolerate some image processing operations for restoration, enhancement, or transmission purposes, at the same time, the image content should be protected from any significant changes [56].

The first study of authentication in cryptography was by Friedman [48, 49] where he discussed its application to computing an associated cryptographic signature to create a "trustworthy camera". Modification of even one bit (e.g., due to noise, compression, quantization, etc.) of one pixel of the whole image will no longer match the signature, as a result, even insignificant tampering can be detected. However, this signature is considered as a metadata that must be transmitted along with the image, in a header field of a particular image format. Nonetheless, removing this header field is easy by just coping the image to another file, the signature will be lost, and the image can no longer be authenticated. Thus, watermarking is the solution to this issue, by embedding the signature into the image, to ensure that the signature stays with the image. It also gives the possibility to detect any occurred tampering, since any changes made to the image will also affect the watermark [33, 91].

Digital watermarking techniques in their early days are designed for copyright protection purpose, based on the steganography or information hiding concept, a watermarking scheme is assumed to be secure if a watermark is invisible and could resist common processing tools [123]. It is designed to survive attacks and to be robust. In contrast, for content authentication and integrity verification, any modifications in an image should be detected

as well as localizing the changed areas, the embedded watermark should be fragile or semi-fragile to attacks [129]. Here, the watermark encodes information, that identifies the source or producer, required to determine that the content is authentic. It must be designed to be destroyed with any alteration creating a mismatch between the content and the watermark. If the extracted watermark properly matches the content, so it can be assured to authentic [107]. On the other hand, if the marked image is manipulated, the watermark extraction is not possible, thus, the image is not trustworthy.

To address the above-mentioned problem, this chapter starts with verifying and formulating the problem in the understanding of watermarking security and robustness. Considering the overall status of their consideration in watermarking literature, their requirements in the signal and image processing context are distinguished. Also, this chapter provides an analysis of individual watermarking schemes, to demonstrate the need for a dedicated security analysis paradigm for watermarking. General weaknesses and potential threats of the schemes are discussed.

3.2 A Formal Generic Watermarking Model

For an image content authentication watermarking scheme, not considering properly the required properties expose it to security flaws. Therefore, a systematic development of watermarking schemes is essential. The watermarking scheme must be expressed mathematically with operation determination. Where watermarking scheme objectives and properties are well identified with their explicit considerations.

A formal study of watermarking schemes has several benefits including providing a rigorous analysis of the required watermarking properties, the readiness for implementation, and avoiding the ambiguities and misconceptions. However, most of the existing developed watermarking schemes are rather informal, and usually ignores the application scenarios, which leads to a confusion in choosing a suitable scheme.

To address this problem a formal generic model is required, in which digital watermarking properties are well-defined, aiming at giving the general concept of watermarking and identifies its possible variants for different applications. In other words, determining the required inputs, outputs, and component functions of a watermarking scheme helps define the required properties and design criteria for different objectives, which helps well characterizing a watermarking scheme [116, 115].

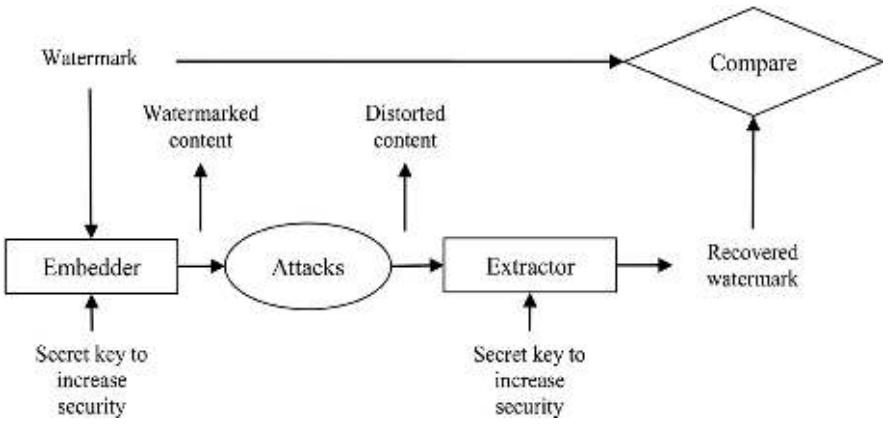


FIGURE 3.1: Fundamental components of digital image watermarking.

An authentication watermarking system is similar to any watermarking system, it should consist of three fundamental components (functions) as shown in Fig.3.1: watermark generation, $G(.)$ that generates a suitable watermark, embedding, $E(.)$ which embeds the authentication watermark in the host data, and detection, $D(.)$ which detects and extracts a watermark from a given image. The formal description of these functions in an image authentication watermarking application are described below [115, 154]:

3.2.1 Authentication Watermark Generation $G(.)$

According to the objectives of a watermarking application, this function generates an appropriate watermark that can be a message, m, a noise n, or other image data, etc. In an advanced application, watermark properties depend upon the watermarking objectives. Not well defining those properties may result in technical flaws and security issues. For an image application, the generation function, $G(.)$, can take as input the image data, I, and message m and/or other image data, and outputs a watermark W [115].

We should notice that the watermark can be generated from the host image itself and embedded into it, in what we call it self-embedding scheme, this method is firstly proposed by Fridrich and Goljan [46] to protect the image content. The main idea of this method is to compress the image then embed it into its pixels [94].

3.2.2 Authentication Watermark Embedding $E(.)$

Considering where and how the watermark is embedded to satisfy the different requirements of the host image (We will discuss these requirements in detail in Section 3.3). The embedding algorithm $E(.)$ takes as inputs the host image (I), the embedding security key (K), and the watermark (W), and generates a watermarked image (I_w), the outputs are the security key (K_{pc}) and the watermarked image (I_w). The watermarked image is then stored or transmitted via the communication channel. If any changes whether malicious or not occur, then the digital content is said to be attacked[1]. Thus, a special attention has to be paid to the kind of attacks in Section 3.4 as they can help to develop better watermarking techniques [120, 115, 129, 153].

In order to combine a watermark W with a host image (I), a watermark (W) that contains the watermarking information, a security key (K_pr) and an encoding algorithm (E) are needed to create a watermarked image (I_w) (See Algorithm 5). The private key (K_pr) controls the generation of watermark sequence or selects the location for embedding. α is a tuning parameter determining the strength of the watermark to ensure the invisibility. It can be a constant or a function proposed by HVS. The watermark is considered to be robust if it is embedded in such a way that the watermark can survive even with severe distortions [153, 94].

Algorithm 5 Authentication Watermark Embedding

1: **Input: Cover image** (I)**, Watermark image** (W)**, Security key** (K_{pr})**, Invisibility parameter** α
2: **Output: Watermarked image** (I_w)**,**
3: $I_w = E(I, W, K_{pr}, \alpha)$

[1]This term is mainly used when a third person uses common signal processing operations or some intentional attacks to remove the watermark.

3.2.3 Authentication Watermark Extraction and Verification $D(.)$

To declare whether the content is authentic or not, a watermark extractor $D(.)$ involves a two-step process. Watermark extraction is the first step that applies extracting algorithm $D(.)$ over the possibly attacked image $I_w^{'}$ to detect the presence or extract the watermark from it, if the watermarked image I_w does not undergo any change during the transmission, then the watermark can be easily detected and extracted. In the second step, the detected and extracted watermark W_e from a suspected watermarked image $I_w^{'}$ is analyzed and compared to decide if the image is authentic or not [153].

Some algorithms with detection and restoration abilities can localize tampered area and recover them [120, 94].

The general description of watermark extraction and verification of authentication watermark is shown in Algorithm 6. The internal design of the watermark extraction $D(.)$ can vary, but it generally takes as inputs the questionable watermarked image data $I_w^{'}$, the public key K_{pc} corresponding to K_{pr} while original image I and watermark W are optional[2]. If the correspondence similarity (more details discussed in Section 3.4) between the extracted and original watermark is less than a predefined threshold, the modification of marked image is acceptable and the image's content is authentic, or the marked image is unauthentic [153].

Algorithm 6 Authentication Watermark Extraction

1: **Input: questionable marked image** $(I_w^{'})$**, Security key** (K_{pc})**, Invisibility parameter** α
2: **Output: Extracted watermark** (W_e)**, Decision**
3: $W_e = E(I_w^{'}, K_{pc}, \alpha)$

3.3 Watermarking Requirements

Before presenting and discussing authentication methods, it is important to outline the criteria of a watermarking system.

According to the given application, the watermarking system requires specific properties; hence, selecting one set of properties satisfied by all watermarking systems is not possible. Several papers have described the characteristics of watermarks [138, 142, 56, 30, 123, 126, 134, 94], robustness, fidelity, computational cost, tamper resistance, and false positive rate are the most important properties. However, in practice designing a watermarking system that satisfies at all of these properties is practically impossible. Thus, it is necessary to make tradeoffs between them, and those tradeoffs must be chosen with careful analysis of the application. In addition, the application can affect the very definition of a property. In the following subsections, we look at each of the five properties listed above and discuss how its importance and definition varies with the application [33, 68, 153].

3.3.1 Security

Describes whether the embedded watermark can not be forged or manipulated without damaging the host image by targeted attacks with an assumption of knowing the embedding and the extraction algorithms except the key, also, at least one watermarked data is known. The basis of watermarking security is similar to the encryption security, it should lie on Kirchhoff's assumption where the encryption method is known to the unauthorized

[2]Using host image I and the original watermark W during extraction process depends on the watermarking types (i.e. blind or non-blind watermarking)

party. Meaning that the watermarking security must depend on the choice of the embedding key [142]. Such systems must be able to protect the authentication data against any falsification attempts, the watermarking key is private and should be difficult to deduce from the extraction information, moreover, inserting a watermark by unauthorized parties should be also difficult. Thus, improving the algorithm security requires enlarging the embedded space, increasing the size of the keys, and split it into small parts of cover image [153, 94, 142, 56, 145]. Thus, without knowing embedding algorithms, unauthorized parties can neither access nor remove the watermark [9].

3.3.2 Robustness (Tolerance)

Once the watermark is embedded in the original content, necessarily, it undergoes several distortions when it is encoded, decoded, and distributed through the Internet. These distortions may or may not significantly interrupt the watermarked signals. Accordingly, robustness [3] refers to the capability of detecting the inserted watermark after attacks and processing operations. In other words, the ability of a watermark to survive simple processing and to resist normal processing [9].

In many applications, robustness to all possible processing is excessive and unnecessary. For example, in television and radio broadcast monitoring, a watermark should survive analog transmission, lossy compression, and some small amount of vertical and horizontal translation. While it is needless to survive scaling, rotation, high-pass filtering, ... [33, 68, 142, 134].

In some cases, robustness may be completely irrelevant, or even undesirable. For example, covert communication doesn't need robust watermarks at all, especially if the host media is diffused digitally without compression. While, an authentication watermark that indicates whether an image has been altered or not, should be fragile. On the other hand, in applications where the signal processing between embedding and extraction is unpredictable, the watermark robustness is required to every possible distortion. This is the case for ownership proof and identification, copy control, and fingerprinting [33, 68, 142, 134].

Noting that it is impossible for a watermarking system to be robust against all signal processing operations whereas the requirement is application subordinate and dependent. For the digital watermarking of images, an efficient watermarking system is designed to resist against noise addition, filtering processing, and geometrical transformations such as scaling, translation, and rotation, and also $JPEG$ compression [153].

3.3.3 Tamper Resistance

It refers to the resistance of a watermarking system to hostile attacks. Depending on the application, certain types of attacks are more important than others. In fact, there are several applications in which the watermark has no hostile attacks, and tamper resistance is irrelevant [33, 68].

3.3.4 Invisibility (Fidelity or Imperceptibility)

The watermarked image must be perceptually identical to the original one under normal observation and the quality of marked image should not be degraded [68, 94, 142, 134]. Thus, the watermark should not be noticeable by human eyes nor should the watermark degrade the watermarked image quality. Moreover, only authorized person could access or detect it using specific operations. Generally, such watermarks are used for content or author authentication and for detecting illegal copies of the data [33, 9].

[3] This property is valid just for algorithms that provide a selective authentication service [134, 56, 145].

There are two main reasons to keep the watermark imperceptibility. Firstly, the primary purpose is to kept the presence or absence of a watermark indistinguishable from the host media, thus, if the watermarked media is distorted that its value is lost. In addition, the knowledge information about the watermark, its existence, and its precise location may expose it to illegal use such as substituting, distorting or removing the watermark [153].

3.3.5 Capacity (Data Payload)

Capacity is known as the number of data can be embedded [142, 9]. Under the condition of imperceptibility as well as the requirements of robustness, the capacity relies on the size of the original data. For images, the capacity refers to the number of embedded bits into host image pixels. It simply means that how much amount of information, could be inserted in the image [9].

Thus, obtaining a higher capacity is usually at the expense of either imperceptibility or, robustness strength or both [142]. In a word, the fewer bits number included in a watermark, the less false positives, and the larger computational complexity [153].

We should notice that the conditions of robustness, imperceptibility, and capacity are conflicted and limited by each other. Increasing the watermarking robustness decreases the watermark perceptibility. On the other hand, to respect the condition of imperceptibility, a watermark would be inserted with the minimum possible of modifications to avoid the watermark detection. Similarly, increasing the data payload by increasing the number of samples allocated to each hidden bit, counterbalanced by a loss of robustness. Accordingly, it is impossible to satisfy all these three requirements simultaneously for any watermarking scheme. As a result, a good trade-off among these requirements has to be achieved[4] [153, 142].

3.3.6 Computational Cost

This requirement depends on the method used for watermarking, meaning that if the watermarking method is more complicated, then it contains a complex algorithm that implies the use of more software and hardware, which by the way increases the computational cost and vice versus [9].

However, each watermarking applications require different speeds of the embedders and detectors. For example, in broadcast monitoring, both embedders and detectors request at least real-time response, to avoid slowing down the media production schedule, while keeping up the real-time broadcasts. On the other hand, in an ownership proof application, since it could find a watermark, the speed of the detector is not important even if it takes days. Furthermore, different applications need different numbers of embedders and detectors. Broadcast monitoring typically involves a few embedders and perhaps several hundred detectors. Conversely, copy control applications may require a handful of embedders and millions of detectors. In general, for a given application, the more device needed, the less cost [33].

3.3.7 Computational Complexity

It is defined as the amount of time taken by a watermarking algorithm to insert and extract a watermark. Thus, for the robust protection and validity of the watermark requires the computational difficulty. On the other hand, real-time applications require both speed and

[4]Many literature researches have been done to find a compromise between these three requirements [67, 119] where authors consider the texture, luminance, corner, and edge information of an image to create a mask and decrease the invisible effect of embedding the watermark [129].

efficiency. Thus, an authentication system must use neither slow nor complex real-time algorithms [145, 56, 9].

3.3.8 False Positive Rate

A false positive rate refers to the detection of a watermark in a host image that does not actually contain that watermark, that could happen in a precondition amount of runs of the detector [33, 68, 153].

For a watermarking-based authentication application more requirements that are essential for any authentication system are needed. Thus, an effective system must satisfy in addition the following requirements [145]:

3.3.9 Sensitivity (Detect Tampering)

This is requirement is considered as the main property to reliably test image's authenticity, an authentication system must be able to detect any tampering in a marked image [94, 134, 145, 56].

3.3.10 Localization of Altered Area (Identification of Manipulated Area)

The authentication system must be able to locate precisely any illegal alteration made to the image and verify the authenticity of other areas and also estimate the kind of occurred modification [94, 134, 145, 56].

Many authentication methods based on watermarking have the ability to identify corrupted areas of an image and verifying that the remainder of the image has been unchanged. Thus, localization is valuable because the knowledge of where an image has been modified can be used to deduce :1) the tampering motive, 2) possible candidate adversaries and 3) whether the alteration is legitimate [30].

There are two closely related approaches to localization. The first, block-wise authentication, splits an image into contiguous blocks and inserts an authentication watermark into each block independently. The second, sample-wise authentication, is an extreme case of block-wise authentication in which each block is reduced to the size of one sample [30].

3.3.11 Accuracy

Watermark detection is accurate if the false negatives, false positives, and the non-detection of the damaged image are reduced [68].

3.3.12 Reconstruction of Altered Regions (Recovery)

It refers to the ability of partially or completely restore the image areas that were altered or destroyed in order to know what was the original content of the manipulated areas [134, 56, 145].

There are two main restoration strategies, exact and approximate restorations. An exact restoration aims at restoring the image to its original state to a perfect copy rejecting even a single bit in error. While the recent concept of approximate restoration restores an image and accept non-significant differences between the restored and original ones. The approximate restoration is further divided into two approaches. In the first, additional information is inserted in the image to use it in the restoration. In the latter, to determine how the image has been modified, an investigation of the watermark is used to invert the distortion [30].

In addition, some technical features must be taken into account:

- Portability: The authentication system must be able to carry the signature with the protected image during any transmission, storage or processing operation [56, 145].
- Extraction mode: describes whether an authentication data is dependent or not on the image, a full-blind or a semi-blind extraction mode is required. Thus, a non-blind extraction mode is irrelevant for an authentication service since the original image is required [134].
- Asymmetrical algorithm: unlike copyright protection, an authentication service involves an asymmetrical watermarking (or encryption) algorithm (i.e., only the author of an image can secure it, but any user must be able to check the content of an image) [134].
- Invertibility: refers to the possibility of producing the original data during the watermark extraction [115, 142].

3.4 Attacks Against Watermarking System and Evaluation

A watermarked object is likely to be subjected to certain manipulative processes before it reaches the receiver. The availability of a wide range of image processing software made it possible to perform attacks on the robustness of the watermarking systems. In the watermarking context, an attack can be roughly defined as any processing that impairs or misleads the watermark detector, aiming at preventing the watermark from performing its intended purpose [65, 115].

An adversary that makes such attempts can be of different capabilities (e.g., can have different inputs or can access the watermarking functions). In practice, it is quite reasonable to assume the capabilities of expected adversaries in modeling attacks. For example, an adversary knowing nothing may assume an image is watermarked and may want to remove the watermark by applying a distortion attack. Having access to the embedding function, an adversary can also find and exploit the weakness of the detection function in applying different attacks. Thus, critical security issues arise with an adversary that have more capabilities [115].

We discuss various possible attacks on the watermarking security that can be mainly divided into two categories: active or intentional and passive or unintentional attacks that a watermarked image is exposed to. An active attack intends to alter the watermarking resources or to affect their operation, by removing the watermark or make it undetectable. This category of attack is a serious problem for many applications in which the purpose of the watermark is defeated when it can not be detected. Such applications include: the copy control, proof of ownership and owner identification, and fingerprinting. However, it is not critical for covert communication or authentication [33].

On the other hand, a passive attack is not attempting to remove the watermark, but it simply trying to determine whether a watermark is present, know, or misuse the watermarking information. In most applications, we are not concerned with this type of attack. In fact, advertising the watermark presence can serve as a deterrent. But in the case of covert communication, our primary interest is in preventing this type of attack [33, 153, 142, 65, 115, 138, 75].

In a first classification these attacks, either intentional or unintentional, can be divided into two main categories: signal processing and geometric distortion attacks. Table 2.2 shows some examples for each category [138, 59].

TABLE 3.1: First classification of attacks against watermarking systems [111]

Signal processing attacks	Geometric distortion attacks
Compression (JPEG-like)	Cropping
Color manipulation (intensity, gamma correction, component adjustments)	Rotate
Noise (adding noise, de-noise, replace bit-planes)	Scaling
Filtering (high pass, low pass, Gaussian, and sharpening)	Translation
Scanning	Remove column/row
Averaging	-

3.4.1 Passive (Unintentional) Attacks

In this category, the attacker is not attempting to remove the watermark but simply trying to determine its presence. In most applications, we are not concerned with this type of attack. In fact, aiming at preventing attacks, the presence of the watermark should be advertised. But in the case of covert communication, the primary interest is in preventing this type of attack [33] where the simple knowledge of the presence of the watermark is often more than one wants to grant [142].

3.4.2 Active (Intentional) Attacks

In this type of attacks, the adversary attempts intentionally to remove the watermark or make it undetectable. This is a critical issue for many applications including owner identification, proof of ownership, copy control, and fingerprinting, in which the purpose of the watermark is defeated when it can not be detected. Though, it is not a serious problem for covert communication or authentication [33, 142].

Active attacks are classified into four different types [5], such as geometric (desynchronization) attacks, removal and interference attacks, security (cryptographic) attacks, and protocol attacks as given in Figure 3.2 [153, 142, 65, 115, 138].

In what follows, the different attacks described below are defined depending on the available inputs to the adversary.

Removal and Interference Attacks

These attacks exploit the fact of adding the watermark in the host image as an additive noise signal. Thus, by estimating the watermark, removal attacks separate it out and remove it. On the other hand, interference attacks add additional noise to the watermarked image. These types of attacks intend to completely remove [6] the watermark without the watermark embedding key [142, 61, 129].

[5] Besides the aforementioned types, there is another class called estimation-based attacks, which estimates the watermark data or the original image using stochastic methods [33]. Depending on the estimation way this attack can be classified as desynchronization, removal, or protocol [112].

[6] The mentioned attacks don't succeed completely in removing the watermark from the watermarked image, but significantly affects the watermark [61].

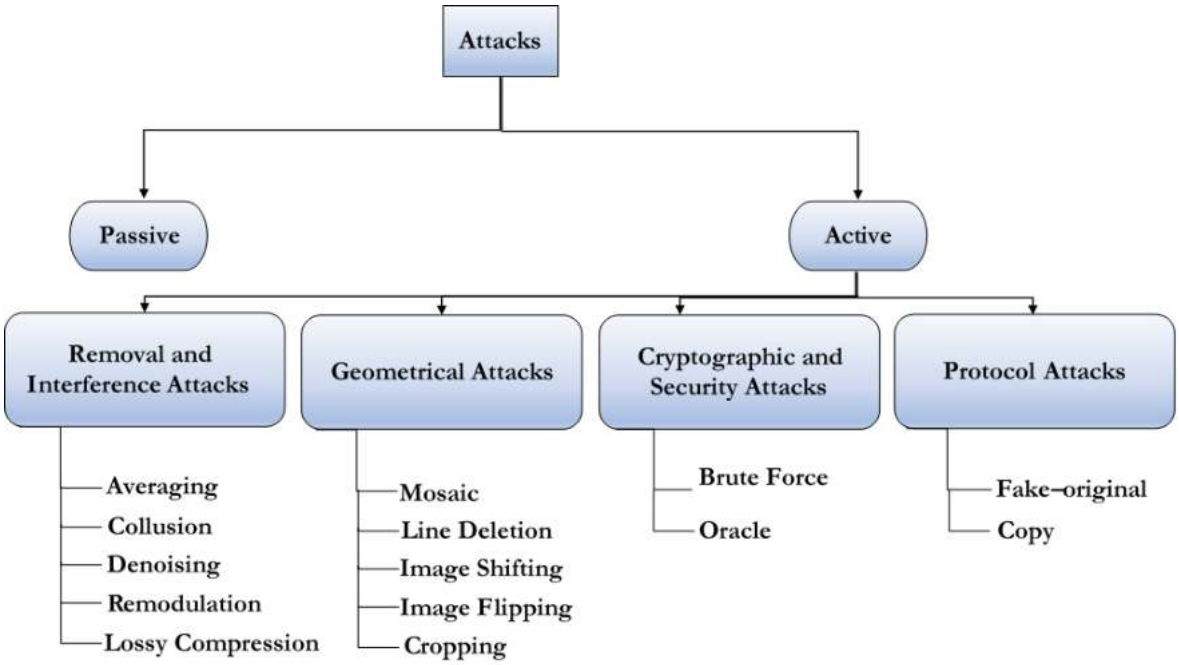


FIGURE 3.2: Second classification of attacks against watermarking systems

Some of the attacks included in this category are denoising, quantization (e.g. for compression), remodulation, collusion attacks, averaging, and noise storm [153].

1. **Averaging Attacks:** In averaging attack, the attacked data is evaluated using many samples of a precondition data-set log in each time with a different watermark or a different key. However, the watermark can not be retrieved if the amount of the inserted data is sufficiently huge, supposing that it will output zero on average [153].

2. **Collusion Attacks:** In this attack, an adversary uses several copies of the watermarked image each signed with a different watermark, to remove the watermark and construct an unwatermarked copy by averaging all the copies or a little part from each of the copies. Thus, the collusion attack occurs when numerous examples of the same data are attainable [7] [142, 65, 61, 33]. In an application such as fingerprinting, resistance to collusion attacks can be critical, in such application different watermarks are inserted in each copy of a media piece. However, this type of attacks is not widely spread because the number of required copies to obtain varies significantly from application to application [33, 142, 65, 61, 33].

3. **Denoising and Lossy Compression Attacks:** This category is relatively broad to common image processing operators such as image denoising [8], lossy compression, and quantization. Therefore, compression is a popular scheme for attacking watermarked images or audio, where attackers may compress the images to remove the watermark. Therefore, both denoising and lossy compression can importantly reduce the capacity of the watermarking [153].

[7] A number of authorized recipients of the image come together or many copies of the watermarked image (each signed with a key) are obtained by the attacker, which is an unlikely prospect [33]

[8] Image denoising, also known as filtering.

4. **Re-modulation Attacks:** The first systematic re-modulation of this attack was demonstrated in [77]. Where the watermark was estimated by subtracting from the host watermarked image to its median filtered version. The estimated watermark was also truncated, high pass filtered, then subtracted from the watermarked image [153].

Geometrical Attacks

Geometric attacks[9] are specific to images and videos. Compared to the removal attacks, geometric deformation attacks do not intend to remove the inserted watermark but distort the watermark which can render for all objectives and purposes of any watermarking application useless. For distortion of it, through temporal or spatial transformations of the watermarked contents. In other words, geometric attacks do not remove the watermark, rather than they manipulate the watermarked image in such a way that the detector can not detect the watermark data that loses synchronization with the inserted information [153, 142, 65, 61].

This type of attack includes the image processing manipulation such as flipping, cropping, scaling, rotation, translation, column/line removals, warping...[153, 142, 65, 61]. For example the effect of cropping leads to the blurring of the spectrum. So, no synchronization is needed when embedding the watermark in the magnitude, that are normalized coordinates [122], in the other hand, rotation in the spatial domain causes the same rotation in the frequency domain [144], resulting in cyclic shifts of extracted signal that can be detected by exhaustive search [122]. While scaling in the spatial domain causes inverse scaling in the frequency domain. Which results amplification of extracted signal that can be detected by a correlation coefficient. Furthermore, translation of an image has no result on extracted signal [126].

The recent trend in digital image watermarking can overcome this type of attacks using transform invariant domain like Fourier-Mellin transform or uses auto-covariance function (ACF) to design special watermarks [153, 142, 65, 61].

Line deletion, image shifting, and mosaic attacks are the main examples of geometric attacks:

1. **Line Deletion and Image Shifting:** To destroy the watermark, an attacker removes the entire pixels line or may change the watermarked image horizontally or vertically. When embedding watermarks in the DCT or the VQ domains, image shifting may result in desynchronization of the watermarked images [153].

2. **Mosaic Attack:** The mosaic attack is another example of the geometric attack that does not attempt to remove the watermark, but rather it aims to interrupt the watermark detection, splitting the image into small fragments. In a word, by dividing the watermarked image into several parts and rearrange it, a new watermarked image is constructed in which the watermark detector will fail to provide desired results [153, 65].

We should notice that the aforementioned categorization makes it potential to have an understandable separation between the different kinds of attacks, it is unavoidable to the reminder that a malicious adversary usually uses not only a single attack, but rather a combination of several attacks at the moment. Such a probability is estimated in the Stirmark benchmark, where all geometric transformations are practically accompanied by a lossy compression attack [153].

[9] These types of attacks are also called synchronization attacks or transformations attacks.

Cryptographic and Security Attacks

The aforementioned two categories of attacks, removal and geometric, do not breach the security of the watermarking algorithm. While, cryptographic attacks are quite equivalent to the attacks applied in cryptography, dealing with the cracking of the security. The main intention of this type of watermarking attack is to understand the security measure taken while embedding the watermark in the digital image and finding out a process to remove or insert another misleading watermark. In particular, the knowledge of the watermarking algorithm by an attacker provide him the capability of performing modifications that rend the watermark invalid or to estimate and change the watermark [142]. This kind of attacks is seriously forced attacks that use exhaustive searches to discover the secret information. Thus, it is critically significant to use keys with a safe length to resist this kind of attacks. The two main examples of cryptographic attacks are brute force and oracle attacks [153, 142, 61].

1. **Brute Force Attack:** One example of this attack is finding the secret watermarking key by exhaustive research [142, 61].

2. **Oracle Attack:** This attack creates a non-watermarked image when a public watermark detector is attainable. These attacks are similar to the attacks used in cryptography. However, these types of attacks have high computational complexity [153, 142, 65, 61].

Protocol (Ambiguity) Attacks

Protocol attack was introduced by Craver et al. [37]. This attack does neither aim at destroying the embedded watermark nor at disabling its detection. Rather than, it takes advantage of semantic deficits of watermark implementation to attack the definition of the watermarking applications. In a word, protocol attacks create ambiguity of the watermarked data by attacking the entire concept of the watermark scheme in reversible watermarking. Consequently, a robust watermark must not be invertible or to be copied [142].

The solution to protocol attacks is to use a one-way function to make the watermark signal dependent [61]. Accordingly, applications such as copyright protection require a non-invertible watermarking scheme. In fact, in an invertible watermarking scheme, an adversary can also include his watermark in the watermarked image, which can produce an ambiguity with respect to the authentic ownership of the contents [153, 65].

Some examples of such attacks are IBM attack and copy Attack [65].

1. **IBM Attack:** This attack[10] re-watermark the watermarked image to make the original owner watermark indistinguishable. In other words, in this attack, an adversary produces a fake watermarked image to discredit the authority of the watermark by embedding several additional watermarks, so that it is not clear which was the first confident watermark. In some deadlock attacks, a fake original image is created to produce the same results through the detector as that of the real original image [65, 129].

2. **Copy Attack:** Copy attack is another type of protocol attacks in which attacker estimate the watermark from the watermarked data and put it into some other digital data which is known as target data. In this attack, the attacker subtracts its own watermark from the watermarked image and claim the ownership thus creating ambiguity

[10] This attack is also known as the fake–original attack, inversion attack, or deadlock attack.

with respect to the true ownership of data [61]. The copy attacks can be term as successful if and only if it a valid watermark can be retrieved from the target without the knowledge of the watermarking key [61, 142].

In authentication systems, the common objective of malicious attacks is not to eliminate the watermarks, but to invalidate them and trick the authentication system, in other words, to show that an image as authentic even though its content has been modified (or sometimes, the opposite). For example, an attacker may bring a known valid watermark from a marked image as the watermark for another. Consequently, the detector announces it as authentic. This type of attack can also be performed on the same image: the mark is first removed, then the image is modified, and finally, the mark is embedded again [94].

Some of these attacks look insignificant and easy to avoid; nevertheless, take them into consideration when developing an authentication algorithm is a very important [94]. Our aim in this section is to show some of the most frequent attacks that an image authentication system can endure.

- One of the most frequent attacks against fragile watermarking systems is the alteration of the protected image without changing the embedded watermark, or even more common, trying to produce a new watermark that will be considered as authentic. For example, in a fragile watermarking that ensures the integrity of an image independently of its content, the watermark is embedded in the LSB of its pixels. Manipulating the image without considering the affected bits, will also affect the watermark and therefore the attack will be detected. While, altering the image without modifying the LSB; the watermark will remain untouched, and the authentication process will not distinguish any falsification [94].

- In general, when the integrity of an image is based on an independent watermark of image content, it is possible to design an attack that could copy a valid watermark of one image into another image. By doing so the second image becomes protected even though the second image is false. This attack can even be performed over the same image by extracting the watermark from the image; then manipulate the image, and finally re-embed the watermark in the manipulated image.

- Following the same concept, Fridrich et al.[43] propose a collage attack to create a falsified image from parts of a group of images protected by the same authenticator using the same watermark and the same key. With no prior knowledge about the inserted binary watermark, or the secret key. Its idea is relatively easy since it replaces each pixel of the altered image by the closest pixel value of equal coordinates of the images in the base. The main difficulty of this attack is the achievement of an images database rich enough to produce a good visual quality falsified image.

- Another classic attack, well known by the security community, is brute force attack. This kind of attack attempts to determine the secret key used to generate the watermark. Once the secret key has been discovered, it is very easy to fake a watermark of the protected image. Thus, the only solution to counter this attack is using long keys that require a high computational cost to deter an adversary from trying to discover these keys [94].

- In [128] Radhakrishnan et al. propose an attack against the image authentication system $SARI$ [11] [86]. The authors show that the image digest of the $SARI$ system is not protected under certain conditions. Specifically, when the digest of multiple numbers

[11] Self-Authentication-and-Recovery Images is a demos and test software, it is a semi-fragile watermarking technique that can be founded at http://www.ctr.columbia.edu/sari [84].

of images is generated using the same secret key and the attacker own this image digests, he is able to cause arbitrary images to be authenticated. Consequently, to overcome this attack, the authors propose several countermeasures [94].

- In forgery attacks the hacker attempts to embed a new valid watermark rather than remove it. This is the main security concern in authentication applications, since, if hackers can embed valid authentication watermarks, they can cause the watermark detector to accept bogus or modified media. In addition, as pointed out by Craver et al [36], this type of attack is a serious concern in proof of ownership [33].

3.4.3 Performance Evaluation and Metrics of Watermarking Systems

Performance evaluation is a very important part when designing any watermarking system. Thus, validating any enhancements requires measures to do so. On the other hand, applying watermarking to an application need to identify the most appropriate system. Thus, a way of evaluating and comparing the different systems is required to find out how much a watermarking system is effective. These measures may also lead to ways of optimizing various properties [138, 30].

Before we can evaluate a watermarking system, we need to have some idea of what makes one system better than another, or what level of performance would be best. If we are interested in using a watermark for some specific application, our evaluation criteria must depend on that application. For example, if we are evaluating an image watermark for use in copy control, we might want to test its robustness to small rotations. However, such robustness might be irrelevant for broadcast monitoring, because rotations are unlikely to occur during normal broadcasting, and we might not be concerned with security against active attacks [30].

Furthermore, the performance of a watermarking algorithm against attacks reflects its quality. However, the performance against attacks and satisfying specific characteristics are two mutually conflicting requirements. Moreover, if we are interested in testing the merit of a new watermarking system in comparison to existing systems, we have more flexibility in choosing our test criteria [30]. Many watermarking algorithms and metrics have been developed for their comparison so that we can decide to use adequate algorithms. Unfortunately, several algorithms use their own designed evaluation systems, which restricts the comparison of each other [153, 65].

In general, image watermarking systems should be tested on a large number of images drawn from a distribution similar to that expected in the application. For example, we would not necessarily expect an algorithm that was tuned to natural scene images to be ideally suited for X-ray images, satellite photos, or animation frames. If a system is being tested without considering a specific application, the image is tested on should be representative of a typical range of applications [30].

Therefore, imperceptibility and robustness are the main properties that are evaluated for any watermarking scheme. Two groups of tests are needed; the first is to evaluate the quality of the original image that should not be affected by the presence of the watermark, while the second group of tests measures the correctness of the extracted watermark by evaluating the impact of different attacks (intentional or unintentional) upon the watermarked image [138, 140].

Imperceptibility Evaluation of Watermarked Image

Inserting a watermark into a cover image implies the occurrence of distortions on that image. To quantify the perceptual similarity several measures are used including: mean square

error (MSE), signal to noise ratio (SNR), weighted or peak SNR ($WPSNR$ or $PSNR$), structural similarity index ($SSIM$), correlation quality (CQ), mean or weighted $SSIM$ ($MSSIM$ or $WSSIM$), normalized cross-correlation (NCC), ... However, currently, there is no approved globally and effective measures for visual quality. Furthermore, not all the measures give similar estimation [154]. Therefore, the perceptual similarity is defined using several measures $PSNR$ and $SSIM$ as an example.

In this section, the most important metrics in a watermarked system are explained, those metrics are used in different literature [57, 105, 105].

In these equations, I is the original image, $I^{'}$ is the watermarked or the recovered image, and N, M denote the height and width of the image respectively.

1. **Mean Square Error (MSE)**: This measurement estimates the average of the squares of the differences in the pixel values (errors), between host image and watermark image [138, 140], its equation is given by:

$$MSE = \frac{1}{3 \times N \times M} \times \sum_{i=1}^{3} \sum_{x=1}^{N} \sum_{y=1}^{M} \left[I_i(x,y) - I_i^{'}(x,y) \right]^2 \tag{3.1}$$

2. **Peak Signal to Noise Ratio (PSNR):** The $PSNR$ is used to evaluate the difference between the original image and the watermarked image with respect to the noise. PSNR is defined as in Equation 3.2. Generally, the value of PSNR should be greater than 30 dB, thus, higher PSNR values indicate the more similarity between the original and watermarked image [138, 25, 111].

$$PSNR = 10 \log \frac{255^2}{MSE} \tag{3.2}$$

3. **Weighted Peak Signal to Noise Ratio (WPSNR):** This metric uses a weighting factor called noise visibility function (NVF) in calculating PSNR which is defined by equation 3.3 [105], the NVF (Eq.3.4) is a texture masking function used to determine the amount of an image texture using the Gaussian model.

$$WPSNR = \frac{255^2}{NFV \times MSE} \tag{3.3}$$

$$NVF = NORM \frac{1}{1 + \delta^2 \times Block} \tag{3.4}$$

Where δ is the luminance variance for the 8×8 block. The $NORM$ function normalize the NVF values between 0 and 1. If image regions are textured or edge region, the NVF is closer to 0, otherwise, it is closer to 1 [111].

4. **Structural Similarity** $(SSIM)$ **Index:** This measurement gives a similarity measure between two images, it is designed to overcome inconsistent with human perception founded in traditional methods [25]. $SSIM$ takes a value between -1 and 1, where the value 1 indicates the total similarity of the two images [111].

$$SSIM(I, I') = \frac{(2\mu_I \times \mu_{I'} + c_1) \times (2 \times cov_{I'} + c_2)}{(\mu_I^2 + \mu_{I'}^2 + c_1) \times (\sigma_I^2 + \sigma_{I'}^2 + c_2)} \tag{3.5}$$

Where μ_I and $\mu_{I'}$ are the averages of I and I', respectively; σ_I^2 and $\sigma_{I'}^2$ are the variances of I and I', respectively; $cov_{I'}$ is the covariance of I'; to stabilize the division

with weak denominator the two variables $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$ are used; $L = 2^{number of bits per pixel} - 1$ the dynamic range of the pixel values [25, 111].

5. **Image Fidelity (IF):** The IF determines the similarity between the watermarked and original image. The higher the IF, the less visibility of the watermark [59].

$$IF = 1 - \frac{\sum_{x=1}^{N}\sum_{y=1}^{M}\left[I(x,y) - I^{'}(x,y)\right]}{\sum_{x=1}^{N}\sum_{y=1}^{M} I(x,y)^2} \tag{3.6}$$

6. **Correlation** *Corr***:** The *Corr* can be used to measure the compatibility between two sequences S_c and S_a. It can take a value in the range of -1 to 1, with -1 representing a direct, negative correlation, 0 representing no correlation, and 1 representing a direct, positive correlation.

$$Corr(S_c, S_a) = \frac{\sum[(S_c(i) - \overline{S_c}) \times (S_a(i) - \overline{S_a})]}{\sum\sqrt{[(S_c(i) - \overline{S_c}) \times (S_a(i) - \overline{S_a})]^2}} \tag{3.7}$$

Where $\overline{S_c}$ and $\overline{S_a}$ are the mean of sequences S_c and S_a respectively.

Robustness Evaluation of Extracted Watermark

The following metrics can be used for binary sequences to measure the consistency between the inserted and the extracted watermarks. In these equations, $W(i,j)$ and $W^{'}(i,j)$ represent the original and the extracted watermark respectively.

1. **Correlation Coefficient (CRC):** The *CRC* can be used to measure the compatibility between the original and extracted watermark. It can take a value in the range of 0 and 1 [59].

$$CRC = \frac{\sum_{i=1}^{3}\sum_{x=1}^{N}\sum_{y=1}^{M}\left[W_i(x,y) \times W_i^{'}(x,y)\right]}{\sqrt{\sum_{i=1}^{3}\sum_{x=1}^{N}\sum_{y=1}^{M}\left[W_i(x,y) \times W_i^{'}(x,y)\right]^2}} \tag{3.8}$$

2. **Bit Error Rate (BER):** The *BER* is a useful metric for a binary sequence watermark. If the value of BER is zero, the presence of the watermark is approved; while, if it is one, it means the absence of the watermark [126, 59]:

$$BER = \frac{FP}{NB} \tag{3.9}$$

Where FP is the number of incorrectly decoded bits and NB is the total number of original watermark bits.

3. **Normalized Correlation (NC):** Generally, the *NC* can take a value in the range of 0 and 1. If the NC value is closer to 1, the tow images are getting more similar.

$$NC(W, W^{'}) = \frac{\sum_{i=1}^{3}\sum_{x=1}^{N}\sum_{y=1}^{M}[W(x,y) \times W^{'}(x,y)]}{\sqrt{\sum_{i=1}^{3}\sum_{x=1}^{N}\sum_{y=1}^{M}[W(x,y)]^2} \times \sqrt{\sum_{i=1}^{3}\sum_{x=1}^{N}\sum_{y=1}^{M}[W^{'}(x,y)]^2}} \tag{3.10}$$

Tamper Detection Accuracy and Content Restoration

To evaluate the detection accuracy performance several metrics are used such as tamper detection rate R_{TD} and false alarm rate R_{FA} defined in Eq.(3.11) and Eq.(3.12) are calculated with different tampering ratio R_T that is calculated using Eq.(3.13).

$$R_{TD} = \frac{num_d}{num_m} \times 100\% \tag{3.11}$$

$$R_{FA} = \frac{num_{fd}}{(N \times M \times 3) - (num_m \times n \times m)} \times 100\% \tag{3.12}$$

$$R_T = \frac{num_m \times n \times m}{(N \times M \times 3)} \times 100\% \tag{3.13}$$

Where num_m is the number of actually altered blocks, and num_d is the number of altered blocks which are detected, num_{fd} is the number of false detected pixels, $n \times m$ is the block size. Lower R_{FA} values mean that the tamper detection accuracy is much better [79].

3.5 Current State of Watermarking Models

Image authentication systems can be categorized in several ways, they can be divided according to the data authentication storage mode into watermark or external signature, and also according to whether they ensure strict integrity or content (selective) authentication [134].

Strict authentication is devised for applications that don't tolerate alterations in the protected image, those methods are further divided into conventional cryptography and fragile watermarking techniques. In contrast, other applications tolerate some image processing operations, thus selective authentication is used especially in these cases [109]. Selective authentication methods can be further divided into digital signature-based algorithms, semi-fragile watermarking, and telltale watermarking (See Figure 3.3) [56, 159].

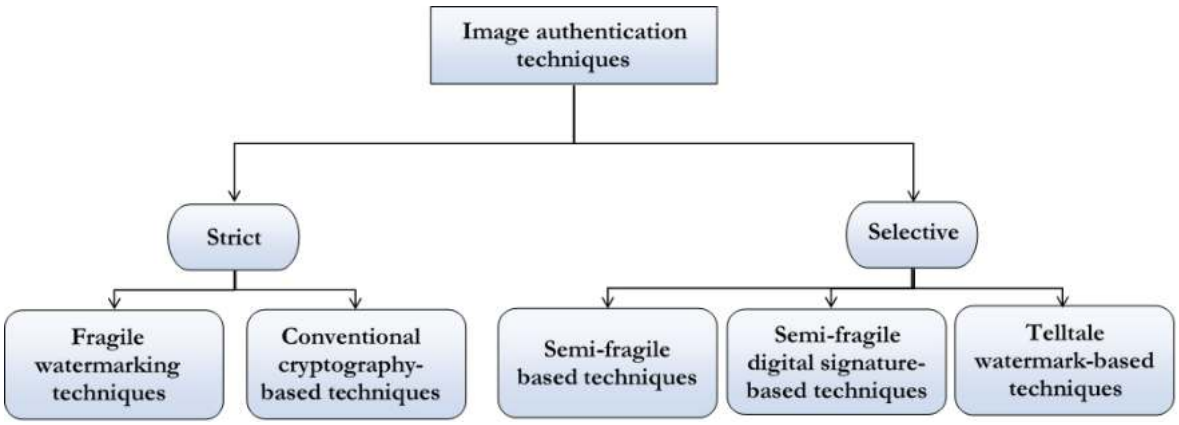


FIGURE 3.3: Image authentication techniques classification

3.5.1 Strict (Exact) Authentication Image Techniques

In many circumstances, alterations to content serve legitimate purposes. However, in other cases, these modifications may be intentionally malicious or may inadvertently influence the image interpretation. There are several applications that need such service, where we must be certain that an image has not been altered, there is a need for verification or authentication

of the integrity of the content. For example, an unintentional modification to an X-ray image might result in a misdiagnosis, while maliciously alter photographic evidence in a criminal trial can result in either a wrong conviction or acquittal [30].

Thus, in such applications, the main task is to verify that an image has not been altered at all since it left a trusted party. If only one pixel or even one bit of the protected image has been changed, that image is considered as unauthentic [56]. By insisting on a perfect copy, we avoid any need to design algorithms that can differentiate between acceptable and unacceptable alterations. Each image is just a collection of equally important bits [30].

According to the used techniques, strict image authentication methods can be further divided into two main categories: conventional cryptography-based methods and fragile watermarking-based methods [30, 134]. The first involves embedding cryptographic signatures, which become invalid when the image is modified. While the second involves generating watermarks specifically designed to become undetectable when the image is modified, such watermarks are sensitive to both malicious and incidental attacks [30].

Conventional Cryptography-based Image Authentication Techniques

Conventional cryptography was developed to solve the problem of message authentication and had great progress since its appearance. It also used to verify the authenticity of a digital image using a signature system [88]. Cryptography-based image authentication methods use hash functions to calculate a message authentication code (MAC) from images (See paragraph 5.4.2) [101, 135, 137]. Then, the resulting hash is encrypted using the sender's secret private key and then appended to the image [162]. At the receiver end, the received image is hashed, then, the appended hash is extracted and decrypted using the public key. Finally, the calculated hash and the extracted one are then compared [56, 88].

Digital signatures-based authentication is quite established and still be used. However, the transmission of this signatures requires an extra bandwidth as meta-data or establishing a separate secure channel [56, 91]. Which exposing this meta-data to be lost, especially if the content undergoes a variety of format changes. Thus, watermarking can reduce the risk that the authentication signature is lost by embedding the signature within the cover image. This permits format conversions to occur without the risk of losing the authentication information. Thus, representing the authentication signature as a watermark may allow the system to work with legacy systems that would otherwise have to be redesigned in order to guarantee that the header information is preserved [30].

On the other hand, the watermark embedding alters the images, causing the subsequent authentication test to fail. To avoid this failure, the image is divided into two parts, a first part to be authenticated and another to be modified to contain the watermark. A simple example is to partition an image such that the LSB plane holds the authentication signature computed from the remaining bits of the image [30].

Algorithms based on conventional cryptography show satisfying results for strict image authentication with high tamper detection. Localization performances are not very good but may be acceptable for some applications [56].

Recently, many teams of researchers have published papers where they try to use hash functions that tolerate some desired manipulations such as compression, and histogram equalization. These methods, however, can not tolerate a combination of several allowed manipulations in the same image. Moreover, they are vulnerable to attacks against the hash functions [56].

Moreover, the localization of changes can be easily lost if more than one region of the image was corrupted. Wolfgang and Delp in [170] propose another approach to overcome this problem. That enables tamper localization. However, these techniques are not able to restore image regions that were tampered [56].

Fragile Watermarking-based Authentication Image Techniques

Until now, we have considered fragility undesirable, seeking instead to design robust watermarks that can survive many forms of distortion. However, fragility can be an advantage for authentication purposes.

A fragile watermark is simply a specific watermark so that any try to modify the content of an image will also modify the watermark itself. If a very fragile watermark is detected in an image, we can assume that the image has probably not been altered since the watermark was embedded [134, 30].

In a typical fragile watermarking-based scheme, the image to be protected is divided into two parts: the first one consists of the most significant perceptual information, and another part containing the less descriptive data, in which the watermark can be embedded without significantly altering the image content [19].

The insertion of check-sums into the LSB of the image was one of the first techniques that are used to detect the tampering in the image [134, 30]. Although, forgery is a simple matter of copying the LSBs from the authentic image to the tampered cover image. Thus, the fragility of a watermark provides only limited authentication capabilities. In other words, fragile watermarks indicate that an image has not been inadvertently modified, the use of predefined patterns can not guarantee that no one has intentionally tampered the image. This is because an adversary can easily falsify the fragile watermark if it is not dependent on the cover image [30].

Thus, strict authentication methods, including fragile watermarking and conventional cryptography methods, provide satisfying results, even though localization and reconstruction performances require more investigations to be enhanced [56, 91].

3.5.2 Selective Image Authentication Techniques:

Strict (exact) authentication is appropriate in many applications. However, in an image or an audio clip, a change of a couple bits rarely makes a difference of any importance. In fact, an image is compressed to save memory space; or enhanced and restored for better perceptual quality or even converted to other formats, with the express intention of making no change in its perceptual quality [56, 30, 91].

In many applications, the perceptual similarity between images suggests that the compressed version is authentic. Thus, strict authentication is not more appropriate in these cases. Which required developing tools that can perform selective authentication, in which only significant changes cause authentication to fail. Thus, selective authentication provides some kind of robustness against specific and desired manipulations that alter pixel values without changing the image content, using image content signatures or semi-fragile watermarking-based techniques [56, 30, 91].

The central issue here is deciding what types of transformations or distortions are significant enough to cause authentication to fail. We then discuss three basic approaches to building such systems. The first two approaches, semi-fragile watermarks, and semi-fragile signatures. The third approach, telltale watermarks, is potentially more interesting, as it points toward systems that can identify the specific transformations that an image has undergone, rather than simply saying whether or not the image has been significantly altered [30].

Semi-fragile Watermarks-based Image authentication Techniques:

A semi-fragile watermark describes a watermark that is unaffected by legitimate distortions but destroyed by illegal distortions, providing a selective authentication [129].

Creating a semi-fragile watermark is similar to creating a robust watermark, especially when the perceptibility is the main requirement to distinguish between legitimate and illegitimate distortion. The robustness should ensure that the watermark survives manipulations while the cover image doesn't lose its value. Beyond that point, surviving is not any more important [129].

Semi-fragile watermark in the other hand should survive manipulations up to that point where the image loses its value, and it should not survive beyond that point. This requirement is often achieved by carefully modifying robust watermarks to be destroyed at a defined distortions level [129]. Several proposed semi-fragile systems are examples of this approach [129, 30, 89, 132, 178].

Generally, the semi-fragility requirement is achieved by taking advantages of transformed coefficients properties of the image or the relationships among them. By quantizing or adjusting these coefficients the watermark is embedded in a way that content-preserving operations doesn't affect these properties while they are modified by malicious manipulations [129].

More difficulties arise when designing a semi-fragile watermark with a more specific list of legitimate distortions, as in the case for medical and legal applications, where the watermark should resist certain manipulations, and accept others, even with negligible perceptual distortions. The semi-fragile watermark should be designed with a specific list of legitimate distortions. Moreover, no optimal criteria for maintaining low false positive and false negative rates are currently in existence. Consequently, these two issues prevent the application of semi-fragile watermarking in legal and national security issues [129, 30].

Semi-fragile Digital Signatures-based Image Authentication Techniques:

Semi-fragile watermarks, like their fragile counterparts, are often not secure against malicious tampering such as copy attacks. In addition, semi-fragile watermarks are only able to authenticate those properties of an image they are embedded within. For example, embedding the authentication watermark in the high-frequency, while avoiding because the low-frequency coefficients, expose the watermark and the system to fail, especially when an adversary changes low frequency and leaves the high frequency untouched [30]. On the other hand, using secure hash functions for digital signatures are also susceptible to failed authentication due to the avalanche effect, caused by the modification of only one or more bits of an image due to noise, quantization, or compression [56, 91].

These problems can be addressed by identifying features of an image that are invariant to legitimate distortions, but not to illegitimate distortions, and using them to construct a signature [141, 148]. Because the signature is unaffected by legitimate distortions but changed by others, we refer to it as a semi-fragile signature [56, 91, 30].

Like the cryptographic signatures used for exact authentication (Section 3.5.1), a semi-fragile signature can be embedded as a watermark. However, in this case, the watermark can not be fragile, because it must be able to survive any legitimate distortion. A properly designed semi-fragile watermark might complement the signature. That is, although both the signature and the watermark are designed to survive legitimate distortions, they might be fragile against different sets of illegitimate distortions [30].

There are at least two advantages of using semi-fragile signatures. First, in such a system, each image has a different watermark embedded. This means that an adversary can not make a forgery appear authentic by performing a simple copy attack. Second, the signature can be based on properties of the image that we can not change without causing unacceptable fidelity problems [30].

Due to the urgent need of image content signatures in several applications, researches are now more focused in this area, perceiving an increasing number of proposed solutions.

Nevertheless, more sophisticated solutions that can resist to combinations of several modifications are still to be found. Results are satisfying, but the problem is far from being solved [56, 30].

Telltale Watermarks-based Image Authentication Techniques

The above-mentioned two semi-fragile approaches are appropriate for applications where distinction between legitimate and malicious distortions is clear. However, in some applications this distinction is not clear, might change over time, or differ from place to place. Thus, rather knowing if an image has been modified or not, it is desirable to know how it has been modified. One possible solution is to investigate how a known, inserted watermark has been corrupted. This type of watermarks are called telltale watermark [30].

In [178, 74] use quantization watermarking system to embed the watermark in the wavelet domain. Then, at the end of discovering interrupted bits in each band various subsets of the wavelet decomposition are examined. Each distortion affects different sub-bands of wavelet coefficients, from this view point an assumption on how the image has been modified [30]. For example, changing low-frequency coefficients while high-frequency coefficients are untouched indicate the application of a high-pass filter to the image.

Telltale watermarks can differentiate between a wide variety of different distortions. However, research in this area is still in its beginning [30].

Challenges for Selective Image Authentication Techniques

Despite the wide interest the digital watermarking had seen, it still a tough issue to design an effective authentication watermark scheme especially for sensitive application such as medical image system and law enforcement. Moreover, satisfying requirements of detection accuracy, tamper localization, and good watermark imperceptibility, without requiring explicit knowledge of the original image is tougher.

A first design challenge for selective image authentication watermarking scheme is the unavailability of the original image for authentication verification. In practical applications, the original image generally has a much larger magnitude than the tolerable legitimate distortions of the channel. Thus, with the unavailability of the original image makes it hard to distinguish legitimate distortions from illegitimate ones [159].

Therefore, the real difficulty of these schemes is the achievement of a good assessment between robustness against minor image distortions and fragility to malicious attacks. This problem is related to the problem of image semantic content definition and the ambiguity of which changes must be tolerated and which are not tolerated. In other words, the authenticator should detect only perceptible or misinterpretation alterations. Consequently, developing an appropriate selective image authentication scheme requires the distinction between manipulations that change the image content and those that preserve it [133]. For many distortions, the correct classification might seem obvious. However, technically it is not easy to realize this distinction since it could change with images, applications and even within a single image [56]. For example, high-quality lossy compression should probably be legitimate, in that it has essentially no perceptible effect. On the other hand, substantial editing should probably be illegitimate, in that it can completely change the interpretation of the image. Other distortions, however, are not as obvious.

Even though, many literature try to propose innovative requirements and features to describe the image content and detect content modifications. Accordingly, a basic rule to classify these distortions is to consider the interpretation of the image when they are used as intended. Any distortions that will not change those conclusions should be legitimate. Table 3.2 categorize several image processing operations into operations that preserve image

TABLE 3.2: Classification of image Manipulations according to the image content preservation.

Manipulations that preserve the image content	Manipulations that change the image content
Transmission error	Addition of objects to the image
Transmission noise	Position change of objects in the image
Storage error	Change in image characteristics (texture, edges, colors...)
Quantization and compression	Change in image background (daytime...)
Geometrical transformations (rotation, scaling...)	Change of luminance conditions (shadows...)
Enhancement techniques (spatial and frequency filtering, histograms and gray level processing...)	cropping
Restoration techniques (de-noising, deconvolution...)	replacement
Image formats conversion	Deletion of objects from the image

content and manipulations that alter the image content [56, 40]. Thought, in a critical field such as medical imaging and criminal court cases, therefore, this classification should be based on controlled studies, rather than on subjective judgment [30].

Given such inconsistent requirements, an effective selective authentication system would first distinguish the distortions applied to an image into legitimate and illegitimate distortions. This is the aim of telltale watermarking, as discussed in Section 3.5.2. Otherwise, if system requirements are consistent and can be recognized a prior, so an alternative system is used where the watermark is only robust to the legitimate distortions which is the aim of semi-fragile watermarks and semi-fragile signatures [30], as discussed in Sections 3.5.2 and 3.5.2.

3.6 Summary of Different Watermarking Schemes

Designing an appropriate general model for authentication watermarking is a fundamental need. This section reviews different models of emerging image authentication techniques described in the literature and thoroughly considers a set of selected criteria to study them. Considering the category of watermarking, watermark type and its dependency on the host image, the authentication service type, and capabilities of localization and reconstruction.

We summarize the different methods in Tables.3.4 and 3.3 below. The category of each, as well as the type of authentication data used, the authentication data support, the objectives regarding integrity (i.e., strict or content), and whether the method offers a possible localization and/or reconstruction of the areas tampered with. Localization and reconstruction capabilities are reviewed to investigate the applicability of those schemes. Furthermore, the watermark dependency to the host image is also investigated to determine how the embedding distortion contributes both in detection and reconstruction of the altered area.

Authentication service type	Objective	Method	Watermark	Watermark dependency	Localization capabilities	Reconstruction capabilities
Strict	Conventional Cryptography	Rey and Dugelay [133]	Lines and columns hashes	Yes	With ambiguity	NO
		Xie et al. [175]	Block hashes	Yes	Yes	NO
		Friedman [48]	Hash of the image	Yes	NO	NO
	Fragile watermarking	Baldoza and Sieffert [8]	Check sum	Yes	Yes	NO
		Fridrich [45]	DCT	Yes	Yes	Yes
		Fridrich, Goljan, and Baldoza [47]	Predefined logo	NO	Yes	NO
		Byun et al. [20]	Color components	Yes	Yes	NO
		Chan and Chang [21]	Parity check data	Yes	Yes	Yes
		Lin, Yang, and Chang [83]	Patient data	NO	YES	NO
		Al-Qershi and Khoo [5]	Electronic patient report	NO	YES	YES
		Memon et al. [105]	LSBs of ROI	YES	YES	YES
		Memon and Gilani [104]	Patient's information, Hash of original image	YES/NO	YES	NO
		Zain and Clarke [180]	Hash of ROI	YES	NO	NO
		Liu [92]	Local mean of blocks	YES	YES	YES
		Lin, Chen, and Chiu [81]	Error correction coding of the image blocks	YES	YES	YES

TABLE 3.3: Summary of Strict watermarking techniques used in relevant studies

Authentication service type	Objective	Method	Watermark	Watermark dependency	Localization capabilities	Reconstruction capabilities
Selective	Semi-fragile watermarking	Tirkel et al. [158]	m-sequences	NO	With some ambiguity	NO
		Wolfgang and Delp [169]	m-sequences	NO	With some ambiguity	NO
		Chen, Wang, and Zhou [24]	Random-noise	NO	Yes	NO
		Paquet [118]	Random-noise	NO	Yes	NO
		Kostopoulos, Skodras, and Christodou-lakis [71]	Luminance	Yes	Yes	Yes
		Kostopoulos, Gilani, and Skodras [70]	Luminance	Yes	Yes	Yes
		El-Din and Moniri [42]	Block similarity	Yes	Yes	NO
		Lee, Jang, and Yoo [78]	Predefined mark	Yes	Yes	NO
		Lin, Liu, and Lin [80]	Generated from host image	Yes	Yes	NO
		Ajala Funmilola, Ojebamigbe Victoria, and Adegoke Benjamin [4]	Patient data	NO	Yes	Yes
	Telltale watermarking	Kundur and Hatzinakos [73]	DWT	NO	Yes	NO
		Kundur and Hatzinakos [72]	DWT	NO	Yes	Yes
		Kundur and Hatzinakos [74]	DWT	NO	Yes	NO
		Abdel-Aziz and Chouinard [1]	DWT	-	-	-

TABLE 3.4: Summary of Selective watermarking techniques used in relevant studies

By analyzing the above table (Table.3.4), we can conclude that fragile watermarking and digital signature systems can generally provide only strict authentication. However, compared to signature systems, watermarking schemes provides several advantages at the expense of modifying the host image to embed the watermark directly [103]. Thus, the authenticator does not require additional data for verification. Unlike digital signatures which need to be appended as meta-data or establishing a separate secure transmission channel, fragile watermarking imperceptibly embeds the authentication data to be more difficult to remove than a digital signature. Also, digital signature systems do not exploit the image characteristics and consider it as an arbitrary bit stream. Which why digital signature systems can detect the modified image but can not identify the alterations. Watermarking systems can recover modified areas of a watermarked image and which areas have not, as well as estimate the nature of modification [88]. However, we also should notice that only few schemes are currently able to reconstruct, even partially, the tampered areas of an image [134].

We can also conclude from Table.3.4 that many literature try to propose innovative requirements and features to describe the image content and detect content modifications. Thought, in a critical field such as medical imaging and criminal court cases, therefore, this classification should be based on controlled studies, rather than on subjective judgment [30].

Given such inconsistent requirements, an effective selective authentication system would first distinguish the distortions applied to an image into legitimate and illegitimate distortions. This is the aim of telltale watermarking. Otherwise, if system requirements are consistent and can be recognized a prior, so an alternative system is used where the watermark is only robust to the legitimate distortions which is the aim of semi-fragile watermarks and semi-fragile signatures-based schemes [30].

It can be inferred from Table.3.4 that almost all models do well on detecting modified areas and they offer satisfactory localization capabilities while restoration performances still need to be improved.

3.7 Analyze of Future Watermarking Directions

Watermarking technology is still in the progress and its future seems bright. The watermarking technology becomes increasingly important since it attracts more interest of businesses who wish to expose their digital products (images, music, movies, and books...) on the Internet. However, multimedia copies of illegally distributed content over the Internet could reach billions of dollars a year. With such high risks, entertainment and other multimedia companies are involved to keep pushing for a secure technology that they willing to pay for, to track and reduce copyright violation and avoid their loses. Many companies have already been active in this research field, consequently, lots of developments and improvements have been made in the last decades which make it interesting to discover how these developments and adoption plays out. Microsoft is one of these companies that has developed a prototype system that embeds a watermark in audio files to limit unauthorized playback. In future versions, such technology could be useful for Windows operating system as a default playback mechanism. If music industries begin to embed watermarks in its audio files, Windows would refuse to play illegally obtained copyrighted music released after an indicated date. As well, Microsoft Research has used graph theory to invent a separate watermarking system to hide watermarks in software. In fact, security technology is vulnerable to attacks. However, combining the technology with proper legal enforcement, industry standards and respects of the privacy of the intellectual property, endorse digital watermarking users and content creators to trust the Internet more.

Despite the development and improvement in the field of digital image watermarking, designing a watermarking system that satisfying all requirements at the same time is not

possible. However, a number of different approaches and incremental research have been investigated to find a satisfactory compromise. A revolution is probably not imminent, but improvement in this area would lead to significantly more satisfactory systems. Nevertheless, the ability of a designed watermarking system to allows public watermark detection while preventing an attacker from eliminating the watermark remain an open question. Many researchers attempting to designing public secure watermarking systems, but all appear vulnerable to attack. Standardization, set of precise, and accurate requirements are the main prerequisites that the majority of watermarking systems lack. Also, the lack of agreement on the definition of a common benchmark for method comparison and on the definition of the performance related concept is the third aspect for this hindering. Several challenges should be overcome to retrieve the wanted aim, therefore, researcher's efforts have already been focused to achieve the aimed results. Therefore, some of the advances are in their beginning, and many interesting researches remain to be done. Finally, it is expected that this effort lead at the end to widely accept the digital image watermarking as legal evidence of ownership.

3.8 Conclusion

This chapter has reviewed the different watermarking theoretical aspects and their practical interpretations for several applications including image authentication application. Thereby, the formal generic watermarking model, the formal definitions of main watermarking requirements, and possible attacks have been investigated. More precisely: the state of the watermarking model is substantially reviewed with a concise comparison. Despite having the main requirements, few researchers define the necessary properties for their model, while others do not. Existing models have some common limitations including: 1) incomplete consideration of inputs, outputs, and functions, 2) lack of definitions for the watermarking properties, and 3) incomplete realization of the application scenarios. Moreover, the formal definition of a set of fundamental watermarking requirements (robustness, invisibility, tamper resistance...) has been presented. This chapter has presented a set of watermarking attacks giving their general context for image applications. Considering the different capabilities of adversary have been illustrated. This includes a number of active attacks (e.g., interference, removable, protocol attacks...) and passive attacks. Also, this chapter reviews the different literature image authentication models. Almost all models detect modified areas and they offer satisfactory localization capabilities, unfortunately, restoration performances still need to be improved. Finally, this chapter has summarized the future of watermarking as promising security technique and discussed facts and latest findings and their significance in this field and pointed out the necessary tasks to be followed up in the research presented in the next chapters of this thesis.

Addressing the identified gap in watermarking literature, in the following part, our contributions are explained. A set of fundamental watermarking properties and application scenarios are considered, for digital images. With the aid of some practical examples, the uses of the properties addressing a few hidden assumptions in current practice are also shown.

Part II

Contributions

Chapter 4

A Bayer Pattern-based Fragile Watermarking Scheme for Color Image Tamper Detection and Restoration

4.1 Introduction

Aiming at developing a novel authentication approach of digital watermarking, the research presented in the former chapters has led to a new self-authentication model. The target application background aims at achieving different security properties such as authentication and integrity verification of images.

As discussed in the chapter 3, the watermarking has the ability to attain this requirement. The presented self-authentication model has been designed to overcome the limitations of the existing models for attaining the required security properties of digital images.

As discussed in the previous chapter, the detection accuracy and restoration capabilities are the major considerations for the fragile watermarking based image authentication approaches. The majority of the existing schemes are able to detect tampered areas. Unfortunately, most of them do not provide the restoration ability [99, 22, 44, 93, 171, 85, 82]. Moreover, if those techniques can recover the tampered area, they do not perfectly recover the damaged data [21]. Moreover, most research efforts in the area of digital watermarking for image authentication are focused primarily on the gray-scale image. The color image watermarking methods for authentication were primarily developed by applying the gray image watermarking methods to color channels separately. However, the design of a gray-level image authentication scheme did not take advantage of the color images characteristics; it must be modified or even redesigned to achieve the color image authentication requirement [81, 92].

The proposed digital watermarking-based scheme thus can help overcome these limitations and can ensure the security properties of the standard cryptographic tools by embedding the watermark imperceptibly. Our scheme satisfies the requirements of self-embedding, tamper detection, and restoration abilities.

The practical development of an authentication scheme on the basis of the self-authentication model, however, entails several tasks, as pointed out in the previous chapter. These tasks include, for example, choosing the appropriate watermark encryption schemes, mapping function, finding suitable feature extraction, and watermark embedding schemes, etc.

The reason for considering this task and the challenges in pursuing it, to follow up this research are described below.

The method, analysis, and experiments are then briefly introduced that will be presented in the subsequent sections of this chapter.

4.2 Considerations Before Applying the Self-Authentication Model

One main challenging issue in developing a suitable authentication scheme is to find an efficient watermark embedding scheme. The choice of developing a novel approach, however, as distinct from the other existing mentioned above schemes has been made by considering the following factors.

1. **Finding an appropriate way of proving the authentication of a digital image without the use of extrinsic data or a second image:** So, there no need to use two types of inserted data (authentication and recovery data). Thus, the capacity of insertion is decreased. This request can be answered using a self-embedding scheme.

2. **Finding an appropriate watermark generation method:** Choosing the appropriate arrangement of watermark pixels to reduce the capacity of insertion. This can be readily accomplished by identifying a method to interpolate the color value of the pixels of the same color in the neighborhood. For example, a green pixel provides an exact measurement of the green component. While the red and blue components of this pixel are obtained from the neighbors. Thus, to achieve this requirement a *CFA* can be used.

3. **Finding a suitable mapping function:** A security level and computational complexity are required criteria for the watermark generation, insertion, and extraction schemes. This may, however, require verification of the overall watermarking performance. A mapping function can be readily chosen from the literature considering the properties of the key (e.g., length, type public or private,...). A *TA* based permutation can be used to avoid the possible weaknesses of a small key space and guaranty a security level.

4. **Finding a suitable extraction function:** This task includes a further study of:

    - The verification of the dependence on the input image (blind extraction), watermark regeneration, and computational complexity, etc.
    - The adaptation of the chosen extraction function for the self-authentication objectives.
    - Ensuring a security level, to avoid any properties that may cause security weaknesses.

Therefore, aiming at developing a suitable embedding scheme, several challenges are discussed below.

4.3 Challenges for Developing an Embedding Scheme

Applying the self-embedding scheme for image authentication seems to be an efficient and appropriate alternative. Although, the computational efficiency is required to compensate for the computational cost of the watermark generation scheme. Furthermore, the special properties of strict authentication requirements should be considered. As well, the compromise between the embedding distortion level, tamper localization, and recovered image quality, which represent a serious problem to the majority of current fragile watermarking schemes with restoration ability.

The high embedding capacity, low level of embedding distortion, tamper localization, and recovered image quality are the major challenges in achieving the efficiency and suitability of an embedding scheme, those challenges are specified in the section below.

1. A relatively high embedding capacity is required to provide large size watermarks that help achieve different objectives. But, usually the higher the required capacity, the higher the embedding distortion [127]. However, a lower level of embedding distortion is required to ensure the invisibility and the protection requirements of the watermark image.

2. On the other hand, the accuracy of tamper localization requirement implies increasing watermarking capacity to further improve tamper detection, which subsequently, decreases the level of embedding distortion.

3. The ability of tamper recovery, to restore the tampered regions with a good visual quality more bits to represent the effective information are used, thus the level of embedding distortion is subsequently affected [179].

In summary, to overcome the compromise between all the above requirements and develop an efficient and appropriate watermark embedding scheme for images authentication, the following requirement should be addressed simultaneously: i) the computational efficiency, ii) the high embedding capacity with minimum distortion, iii) the high detection accuracy, and iv) the high tamper recovery.

The rest of this chapter will deal with a developed fragile watermarking scheme for color image tamper detection and restoration and presents its development, computational analysis, and performance evaluation. This starts with verifying the achievement of abovementioned requirements. Then developing a spatial domain embedding scheme, that uses the least significant bit of the pixels (LSB), to preserve the good perceptual quality. This would further allow the minimization of the legal and ethical impact and achieve an image-content-independent capacity with low computational complexity.

To validate the suitability of the embedding scheme, analysis and experiments are conducted and the performance of the developed embedding scheme is compared with similar schemes [81, 92, 23].

4.4 Proposed Bayer Pattern-based Fragile Watermarking Scheme

In this section, a watermark generation process is developed that would effectively help consider the computational aspects in watermark embedding while keeping the distortion at lower levels. Additionally, a general approach for self-embedding is investigated in light of the strict authentication properties and requirements.

The main idea of our proposed method is to decrease the amount of the authentication and recovery data that modifies the host color image, which allows embedding several copies of a reduced watermark into the LSB bits of the host color image. Although, using the least significant bits of pixels is demonstrated for watermarking, not only to minimize the embedding distortion, but to also ensure continuous protection with much reduced computational complexity.

4.4.1 Features of the Proposed Scheme

The proposed scheme offers several features in addressing the contradictory embedding problems and limitations of existing schemes. The main features of the proposed scheme are described below.

1. **Self-embedding and high capacity:** Since our proposed method is a self-embedding-based scheme the color host image is reduced to a gray-scale image, then inserted as a watermark. To avoid the existing low capacity problems, our proposed scheme uses

the same data as authentication and recovery data. In other words, the authentication and recovery data are not separated and used as one to localize and restore the damaged regions which allows embedding fewer data and thereby less modified LSB bit. Moreover, this scheme selects the suitable number LSB of the pixels to keep the distortion in the embedding region at the lowest level.

2. **Reversible watermark generation process:** To more decrease the capacity of insertion, this scheme provides a watermark generation method using the Bayer color filter array CFA. Exploiting the correlation between neighboring pixel intensities to capture for each pixel a single color channel, green, red, or blue, the proposed method converts a color full image to a gray-scale image, then, by demosaicking the gray-scale image we obtain a color full image which is highly correlated (close to 1) with the original one (See Table.4.2).

3. **Flexible capacity ability:** The proposed scheme attempts to minimize this watermark by calculating the average of every $n \times m$ neighboring pixel, which introduces an adaptive capacity for increasing size payload. The selection of the $n \times m$ block size parameter must be made as a tradeoff, between the watermark invisibility and the recovered image quality, which means that the quality of the recovered image increases by decreasing $n \times m$ block size of the host image and the invisibility of the watermark decreases (See Table.4.2); since we use more LSB bits to embed the watermark and reversely. In other words, each averaged 8 bits value within the $n \times m$ non-overlapped block is embedded into $n \times m$ pixels of the host image. This would offer flexibility to increase the embedding capacity when required and to maintain a minimum distortion in the embedding region.

4. **Security enhancement and robustness:** As mentioned earlier, any particular security properties that a watermarking scheme requires, can be obtained by deploying a suitable cryptographic technique. The proposed scheme increases the security by permuting the payload using three different keys and inserting it in three different places in the host color image. Here, a watermark could be secure using TA permutation, which allows the watermark to remain embedded to provide both security and robustness enhancement.

5. **Localization accuracy and restoration ability:** In our scheme, the reduced watermark is inserted in three different places in the host color image. To decide if a block is tampered or not, the extracted values are compared; if no alterations arise, the three values are equal, however, if there is any alteration we use the majority voting technique to decide which value is altered and which ones are not. Then, at the restoration process, the extracted values considered as unchanged are used to reverse the reduction process and reconstruct a high correlated image with the original one, to recover tampered regions.

In support of the above features, the computational analysis and experimental results will be presented in Section 4.5 before that, the implementation and applicability of the proposed scheme. Thus, to be well understood the proposed scheme, its background is covered in the following section.

4.4.2 Preliminaries and Background of the Proposed Scheme

To build up the proposed method we shall introduce two main components: Bayer Color Filter Array (CFA) that is used to reduce the host color image and recover the damaged

areas, while Torus Automorphism (TA) permutation is used to permute the inserted data to improve the security.

Color Filter Array (CFA)

For each pixel location, a color image requires, at least, three colors components, the often used are red (R), green (G), and blue (B) components. However, this would be difficult and expensive to implement. Exploiting the correlation between neighboring pixel intensities to capture for each pixel a single color channel, green, red, or blue is called Color Filter Array (CFA) [130].

Fig. (4.1. (b)) shows the Bayer color filter array, the most frequently used pattern. Since the green channel elements contribute most to luminance signals of a color image, the density of the green components is twice than that of the red and blue color components providing the best visual quality of the image [121].

The process of interpolating or estimating the missing pixels in a CFA image to reconstruct the original true color image is called demosaicking. At each location, two missing colors must be interpolated from neighboring pixels to get a high fidelity color image. A variety of methods for this interpolation are available [121, 130, 172].

The Bayer color filter array is used at the insertion process to convert a color full image to a gray-scale image, then, by demosaicking the gray-scale image we obtain a color full image which is highly correlated (close to 1) with the original one.

We should notice that using luminance channel instead of using a color filter array doesn't allow us demosaicking the luminance component to a high correlated color full image. Thus, the pixels of obtained CFA image are rearranged so it allows estimating the missing pixels, unlike the luminance component pixels which are not correlated.

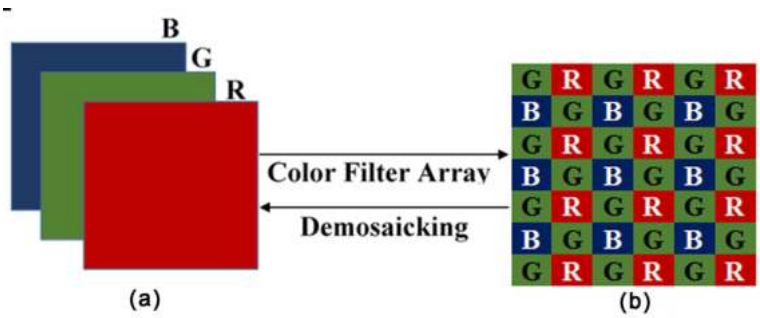


FIGURE 4.1: (a): R, G, and B components of color images, (b): Bayer color filter array pattern [121] used to reduce color images to gray-scale ones by capturing at each pixel location a single color channel, green, red, or blue

Pseudo Random Permutation (Torus Automorphism TA**)**

According to [163], Torus automorphism permutation is two dimensions pseudo-random function that assign the row and column index of a block to those of another. The Torus Automorphism expression is defined as follows:

$$\begin{pmatrix} i' \\ j' \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ k & k+1 \end{pmatrix} \begin{pmatrix} i \\ j \end{pmatrix} \mod S \tag{4.1}$$

Equation (4.1) indicates that each bit of the watermark at position (i, j) will be moved to a new position (i', j'), S is obtained from the size $P \times Q$ of the watermark, while k is randomly chosen by the user. For scrambling or descrambling the watermark the secret keys S and k

values are needed which ensures a cryptographic protection against malicious attacks, even though the proposed method is still blind.

The proposed method is composed of four processes: watermark generation process, embedding process, extraction process, and the detection and restoration process. The detailed steps of the proposed method are shown in Fig.(4.2), Fig.(4.4), and Fig.(4.5), and they will be described in the incoming processes:

4.4.3 Watermark Generation Process

Since our proposed method is a self-embedding one, the original color host image is reduced and used as a watermark, detailed steps are shown in Fig.(4.2).

1. Firstly, the $N \times M$ color host image I is decomposed into three components R, G, and B.
2. Then, each component is decomposed into $n \times m$ non-overlapped blocks, where the block size $n \times m$ is selected depending on a compromise between the watermark invisibility and the recovered image quality as will be shown later in the Subsection 4.5.1.
3. The watermark pixels are calculated as the pixels mean of every $n \times m$ non-overlapped block within the host image, the obtained watermark is a reduced image from the original one with size of $3 \times \frac{N}{n} \times \frac{M}{m}$.
4. To reduce the color watermark to a gray-scale image the Bayer pattern is used, at each pixel location of the image only one color sample is captured thus, the obtained watermark is of size $\frac{N}{n} \times \frac{M}{m}$.
5. To more improve the security and enhance the robustness, the gray-scale watermark is decomposed into four sub-images $W_{1,2,3,4}(i,j)$ where $i \in [1, \frac{M}{2\times m}]$ and $j \in [1, \frac{N}{2\times n}]$.
6. Using Torus Permutation, each watermark sub-image $W_{1,2,3,4}$ is permuted three times with three different keys k_1, k_2, and k_3 to get three different permuted copies of each sub-image.
7. Finally, each permuted sub-image is converted to a binary sequence.

4.4.4 Embedding Process

At this phase, the reduced watermark image is inserted in the host color image, detailed steps of the proposed embedding process are shown in Fig.(4.4) and described as follows:

1. Firstly, the color host image is decomposed into three components R, G, and B and each component is decomposed into four sub-images $R_d(i,j)$, $G_d(i,j)$, and $B_d(i,j)$ where $d \in \{1,2,3,4\}$ and $i \in [1, \frac{N}{2}]$ and $j \in [1, \frac{M}{2}]$.
2. To increase the robustness and enhance the security, a permutation Pr is used to insert each permuted watermark sub-image in a sub-image of the host color image. Fig.(4.3) shows an example of the permutation Pr used to embed the watermark sub-image W_1 permuted with the keys k_1, k_2, k_3 in the host image sub-images R_1, R_2, R_3 respectively.
3. Every eight bits of the watermark image are inserted into the LSB bits of the host image in a block of $n \times m$, the fewer LSB bits are used to insert the watermark, the higher quality of the watermarked image can be achieved.
4. Then watermarked R, G, and B components are merged to obtain the watermarked image.

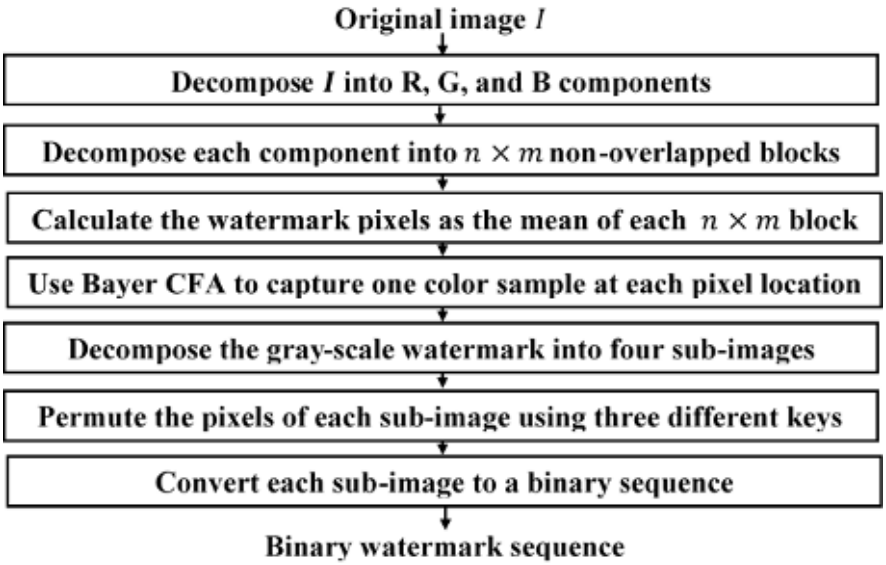


FIGURE 4.2: Flowchart of the watermark generation process

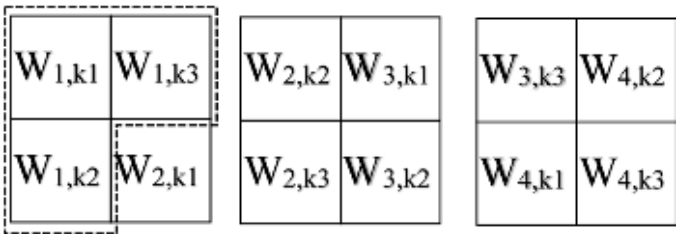


FIGURE 4.3: Block diagram positions example of the W_1, W_2, W_3 and W_4 watermark sub-images embedding: Depending on the permutation Pr; the W_1 watermark sub-image is permuted with the keys k_1, k_2 and k_3 and inserted in the host image sub-images R_1, R_2, and R_3 respectively (and the same process is repeated with watermark sub-images W_2, W_3 and W_4)

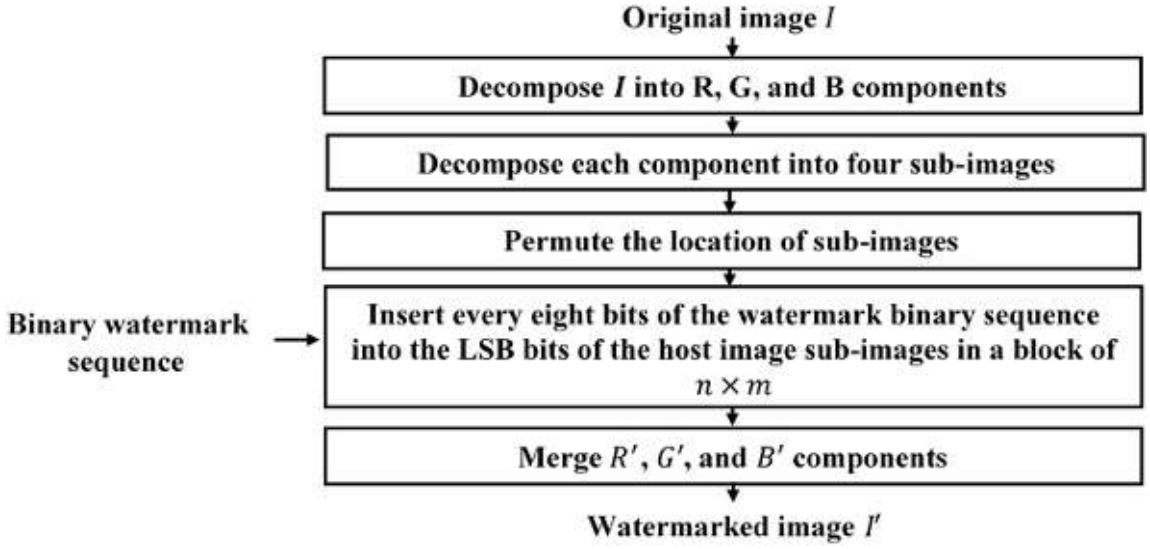


FIGURE 4.4: Flowchart of the embedding process

4.4.5 Extraction Process

At this phase, the insertion process is reversed to extract the watermark.

1. Firstly, the probably modified watermarked $N \times M$ image I', is divided into R, G, and B components.
2. Then every component is decomposed into four sub-images $R_d(i,j)$ where $d \in \{1,2,3,4\}$, $i \in \left[1, \frac{N}{2}\right]$, and $j \in \left[1, \frac{M}{2}\right]$.
3. Each sub-image is decomposed into $n \times m$ non-overlapped blocks, then the LSB bits are extracted.
4. For each $n \times m$ non-overlapped block, the extracted 8 bits are converted to obtain a watermark pixel value.
5. The Torus permutation is reversed by the same keys k_1, k_2, and k_3 used during the embedding process, thus three copies of each sub-image are reconstructed.

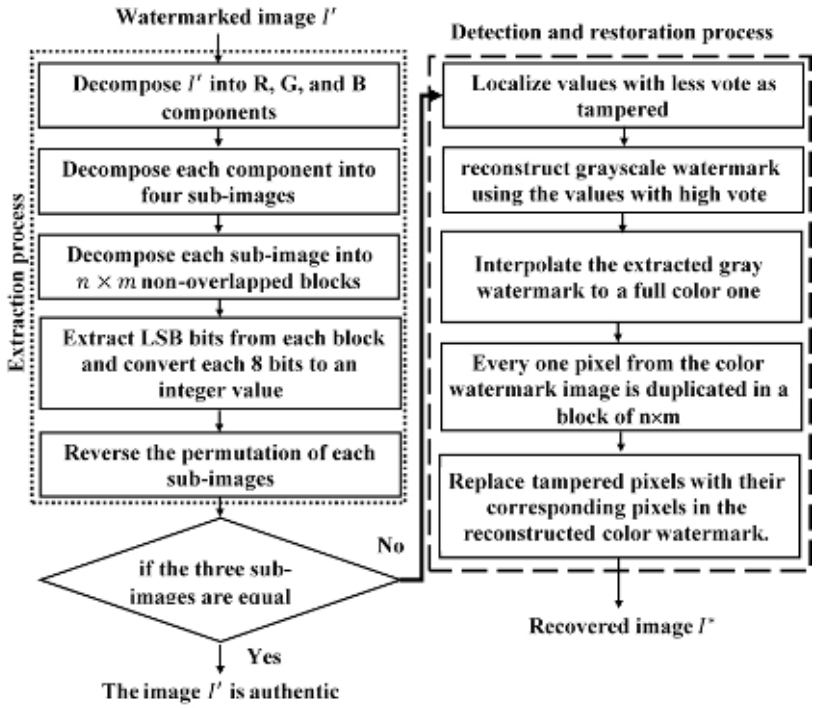


FIGURE 4.5: Flowchart of extraction process: enclosed in the punctuated line beside the process of the detection and restoration enclosed in dashed line

4.4.6 Detection and Restoration Process

At this phase, the majority voting technique described in the Algorithm 7 is applied to decide if the image has tampered or not. The three extracted sub-images are compared. If every extracted three values in the same location are equal, the image is considered as authentic and the process is ended. In the other case, the values with the highest vote are selected to reconstruct the watermark and the value with lowest vote is marked as tampered and the reconstruction process is launched.

1. Using the extracted values with the highest vote, the gray-scale watermark image with size of $\frac{N}{n} \times \frac{M}{m}$ is reconstructed.

2. To obtain a good quality color watermark image from the gray-scale one using the Bayer pattern, the gray-scale image is interpolated to a $3 \times \frac{N}{n} \times \frac{M}{m}$ size image.

3. To obtain a watermark image with the same size of the tampered one, every one pixel from the reconstructed watermark is duplicated in a block of $n \times m$, which has a good quality and a high correlation with the original one, this image will be used to recover the tampered regions.

4. Finally, the pixels marked as tampered are replaced with their corresponding pixels in the reconstructed watermark.

Algorithm 7 Majority vote decision

Input: Watermark sub-images $W_{1,k1}$, $W_{1,k2}$, $W_{1,k3}$
Output: Recover sub-images W
for $l = 1$ to $M/(2 \times m)$ **do**
  for $r = 1$ to $N/(2 \times n)$ **do**
    $w1 \leftarrow W_{1,k1}(l,r)$ {$W_{1,k1}$ is the extracted watermark sub-image from the R_1 sub-image}
    $w2 \leftarrow W_{1,k2}(l,r)$ {$W_{1,k2}$ is the extracted watermark sub-image from the R_2 sub-image}
    $w3 \leftarrow W_{1,k3}(l,r)$ {$W_{1,k3}$ is the extracted watermark sub-image from the R_3 sub-image}
    if $(w1 \neq w2)$ **then**
      if $(w1 \neq w3)$ **then**
        if $(w2 \neq w3)$ **then**
          $W_{1,k1}(l,r) \leftarrow 255$ {Marked as Tampered}
          $W_{1,k2}(l,r) \leftarrow 255$
          $W_{1,k3}(l,r) \leftarrow 255$
        else $\{w_2 = w_3\}$
          $W_{1,k1}(l,r) \leftarrow 255$ {Marked as Tampered}
          $W(l,r) \leftarrow w2$ {Recover the Tampered pixel in the new sub-image W}
        end if
      else $\{w1 = w3\}$
        $W_{1,k2}(l,r) \leftarrow 255$ {Marked as Tampered}
        $W(l,r) \leftarrow w1$ {Recover the Tampered pixel}
      end if
    else $\{w_1 = w_2\}$
      if $w2 \neq w3$ **then**
        $W_{1,k3}(l,r) \leftarrow 255$ {Marked as Tampered}
        $W(l,r) \leftarrow w1$ {Recover the Tampered pixel}
      end if
    end if
  end for
end for
return W

4.5 Experimental Results

In this section, the experimental results are conducted to validate the performance of the proposed scheme. A variety of experiments have been carried out to evaluate the performance of the proposed scheme, and to compare its performance with that of similar schemes namely Lin, Chen, and Chiu [81], Liu [92], and Chen, Tang, and Hsieh [23]. Different perceptual and computational aspects have been considered for performance evaluation.

The performance of the proposed method is tested using a wide variety of color images in the $CVG-URG$ database [38] (e.g.,"Boat", "House", "Lena", "Airplane", "Girl" and "Woman", etc.) and of different file formats (e.g. JPG, BMP, etc.). Image sizes are 512×512 and 256×256 pixels, and image bit-depths are of 24-bit. All necessary simulations were carried out in $C++$ using *Micsoft Visual Studio 2008* with the use of *"OpenCV"* library and $MATLAB$ $(R2012a - 7.140.739)$ using an *Intel Core i*3 3.2GHz CPU.

To evaluate the visual quality of both watermarked and recovered images compared with the original ones, we use the Peak Signal to Noise Ratio $(PSNR)$ defined in Eq.(3.2), besides the use of the structural similarity $(SSIM)$ index defined in Eq.(3.5), which is a metric to measure the similarity between two images. It is designed to overcome inconsistent with human perception founded in traditional methods [25].

In general, when the $PSNR$ values are greater than $30(dB)$ the visual quality is acceptable [25].

To evaluate the detection accuracy performance of our method the tamper detection rate R_{TD} and false alarm rate R_{FA} defined in (3.11) and (3.12) are calculated with different tampering ratio R_{T}, that is calculated using Eq.(3.13). Lower R_{FA} values mean that the tamper detection accuracy is much better [79].

4.5.1 Parameters Selection and Payload

As discussed in Sect.4.4.3, each component of the host image is decomposed into $n \times m$ non-overlapped blocks. The selection of the $n \times m$ block size parameter must be made as a tradeoff, between the watermark invisibility and the recovered image quality, which means that the quality of the recovered image increases by decreasing $n \times m$ block size of the host image and the invisibility of the watermark decreases; since we use more LSB bits to embed the watermark and reversely. In other words, each averaged 8 bits value within the $n \times m$ non-overlapped block is embedded into $n \times m$ pixels of the host image.

Table.4.1 shows the results of using 2×2 and 2×4 block sizes. In case of using the 2×2 block size, the $PSNR = 44.25(dB)$ and $SSIM = 0.98$ while the $PSNR$ value is $51.14(dB)$ and $SSIM$ is 1 when using 2×4 block size. In the first case, the reduced watermark from the host image is of size $\frac{N}{2} \times \frac{M}{2}$. The 8 bits value is embedded into 4 pixels, which means that 2 bits are used to embed the watermark for each pixel, by result the watermarking invisibility is decreased and the recovered image quality is increased to reach a $PSNR$ mean of $39.42(dB)$. In the other hand, if the block size is set to 2×4 the watermark will be of size $\frac{N}{2} \times \frac{M}{4}$, the visual quality of the watermark is degraded since each watermark value is calculated as the mean of every 2×4 pixels, the 8 bits value is embedded into 8 pixels, which means that 1 watermark bit is embedded in each pixel of host image. Thus, the watermark invisibility is increased and the recovered image quality is decreased to reach an average $PSNR$ value of $36.25(dB)$.

Subsequently, using blocks of size 4×4 means that every 8 bits of the watermark will be inserted in a 4×4 block. Thus, the capacity of insertion is higher. However, the quality of the recovered image is lower.

In addition, results in Table.4.1 show that block sizes affects the false alarm ratio, meaning that altering a single pixel affects the whole block which will be subsequently marked

as tampered, as can be seen the R_{FA} mean is equal to 0.17 when using blocks of size 2×2, while it is equal to 0.25 with blocks of size 2×4. Thus, using smaller block sizes provides lower false alarm ratios R_{FA}.

From the above-mentioned results, we can conclude that increasing the $n \times m$ block size, for different tampering ratios, increases both of the watermarked image quality and the false alarm ratio and by results, the recovered image quality is decreased.

TABLE 4.1: The Performance of the Proposed Scheme with 2×2 and 2×4 Block Sizes.

Block size$n \times m$	$R_{\mathrm{T}}^{a}(\%)$	$R_{\mathrm{TD}}^{b}(\%)$	$R_{\mathrm{FA}}^{c}(\%)$	PSNR of water-marked image (dB)	SSIM of water-marked image	PSNR of recovered image (dB)	SSIM of recovered image
2×2	5 (%)	100 (%)	0.00 (%)	44.25 (dB)	0.98	47.30 (dB)	0.99
2×4	5 (%)	105 (%)	0.09 (%)	51.14 (dB)	1	45.38 (dB)	0.99
2×2	10 (%)	100 (%)	0.14 (%)	44.25 (dB)	0.98	41.53 (dB)	0.98
2×4	10 (%)	104 (%)	0.28 (%)	51.14 (dB)	1	36.05 (dB)	0.97
2×2	15 (%)	100 (%)	0.00 (%)	44.25 (dB)	0.98	38.11 (dB)	0.96
2×4	15 (%)	101 (%)	0.18 (%)	51.14 (dB)	1	36.68 (dB)	0.95
2×2	20 (%)	100 (%)	0.22 (%)	44.25 (dB)	0.98	35.89 (dB)	0.94
2×4	20 (%)	101 (%)	0.44 (%)	51.14 (dB)	1	32.32 (dB)	0.93
2×2	25 (%)	101 (%)	0.14 (%)	44.25 (dB)	0.98	34.27 (dB)	0.93
2×4	25 (%)	100 (%)	0.13 (%)	51.14 (dB)	1	30.80 (dB)	0.91

[a] R_{T}–Tamper ratio

[b] R_{TD}–Tamper detection ratio

[c] R_{FA}–False alarm ratio

The number of watermark bits that can be embedded in the host image in terms of the number of bits per pixel (bpp) is known as the payload [30].

In our method, the size of the watermark is a function of the host image size $N \times M$ as well as the selected block size $n \times m$ and the number of LSB bits used to embed the watermark. Thus, one watermark copy of size $\frac{N}{n} \times \frac{M}{m}$ is embedded in each host image component.

At least 1 watermark bit is embedded in 1 host image pixel, so, 1 watermark pixel value is embedded in 8 host image pixels in a block of 2×4. Thus, the $\frac{N}{n} \times \frac{M}{m}$ watermark is embedded in each $N \times M$ host image component. Thereby, the minimal inserted watermark is of size $\frac{N}{2} \times \frac{M}{4}$ for each host image component.

In the incoming experiments, the host image is of sizes 512×512 or 256×256 , to provide much less false alarms, we use small blocks of size 2×2 pixels, in each block 2 LSB bits of the watermark are embedded in each pixel. Thus, the size of the watermark is to be $\frac{512}{2} \times \frac{512}{2}$ or $\frac{256}{2} \times \frac{256}{2}$bits with a payload of 2 (bpp).

For illegal extraction, the security of our proposed method broadly lies on the Torus permutation and the embedding keys k_1, k_2, and k_3 and partially on the selected permutation Pr of the host image sub-images.

4.5.2 Performance of the Bayer Color Filter Array

Our proposed method uses the Bayer color filter array at the insertion process to convert a color full image to a gray-scale image, then, by demosaicking the gray-scale image we obtain a color full image which is highly correlated (close to 1) with the original one. Eq.3.10 is used to calculate Normalized Correlation (NC) values (See Table.4.2). Generally, the NC can take a value between 0 and 1. If the NC value is closer to 1, the two images are getting more similar.

We should notice that using luminance channel instead of using a color filter array doesn't allow us demosaicking the luminance component to a high correlated color full image, as shown in Table.4.2, the Structural Similarity index $SSIM$ values, calculated using Eq.3.5 , are very low. Thus, the pixels of the obtained CFA image are rearranged so it allows estimating the missing pixels, unlike the luminance component pixels which are not correlated.

TABLE 4.2: The Performance of the Bayer color Filter Array to convert a color full image to a gray-scale image: Structural Similarity index (SSIM) and Normalized Correlation (NC) values calculated between the original and demosaicked images.

Image	Size	SSIM	NC
Lena	512×512	0.94	1.00
House	512×512	0.93	1.00
Woman	256×256	0.97	1.00
Girl	256×256	0.94	1.00
Luminance component	512×512	0.62	0.90

4.5.3 Performance Evaluation

The performance of the proposed scheme has been evaluated in terms of the quality of both watermarked and recovered images and detection accuracy. For this, we separated the tests

		
PSNR =44.14	PSNR =44.22	PSNR =44.25
SSIM = 0.99	SSIM = 0.99	SSIM = 0.98
		
PSNR =43.97	PSNR =44.43	PSNR =44.30
SSIM = 0.98	SSIM = 0.97	SSIM = 0.98

TABLE 4.3: "Boat", "House", "Lena", "Airplane", "Woman" and "Girl" Watermarked Images and their Corresponding PSNR and SSIM Values

into two parts: the first is to analyze the property of imperceptibility and the second is the evaluation of the detection accuracy and the restored image quality under several attacks.

Quality of the Watermarked Image

The modification of the LSB bits of the host image induces image alteration. Thus, using more LSB bits to embed the watermark affect the quality of the watermarked image and consequently results in more degradation in the watermarked image. Our proposed method uses only 2 LSB bits to embed both authentication and recovery data.

Table.4.3 shows the watermarked images after the embedding phase. The average $PSNR$ value between the original and the watermarked images is $44.2495(dB)$ and $SSIM = 0.98$ with a high visual quality of the watermarked images and a good watermark invisibility. It means that a large amount of the watermark image (the reduced content of the host color image) can be effectively embedded into the color image by the proposed watermarking scheme without a perceptible degradation of the visual quality.

Tamper detection and content restoration

To evaluate the accuracy of the proposed tamper detection and the quality of the recovered image, different attacks are performed on the watermarked images, then detecting and restoring the tampered regions of the watermarked color images. In the experiments, cut, collage and hybrid attacks are performed with different proportion of tampered regions, by adding, for example, another airplane in the "Airplane" image or a hat in the "Girl" image or even moving the sky from the "House" image as shown in Table.4.5.

For the restoration ability, the size of tampered areas, the detection accuracy of those tampered blocks, besides, the complexity of the image content highly affect the quality of the recovered image. The last two columns of Table.4.4 show the $PSNR$ and $SSIM$ values of recovered images from attacked images. The worst $PSNR$ value is $34.33(dB)$ which remain high and unnoticeable by the human eye proving the high quality of recovered images. In addition, the low standard deviation values of the recovered image $PSNR$ show the high quality of recovered images despite the high tamper ratio and the low $PSNR$ of attacked images, in other words, the standard deviation is equal to 4, meaning that the $PSNR$ average value of recovered images could increase by $4(dB)$ to reach $44(dB)$ or decrease by $4(dB)$ to

reach $36(dB)$, which remain high and shows that $PSNR$ values of the recovered image are close to each other despite the high tampering ratio.

TABLE 4.4: The Performance of the Proposed Scheme Under Cut, Collage and Hybrid Attacks Performed with Different Tampering Ratios

Attacked image	Size	num_{D}^{a}	num_{M}^{b}	$R_{\mathrm{TD}}^{c}(\%)$	$R_{\mathrm{FA}}^{d}(\%)$	$R_{\mathrm{T}}^{e}(\%)$	PSNR of Attacked image (dB)	SSIM of Attacked image	PSNR of Recovered image (dB)	SSIM of Recovered image
House	512×512	330	330	100	0.02	0.2	40.39	1.00	52.22	1.00
Airplane	256×256	228	228	100	0.11	0.5	37.63	1.00	41.28	1.00
House	512×512	3754	3750	100	0.06	1.9	26.62	0.98	43.14	0.99
Fruits	256×256	1447	1444	100	0.22	2.9	26.81	0.97	42.66	1.00
Airplane	512×512	6918	6910	100	0.15	3.5	25.29	0.97	43.82	0.99
Girl	256×256	2058	2049	100	0.25	4.2	31.50	0.98	41.66	0.99
Boat	512×512	9094	9072	100	0.24	4.6	23.48	0.96	38.43	0.99
Lena	256×256	4104	4104	100	0.13	8.3	14.11	0.94	42.46	0.99
House	512×512	20420	20425	100	0.28	10.4	18.28	0.91	37.61	0.98
Girl	256×256	6636	6623	100	0.67	13.5	20.60	0.88	39.25	0.98
House	512×512	31001	31007	100	0.43	15.8	16.52	0.86	36.19	0.97
Woman	256×256	8482	8402	100	1.37	17.1	20.18	0.86	37.14	0.97
House	512×512	34755	34800	100	0.93	17.7	28.47	0.96	40.01	098
Lena	256×256	12286	12288	100	0.00	25.0	09.60	0.81	34.62	0.94
Average				100(%)	0.35(%)	-	24.25(dB)	0.93	40.73(dB)	0.98
Standard Deviation							8.31	0.06	4.17	0.01

[a] num_{D}–Number of identified altered blocks [b] num_{M}–Number of actually altered blocks

[c] R_{TD}–Tamper detection ratio [d] R_{FA}–False alarm ratio [e] R_{T}–Tamper ratio

TABLE 4.5: The Performance of the Proposed Method Under Cut, Collage and Hybrid Attacks Performed with Different Proportion of Tampered Regions

Attacked images					
					
PSNR= 19.91 SSIM= 0.91	PSNR=20.61 SSIM= 0.88	PSNR= 20.18 SSIM= 0.86	PSNR=28.47 SSIM= 0.96	PSNR=18.28 SSIM= 0.91	PSNR= 24.03 SSIM= 0.95
Tamper localization					
					
R^a_{TD}=100% R^b_{FA}=0.37% R^c_T=11%	R_{TD}=100% R_{FA}=0.67% R_T=13%	R_{TD}=101% R_{FA}=1.37% R_T=17%	R_{TD}=100% R_{FA}=0.93% R_T=18%	R_{TD}=100% R_{FA}=0.28% R_T=10%	R_{TD}=100% R_{FA}=0.89% R_T=16%
Recovered images					
					
PSNR= 37.29 SSIM= 0.98	PSNR= 39.29 SSIM= 0.98	PSNR= 37.14 SSIM= 0.97	PSNR=40.01 SSIM= 0.98	PSNR=37.61 SSIM= 0.98	PSNR= 35.12 SSIM= 0.97
Close-up view of part of original image					
					
Close-up view of part of the recovered image					
					

[a] R_{TD}–Tamper detection ratio
[b] R_{FA}–False alarm ratio
[c] R_T–Tamper ratio

Tables.4.4 and 4.5 summarize the *PSNR* and *SSIM* values, tamper detection and false alarm ratios of the recovered images with different tampering rate in [0.2%, 25%].

It can be found from Table.4.4 and Fig.4.6 that although the high tampering ratio in the attacked images "Airplane", "Women", "Girl", "House" and "Boat", tampered areas were perfectly identified by the proposed method, even if 25% of the watermarked image has tampered. The average value of tamper detection ratio is 100% and the higher false alarm ratio is 1.37%, which means that our proposed method can localize the tampered regions under high tampering ratios can achieve 25% with high accuracy and a low false alarm

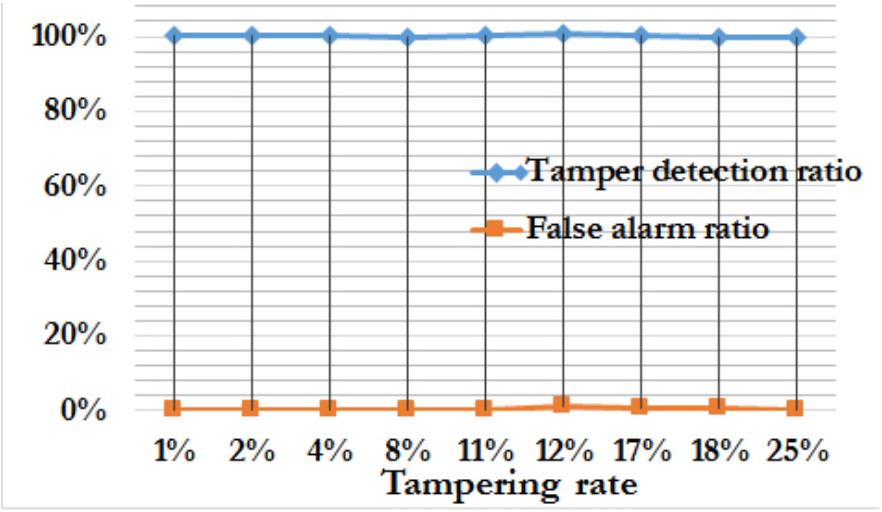


FIGURE 4.6: Tamper detection and false alarm ratios of the proposed method under different tampering ratios

ratio.

4.5.4 Performance Comparison

In this section, we compare our proposed scheme with other digital watermarking schemes in terms of imperceptibility and tamper detection and restoration ability.

Schemes Properties Comparison

The different advantages provided by our proposed method are highlighted in Table.4.6, as well the comparison with several related image authentication schemes.

From the comparison results in Table.4.6, we can deduce that contrary to many related image authentication schemes that either lacking the ability of tamper recovery or uses only gray-scale images [99, 22, 44, 93, 171, 85, 82] or even fail in tamper detection [21], our proposed method outperforms those related schemes by exploiting the color image characteristics to embed the watermark and to recover the altered areas, moreover, it has a higher tamper detection accuracy besides the restoration ability proved in the incoming experiments.

Quality of the Watermarked Image

To validate the visual quality superiority of the watermarked images with our proposed method, a comparison with the related image watermarking schemes is shown in Table.4.7, except Qin *et al.* scheme [127] that changes only one LSB bit and uses gray-scale image as a watermark, $PSNR$ values of the watermarked images with the proposed method are better than that of the image obtained by the methods of the state of art that modifies 2 or more LSB bits, which proves the visual quality superiority of the proposed method.

Unlike the state of art schemes that needs more bits to embed the authentication data besides the restoration information which affects the invisibility of the watermark, our proposed method shows a high visual quality due to the reduction process that takes the advantage of the color filter array to reduce the size of the authentication and restoration data besides the use of the Torus permutation to increase the security and spread the watermark bits within the host image.

TABLE 4.6: Comparison of the Related Work and the Proposed Method Depending on the Schemes Properties, Including the Using of Color Image Format, the Tamper Detection Accuracy, their Tamper Detection and Restoration Abilities

Scheme	Color image format	tamper detection accuracy	Tamper localization	Tamper Restoration
Lu, Xu, and Sun [99]	No	No	Yes	No
Chang, Hu, and Lu [22]	No	Fail[a]	Yes	No
Fridrich [44]	No	Yes	Yes	No
Liu et al. [93]	No	Yes	Yes	No
Wong and Memon [171]	No	Yes	Yes	No
Lin and Chang [85]	No	No	Yes	No
Lin and Yang [82]	No	Yes	Yes	No
Chan and Chang [21]	No	Fail	Yes	Yes
Zhang et al. [182]	No	Yes	Yes	Yes
Lin, Chen, and Chiu [81]	Yes	Yes	Yes	Yes
Liu [92]	Yes	Yes	Yes	Yes
Wang and Chen [164]	Yes	Yes	Yes	Yes
Chen, Tang, and Hsieh [23]	Yes	Yes	Yes	Yes
Lin, Yang, and Xu [90]	Yes	Yes	Yes	Yes
Proposed method	Yes	Yes	Yes	Yes

[a] mentioned in [124]

TABLE 4.7: Objective Quality of the Watermarked Images and Embedding Capacity: Comparisons of Several Embedding Watermarking Schemes

Watermarking schemes	Average PSNR values of watermarked images (dB)	Volume of embedded data
Zhang et al. [182]	37.8500	3 LSB
Wang and Chen [164]	37.9167	3 LSB
Liu [92]	39.6850	2/3 LSB
Chen, Tang, and Hsieh [23]	40.4775	2/3 LSB
Yang and Shen [176]	40.7250	3 LSB
Lo and Hu [95]	41.2308	-
Lin, Yang, and Xu [90]	44.1535	2 LSB
Qin, Chang, and Chen [127]	51.1250	1 LSB
Proposed method	44.2495	2 LSB

Tamper detection and content restoration

To further demonstrate the advantages of our proposed method, a comparison is performed with several schemes, since these schemes and ours have similar abilities, including the use

of color images, the tamper detection, and restoration as it is shown in Table.4.6, besides the volume of embedded data. These schemes used for the comparison are:

1. Lin, Chen, and Chiu [81] scheme: a color image authentication fragile scheme that uses error correction coding and Bayer pattern. The authentication data was generated as BCH code of the image blocks, to achieve the requirement of tamper recovery the Bayer pattern technique is used to decrease the amount of block recovery data and enhance the quality of watermarked images.

2. Liu [92] scheme: a color image authentication scheme was proposed. The proposed self-embedding fragile watermarking scheme had the ability to detect and recover tampered areas. The luminance component of the host image is used to generate the authentication data by calculating the local mean of blocks, then embedding it into the color image. For tamper detection, besides the use of the dual option parity check technique, morphological operations are used to detect tampered regions.

3. Chen, Tang, and Hsieh [23] scheme: a color images authentication scheme with recovery ability was proposed. The details of the luminance channel are encoded using the differential coding technique and embedded into the image as the recovery data. For more protection error correcting codes technique and duplication are used.

Our proposed method is compared with Lin's method [90], where more attacks are performed under tampering rate can achieve 25%, the obtained $PSNR$ results in Table.4.8 with an average of $42.30(dB)$, prove the superiority and efficiency of our proposed method compared with Lin's method where the average $PSNR$ is $37.74(dB)$.

Therefore, recent researches don't focus on high tampering rates attacks since it is a fragile watermarking scheme. Thus, we further compare the visual quality of the watermarked image with different tampering rates with our proposed method and [92] and [23] schemes where the shown experiments are made with tampering rate less than 3%.

According to Table.4.9, the average of the $PSNR$ values of the restored image obtained with our proposed method is $44.21(dB)$ which is much higher with $4(dB)$ than Chen *et al.* scheme [23] where the average is $40.33(dB)$ and greater than Liu's scheme [92] with $6(dB)$ where the $PSNR$ average is $38.18(dB)$, the reason is that the block size used in the proposed method is smaller, besides the high quality of the interpolated extracted watermark, which proves the effectiveness and superiority of our proposed method compared with [92] and [23] schemes.

The experimental results above-mentioned illustrate: the high quality of the watermarked image with an average $PSNR = 44.25(dB)$ and an average $SSIM = 0.98$, the accuracy of tamper detection, the high quality of restored color images and show the superiority of our method compared with similar schemes.

TABLE 4.8: Restoration Performance Comparison of the [90] Color Image Related Work and the Proposed Method Under Different Tampering Ratio: Calculated PSNR Values Between the Original and Recovered Images

Tampering ratio R_T	1.6(%)	3.1(%)	6.3(%)	12.5(%)	25.0(%)	Average
Lin, Yang, and Xu [90]	$43.90(dB)$	$40.47(dB)$	$37.92(dB)$	$34.95(dB)$	$31.44(dB)$	$37.74(dB)$
Proposed method	**52.79**(dB)	**42.66**(dB)	**42.46**(dB)	**39.25**(dB)	**34.33**(dB)	**42.30**(dB)

TABLE 4.9: Restoration Performance Comparison of [92] and [23] Color Image Related Work and the Proposed Method Under Different Tampering Ratio: Calculated PSNR Values Between the Original and Recovered Images

Attacked image	Tampering ratio R_T	[92]	[23]	Proposed method
Airplane	0.46(%)	38.55	40.5	**41.28**
School	0.93(%)	38.46	40.03	**40.12**
Light house	1.67(%)	38.11	40.68	**52.79**
Fruits	2.94(%)	37.59	40.09	**42.66**
Average		38.18 (dB)	40.33 (dB)	**44.21** (dB)

4.6 Conclusion

This chapter has presented a novel self-embedding fragile watermarking scheme based on Bayer pattern, for being used in the color image authentication with tamper detection and restoration abilities.

Aiming at developing a watermarking scheme in light of the self-authentication model, the choice of developing a new, efficient embedding scheme has been justified and the relevant challenges have been specified.

The analysis and experiments of the developed scheme have been presented, and its efficiency and suitability for images is validated. Moreover, the analysis and performance comparison of similar schemes ([81, 92, 23]) is performed. Specifically:

- A performance comparison with similar schemes has been reviewed in light of the specified challenges. Thereby, the need for an efficient authentication scheme have been identified.
- Thereby, the existing authentication schemes have been shown lacking the necessary consideration of the detection accuracy and the restoration ability.
- Considering the computational aspects and capacity of the proposed scheme has been conducted and compared with similar schemes.
- Considering the perceptual aspects, the performance of the embedding scheme has been evaluated and compared with the existing schemes.

Consequently, a fragile, blind, irreversible, LSB-based (spatial domain) watermark self-embedding scheme has been developed. Its implementation, technical properties, and features are described to demonstrate its efficiency.

Experimental results are concluded to prove the effectiveness of the proposed method and show the major contributions of this paper includes: firstly, the simple structure of the self-embedding watermark information for color images that is visually indistinguishable with $PSNR$ values greater than 44 (dB). Secondly, the exploitation of the color images characteristics, to reduce the watermark to a gray-scale image using the Bayer pattern and Torus Automorphism to perform tamper detection with high detection accuracy reaches up 100% and low false alarm ratio less than 2%; and finally, a method of recovering the tampered image by interpolating the extracted gray-level watermark has also been devised for the quality improvement of the restored color image which can achieve 34 (dB) with tampering rates of 25%.

Finally, it has been concluded, with the facts from the observations, analysis and experimental results, that the proposed image authentication method satisfies three principal

features: sensitive tamper detection, tamper localization accuracy, and high-quality tamper restoration. So, it can find a niche area of application where the security becomes a serious problem such as in medical images and in law enforcement. Furthermore, the proposed scheme is more suitable and efficient and has a great promise to address all the major limitations of existing schemes.

Chapter 5

A Novel Cholesky Decomposition-based Scheme for Strict Image Authentication

5.1 Introduction

As discussed in previous chapters; in some cases, modifications may be intentionally malicious or may inadvertently influence the interpretation. Thus, strict authentication has already demonstrate its necessity for several applications, where we must be certain that an object has not been altered "at all" since it left a trusted part, even minor modification is not allowed to verify or authenticate the content integrity. Thus, strict authentication is devised for applications that don't tolerate alterations in the protected object.

According to the techniques used, strict image authentication methods can be divided into conventional cryptography-based methods and fragile watermarking-based methods [30, 134]. The first involves hash functions to generate cryptographic signatures, which become invalid when the content is modified. While the second involves generating watermarks specifically designed to become undetectable when the cover object is modified, such watermarks are sensitive to both malicious and incidental attacks [30]. In this chapter, our attention is restricted only on conventional cryptography-based methods, moreover, considering that the basic realization of a model may be valid for other multimedia applications, our focus in this chapter is more precisely on the digital image applications.

Digital signatures-based authentication is quite established and still be used especially for sensitive application such as medical, military, and law-court images. Embedding the signature within the cover image can reduce the risk of losing the authentication signature. However, altering the cover image to embed the watermark causes the subsequent authentication test failure. Moreover, the watermark undergoes the same manipulations as the host image which could also expose the whole process to fail. Thus, a novel strict model that gains several benefits is desirable.

Most of the existing strict schemes that use hash functions are able to detect image tampering. Unfortunately, most of them do not take into account the cover object type, thus, using an image as a cover object in not similar when using other types, moreover, most of these schemes do not profit the image properties when hashing them and do not distinguish them from other types of cover object, which implies that the computational complexity and by the way the computational cost increases by increasing the image size. On the other hand, the security of these schemes relies basically on symmetric keys, which expose them to brute force attacks. Furthermore, as it discussed in previous chapters, most research efforts in the area of image authentication are focused primarily on the gray-scale image. However, the design of a gray-level image authentication scheme did not take advantages of the color

image characteristics; it must be modified or even redesigned to achieve the color image authentication requirement.

The research presented in this chapter has led to a new strict authentication model to achieve different security properties such as authentication and integrity verification of color images. The proposed Cholesky decomposition-based scheme thus can help to propose a solution to existing schemes limitations and can ensure the security properties using some mathematics and cryptographic tools. Our scheme satisfies the requirements of tamper detection, content integrity, and owner authentication. Thus, the main focus of this model is the efficiency against any attempts to alter the image intentionally or unintentionally.

In fact, developing a strict authentication scheme requires several tasks, as pointed out in the previous chapter. These tasks include, for example, choosing the appropriate signature, the adequate transform, and choosing appropriate keys, etc.

The reason for considering this task and the challenges in pursuing it, to follow up this research are described below. The method, analysis, and experiments are then briefly introduced that will be presented in the subsequent sections of this chapter.

5.2 Considerations Before Applying the Cholesky-based Model

For strict authentication schemes application, the main challenging issue is to develop a reliable scheme, that ensures the integrity and the authenticity of the image. Thus, developing a novel approach should be based on the following considerations.

1. **Satisfying the integrity requirement of the content integrity and authenticity of the owner of a digital image according to strict authentication requirements:** The first strict authentication requirement is to prove that an image has not modified at all, also, to improve the scheme's effectiveness, it is unnecessary to use two separated schemes, one to authenticate the sender and the second to verify the integrity. The integrity of content and the authenticity of the sender can be verified using the same scheme.

2. **Retrieving an appropriate image authentication data:** Extracting the appropriate authentication data that can describe the whole content in fewer data may reduce it, this consideration seems to be obvious, however, it is not evident to extract the most representing data of an image while preserving the requirement of strict authentication.

3. **Reducing the computational complexity:** Computational complexity is required criteria for such schemes, especially when applying the hash function on digital images which may influence the overall performance and increases the computational complexity.

4. **Finding a suitable encryption function:** A security level should be achieved. A cryptographic function can be readily chosen from the literature considering the properties of the key (e.g., length, type public or private, etc.). In addition, this function can also serve as a tool to prove the sender authenticity.

Therefore, aiming at developing a suitable embedding scheme, several challenges are discussed below.

5.3 Challenges for Developing a Digital Signature-based Scheme

When designing any authentication model, the application should be considered to determine the several challenges that should this model overcome. Several limitations represent serious challenges to the majority of existing schemes are specified in the section below.

1. In strict authentication schemes, applying the hash function for an image seems to be effective. Although the efficiency is required to compensate for the computational cost when hashing the image, this issue is more and more critical with larger image size. As well, the encryption function used to ensure a security level, that can more enlarge this problem.

2. An efficient authentication data generation is required to provide reduced data size while ensuring the strict authentication requirement and the content integrity.

3. Moreover, using the Cholesky decomposition to satisfy the integrity requirement implies the use of positive definite matrices, which is not always the case.

Thus, to overcome all the above challenges and design an efficient digital signature-based scheme for image authentication, the following requirement should be considered: i) the rejection of any modification, ii) the reduction of the computational efficiency, iii) the reduction of authentication data, iv) the verification of the content integrity, and v) the sender authenticity.

The following sections of this chapter describe the developed digital signature-based scheme for color image and present its development, performance evaluation, and its applicability.

5.4 Proposed Cholesky Decomposition-based Scheme

Our main issue in the proposed scheme is to address both the authentication and integrity of images. The use of Cholesky decomposition properties is the major difference that differentiates our proposed scheme from existing state of the art approaches

The goal of our proposed scheme is to detect any modification in the digital images. If the image is untampered, the recalculated authentication data and the attached one will be equal. If the image is tampered, the two extracted and attached authentication data will not be equal and the image considered as non-authentic.

5.4.1 Features of the Proposed Scheme

As discussed in previous section (Section 5.3), the emerging evolution of image processing tools imposes new challenges for researchers to design effective authentication schemes. The existing schemes should be further improved to meet new requirements including the use of large size authentication data and tolerating more image processing operations without compromising security or the computational complexity or even the scheme performance.

The compromise between these requirements leads to a new generation of image authentication schemes that use new techniques to combine the cryptography and digital signature techniques [96].

The proposed formal scheme offers several features in addressing the above requirements to overcome limitations of existing schemes, these major features are described below.

1. **Strict image authentication:** To ensure the requirement of strict authentication where no changes are tolerated to the host image a cryptographic hash function, a public key cryptosystem, and Cholesky decomposition properties are used.

2. **Modifications control:** Correlated parameters of the Cholesky decomposition allows the control of any modification arises the factorized matrix; this means that modification that changes the value of one parameter will also partially affect the value of the other.

3. **Reduced size of the authentication data:** The diagonal of the Cholesky decomposition matrix is used instead of using the entire matrix. In addition, the parameters correlation degree is indicated using the size of the diagonal entries, providing the use of the principal diagonal instead of using the whole matrix.

4. **Security and robustness:** The several requirements of security and robustness of the proposed signature scheme is achieved by using the hash function and RSA cryptosystem that encrypts the obtained sequence before attaching it with the host image.

5. **Low computational complexity:** The authentication data generated is sufficiently short, enabling the fast creation of the digital signature by applying the hash function . In addition, the hash function is very sensitive to any modifications made on the images, which means that changing a single bit in an image may result a different hash value.

6. **High embedding capacity:** The authentication data used to identify the image is associated with it in a separate file, subsequently, the embedding capacity of those techniques is higher than the embedding capacity of watermarking-based schemes [96].

Thus, to be well understood the proposed scheme, its background is covered in the following section.

5.4.2 Preliminaries and Cryptographic Background of the Proposed Scheme

Over the ages, cryptography has attracted the attention as a tool to meet some of the information security requirements. Cryptography is the study of mathematical tools and techniques used in information security to provide a security service including confidentiality, data integrity, and entity authentication [106]. Enabling significant information to be stored or transmitted over non-secure networks, so that only authorized recipients can read it [56].

Message authentication techniques such as hash functions, private or public key systems and digital signatures are also used in image authentication and integrity systems [56]. There are several tools that are used in image authentication algorithms. The most important one is based on the hash function which afford an effective and secure information processing [56, 53], and the RSA cryptosystem which is one of the most used public key cryptosystem.

Hash function

Hash functions are known with the high sensitivity to any small modification in the image pixels or even in a bit in the image. In consequence, the image is classified as manipulated when just only one bit of this image is changed; this is very severe for most of applications [56].

In modern cryptography, the cryptographic hash function is considered as one of the most frequently used primitives [106]. A hash function interpret an arbitrary finite length input message to a fixed length output called the hash value or digest; in other words, it is the compact representation of the input message [56, 106, 53].

To achieve the security (cryptographic) requirement, a typically chosen hash function must satisfy at least, the following properties [136, 53]:

- For any given hash value y of H, it is "computationally impossible" to obtain a message m such that $H(m) = y$. It must be difficult to reverse H from y to get an m corresponding to y.

- For any given message m, it is "computationally impossible" to obtain any message $m^{'}$ such that $m \neq m^{'}$ and $H(m) = H(m^{'})$.

- It is "computationally impossible" to obtain any two messages m and $m^{'}$ such that $m \neq m^{'}$ and $H(m) = H(m^{'})$.

Rivest, Shamir, and Adleman (RSA) Public Key Cryptosystem

One of the most used public key cryptosystem is the RSA cryptosystem. It is used to provide both secrecy and digital signatures, The security of this cryptosystem is ensured by the integer factorization intractability [106].

The RSA key generation, encryption and decryption algorithms are described as follows [106]:

Key Generation Algorithm Each entity generates an RSA public key and a corresponding private key and do the following:

1. Generate two large distinct random primes p and q each roughly the same size.
2. Compute $n = p \times q$ and $\phi(n) = (p-1) \times (q-1)$.
3. Randomly choose the integer e such that $\gcd(e, n) = 1$.
4. Using the extended Euclidean algorithm (see [106]), Calculate the unique integer d, $1 \prec d \prec \phi(n)$, such that $e \times d \equiv 1\,[\phi]$.

The public key is (e, n) and the private key is d.

Encryption Algorithm At the sender end, A encrypts a message m for the receiver B.

1. Obtain the authentic public key (e, n) of the receiver B.
2. Represent the message as an integer m in the interval $[0, n-1]$.
3. Compute the encrypted message $m_c \equiv m^e\,[n]$.
4. Send m_c to B.

Decryption Algorithm B should use the private key d to recover the plain text m from the encrypted message m_c.

1. Compute $m \equiv (m_c)^d\,[n]$.

The proposed scheme consists of two processes: authentication data generation process and authentication verification process, which are described in the following subsections.

5.4.3 Authentication Data Generation Process

In this process the authentication data that will be used to verify the authenticity of the protected image is generated, considering the several requirements and challenges issues previously mentioned. Thus, the performance of the whole authentication process depends on this process. To ensure the strictness of the authentication the hash function is used, and since the protected object is an image, we profit their properties to apply the hash function on the diagonal coefficients of the Cholesky decomposition instead of hashing the whole image, reducing size of the data to be hashed from $N \times M$ to N which is M times fewer.

Let the original image I be a color image with the size of $N \times M$, in what follows, the original image I is represented as an $N \times M$ matrix. Detailed steps are depicted in Fig.5.1 and Algorithm 8, and described as follows:

1. Cholesky decomposition is unique only if the matrix to be factorized is a positive definite matrix (See Chapter 2). Thus, to overcome this problem and obtain a positive definite matrix from the host image matrix I, it is firstly multiplied by its transpose.

$$I_m = I \times I^t$$

2. Using the Cholesky decomposition, the obtained positive definite matrix I_m is decomposed into two Matrices L and L^t.

$$I_m = L \times L^t$$

3. Principal diagonal D of the L matrix is obtained.

4. The hash function is applied on the obtained principal diagonal D.

$$H = Hash(D)$$

5. Using the RSA to generate a sender private key K_{pr}, then using this key, the hashed sequence is to signed and attached with the original image. The RSA is used to protect the signature and the verify the sender authenticity.

So, an authentication data sequence $Sign$ (digital signature) can be transmitted electronically with the original image.

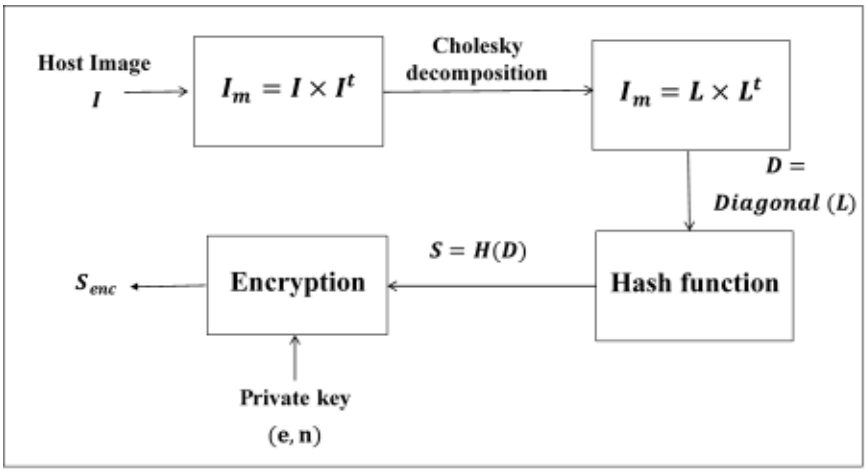


FIGURE 5.1: Authentication data generation process

Algorithm 8 Authentication Data Generation Algorithm

1: **Input: Original image** I**, Private key** K_{pr}
2: **Output: Digital signature** $Sign$
3: $I_m \leftarrow I \times I^t$
4: $[L, L^t] \leftarrow Chol(I_m)$
5: $D \leftarrow Diag(L)$
6: $H \leftarrow Hash(D)$
7: $Sign \leftarrow Encrypt(H, K_{pr})$

5.4.4 The Verification Process

As shown in Fig.5.2 and Algorithm 9, to verify if the received image has tampered or not, the receiver computes the authentication data from the received image $I^{'}$, following the same steps used to generate the digital signature. Then, applying the public key K_p on the authentication data $Sign$ that was appended to the received image, allows the authentication of the sender. Finally, to decide if an image is authentic, the verified attached signature H and the calculated one $H^{'}$ are then compared.

If the two recalculated $H^{'}$ and verified attached digital signatures H are equal the image is considered as authentic, otherwise, the received image is not authentic.

Algorithm 9 Authentication Verification Algorithm

Input: Transmitted image $I^{'}$**, Public key** K_p**,** $Sign$
Output: Decision
$I_m \leftarrow I^{'} \times I^{'t}$
$[L, L^t] \leftarrow Chol(I_m)$
$D \leftarrow Diag(L)$
$H^{'} \leftarrow Hash(D)$
$H \leftarrow Verify(Sign, K_p)$
if $H = H^{'}$ **then**
    $Decision \leftarrow Authentic$
else
    $Decision \leftarrow Not\ authentic$
end if

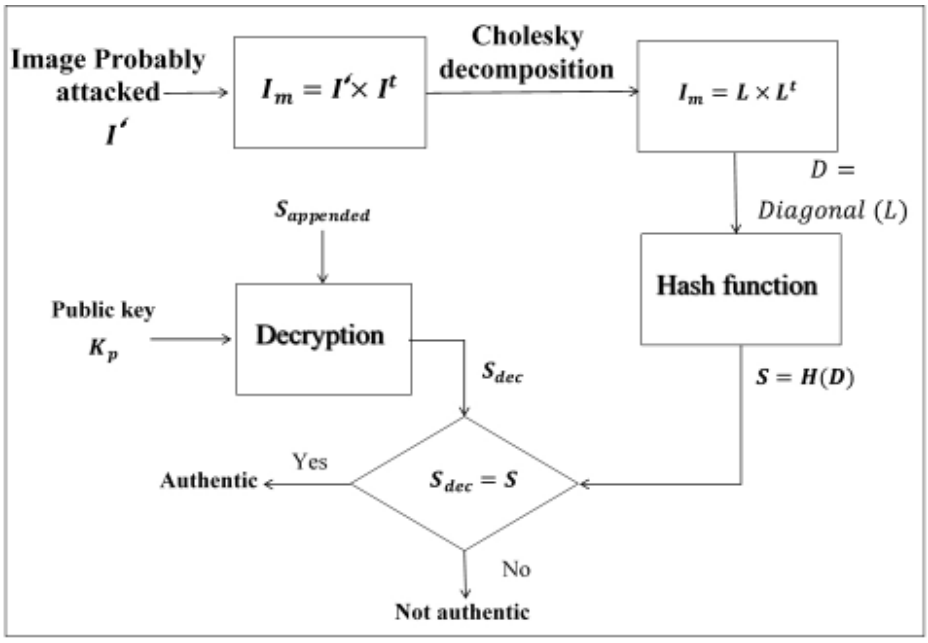


FIGURE 5.2: Authentication data verification process

5.5 Experimental Results

In our experiments, the testing programs are implemented using $C++$. Besides the use of *Cryptoc++* library to call *SHA256* hash function and *RSA* cryptosystem.

The performance of the proposed Cholesky decomposition based scheme is tested using a wide variety of images in the $CVG - URG$ database such as "Baboon", "Airplane", "Girl", "House", "Lena" and "Peppers" with the size of 256×256 pixels as shown in Fig (5.3), beside the use of medical images databases such as: Medical Image Sample [29].

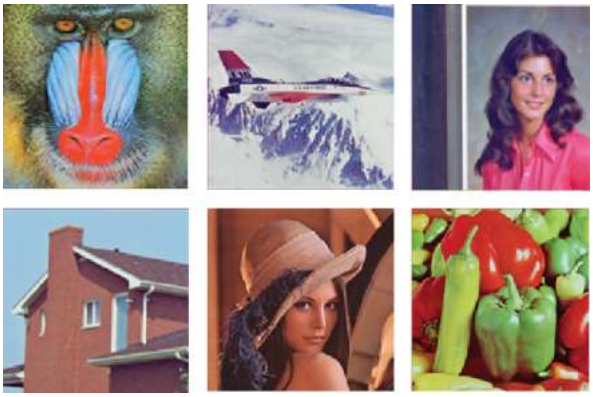


FIGURE 5.3: Test images.

In order to evaluate if the two authentication data sequences (the attached and recalculated) are equal or not, we use correlation metric *Corr* (See Chapter 3 and Eq.3.7).

Generally, the correlation can range from -1 to 1, where -1 represents a direct, negative correlation, 0 representing no correlation, and 1 represents a direct, positive correlation. If it is closer to 1, the two sequences are getting more similar. In our tests, if the correlation is equal to 1 the image is considered as authentic, in the other case the image is considered as tampered.

Moreover, the peak signal to noise ratio $(PSNR)$ (See Chapter 3 and Eq.3.2) in units of (dB), is used to evaluate perceptual distortion between the original and modified images. Generally, the larger the $PSNR$ value is, the more invisible the distortions are.

Our proposed method uses the Cholesky decomposition to extract the most representative data, then, reduce it as possible as we can, while preserving the strict authentication requirement and without affecting the basic idea of not allowing any modification at all.

Thus, to prove the effectiveness of the proposed scheme, the correlation values are calculated between the authentication data attached with original images and the recalculated ones. Furthermore, to demonstrate the advantages of our proposed method different attacks are performed.

As well seen from Table.5.1, without any modification made on the protected image the two sequences are similar and the correlation is equal to 1.

On the other hand, changing only one bit from a pixel of the whole image changes the recalculated sequences and the correlation is different from one and closer to 0 despite the slight modification made.

For better illustration, results are also demonstrated under Salt and Pepper noise with a low variance equal to 0.0001, obtained results in Table.5.1 shows the effectiveness of the proposed method, where the correlation values remain low despite the $PSNR$ high values greater than $46(dB)$ which demonstrate that the slight modification made can't be perceptible by the human eyes but it can be detected using our proposed scheme.

Correlation values calculated between the authentication data attached to original images and the recalculated ones, shown in Table.5.1, prove that the proposed method is effective. Meaning that even with using only the diagonal coefficients of the Cholesky decomposition instead of using the whole image, the strict requirement is guaranteed, which prove the possibility of reducing and using only few parts of the image and preserving the integrity of its content.

We should notice that the Cholesky decomposition is not in wide use for digital signatures and watermarking applications. Unfortunately, we have not found similar schemes in literature to compare them with the proposed scheme.

TABLE 5.1: Calculated Correlation Between the Appended Sequence $S_{appended}$ and the Calculated One $S_{calculated}$ and the $PSNR$ Values Calculated Between the Original and the Modified Images

Attacks	-	Baboon	Plane	House	Lenna	Pepper
Without attack	Corr	1	1	1	1	1
Changing One bit	Corr	0.2996	-0.0181	-0.0921	0.0332	0.3067
	PSNR (dB)	96.2956	96.2956	96.2956	96.2956	96.2956
Changing One pixel	Corr	0.4038	-0.0955	0.2234	0.2353	-0.1673
	PSNR (dB)	62.8536	53.2498	51.8934	62.1441	55.1575
Salt and pepper variance 0.0001	Corr	-0.0361	0.1674	-0.1193	0.3033	0.1881
	PSNR (dB)	50.2796	48.7065	51.9702	48.6796	46.678

5.6 Applicability of the Proposed Scheme

A fragile, irreversible Cholesky digital signature-based scheme has been developed for being used for strict authentication.

Our proposed scheme can find a main application in healthcare systems, as an effective tool for protecting sensitive information such as military and medical images, the secure sharing and control of those images is needed. These images must be protected against any incidental or malicious manipulations, which could affect the judgments based on these images, especially in sensitive information applications.

The main properties of the proposed scheme are described below to demonstrate its applicability for medical image.

1. **Strict authentication:** A strict authentication doesn't tolerate any changes in the protected image, not even one bit of the whole image, it becomes invalid even for the smallest modification. For medical image, which are assumed to be trusted, these requirements become critical and challenging issues.

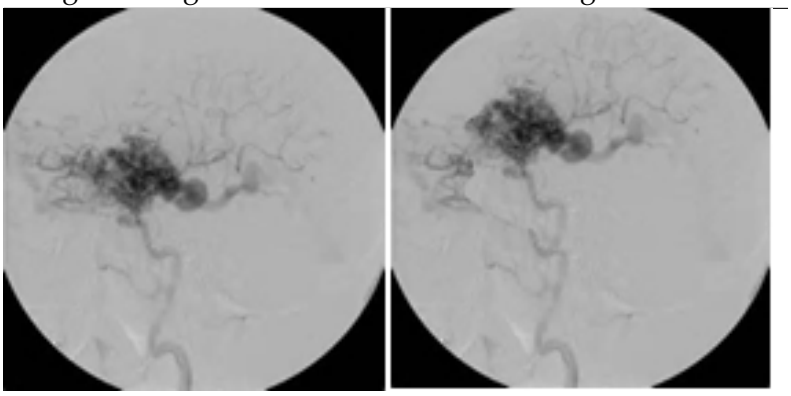
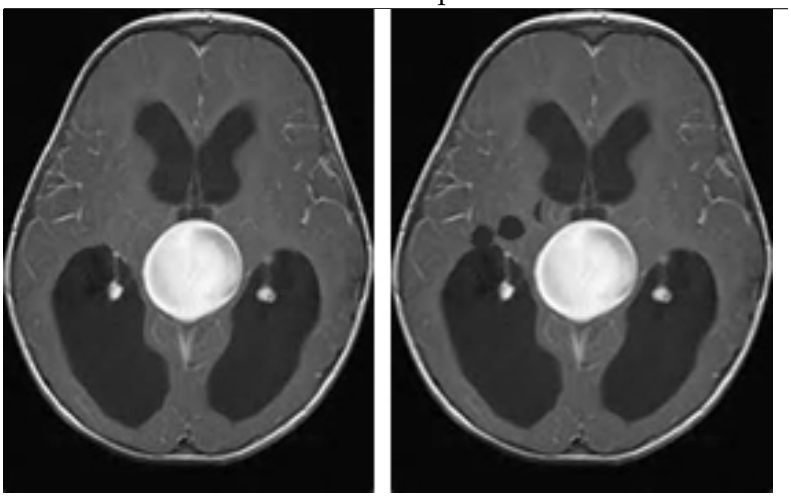
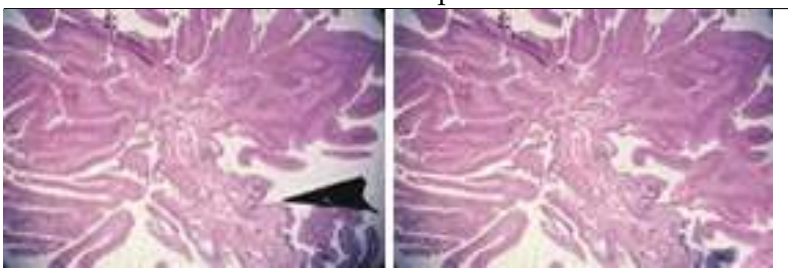
2. **A high embedding capacity:** Which is main requirement for medical images that preserves the content and the security. Moreover, in the proposed method, the authentication data is appended with the protected image, which decreases the embedding capacity.

3. **Reduced size of the authentication data:** Most of the existing schemes make use of the whole image to generate the authentication data, the proposed scheme does not operate on the whole image, rather it uses the diagonal of the Cholesky decomposition matrix, which minimizes the embedding capacity and the embedding time by avoiding complexities unlike many other schemes.

4. **Secrecy:** To achieve several requirements of security and robustness of the proposed signature scheme a hash function and RSA cryptosystem are used to sign the obtained sequence before attaching it with the host image.

The above-mentioned properties demonstrate the applicability of the proposed scheme in healthcare systems for protecting the sensitive information, the secure sharing and the control of those images which should be trusted.

TABLE 5.2: An Example of Application: Calculated Correlation Between the Appended Sequence $S_{appended}$ and the Calculated One $S_{calculated}$ and the $PSNR$ Values Calculated Between the Original and the Modified Images.

Original images	Attacked images
	
PSNR =21.1895 (dB) $Corr(S_{appended}, S_{calculated})$=-0.0968 Decision:Tampered	
	
PSNR= 36.423 (dB) $Corr(S_{appended}, S_{calculated})$=0.2590 Decision:Tampered	
	
PSNR= 19.6475 (dB) $Corr(S_{appended}, S_{calculated})$=0.06965 Decision:Tampered	

To illustrate the efficiency of our proposed method in the medical images an example is shown in Table.5.2.

From the obtained results of the correlation in Table.5.2, we can conclude that any modification in the original images can be detected using our proposed scheme, even if the modifications can't be noticeable by the human eyes (e.g., doctors, other medical professionals).

5.7 Conclusion

This chapter has captured a systematic (i.e mathematical) aspects of the digital signature with their practical applications. The two main components have been considered for the authentication data generation and verification. Thereby, a new strict authentication model, formal definitions of fundamental properties, and possible attacks have been presented. More specifically:

- A novel Cholesky decomposition-based scheme for strict image authentication has been proposed. By exploiting the Cholesky decomposition properties beside the use of cryptographic and mathematics tools our proposed scheme shows effectiveness in detecting tampered images.
- The lack of developing of digital signature-based schemes in the literature are reviewed besides the need for a model that consider the image properties, in addition, the common limitations of existing schemes have been identified: regardless of image characteristics when hashing an image.

In this chapter, research contributions are presented in three main parts: challenges for developing a digital signature-based scheme (Section 5.3), definitions of the proposed Cholesky decomposition-based scheme and fundamental properties (Section 5.4), and the experimental results and possible attacks (Section 5.5).

The definition of major properties and features including: i) strict image authentication, ii) modifications control, iii) reduced size of the authentication data security and robustness, iv) low computational complexity, and v) high embedding capacity.

The given definitions help contextualize the usual literal meaning of the properties for the image applications. Thus, the proposed scheme can find its application in several areas, where sensitive information needs to be protected such as various medical imaging modalities and services (e.g. radiology, surgery, etc.).

Addressing this application an example has been shown. Moreover, possible attacks have been defined to show the suitability of the proposed scheme for enhancing robustness and the security of protected images providing an authentication service.

Finally, this chapter has reviewed the future works that turn around using the Cholesky decomposition properties to detect and restore tampered areas besides the exploitation of medical images properties (e.g., region of interest (ROI), region of non-interest ($RONI$), etc.) to embed the authentication data into the host image.

Chapter 6

A blind Dual-Color Images Watermarking Based on IWT and Sub-sampling

6.1 Introduction

As discussed in previous chapters, the protection of digital contents from illegal use has been receiving more and more attention in recent years, digital watermarking demonstrates its effectiveness as a powerful solution to fulfill this need. The main feature of digital watermarking is to allow for imperceptibly embedding a watermark in the original host image, once created the watermark can be detected or extracted for the purpose of owner identification or/and integrity verification of tested data [110, 39].

Blindness, robustness, and embedding capacity are three contradictory requirements (See Section 3.3). Thus, designing a model that satisfies all of these requirements at the same time is not possible. Subsequently, finding a good compromise between these requirements is challenging issues. On the other hand, color images are being widely used on the Internet, it is usually termed as dual-color image watermarking when embedding color watermark image into color host image, which is one of the most challenging issues in robust image watermarking since the color image has greater amount data to be embedded.

Generally, the existing image watermarking schemes use a binary or gray-scale image as watermark. Moreover, most of these schemes achieve the desired level of robustness against attacks at the expenses the blindness requirement, which is not practical in several applications. It is a challenging problem to design a blind dual-color image watermarking scheme.

Motivated by the above limitation, a blind watermarking method based on IWT and sub-sampling is proposed for embedding color image watermark into a color host image in the frequency domain. In which we are interested in three major features: The first one is to use the integer wavelet transform (IWT), which can map integer to integer without the rounding error and can further concentrate the energy of each sub-image to obtain a good imperceptibility. The second one is to embed color image watermarking into a color host image. This is motivated by the fact that compared with gray-scale watermarking; the digital color image has more amounts of data. The last but not the least is that the proposed watermarking is blind detection while preserving the good robustness against attacks.

The reason for considering this task and the challenges in pursuing it, to follow up this research are described below. The method, analysis, and experiments are then briefly introduced that will be presented in the subsequent sections of this chapter.

6.2 Considerations Before Applying the Dual-color Image Watermarking Model

Aiming at developing an appropriate model, that is different from the existing literature methods and overcome their limitations, several considerations must be taken into account, which are described in what follows:

1. **Finding an appropriate technique of embedding the watermark:** To ensure the requirement of robustness, frequency techniques seem to be obvious choice, however, several transformations existing in the literature each one differs from others with its advantages and inconveniences (See Chapter 2). Moreover, this transformation must consider the properties of an image and its value's type.

2. **Finding an appropriate compromise between the capacity of embedding and watermark invisibility:** Choosing the appropriate locations to embed the color image watermark, with its significant amount, without affecting the host color image quality. This can be readily accomplished by finding a method to embed the watermark while regarding the host image properties and textures.

3. **Choosing an appropriate mapping function:** Thought, ensuring the security and the computational complexity of the designed scheme are required for embedding and extracting the watermark.

4. **Finding a suitable extraction function:** Extracting the watermark in a blind manner while preserving the robustness is a difficult task. That's why several existing methods use the original image or a part of it to extract the watermark.

Therefore, developing the aimed model requires to overcome the several challenges discussed below.

6.3 Challenges for Developing an Embedding Scheme

To design an efficient and appropriate watermarking scheme for color image protection that satisfies the requirements of high robustness, high invisibility, and low capacity is depending on finding a good compromise between them. Thus, achieving this purpose requires overcoming several challenges that are cited below:

1. Using a color image as a watermark requires a relatively high embedding capacity since color images contain more amount of data. Achieving this purpose is at the expense of the invisibility of the watermark. However, a lower level of embedding distortion is required to ensure the invisibility and the protection requirements of the watermark image.

2. On the other hand, robust watermarks require low invisibility, visible watermarks are more exposed to attacks, it is obvious that invisible watermarks are more secure than visible ones. Thought, decreasing the distortion level implies increases the watermarking capacity.

3. The robustness is the main criterion that should be achieved by the designed model. In fact, the resistance to all attacks is not possible, instead, the watermark should survey the common image processing manipulations.

In summary, to overcome the above-mentioned challenges a compromise between all the above requirements should be considered to develop an efficient, robust watermark embedding scheme, the following requirement should be addressed simultaneously: i) the high embedding capacity, ii) the low embedding distortion, and iii) the good robustness against attacks.

In the rest of this chapter, a dual color image watermarking scheme is developed, computational analysis, and performance evaluation is reviewed, starting with verifying the features and the components of the developed scheme. Then, experimental results are conducted to testify the validity and the suitability of the proposed scheme, moreover, its performance is compared with similar schemes.

6.4 Proposed Blind Dual-color Images Watermarking

In this section, a blind watermarking algorithm for embedding color image watermark into color host image is proposed. Its main features is to transform the host image via IWT after sub-sampling it, obtained sub-bands are used to embed the watermark, exploiting the advantage of high correlation between LL sub-band coefficients.

6.4.1 Features of the Proposed Scheme

The proposed dual-color image watermarking scheme offers several features in addressing the contradictory requirements and challenges. The main features of the proposed scheme are described below.

1. **Consideration of the image properties:** Since the proposed model is a dual color image watermarking scheme, the image properties should be taken into account when designing this scheme, including:
   - Considering the image integer value type to find the appropriate technique and avoid transforms that deal with real or complex values type (See Chapter 2), by the way, the rounding errors are avoided, which why the IWT transform offers an appropriate choice.
   - Considering the image edge when embedding the watermark, where, an edge mask is required to select the best location to embed the watermark without affecting the visual quality of the host image.
   - Choosing the appropriate transform coefficient to be modified while ensuring acceptable robustness and significant invisibility.
   - Considering the image redundancy to extract the watermark in a blind manner.
2. **High embedding capacity:** Embedding the watermark values in the LL coefficient of the IWT transform offers relatively high resistance against attacks, high invisibility beside the significant capacity.
3. **Security enhancement and robustness:** A particular security level can further improve the watermarking scheme, it can be obtained by using a suitable cryptographic technique. The proposed scheme increases the security by permuting the watermark sequence before embedding it.
4. **Robustness:** In our scheme, robustness against several attacks is achieved. The high quality of the extracted watermark especially under the cropping, salt and pepper noise and compression attacks proves the robustness and the effectiveness of the proposed method. This good robustness is offered by the use of image sub-sampling.

Well understanding the proposed scheme requires to cover its background in the following section.

6.4.2 Preliminaries and Background of the Proposed Scheme

To increase the robustness and the embedding capacity of watermarking techniques, many researches have been done [152, 169, 31] to take advantages of the perceptual properties of the human visual system (HVS). Thus, developing and improving accurate human vision model has the main role in designing a perceptual mask that can be used to better embed the watermark information while preserving the invisibility, the robustness and, security requirements [129].

Image Sub-sampling

Since the pixel values of any image have a good correlation with their pixel values neighboring, the sub-sampling allows as mapping the pixel values of any image into four highly correlated sub-images.

Given the image V of size $N \times M$, $N = 0...n-1, M = 0,..,m-1$, the sub-sampling process is as follows:

$$\begin{aligned} V_1[k_1;k_2] &= V[2k_1;2k_2] \\ V_2[k_1;k_2] &= V[2k_1+1;2k_2] \\ V_3[k_1;k_2] &= V[2k_1;2k_2+1] \\ V_4[k_1;k_2] &= V[2k_1+1;2k_2+1] \end{aligned} \tag{6.1}$$

For $k_1 = 0,..., \frac{N}{2}-1, k_2 = 0,..., \frac{M}{2}-1$, $V_i[k_1,k_2], i = 1,2,3,4$ are four sub-images obtained by sub-sampling Fig.6.1. Since the sub-images V_i's are highly correlated, it is expected that the IWT coefficients of different sub-images are approximately equal, $Y_i \approx Y_j$, for $i \neq j$ except edge, where Y_i and Y_j denote that the IWT of the sub-image V_i and V_j [98, 161].

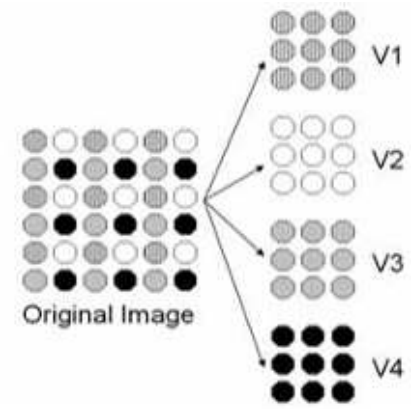


FIGURE 6.1: Sub-sampling image into four sub-images

Edge Masking

As it discussed, the IWT sub-images have approximately same coefficients at the same spatial location except edge areas [110, 98, 161]. To deal with this problem, the proposed method uses an edge mask as in Fig.6.2. It is used to consider the edge of local image characteristics [110]. The edge mask is constructed according to the pixel-wise relationship between the selected sub-band and neighboring two sub-bands.

Once we select a sub-image to compare with for later watermark extraction, we have to select two neighboring sub-images to embed the watermark. Using the edge mask we select just an horizontal or vertical neighboring sub-image for be modified.

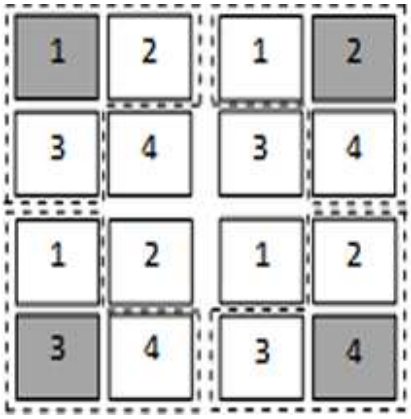


FIGURE 6.2: Selecting neighboring sub-images (two white blocks enclosed in dashed line are neighboring sub-images of gray block sub-image)

In support of the above features, components of the proposed scheme are presented in Sections 6.4.3 and 6.4.4, experimental results and comparison with similar schemes will be presented in Section 6.5.

The proposed method is composed of two main component: watermark embedding and extraction processes. The detailed steps of the proposed method are shown in Fig.(6.4) and Fig.(6.5), and they will be described in the incoming processes:

6.4.3 Watermark Embedding Process

1. Firstly, the 24-bit $P \times Q$ color watermark image W is divided into R, G, and B components; every component is encoded to $1D$ sequence, then in order to further remove the space correlations between the watermark coefficients and enhance the robustness of watermarking the $1D$ sequence is randomly permuted by a secret key K.

2. Next, the $24-bit$ $M \times N$ color host image I is divided into R, G, and B components, every component is decomposed into four sub-images (See Eq.6.1), each of the four sub-images is transformed via IWT to obtain four LL_i's sub-bands, one of which is selected for embedding a watermark.

3. Then, one pair of coefficients from two different sub-images situated in the same LL sub-band location is used to insert one watermark value, according to the edge mask (See Fig.6.3 shows an example of embedding procedure, where (a)-(c) represent the LL sub-bands of Y_1 and two neighboring LL sub-bands of Y_2 and Y_3 respectively.

4. According to the mask the LL_j sub-band values are modified according to the following equation:

$$LL_{\text{Mask(i)}} = LL_i + \alpha W_i \tag{6.2}$$

Where LL_i is the chosen sub-band to compare with, $Mask(i) = \{H, V\}$, H and V are horizontal or vertical neighboring respectively, the positive constant α is known as watermark strength control variable, the choice of α is a tradeoff between image distortion (watermark invisibility) and detection accuracy (robustness).

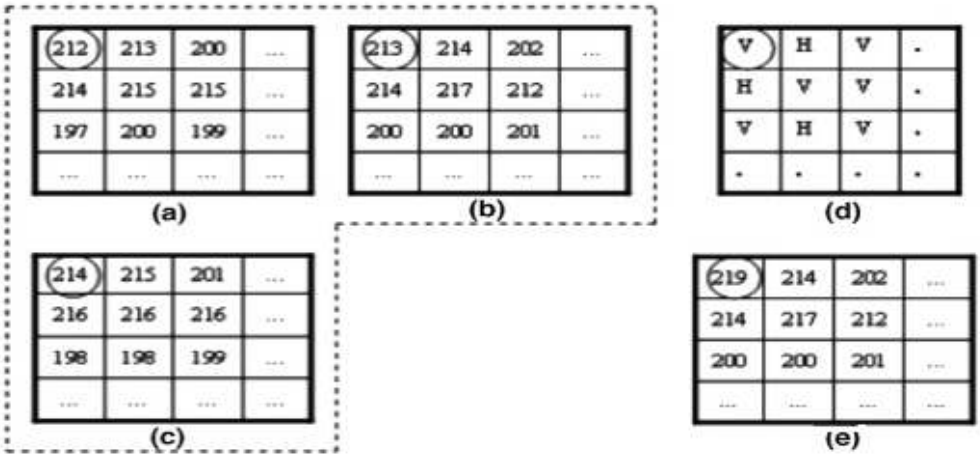


FIGURE 6.3: Watermark embedding example:(a) LL_1 sub-band of Y_1, (b) LL_2 sub-band of Y_2, (c) LL_3 sub-band of Y_3, (d) the edge mask, (e) the modified coefficients

5. Once the watermark is embedded, every sub-image is firstly transformed by the inverse IWT, and then, the inverse sub-sampling is applied to obtain the watermarked image I_w. Embedding steps are described in Fig.6.4.

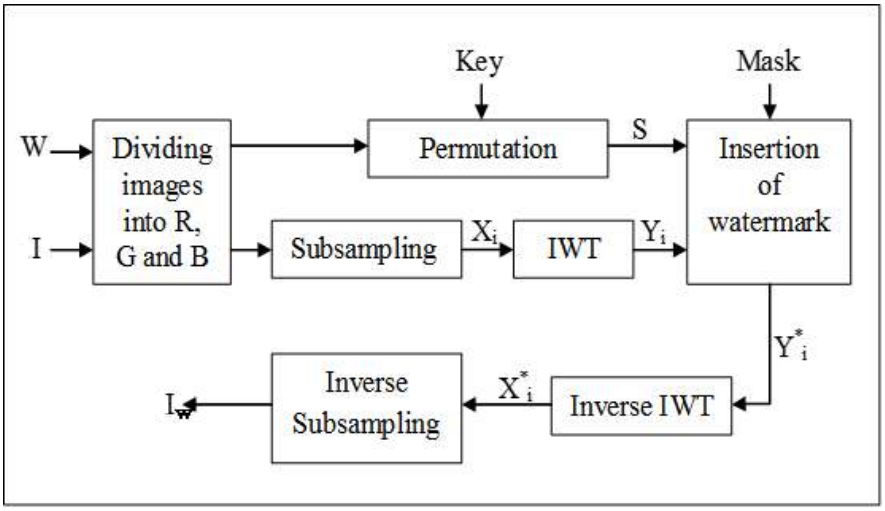


FIGURE 6.4: Watermark embedding process

6.4.4 Watermark Extraction Process

The procedure of watermark extraction is the inverse operation of the embedding process Fig.6.5 depict the detailed steps.

1. The watermarked image $I^{'}$ (probably attacked) is firstly sub-sampled into four sub-images.

2. Each of sub-image is transformed via IWT. And the same watermark insertion order sequence is required to select the chosen sub-band LL_i and two neighboring LL_j and LL_k sub-bands.

3. To recover the watermark Eq.6.3 is performed:

$$W_i^{'} = (LL^{*}_{\text{Mask(i)}} - LL^{*}_{\text{i}}) / \alpha \tag{6.3}$$

Where $W_i^{'}$ is the extracted watermark value, $Mask(i) = H, V$ is the edge mask value, $LL^{*}_{\text{Mask(i)}}$ is the neighboring value of the watermarked image, chosen according to the edge mask value. LL^{*}_{i} is the chosen sub-band to compare with.

4. Then the next process is to inverse the permutation for extracting a visually recognizable watermark.

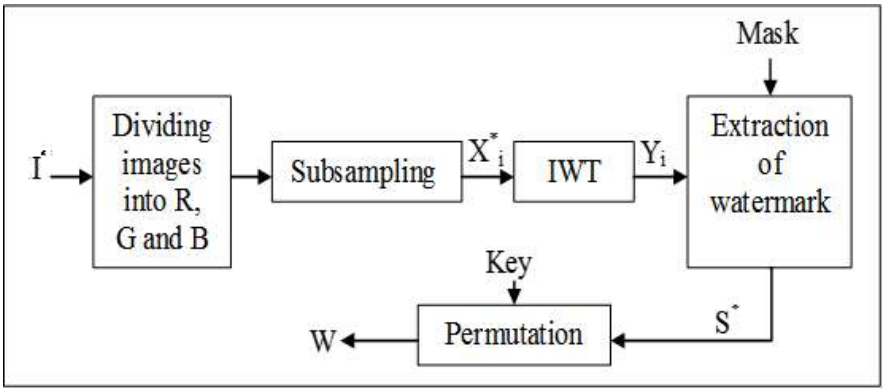


FIGURE 6.5: Watermark extraction process

6.5 Experimental Results

A set of $24-bit$ color images in the $CVG-URG$ database [38] with size of 512×512 are selected as original host image, and two $24-bit$ color images with size of 64×64 are used as original watermarks Fig. 6.6.

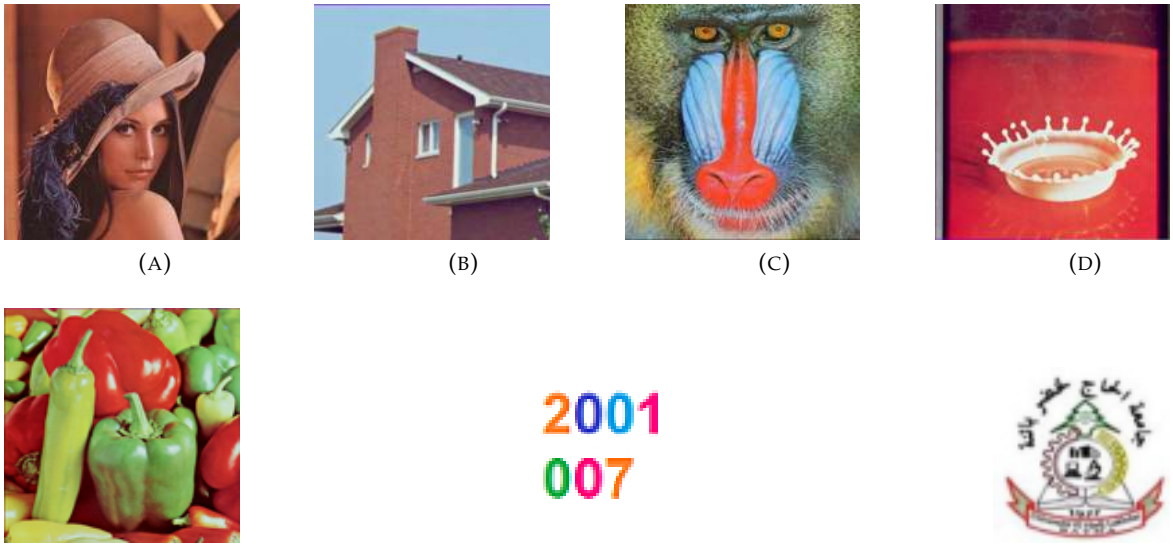
6.5.1 Parameters Selection and Payload

The strength of watermark $\alpha = 0.05$ chosen according to a tradeoff between image distortion and robustness as it shown in Fig.6.7, in other words, when the strength of watermark α increase the invisibility of watermark decrease and conversely.

For the imperceptible capability, we make use the peak signal to noise ratio $(PSNR)$ (See Chapter 3) in units of (dB). In general, if the $PSNR$ value is greater than 35 (dB) the perceptual quality is acceptable, that means that the watermark is invisible to human eyes [110, 150].

Besides, the use of the normalized correlation (NC), between the original watermark W and the extracted watermark $W^{'}$, to measure the robustness of watermarking. Generally, the NC can take a value between 0 and 1. If the NC value is closer to 1, the extracted watermark is getting more similar to the embedded one, which means that the watermarking has strong robustness.

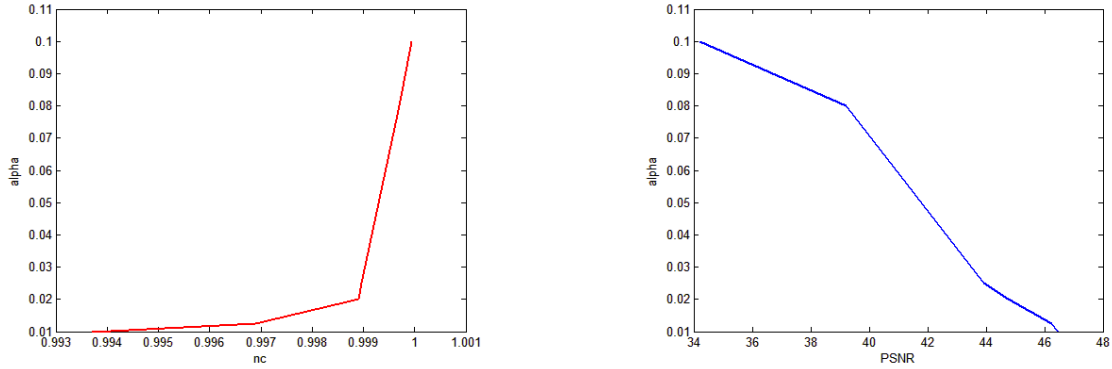
In general, the watermark may be efficient if the NC is more than or equal to 0.750, conversely may be inefficient [151].



(A) (B) (C) (D)

(E) (F) (G)

FIGURE 6.6: Original host images: (A) Lena, (B) House, (C) Baboon, (D) Splash, (E) Peppers, (F), (G) watermarks



(A) NC values (B) PSNR values

FIGURE 6.7: Tradeoff between watermark invisibility and detection accuracy curve

6.5.2 Performance Evaluation

The performance of the proposed scheme has been evaluated in terms of watermarked images quality and robustness against attacks. For this, we separated the tests into two parts: the first is to analyze the imperceptibility and the second is the evaluation of the robustness of watermark under several attacks. Furthermore, the invisibility and robustness of the proposed method are proved; by comparing it with the performance of the similar algorithms such as Chou and Wu [26] and Su et al. [151] schemes.

Quality of the watermarked image

As can be seen from Table.6.1, embedded watermarks are visually transparent and the watermarked images have high $PSNR$ values with an average of $38.8259(dB)$, meaning that the distortions level of the embedded watermark remain low despite the significant amount of data of color image watermark.

On the other hand, our proposed scheme is compared with the dual color image watermarking algorithms in [26, 151], it can be seen from Table.6.1 that the two algorithms have higher $PSNR$ values, despite that the obtained $PSNR$ values with our method still high and acceptable, moreover, our proposed method is the less complicated one, achieving lower computational complexity and lower computational cost, which it is a better compromise since the $PSNR$ values remain greater than $35(dB)$.

Robustness tests

From obtained results in Table.6.1, we can conclude that without any attacks performed on the watermarked images, the extracted watermarks quality remain high with an NC average of 0.9999. Moreover, compared with the exiting method [151], the NC value of extracted watermarks using the proposed scheme are usually higher, which also prove effectiveness of the proposed scheme compared with Su et al. [151] scheme.

On the other hand, to show the robustness of the proposed scheme, various attacks such as compression, cropping, adding noise, scaling and rotation are performed on the watermarked image. The obtained results can be seen in Table.6.2 and 6.3.

From the comparison results obtained in Table.6.2, compared with the NC values obtained with Chou and Wu [26] and Su et al. [151], it is obvious that the proposed method have higher robustness against tested attacks, except rotation attacks where the three schemes fail.

For better illustration, the proposed method is tested against more attacks to show its performance, Tables. 6.3 and 6.4 show obtained results.

From obtained results in Tables.6.3 and 6.4 , in cropping attacks the average of NC values is 0.78, the attack can damage 50% of the image and the NC value still high which equal to 0.70, in noising attacks the average of NC values is 0.80, which is also high since the variance can achieve 25%, however, compression attacks with low ratios can damage the watermark, as it can viewed Table.6.3, while NC values of the extracted watermarks under compression attacks with high ratios are high and can achieve 0.98. Moreover, under scaling attacks NC values are high (0.94), while rotation attacks damage the watermark.

We should notice that despite the considerable damage caused, the proposed scheme have a good resistance, especially against attacks including: the cropping, salt and peppers noise, and compression attacks, because of the use of image sub-sampling and IWT transform, while the failure of extraction after rotation with large angles is due the large changes of the watermarked image caused by the rotation.

6.6 Conclusion

This chapter has presented a blind dual-color images watermarking scheme based on IWT and sub-sampling, for being used in applications that require the robustness as the main criteria such as traceability and copyright protection.

Aiming at developing a robust watermarking scheme several requirements, challenges, and limitations of existing schemes are discussed. The analysis and experiments of the developed scheme have been presented, and its efficiency and suitability for images are validated. Moreover, the analysis and performance comparison of similar schemes (Chou and Wu [26] and [151]) is performed. Specifically:

- A performance comparison with similar schemes has been reviewed in light of the specified challenges. Thereby, the need for an efficient, robust scheme has been identified.

- Thereby, the existing literature schemes have been shown limitations of low resistance against several attacks (rotation, scaling...).
- Considering the computational aspects and capacity of the proposed scheme has been conducted and compared with similar schemes.
- Considering the perceptual aspects, the performance of the embedding scheme has been evaluated and compared with the existing schemes.

Consequently, a robust, frequency, blind, dual color image watermarking scheme has been developed. Its implementation, technical properties, and features are described to demonstrate its efficiency.

Experimental results are concluded to prove the effectiveness of the proposed scheme and show the major contributions of this paper includes: i) the consideration of color images properties to embed the watermark invisibly ($PSNR = 38.82(dB)$), ii) the exploitation of the IWT characteristics, to increase the watermark robustness and decrease the distortions, and iii) a robust method of embedding color image watermark into color host image. Finally, it has been concluded from the analysis and experimental results that the proposed scheme satisfies three principal features: i) the high quality of the watermarked image, ii) the significant robustness against several attacks, and iii) the considerable embedding capacity.

Thus, the proposed scheme can be used in many applications where the robustness is required like traceability and copyright protection. However, the proposed scheme can be more enhanced. Color image characteristics can further be exploited to reduce the watermark size, subsequently, the payload decreases and the imperceptibility increases.

TABLE 6.1: The performance comparison between proposed algorithm and the proposed one in Su et al. [151] without performing attacks on the watermarked image

Su et al. [151]		Proposed scheme	
Watermarked image PSNR (dB)	Extracted watermark (NC)	Watermarked image PSNR (dB)	Extracted watermark (NC)
			
48.6235	**0.9997**	38.9599	0.9995
			
40.3840	0.9960	38.4909	**0.9999**
			
46.4410	0.9984	39.1033	**1.000**
			
43.4647	0.9983	38.0118	**0.9999**
			
45.5024	**1.000**	39.2698	**1.000**
			
46.4709	0.9969	39.1197	**1.000**

TABLE 6.2: The performance comparison between proposed algorithm and the proposed one in Chou and Wu [26] and in Su et al. [151]

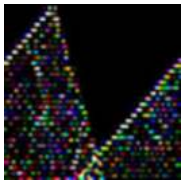
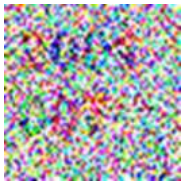
Host image	Attacks	Chou and Wu [26]		Su et al. [151]		Proposed method	
		Extracted watermark	NC	Extracted watermark	NC	Extracted watermark	NC
	Cropping 50%		0.539		0.642		**0.709**
	Scaling 4		0.851		0.946		**0.996**
	Rotation 30		-		-		-
	Gaussian noise (0.002)		0.982		**0.946**		0.864

TABLE 6.3: The results of extracted watermark (NC) under Cropping attacks

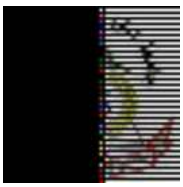
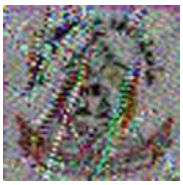
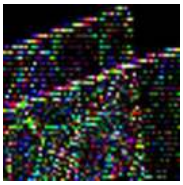
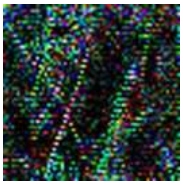
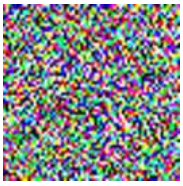
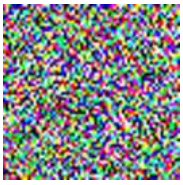
Attacks	Attacked Image (PSNR=39.2169)	Extracted watermark	NC
Without attacks			0.9997
Cropping			0.9440
			0.9349
			0.8415
			0.7142
			0.7054
			0.4651

TABLE 6.4: The results of extracted watermark (NC) under different attacks

Attacks	Attacked Image (PSNR=39.2169)	Extracted watermark	NC
Scaling 4			0.9413
Rotation 0.002			0.7425
Rotation 10			-
Compression ratio 60			0.9848
Compression ratio 10			0.5716
Salt& peppers			0.9880
Gaussian noise variance 0.01			0.7058
Gaussian noise variance 25			0.7058

Chapter 7

Conclusions and Future Research

7.1 Conclusion

The research presented in this thesis aimed to investigate digital watermarking to provide an authentication service for designing, analyzing, and applying it, with a particular focus on medical image security. Moreover, robust watermarking is also reviewed to be applied for traceability and copyright protection application.

One main application of watermarking is strict authentication, it is an outcome of advanced imaging technology. Thereby, with the easy access and distribution, images are subject to growing security threats with legal and ethical complexities. Under this circumstance, while the images security becomes critical, the conventional tools fail to providing the complete and necessary protection. Thus, digital watermarking has thus been demonstrated as a solution to address this issue.

Digital watermarking is a promising technology for addressing multimedia security. However, it has not been widely adopted in practice. Existing watermarking schemes often suffer from technical and security flaws. Validation of the suitability of those schemes for an application becomes more challenging. One main reason for these problems is the compromise between the several watermarking properties (e.g., embedding capacity, robustness, security).

In addressing the above-mentioned gaps, several original contributions have been made and presented in this thesis. Starting with a literature review on digital watermarking schemes, their requirements, and applications (Chapter 2), and theoretical aspects and their practical interpretations for several applications including image authentication application, main watermarking requirements, possible attacks, review of image authentication schemes, and the future of watermarking (Chapter 3). Thereby the research gaps have been verified, which has led this research to come up with the following main contributions:

1. A novel self-embedding watermarking scheme aims to authenticate the content of a watermarked image and to detect any possible alterations and recover damaged area. The novelty of the watermarking model has been verified by a comprehensive comparative study of literature. Thus, the existing schemes have common limitations: lacking the ability of tamper recovery, uses only gray-scale images, or even fail in tamper detection.

   Therefore, the general watermarking principle has formally been conceptualized in this thesis by developing a scheme that improves the performance in terms of imperceptibility and robustness. Additionally, the proposed scheme focuses on three major considerations: i) the invisibility of the embedded watermark, ii) the accuracy of detection, and iii) the high quality of the recovered color image.

   Experimental results are concluded to prove the effectiveness of the proposed method and show the major contributions of this scheme includes: firstly, the simple structure

of the self-embedding watermark information for color images that are visually indistinguishable with $PSNR$ values greater than 44 (dB). Secondly, the exploitation of the color images characteristics, to reduce the watermark to a gray-scale image using the Bayer pattern and Torus Automorphism to perform tamper detection with high detection accuracy reaches up 100% and low false alarm ratio less than 2%; and finally, a method of recovering the tampered image by interpolating the extracted gray-level watermark has also been devised for the quality improvement of the restored color image which can achieve 34 (dB) with tampering rates of 25%.

2. Moreover, in support of developing a strict-authentication scheme for medical image applications, a novel digital signature-based scheme using Cholesky decomposition is developed, which overcomes the limitations of existing schemes. The novelty of the presented counterfeiting attacks has been demonstrated in the light of the existing attacks.

   Experimental results demonstrate that the novel scheme has higher capacity distortion and more efficient detection ability. Changing only one bit of the whole image shows low correlation values can achieve -0.0921 despite the $PSNR$ values that achieving 96.2956 (dB). Thus, being irreversible and Cholesky decomposition based, this novel scheme would help provide strict authentication and detect any modifications and thereby has the potential to be applicable for medical image applications.

3. To avoid the identified security problems, a new blind dual-color image watermarking scheme has been presented, which is different from some existing schemes that use the binary or gray-level image as watermark. A set of requirements and objectives have been outlined. Thereby, new construction of an IWT-based watermarking scheme has been proposed to attain the identified requirements and objectives, and to resist the proposed counterfeiting attacks. The strong correlation property between the LL subbands coefficients of the IWT (Integer Wavelet Transform) is exploited for embedding and extracting watermark in a blind way. Experimental results demonstrate that the proposed scheme has a high robustness against most common attacks such as image compression, cropping, noising, scaling, etc.

The above-mentioned contributions, represent an advance in the area of digital image watermarking and its applications. In summary, contributions, theoretical developments and analysis, findings, and experimental evidence of this thesis represent a comprehensive information source, that can be used for future watermarking schemes researches.

7.2 Perspectives

A number of possible avenues for future research have been identified. These are summarized below.

1. An efficient watermarking scheme may have some relationship with the host image on which it is going to apply. Moreover, its performance or at least the selection of the input parameters must be related to image characteristics.

2. Using the Cholesky decomposition properties to detect and restore tampered areas besides the exploitation of medical image properties to embed the authentication data into the host image.

3. Further studies may be conducted to distinguish the effect of each attack on the watermarked images, so, developed watermarking schemes that could minimize those impacts to perform a better recovery.

4. Further development and validation of the watermarking model for medical image application could be an interesting area for future research.

Part III

Appendices

Appendix A

Peer Reviewed Publications

A.1 Conference Papers:

[166] **Wassila Belferdi**, Lemnouar Noui, Ali Behloul. "A Novel Cholesky Decomposition-based Scheme for Strict Image Authentication". In: Proceedings of the 2nd international Conference on Pattern Analysis and Intelligent Systems PAIS'2016. Khenchla,IEEE Algeria Section.16–17 Nov 2016, pp. 17–22.

[167] **Wassila Belferdi**, Ali Behloul, Lemnouar Noui. "A blind dual color images watermarking based on IWT and sub-sampling". In: Proceedings of the 3rd international Conference on Complex Systems CISC'2014. Jijel.9–10 Dec 2014, pp. 29–34.

[165] **Wassila Belferdi**, Ali Behloul. "Protection of digital watermarking based on SVD against false positive detection vulnerability". In: Proceedings of the 1st international Conference on Advanced Communication and Information Systems ICASIS'2012. Batna University. 12–13 Dec 2012, pp. 22–27.

A.2 Journal Papers:

[168] **Wassila Belferdi**, Ali Behloul, Lemnouar Noui. "A Bayer Pattern-based Fragile Watermarking Scheme for Color Image Tamper Detection and Restoration". In: Multidimensional Systems and Signal Processing 26.4 (2018). DOI: 10.1007/s11045-018-0597-x.

Bibliography

[1] Bassem Abdel-Aziz and J-Y Chouinard. "Performance analysis of a content authentication semifragile watermark". In: *Electrical and Computer Engineering, 2003. IEEE CCECE 2003. Canadian Conference on*. Vol. 3. IEEE. 2003, pp. 2055–2058. DOI: 10.1109/CCECE.2003.1226320.

[2] WA Wan Adnan et al. "A review of image watermarking". In: *Research and Development, 2003. SCORED 2003. Proceedings. Student Conference on*. IEEE. 2003, pp. 381–384.

[3] van den Bos Adriaan. "Positive Semidefinite and Positive Definite Matrices". In: *Parameter Estimation for Scientists and Engineers* (2007), pp. 259–263. DOI: 10.1002/9780470173862.app3.

[4] A Ajala Funmilola, I Ojebamigbe Victoria, and O Adegoke Benjamin. "Medical Image Authentication and Restoration Using Block-Wise Fragile Watermarking and Clustering Approach". In: *International Journal of Science and Engineering Investigations* 7.80 (2018), pp. 91–97.

[5] Osamah M Al-Qershi and Bee Ee Khoo. "Authentication and data hiding using a hybrid ROI-based watermarking scheme for DICOM images". In: *Journal of Digital Imaging* 24.1 (2011), pp. 114–125.

[6] Adnan M Alattar. "Smart images using Digimarc's watermarking technology". In: *Security and Watermarking of Multimedia Contents II*. Vol. 3971. International Society for Optics and Photonics. 2000, pp. 264–274.

[7] Irshad Ahmad Ansari, Millie Pant, and Chang Wook Ahn. "Robust and false positive free watermarking in IWT domain using SVD and ABC". In: *Engineering Applications of Artificial Intelligence* 49 (2016), pp. 114–125. DOI: 10.1016/j.engappai.2015.12.004.

[8] Arnold Baldoza and Michael Sieffert. "Methods for detecting tampering in digital images". In: *AFRL Technology Horizons* 1.1 (2000), pp. 15–17.

[9] Kaur Baljit and Sharma Sonia. "Digital Watermarking and Security Techniques: A Review". In: *International Journal of Computer Science and Technology* 8.2 (2017), pp. 44–47.

[10] Paul Bao and Xiaohu Ma. "Image adaptive watermarking using wavelet domain singular value decomposition". In: *IEEE transactions on circuits and systems for video technology* 15.1 (2005), pp. 96–102.

[11] Mauro Barni, Franco Bartolini, and Alessandro Piva. "Improved wavelet-based watermarking through pixel-wise masking". In: *IEEE transactions on image processing* 10.5 (2001), pp. 783–791.

[12] Mauro Barni et al. "A DCT-domain system for robust image watermarking". In: *Signal processing* 66.3 (1998), pp. 357–372.

[13] Oliver Benedens. *Geometry-based watermarking of 3D models*. Tech. rep. Fraunhofer Inst for Computer Graphics Darmstadt (Germany) Virtual Reality Demonstration Centre, 1999.

[14] Dan Boneh and James Shaw. "Collusion-secure fingerprinting for digital data". In: *IEEE Transactions on Information Theory* 44.5 (1998), pp. 1897–1905.

[15] Laurence Boney, Ahmed H Tewfik, and Khaled N Hamdy. "Digital watermarks for audio signals". In: *Multimedia Computing and Systems, 1996., Proceedings of the Third IEEE International Conference on*. IEEE. 1996, pp. 473–480. DOI: 10.1109/MMCS.1996.535015.

[16] Mahsa Boreiry and Mohammad-Reza Keyvanpour. "Classification of watermarking methods based on watermarking approaches". In: *Artificial Intelligence and Robotics (IRANOPEN), 2017*. IEEE. 2017, pp. 73–76.

[17] Jack Brassil et al. "Electronic marking and identification techniques to discourage document copying". In: *INFOCOM'94. Networking for Global Communications., 13th Proceedings IEEE*. IEEE. 1994, pp. 1278–1287.

[18] Jack T Brassil et al. "Electronic marking and identification techniques to discourage document copying". In: *IEEE Journal on Selected Areas in Communications* 13.8 (1995), pp. 1495–1504.

[19] Sergio Bravo-Solorio and Asoke K. Nandi. "Secure fragile watermarking method for image authentication with improved tampering localisation and self-recovery capabilities". In: *Signal Processing* 91.4 (Apr. 2011), pp. 728–739. DOI: 10.1016/j.sigpro.2010.07.019.

[20] Sung-Cheal Byun et al. "A public-key based watermarking for color image authentication". In: *Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on*. Vol. 1. IEEE. 2002, pp. 593–596.

[21] Chi-Shiang Chan and Chin-Chen Chang. "An efficient image authentication method based on Hamming code". In: *Pattern Recognition* 40.2 (Feb. 2007), pp. 681–690. DOI: 10.1016/j.patcog.2006.05.018.

[22] Chin-Chen Chang, Yih-Shin Hu, and Tzu-Chuen Lu. "A watermarking-based image ownership and tampering authentication scheme". In: *Pattern Recognition Letters* 27.5 (Apr. 2006), pp. 439–446. DOI: 10.1016/j.patrec.2005.09.006.

[23] Chun-Hung Chen, Yuan-Liang Tang, and Wen-Shyong Hsieh. "Color Image Authentication and Recovery via Adaptive Encoding". In: *Computer, Consumer and Control (IS3C), 2014 International Symposium on*. IEEE, June 2014, pp. 272–275. DOI: 10.1109/IS3C.2014.79.

[24] Tao Chen, Jingchun Wang, and Yonglei Zhou. "Combined digital signature and digital watermark scheme for image authentication". In: *Info-tech and Info-net, 2001. Proceedings. ICII 2001-Beijing. 2001 International Conferences on*. Vol. 5. IEEE. 2001, pp. 78–82.

[25] Dae-Jea Cho. "A Study on performance evaluation-Metrics for digital watermarking algorithms". In: *Advanced Science and Technology letters* 78 (2014), pp. 73–76. DOI: 10.14257/astl.2014.78.14.

[26] Chun-Hsien Chou and Tung-Lin Wu. "Embedding color watermarks in color images". In: *EURASIP Journal on Advances in Signal Processing* 2003.1 (2003), 32–40.

[27] Gouenou Coatrieux, Catherine Quantin, and François-André Allaert. "Watermarking as a traceability standard". In: *Studies in health technology and informatics* 180 (2012), pp. 761–765. DOI: 10.3233/978-1-61499-101-4-761.

[28] Aaron S Cohen and Amos Lapidoth. "The Gaussian watermarking game". In: *IEEE Transactions on Information Theory* 48.6 (2002), pp. 1639–1667. DOI: 10.1109/TIT.2002.1003844.

[29] *Collection of DICOM images*. Dec. 28, 2015. URL: https://www.aycan.de/sample-dicom-images.html (visited on 01/01/2019).

[30] Ingemar Cox et al. *Digital watermarking and steganography*. Morgan Kaufmann, 2007. DOI: 10.1016/B978-012372585-1.50005-X.

[31] Ingemar J Cox and Matt L Miller. "Review of watermarking and the importance of perceptual modeling". In: *Human Vision and Electronic Imaging II*. Vol. 3016. International Society for Optics and Photonics. 1997, pp. 92–100.

[32] Ingemar J Cox and Matt L Miller. "The first 50 years of electronic watermarking". In: *EURASIP Journal on Advances in Signal Processing* 2002.2 (2002), 126–132.

[33] Ingemar J. Cox, Matt L. Miller, and Jeffrey A. Bloom. "Watermarking applications and their properties". In: *itcc*. IEEE, 2000, p. 6. DOI: 10.1109/ITCC.2000.844175.

[34] Ingemar J Cox et al. "Secure spread spectrum watermarking for multimedia". In: *IEEE transactions on image processing* 6.12 (1997), pp. 1673–1687.

[35] Ingemar J. Cox et al. *Digital watermarking*. Springer, 2002.

[36] Scott Craver et al. "Resolving rightful ownerships with invisible watermarking techniques: limitations, attacks, and implications". In: *IEEE Journal on Selected areas in Communications* 16.4 (1998), pp. 573–586.

[37] Scott A Craver et al. "Can invisible watermarks resolve rightful ownerships?" In: *Storage and Retrieval for Image and Video Databases V*. Vol. 3022. International Society for Optics and Photonics. 1997, pp. 310–322.

[38] *CVG - UGR - Image database*. Dec. 2, 2016. URL: https://decsai.ugr.es/cvg/dbimagenes/ (visited on 06/05/2018).

[39] Chinmayee Das et al. "A novel blind robust image watermarking in DCT domain using inter-block coefficient correlation". In: *AEU-International Journal of Electronics and Communications* 68.3 (2014), pp. 244–253.

[40] Jana Dittmann, Arnd Steinmetz, and Ralf Steinmetz. "Content-based digital signature for motion pictures authentication and content-fragile watermarking". In: *Multimedia Computing and Systems, 1999. IEEE International Conference on*. Vol. 2. IEEE. 1999, pp. 209–213.

[41] Richard Edlin et al. "Correlated Parameters and the Cholesky Decomposition". In: *Cost Effectiveness Modelling for Health Technology Assessment*. Springer International Publishing, 2015, pp. 119–132. DOI: 10.1007/978-3-319-15744-3_8.

[42] SH El-Din and Mansour Moniri. "Fragile and semi-fragile image authentication based on image self-similarity". In: *Image Processing. 2002. Proceedings. 2002 International Conference on*. Vol. 2. IEEE. 2002, pp. II–II.

[43] Jessica Fridrich, Miroslav Goljan, and Nasir D Memon. "Further attacks on Yeung-Mintzer fragile watermarking scheme". In: *Security and Watermarking of Multimedia Contents II*. Vol. 3971. International Society for Optics and Photonics. 2000, pp. 428–438.

[44] Jiri Fridrich. "Image watermarking for tamper detection". In: *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*. Vol. 2. IEEE, 1998, pp. 404–408.

[45] Jiri Fridrich. "Protection of digital images using self-embedding". In: *Symposium on Content Security and Data Hiding in Digital Media, New Jersey Institute of Technology*. 1999.

[46] Jiri Fridrich and Miroslav Goljan. "Images with self-correcting capabilities". In: *Image Processing, 1999. ICIP 99. Proceedings. International Conference on*. Vol. 3. IEEE. 1999, pp. 792–796.

[47] Jiri Fridrich, Miroslav Goljan, and Arnold C Baldoza. "New fragile authentication watermark for images". In: *Proceedings. International Conference on Image Processing*. Vol. 1. IEEE. 2000, pp. 446–449.

[48] Gary L Friedman. "The trustworthy digital camera: Restoring credibility to the photographic image". In: *IEEE Transactions on consumer electronics* 39.4 (1993), pp. 905–910.

[49] Gary L Friedman. *Digital camera with apparatus for authentication of images produced from an image file*. US Patent 5,499,294. 1996.

[50] Charles Fung, Antonio Gortan, and Walter Godoy Junior. "A review study on image digital watermarking". In: *The Tenth International Conference on Networks*. The Tenth International Conference on Networks. 2011, pp. 24–28.

[51] Teddy Furon and Pierre Duhamel. "An asymmetric watermarking method". In: *IEEE Transactions on Signal Processing* 51.4 (2003), pp. 981–995.

[52] Teddy Furon, Ilaria Venturini, and Pierre Duhamel. "Unified approach of asymmetric watermarking schemes". In: *Security and Watermarking of Multimedia Contents III*. Vol. 4314. International Society for Optics and Photonics. 2001, pp. 269–280.

[53] Praveen Gauravaram and Lars R. Knudsen. "Cryptographic hash functions". In: *Handbook of Information and Communication Security*. Springer, 2010, pp. 59–79. DOI: 10.1007/978-3-642-04117-4_4.

[54] Kaiser J Giri, Mushtaq Ahmad Peer, and P Nagabhushan. "A robust color image watermarking scheme using discrete wavelet transformation". In: *IJ Image, Graphics and Signal Processing* 1 (2015), pp. 47–52. DOI: DOI:10.5815/ijigsp.2015.01.06.

[55] UM Gokhale and YV Joshi. "A semi fragile watermarking algorithm based on SVD-IWT for image authentication". In: *International Journal of Advanced Research in Computer and Communication Engineering* 1.4 (2012).

[56] Adil Haouzia and Rita Noumeir. "Methods for image authentication: a survey". In: *Multimedia tools and applications* 39.1 (2008), pp. 1–46. DOI: 10.1007/s11042-007-0154-3.

[57] Kevin Heylen and Tim Dams. "An image watermark tutorial tool using Matlab". In: *Mathematics of Data/Image Pattern Recognition, Compression, and Encryption with Applications XI*. Vol. 7075. International Society for Optics and Photonics. 2008, p. 70750D.

[58] Nicholas J. Higham. "Analysis of the Cholesky decomposition of a semi-definite matrix". In: (1990). (Visited on 03/03/2016).

[59] Vaishali S Jabade and Dr Sachin R Gengaje. "Literature review of wavelet based digital image watermarking techniques". In: *Int. J. Comput. Appl* 31.1 (2011), pp. 28–35.

[60] Nikil Jayant, James Johnston, and Robert Safranek. "Signal compression based on models of human perception". In: *Proceedings of the IEEE* 81.10 (1993), pp. 1385–1422.

[61] Ningombam Jimson and K Hemachandran. "DFT-based digital image watermarking: A survey." In: *International Journal of Advanced Research in Computer Science* 9.2 (2018).

[62] Ramesh Uttam Kadam et al. "Digital image watermarking". In: *International Journal of Emerging Technologies and Innovative Research JETIR*. Vol. 1. 5 (October-2014). JETIR. 2014.

[63] Andrew B Kahng et al. "Robust IP watermarking methodologies for physical design". In: *Proceedings of the 35th annual Design Automation Conference*. ACM. 1998, pp. 782–787.

[64] Stefan Katzenbeisser and Helmut Veith. "Securing symmetric watermarking schemes against protocol attacks". In: *Security and Watermarking of Multimedia Contents IV*. Vol. 4675. International Society for Optics and Photonics. 2002, pp. 260–269.

[65] Manpreet Kaur, Sonika Jindal, and Sunny Behal. "A study of digital image watermarking". In: *Journal of Research in Engineering and Applied Sciences* 2.2 (2012), pp. 126–136.

[66] Sandeep Kaur and Himanshu Jindal. "Enhanced Image Watermarking Technique using Wavelets and Interpolation". In: *International Journal of Image, Graphics and Signal Processing* 9.7 (2017), p. 23.

[67] Selena Kay and Ebroul Izquierdo. "Robust content based image watermarking". In: *Proc. Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS*. 2001.

[68] Shivani Khurana. "Watermarking and Information-Hiding". In: *International Journal of Computer and Information Technology* 2 (2011), pp. 1679–1681.

[69] Darko Kirovski and Henrique S Malvar. "Spread-spectrum watermarking of audio signals". In: *IEEE transactions on signal processing* 51.4 (2003), pp. 1020–1033.

[70] I Kostopoulos, SAM Gilani, and AN Skodras. "Colour image authentication based on a self-embedding technique". In: *Digital Signal Processing, 2002. DSP 2002. 2002 14th International Conference on*. Vol. 2. IEEE. 2002, pp. 733–736.

[71] I Kostopoulos, AN Skodras, and D Christodou-lakis. "Self-authentication of colour images". In: *Proceedings of the European conference on electronic imaging and visual arts, Florence, Italy*. 2001.

[72] Deepa Kundur and Dimitrios Hatzinakos. "Semi-blind image restoration based on telltale watermarking". In: *Signals, Systems & Computers, 1998. Conference Record of the Thirty-Second Asilomar Conference on*. Vol. 2. IEEE. 1998, pp. 933–937. DOI: 10.1109/ACSSC.1998.751399.

[73] Deepa Kundur and Dimitrios Hatzinakos. "Towards a telltale watermarking technique for tamper-proofing". In: *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No. 98CB36269)*. Vol. 2. IEEE. 1998, pp. 409–413.

[74] Deepa Kundur and Dimitrios Hatzinakos. "Digital watermarking for telltale tamper proofing and authentication". In: *Proceedings of the IEEE* 87.7 (1999), pp. 1167–1180.

[75] Martin Kutter. "Watermarking resistance to translation, rotation, and scaling". In: *Multimedia Systems and Applications*. Vol. 3528. International Society for Optics and Photonics. 1999, pp. 423–432.

[76] John Lach, William H Mangione-Smith, and Miodrag Potkonjak. "Fingerprinting digital circuits on programmable hardware". In: *International Workshop on Information Hiding*. Springer. 1998, pp. 16–31.

[77] Gerrit C Langelaar, Reginald L Lagendijk, and Jan Biemond. "Removing spatial spread spectrum watermarks by non-linear filtering". In: *Signal Processing Conference (EUSIPCO 1998), 9th European*. IEEE. 1998, pp. 1–4.

[78] Seungiae Lee, Dalwon Jang, and Chang D Yoo. "An SVD-based watermarking method for image content authentication with improved security". In: *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*. Vol. 2. IEEE. 2005, pp. ii–525.

[79] Chunlei Li et al. "A novel self-recovery fragile watermarking scheme based on dual-redundant-ring structure". In: *Computers & Electrical Engineering* 37.6 (Nov. 2011), pp. 927–940. DOI: 10.1016/j.compeleceng.2011.09.007.

[80] Chia-Chen Lin, Xiao-Long Liu, and Cheng-Han Lin. "SEMI-FRAGILE WATERMARKING SCHEME FOR CFA IMAGES BASED ON AUTHENTICATION TABLE". In: *Security and Management,.2018 International Conference on* (2018), pp. 195–201.

[81] Chih-Hung Lin, Tzung-Her Chen, and Chun-Wei Chiu. "Color image authentication with tamper detection and remedy based on BCH and Bayer Pattern". In: *Displays* 34.1 (Jan. 2013), pp. 59–68. DOI: 10.1016/j.displa.2012.11.004.

[82] Chih-Hung Lin and Ching-Yu Yang. "Multipurpose watermarking based on blind vector quantization (BVQ)". In: *Journal of Information Hiding and Multimedia Signal Processing* 2.2 (2011), pp. 239–246.

[83] Chih-Hung Lin, Ching-Yu Yang, and Chia-Wei Chang. "Authentication and protection for medical image". In: *International Conference on Computational Collective Intelligence*. Springer. 2010, pp. 278–287.

[84] Ching-Yung Lin and Shih-Fu Chang. "Semifragile watermarking for authenticating JPEG visual content". In: *Security and Watermarking of Multimedia Contents II*. Vol. 3971. International Society for Optics and Photonics. 2000, pp. 140–152.

[85] Ching-Yung Lin and Shih-Fu Chang. "A robust image authentication method distinguishing JPEG compression from malicious manipulation". In: *Circuits and Systems for Video Technology, IEEE Transactions on* 11.2 (2001), pp. 153–168. DOI: 10.1109/76.905982.

[86] Ching-Yung Lin and Shih-Fu Chang. "SARI: self-authentication-and-recovery image watermarking system". In: *Proceedings of the ninth ACM international conference on Multimedia*. ACM. 2001, pp. 628–629.

[87] Ching-Yung Lin et al. "Rotation, scale, and translation resilient watermarking for images". In: *IEEE Transactions on image processing* 10.5 (2001), pp. 767–782.

[88] Eugene T Lin and Edward J Delp. "A review of fragile image watermarks". In: *Proceedings of the Multimedia and Security Workshop (ACM Multimedia'99) Multimedia Contents*. Vol. 1. Citeseer. 1999, pp. 25–29.

[89] Eugene T Lin, Christine I Podilchuk, and Edward J Delp. "Detection of image alterations using semifragile watermarks". In: *Security and Watermarking of Multimedia Contents II*. Vol. 3971. International Society for Optics and Photonics. 2000, pp. 152–164.

[90] Hongwen Lin, Shaoqing Yang, and Linzhou Xu. "Watermark algorithm for color image authentication and restoration". In: *Electronic and Mechanical Engineering and Information Technology (EMEIT), 2011 International Conference on*. Vol. 6. IEEE. 2011, pp. 2773–2776. DOI: 10.1109/EMEIT.2011.6023677.

[91] Chen Ling and Obaid Ur-Rehman. "Watermarking for image authentication". In: *Robust Image Authentication in the Presence of Noise*. Springer, 2015, pp. 43–73.

[92] K.-C. Liu. "Colour image watermarking for tamper proofing and pattern-based recovery". In: *Image Processing, IET* 6.5 (2012), pp. 445–454. DOI: 10.1049/iet-ipr.2011.0574.

[93] Shao-Hui Liu et al. "An image fragile watermark scheme based on chaotic image pattern and pixel-pairs". In: *Applied Mathematics and Computation* 185.2 (Feb. 2007), pp. 869–882. DOI: 10.1016/j.amc.2006.07.036.

[94] Tong Liu and Zheng-ding Qiu. "The survey of digital watermarking-based image authentication techniques". In: *Signal Processing, 2002 6th International Conference on*. Vol. 2. IEEE. 2002, pp. 1556–1559.

[95] Chun-Chi Lo and Yu-Chen Hu. "A novel reversible image authentication scheme for digital images". In: *Signal processing* 98 (May 2014), pp. 174–185. DOI: 10.1016/j.sigpro.2013.11.028.

[96] Der-Chyuan Lou, Jiang-Lung Liu, and Chang-Tsun Li. "Digital signature-based image authentication". In: (2003).

[97] Steven H Low and Nicholas F Maxemchuk. "Performance comparison of two text marking methods". In: *IEEE Journal on Selected Areas in Communications* 16.4 (1998), pp. 561–572.

[98] Wei Lu, Hongtao Lu, and Fu-Lai Chung. "Robust digital image watermarking based on subsampling". In: *Applied mathematics and computation* 181.2 (2006), pp. 886–893.

[99] Zhe-Ming Lu, Dian-Guo Xu, and Sheng-He Sun. "Multipurpose image watermarking algorithm based on multistage vector quantization". In: *Image Processing, IEEE Transactions on* 14.6 (June 2005), pp. 822–831. DOI: 10.1109/TIP.2005.847324.

[100] Budimir Lutovac et al. "An algorithm for robust image watermarking based on the DCT and Zernike moments". In: *Multimedia tools and applications* 76.22 (2017), pp. 23333–23352.

[101] Toshihiko Matsuo and Kaoru Kurosawa. "On parallel hash functions based on block-ciphers". In: *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences* 87.1 (2004), pp. 67–74.

[102] Joceli Mayer and Rafael Araujo Silva. "Efficient informed embedding of multi-bit watermark". In: *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*. Vol. 3. IEEE. 2004, pp. iii–389.

[103] Nasir Memon, Sunil Shende, and Ping Wah Wong. "On the security of the Yeung-Mintzer authentication watermark". In: *PICS*. 1999, pp. 301–306.

[104] Nisar Ahmed Memon and Syed Asif Mahmood Gilani. "Watermarking of chest CT scan medical images for content authentication". In: *International Journal of Computer Mathematics* 88.2 (2011), pp. 265–280.

[105] Nisar Ahmed Memon et al. "Hybrid watermarking of medical images for ROI authentication and recovery". In: *International Journal of Computer Mathematics* 88.10 (2011), pp. 2057–2071.

[106] Alfred J. Menezes, Paul C. Van Oorschot, and Scott A. Vanstone. *Handbook of applied cryptography*. CRC press, 1996.

[107] Matt L Miller et al. "A review of watermarking principles and practices". In: *Digital signal processing in multimedia systems* (1999), pp. 461–485.

[108] Matthew L Miller and Ingemar J Cox. *Applying informed coding, informed embedding and perceptual shaping to design a robust, high-capacity watermark*. 2006.

[109] Fred Mintzer, Gordon W Braudaway, and Minerva M Yeung. "Effective and ineffective digital watermarks". In: *Image Processing, 1997. Proceedings., International Conference on*. Vol. 3. IEEE. 1997, pp. 9–12.

[110] Ho Seok Moon et al. "Expert system for low frequency adaptive image watermarking: Using psychological experiments on human image perception". In: *Expert Systems with Applications* 32.2 (2007), pp. 674–686.

[111] Seyed Mojtaba Mousavi, Alireza Naghsh, and SAR Abu-Bakar. "Watermarking techniques used in medical images: a survey". In: *Journal of digital imaging* 27.6 (2014), pp. 714–729. DOI: 10.1007/s10278-014-9700-5.

[112] Edin Muharemagic and Borko Furht. "Survey of watermarking techniques and applications". In: *Multimedia Watermarking Techniques and Applications* 3.91 (2006), p. 130.

[113] Rongrong Ni, Qiuqi Ruan, and Heng-Da Cheng. "Secure semi-blind watermarking based on iteration mapping and image features". In: *Pattern Recognition* 38.3 (2005), pp. 357–368.

[114] Athanasios Nikolaidis and Ioannis Pitas. "Region-based image watermarking". In: *IEEE Transactions on image processing* 10.11 (2001), pp. 1726–1740.

[115] Hussain Nyeem, Wageeh Boles, and Colin Boyd. "Digital image watermarking: its formal model, fundamental properties and possible attacks". In: *EURASIP Journal on Advances in Signal Processing* 1 (2014), p. 135.

[116] Hussain MD Abu Nyeem. "A digital watermarking framework with application to medical image security". PhD thesis. Queensland University of Technology, 2014.

[117] Ryutarou Ohbuchi, Hiroshi Masuda, and Masaki Aono. "Watermarking three-dimensional polygonal models through geometric and topological modifications". In: *IEEE Journal on selected areas in communications* 16.4 (1998), pp. 551–560.

[118] Alexandre Paquet. "Wavelet packets-based digital watermarking for image authentication". PhD thesis. University of British Columbia, 2002.

[119] Arvind Kumar Parthasarathy and Subhash Kak. "An improved method of content based image watermarking". In: *IEEE Transactions on broadcasting* 53.2 (2007), pp. 468–479.

[120] Dixa Patel et al. "Digital Video Watermarking: A Retrospective". In: *International Journal of Scientific and Engineering Research* 5.12 (2014), pp. 1–7.

[121] Soo-Chang Pei and Io-Kuong Tam. "Effective color interpolation in CCD color filter arrays using signal correlation". In: *Circuits and Systems for Video Technology, IEEE Transactions on* 13.6 (June 2003), pp. 503–513. DOI: 10.1109/TCSVT.2003.813422.

[122] Shelby Pereira and Thierry Pun. "Robust template matching for affine resistant image watermarks". In: *IEEE transactions on image Processing* 9.6 (2000), pp. 1123–1129.

[123] Luis Pérez-Freire et al. "Watermarking security: a survey". In: *Transactions on Data Hiding and Multimedia Security I*. Springer, 2006, pp. 41–72.

[124] Raphael C-W Phan. "Tampering with a watermarking-based image authentication scheme". In: *Pattern Recognition* 41.11 (2008), pp. 3493–3496.

[125] Christine I Podilchuk and Edward J Delp. "Digital watermarking: algorithms and applications". In: *IEEE signal processing Magazine* 18.4 (2001), pp. 33–46.

[126] Vidyasagar Potdar, Song Han, and Elizabeth Chang. "A survey of digital image watermarking techniques". In: *3rd IEEE International Conference on Industrial Informatics (INDIN 2005)*. IEEE. 2005, pp. 709–716.

[127] Chuan Qin, Chin-Chen Chang, and Pei-Yu Chen. "Self-embedding fragile watermarking with restoration capability based on adaptive bit allocation mechanism". In: *Signal Processing* 92.4 (Apr. 2012), pp. 1137–1150. DOI: 10.1016/j.sigpro.2011.11.013.

[128] Regunathan Radhakrishnan and Nasir Memon. "On the security of the SARI image authentication system". In: *Image Processing, 2001. Proceedings. 2001 International Conference on*. Vol. 3. IEEE. 2001, pp. 971–974.

[129] S Radharani and ML Valarmathi. "A study on watermarking schemes for image authentication". In: *International Journal of Computer Applications* 2.4 (2010), pp. 24–32.

[130] C. RajaRao, Mahesh Boddu, and Soumitra Kumar Mandal. "Single Sensor Color Filter Array Interpolation Algorithms". In: *Information Systems Design and Intelligent Applications*. Ed. by Suresh Chandra Mandal J. K.and Satapathy et al. Vol. 340. Springer India, 2015, pp. 295–307. ISBN: 978-81-322-2247-7. DOI: 10.1007/978-81-322-2247-7_31.

[131] Neha Rawat and Rachna Manchanda. "Review of Methodologies and Techniques for Digital Watermarking". In: *International Journal of Emerging Technology and Advanced Engineering, 4 (4)* 237 (2014).

[132] Christian Rey and J-L Dugelay. "Blind detection of malicious alterations on still images using robust watermarks". In: (2000).

[133] Christian Rey and Jean-Luc Dugelay. "Un panorama des méthodes de tatouage permettant d'assurer un service d'intégrité pour les images". In: *Traitement du Signal* 18.4 (2001), pp. 283–295.

[134] Christian Rey and Jean-Luc Dugelay. "A survey of watermarking algorithms for image authentication". In: *EURASIP Journal on Advances in Signal Processing* 2002.6 (2002), pp. 6–9. DOI: 10.1155/S1110865702204047.

[135] Ronald Rivest. *The MD4 message-digest algorithm*. Tech. rep. 1992.

[136] Phillip Rogaway and Thomas Shrimpton. "Cryptographic hash-function basics: Definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance". In: *Fast Software Encryption*. Springer, 2004, pp. 371–388. DOI: 10.1007/978-3-540-25937-4_24.

[137] N Rogier and Pascal Chauvaud. "MD2 is not secure without the checksum byte". In: *Designs, Codes and Cryptography* 12.3 (1997), pp. 245–251.

[138] Lalit Kumar Saini and Vishal Shrivastava. "A Survey of Digital Watermarking Techniques and its Applications". In: *International Journal of Computer Science Trends and Technology* 2.3 (2014), pp. 70–73.

[139] David Salomon. *Data compression: the complete reference*. Springer Science & Business Media, 2004.

[140] Praful Saxena, Shanon Garg, and Arpita Srivastava. "DWT-SVD semi-blind image watermarking using high frequency band". In: *2nd International Conference on Computer Science and Information Technology (ICCSIT'2012) Singapore April*. Vol. 28. 2012, p. 29.

[141] Marc Schneider and Shih-Fu Chang. "A robust content based digital signature for image authentication". In: *Image Processing, 1996. Proceedings., International Conference on*. Vol. 3. IEEE. 1996, pp. 227–230.

[142] Prabhishek Singh and RS Chadha. "A survey of digital watermarking techniques, applications and attacks". In: *International Journal of Engineering and Innovative Technology (IJEIT)* 2.9 (2013), pp. 165–175.

[143] Vassilios Solachidis and Ioannis Pitas. "Watermarking polygonal lines using Fourier descriptors". In: *IEEE computer graphics and applications* 24.3 (2004), pp. 44–51.

[144] Vassilios Solachidis and Loannis Pitas. "Circularly symmetric watermark embedding in 2-D DFT domain". In: *IEEE transactions on image processing* 10.11 (2001), pp. 1741–1753.

[145] Deshmukh Sonal Kokateand Manjusha. "Authentication of colour images by using encrypted png image with data repair capability". In: *SUSTECH 15. Proceedings. International Conference on Technologies for sustainability.* DAV Institute of management, Faridabad. 2015, pp. 1139–1149. ISBN: 978-81-931039-7-5.

[146] Julien P Stern et al. "Robust object watermarking: Application to code". In: *International Workshop on Information Hiding*. Springer. 1999, pp. 368–378.

[147] Harold S Stone. *Analysis of attacks on image watermarks with randomized coefficients*. NEC Research Institute, 1996.

[148] D Storck. "A new approach to integrity of digital images". In: *Mobile Communications*. Springer, 1996, pp. 309–316.

[149] Qingtang Su and Beijing Chen. "Robust color image watermarking technique in the spatial domain". In: *Soft Computing* 22.1 (2018), pp. 91–106.

[150] Qingtang Su et al. "A blind dual color images watermarking based on IWT and state coding". In: *Optics Communications* 285.7 (2012), pp. 1717–1724.

[151] Qingtang Su et al. "Embedding color image watermark in color image based on two-level DCT". In: *Signal, Image and Video Processing* 9.5 (2015), pp. 991–1007.

[152] Mitchell D Swanson, Mei Kobayashi, and Ahmed H Tewfik. "Multimedia data-embedding and watermarking technologies". In: *Proceedings of the IEEE* 86.6 (1998), pp. 1064–1087.

[153] Hai Tao et al. "Robust image watermarking theories and techniques: A review". In: *Journal of applied research and technology* 12.1 (2014), pp. 122–138.

[154] Anastasios Tefas, Nikos Nikolaidis, and Ioannis Pitas. "Image watermarking: Techniques and applications". In: *The Essential Guide to Image Processing*. Elsevier, 2009, pp. 597–648.

[155] Anastasios Tefas and Ioannis Pitas. "Robust spatial image watermarking using progressive detection". In: *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*. Vol. 3. IEEE. 2001, pp. 1973–1976.

[156] Ahmed H Tewfik et al. *Method and apparatus for embedding data, including watermarks, in human perceptible sounds*. US Patent 6,061,793. Sept. 2000.

[157] Manjit Thapa, Sandeep Kumar Sood, and AP Meenakshi Sharma. "Digital image watermarking technique based on different attacks". In: *IJACSA) International Journal of Advanced Computer Science and Applications* 2.4 (2011).

[158] Anatol Z Tirkel et al. "Electronic watermark". In: *Digital Image Computing, Technology and Applications (DICTA'93)* (1993), pp. 666–673.

[159] Archana Tiwari and Manisha Sharma. "Comparative evaluation of semi fragile watermarking algorithms for image authentication". In: *Journal of Information Security* 3.3 (03 2012), p. 189. DOI: 10.4236/jis.2012.33023.

[160] Hung-Hsu Tsai, Yu-Jie Jhuang, and Yen-Shou Lai. "An SVD-based image watermarking in wavelet domain using SVR and PSO". In: *Applied Soft Computing* 12.8 (2012), pp. 2442–2453.

[161] Min-Jen Tsai and Hsiao-Ying Hung. "DCT and DWT-based image watermarking by using subsampling". In: *Distributed Computing Systems Workshops, 2004. Proceedings. 24th International Conference on*. IEEE. 2004, pp. 184–189.

[162] Ramarathnam Venkatesan et al. "Robust image hashing". In: *Image Processing, 2000. Proceedings. 2000 International Conference on*. Vol. 3. IEEE. 2000, pp. 664–666.

[163] George Voyatzis and Ioannis Pitas. "Applications of toral automorphisms in image watermarking". In: *Image Processing, 1996. Proceedings., International Conference on*. Vol. 1. IEEE, 1996, pp. 237–240.

[164] Ming-Shi Wang and Wei-Che Chen. "A majority-voting based watermarking scheme for color image tamper detection and recovery". In: *Computer Standards & Interfaces* 29.5 (July 2007), pp. 561–570. DOI: 10.1016/j.csi.2006.11.009.

[165] Ali Behloul Wassila Belferdi. "Protection of digital watermarking based on SVD against false positive detection vulnerability". In: *Proceedings of the 1st international Conference on Advanced Communication and Information Systems ICASIS'2012*. Batna University. 12–13 Dec 2012, pp. 22–27.

[166] Ali Behloul Wassila Belferdi Lemnouar Noui. "A Novel Cholesky Decomposition-based Scheme for Strict Image Authentication". In: *Proceedings of the 2nd international Conference on Pattern Analysis and Intelligent Systems PAIS'2016*. Khenchla University, IEEE Algeria Section. 16–17 Nov 2016, pp. 17–22.

[167] Lemnouar Noui Wassila Belferdi Ali Behloul. "A blind dual color images watermarking based on IWT and sub-sampling". In: *Proceedings of the 3rd international Conference on Complex Systems CISC'2014*. Jijel University. 9–10 Dec 2014, pp. 29–34.

[168] Lemnouar Noui Wassila Belferdi Ali Behloul. "A Bayer Pattern-based Fragile Watermarking Scheme for Color Image Tamper Detection and Restoration". In: *Multidimensional Systems and Signal Processing* 26.4 (2018). DOI: 10.1007/s11045-018-0597-x.

[169] Raymond B Wolfgang and Edward J Delp. "A watermark for digital images." In: *ICIP (3)*. 1996, pp. 219–222.

[170] Raymond B Wolfgang and Edward J Delp. "A watermarking technique for digital imagery: further studies". In: *International Conference on Imaging, Systems, and Technology*. 1997, pp. 279–287.

[171] Ping Wah Wong and Nasir Memon. "Secret and public key image watermarking schemes for image authentication and ownership verification". In: *Image Processing, IEEE Transactions on* 10.10 (Oct. 2001), pp. 1593–1601.

[172] Jiqing Wu, Radu Timofte, and Luc Van Gool. "Efficient regression priors for post-processing demosaiced images". In: *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, Sept. 2015, pp. 3495–3499. DOI: 10.1109/ICIP.2015.7351454.

[173] Min Wu, Edward Tang, and Bo Lin. "Data hiding in digital binary image". In: *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on*. Vol. 1. IEEE. 2000, pp. 393–396.

[174] Xiaoyun Wu et al. "A secure semi-fragile watermarking for image authentication based on integer wavelet transform with parameters". In: *Proceedings of the 2005 Australasian workshop on Grid computing and e-research-Volume 44*. Australian Computer Society, Inc. 2005, pp. 75–80.

[175] Liehua Xie et al. "Image enhancement towards soft image authentication". In: *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on*. Vol. 1. IEEE. 2000, pp. 497–500.

[176] Chun-Wei Yang and Jau-Ji Shen. "Recover the tampered image based on VQ indexing". In: *Signal Processing* 90.1 (Jan. 2010), pp. 331–343. DOI: 10.1016/j.sigpro.2009.07.007.

[177] Boon-Lock Yeo and Minerva M Yeung. "Watermarking 3D objects for verification". In: *IEEE Computer Graphics and Applications* 19.1 (1999), pp. 36–45.

[178] Gwo-Jong Yu et al. "Mean quantization blind watermarking for image authentication". In: *Image Processing, 2000. Proceedings. 2000 International Conference on*. Vol. 3. IEEE. 2000, pp. 706–709.

[179] Mei Yu et al. "New fragile watermarking method for stereo image authentication with localization and recovery". In: *AEU-International Journal of Electronics and Communications* 69.1 (Jan. 2015), pp. 361–370. DOI: 10.1016/j.aeue.2014.10.006.

[180] Jasni Mohamad Zain and Malcolm Clarke. "Reversible region of non-interest (RONI) watermarking for authentication of DICOM images". In: *arXiv preprint arXiv:1101.1603* (2011).

[181] Jiashu Zhang, Lei Tian, and Heng-Ming Tai. "A new watermarking method based on chaotic maps". In: *Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on*. Vol. 2. IEEE. 2004, pp. 939–942. DOI: 10.1109/ICME.2004.1394356.

[182] Xinpeng Zhang et al. "Reference sharing mechanism for watermark self-embedding". In: *Image Processing, IEEE Transactions on* 20.2 (Feb. 2011), pp. 485–495. DOI: 10.1109/TIP.2010.2066981.

Abstract

A Robust Watermarking Approach for Images Authentication and Traceability

Digital watermarking is a promising technology that has shown the ability to achieve the security and protection of image data.The enormous growth, use, and distribution of these images reveal security threats with legal and ethical complexities. Though, despite the wide interest that has, digital watermarking still not been widely adopted in some applications like image authentication with restoration ability. Existing watermarking schemes often suffer from technical and security flaws. Validation of the suitability of those schemes for an application becomes more challenging. One main reason for these problems is the compromise between the watermarked image quality, tamper localization and recovered image quality, which is a serious problem to the majority of current watermarking schemes with restoration ability.

Addressing these gaps, in this thesis, a number of original contributions have been made. Starting with a comprehensive literature review on digital watermarking schemes and their applications and determine the requirements for watermarking, which has led to the following main contributions:

A novel self-embedding watermarking scheme aims to authenticate the content of a watermarked image and to detect any possible alterations and recover damaged area is developed, the proposed scheme improves the performance in terms of imperceptibility and robustness, focusing on three major considerations: the invisibility of the embedded watermark, the accuracy of detection and the high quality of the recovered color image.

Moreover, in support of developing a strict-authentication scheme for medical image applications, a novel digital signature-based scheme that uses the Cholesky decomposition is developed, which overcomes the limitations of existing schemes. Experimental results prove that the proposed scheme has higher capacity and more efficient detection ability

In addition, aiming at developing a blind dual-color image watermarking scheme, a new watermark embedding scheme is addressed, which is different from some existing schemes that use the binary or gray-level image as watermark. Experimental results demonstrate the stronger robustness of the proposed scheme against most common attacks including image compression, cropping, noising and scaling, etc, that allows this scheme to be applicable for traceability applications.

The presented new self-authentication model would help develop more secure self-authentication scheme with restoration ability. Additionally, the presented Cholesky digital signature-based scheme for medical images and its validation has created a new efficient approach, which can be used for different applications including content authentication and tamper detection. Moreover, the proposed blind dual-color image watermarking scheme offers an efficient robust tool for images traceability.

Key words: Digital watermarking; Strict authentication; Digital signature; Robust watermarking; Fragile watermarking; Traceability.