國立交通大學

資訊管理研究所

碩士論文

以小波轉換為基礎的數位浮水印之影像鑑定與篡改偵測

A Wavelet Transform Based Digital Watermarking for Image Authentication and Tampering Detection

研 究 生：洪筱盈

指導教授：蔡銘箴　博士

中華民國九十三年六月

以小波轉換為基礎的數位浮水印之影像鑑定與篡改偵測

A Wavelet Transform Based Digital Watermarking for Image Authentication and Tampering Detection

研 究 生： 洪筱盈 Student: Hsiao-Ying Hung

指導教授： 蔡銘箴 Advisor: Min-Jen Tsai

國立交通大學
資訊管理研究所
碩士論文

A Thesis
Submitted to Institute of Information Management
College of Management
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master of Business Administration
in
Information Management
June 2004
Hsinchu, Taiwan, the Republic of China

中華民國九十三年六月

以小波轉換為基礎的數位浮水印之影像鑑定與篡改偵測

學生：洪筱盈　　　　　　　　　　指導教授：蔡銘箴老師

國立交通大學資訊管理研究所

摘要

網際網路的迅速擴張及數位科技的長足進步，增加了數位媒體的容易取得，造成媒體所有權擁有的合法性問題。浮水印技術是一種適合做為數位資料所有權識別的工具，它能夠在經由無法察覺的修改而加入隱藏的資訊。

本論文研究以小波轉換為基礎的半易碎型浮水印之影像鑑定與篡改偵測。為達到影像保護的目的，影像在小波分解及量化程序之後，嵌入一個安全的、以小波樹為基礎的二元影像簽章（wavelet tree based binary image signature, *WTS*）；除此之外，被選取的小波係數需經過係數微調的步驟，以增加浮水印的彈性。在影像鑑定過程中，不需要原始未加入浮水印的影像作對照；比較可能受篡改影像的 *WTS* 與經過確認的 *WTS*，即能夠偵測影像是否受到未經許可的篡改。本論文提出的技術不只能夠定位出受到篡改的位置，而且有分辨非故意之偶然修改與惡意篡改的不同；這項技術能夠承受中等品質的 JPEG 壓縮，也能夠有效地偵測出影像的細微篡改。

關鍵字：小波轉換、數位浮水印、影像鑑定、篡改偵測

A Wavelet Transform Based Digital Watermarking for Image Authentication and Tampering Detection

Student: Hsiao-Ying Hung **Advisor: Min-Jen Tsai**

Institute of Information Management
National Chiao-Tung University

Abstract

The rapid expansion of the Internet and the recent advance of digital technologies have sharply increased the availability of digital media. In consequence, watermarking is developed as a suitable candidate for the ownership identification of digital data as it allows the invisible insertion of information with the imperceptible modification.

This thesis is to investigate a wavelet based semi-fragile watermarking technique for image authentication and tampering detection. Image protection is achieved by the insertion of a secret wavelet tree based binary image signature (*WTS*) after wavelet decomposition followed by quantization procedure. In addition, tuning steps is performed for the selected watermarked coefficients in order to increase the elasticity of watermark. During the verification, the original unmarked image is not needed for comparison. The detection of unauthorized tampering within the image is performed by comparison with the possibly modified image's WTS and the authentic one. The proposed technique not only localizes the tampered position, but also has the capability to distinguish incidental modification from the malicious tampering. It stays unaffected by medium JPEG quality compression and also effectively points out the small image modifications.

Keywords: Wavelet Transform, Digital Watermarking, Image Authentication, Tampering Detection

Acknowledgements

本論文得以完成，首先感謝指導教授蔡銘箴老師對於研究題材的開放態度及費心的教導，讓我得以對影像等多媒體的操作、數位浮水印及其相關智財權的保護議題，有深入的涉獵進而研究實作；還有博宇及承龍兩位學弟對於程式撰寫上的幫助，使我的論文能夠順利的完成。修習李素瑛老師的多媒體資訊系統課程，帶領我初入數位多媒體這個豐富且多樣性的領域，進而產生濃厚的興趣；其後，陳玲慧老師的影像處理課程，紮實的課程安排及程式撰寫，讓我更深入瞭解影像處理領域的觀念及操作，使我能夠在論文的研究上更加得心應手。感謝李素瑛老師、劉敦仁老師、許秋婷老師，三位口試委員的對論文的建議及指正，使我的論文內容更加豐富且嚴謹。

兩年充實的碩士生涯很快就結束了，雖然在這個數位浮水印領域我還只是個初學者，相關的研究基礎以及文獻的閱讀研究仍然深感不足，但是交大豐富的研究資源以及教授們循循善誘且深入淺出的教導，讓我能夠十分順利且愉快的學習與研究。感謝研究室伙伴們，有了你們，使我在繁忙的課業與研究外，添加了非常多豐富且愉快的美好回憶；還有千毓、瓊茹、秋雅，我們是商務多媒體研究室漂亮的四姊妹，謝謝妳們的陪伴及患難與共，是我碩士生涯中最珍貴的收穫。

1896

最後，感謝我的家人以及中威，因為有你們的全力支持，才能讓我無憂無慮的進行課業及研究；每次從新竹交大回到台北的家，總是讓我感受到可貴的親情溫暖，你們的呵護與支持，是我學習與研究上的原動力。

洪筱盈

2004年6月28日

謹於交大資管所

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Full Name
CoSIM	Co-Similarity
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
HVS	Human Visual System
IDCT	Inverse Discrete Cosine Transform
IDWT	Inverse Discrete Wavelet Transform
JPEG	Joint Photographic Experts Group
MAC	Message Authentication Code
MSE	Mean Square Error
MSQE	Mean Square Quantization Error
PSNR	Peak Signal to Noise Ratio
TAF	Tamper Assessment Function
TRF	Tamper Response Function
WTS	Wavelet Tree based image Signature

List of Symbols

Symbols	Definition
I	Original Image
I^*	Authentic Image
$\tilde{I}$	Possibly Modified Image
wtType	Wavelet Transform Type
WTS	Wavelet-Tree-based Binary Image Signature
enType	Encryption Type of the *WTS*
Δ	Quantization Step Size
cKey	Coefficient Selection Key
λ	Tuning Strength, usually between 0~0.5
ρ	Fuzzy Strength, usually between 0~0.5

1 Introduction

The rapid expansion of the Internet and the overall development of digital technologies in the past years have sharply increased the availability of digital contents. One of the advantages of digital data is that it can be reproduced without loss of quality. However, it can also be modified easily and unperceptively. In a lot of context, the sudden increase in watermarking interest is most likely due to the increase in concern over copyright protection of content. Internet is an excellent distribution system for digital media because it is inexpensive, eliminates warehousing and stock, and delivery is almost instantaneous. However, content owners also see a high risk of piracy.

This risk of piracy is exacerbated by the proliferation of high-capacity digital recording devices. Using these recording devices and using Internet for distribution, would-be pirates can easily record and distribute copyright-protected material without appropriate compensation being paid to the actual copyright owners. Thus, content owners are eagerly seeking technologies that promise to protect their rights.

1.1 Motivation of the Research

In order to protect the rights of digital content, the first technology content owners turn to is cryptography. However, encryption cannot help the seller monitor how a legitimate customer handles the content after decryption. In other words, cryptography can protect content in transit, but one decrypted, the content has no further protection.

Thus, there is a strong need for an alternative or complement to cryptography: a technology that can protect content even after it is decrypted. Watermarking has the potential to fulfill this need because it places information within the content where it is never removed during normal usage. Decryption, re-encryption, compression, digital-to-analog conversion, and file format changes – a watermark can be designed to survive all of these processes.

Due to watermarking's wide range of applications and high potential, this sub-discipline of communication security has attracted a lot of interest in the last ten years. It has now evolved as an established candidate for copyright protection, ownership identification and fingerprinting systems. Moreover, several commercial applications of watermarking for copy control devices are planned, or are already implemented. For all these contexts, a lot of effort is dedicated to the development of *robust watermarking* schemes that permanently mark the works. On the other hand, the use of *fragile* or *semi-fragile* embedding schemes – ones where the embedded key, that is, the mark, is destroyed by the modification of the work – is much less investigated. Nevertheless, this kind of system shows great promise for content authentication as it allows for the validation of digital data, thus giving it legal value. As digital media are now widely employed and commonly accepted as official documents, protection of their informative content will grow as an important issue, as with the protection of intellectual property in the past years.

1.2 Problem Definition

In many applications, such as courtroom evidence and video security systems, any modification of image, video or audio data must be detected if it cannot be prevented. As digital images are widely available, online or elsewhere, and because they are so easy to modify, some work needs to be one to protect the information they contain. As the number of images increases, the direct storage of unique reference patterns becomes impractical. Moreover, as some images need to be slightly compressed in order to be efficiently stored, authentication systems need to offer flexibility. Unfortunately, many of the approaches previously proposed lack this characteristic, while others require too much user interaction to be truly considered secure for commercial applications.

In this thesis, a novel technique is proposed for the content authentication of digital images, which doesn't need the original image. This approach is able to detect, as well as localize, malicious image alterations, while offering robustness to high quality image compression. The proposed method is based on semi-fragile watermarking technology. It uses the knowledge of characteristics of the human visual system (HVS) to round discrete wavelet coefficients from an images' decomposition to appropriate quantization levels. Tampering detection is performed using each wavelet resolution scale's frequency band verification.

1.3 Organization of the Thesis

The goal of the thesis is to find a watermarking method that can detect, as well as localize, tampering in digital images. So the main objectives of this thesis are:

- To introduce a set of well-defined goals for a image authentication scheme.
- To present an authentication technique which provides more complete information on how the image is modified.
- To demonstrate the potential of tampering detection methods through implementations of the proposed method and existing techniques.
- To provide a comparative study of the strengths and limitations of the proposed and existing image authentication methods.

The earlier work on digital watermarking and image authentication schemes are first reviewed. And then, the development of a semi-fragile watermarking scheme for image authentication is illustrated. In order to make the thesis complete, an overview of image processing techniques and a more extensive background on watermarking technology is needed. The thesis is organized as follows.

In Chapter 2, the specificities of image authentication are introduced. In the first section, the definition of authentication is given. The basis of wavelets analysis and digital watermarking are then described. In that sense, the requirements that such authentication system schemes should fulfill in order to be effective and efficient are draw. Then, the detail specific methods are introduced that have served as bases in the development of the proposed method. Two different families of embedding approach are emphasized: those acting in the spatial domains, and the others, acting in some transform domain. The pros and cons of each are highlighted through the examination of published work.

A novel method that digital watermarking based on the wavelet transform for image authentication and tampering detection is introduced in Chapter 3. At the beginning, some defects of the conventional quantization based approach are discussed. In this context, the following sections will specify the proposed image authentication, including the overview, the watermark embedding process, and the tampering detecting process.

Afterward, the experimental results are shown in Chapter 4. At first, the experimental setup is described in Section 4.1. The parametric inferences about both encoding and decoding process are performed in Section 4.2. After that, a series of watermark resistance experiment are represented in Section 4.5 to 4.5, including incidental distortions, malicious distortion and the complex attacks. Section 4.6 shows the system prototype, and finally the comparison with other conventional quantization-based approach is described in Section 4.7.

To finish, the discussion of the proposed method obtained, as well as future possible research work in this field of digital watermarking for image authentication and tampering detection, are presented in Chapter 5.

2 Image Authentication

Authenticity is among digital document security properties needing attention. In this chapter, Section 2.1 introduces the definition of authentication; the review of wavelet analysis and digital watermarking are summarized in Section 2.2 and 2.3. In Section 2.4, the requirements that such authentication system should fulfill in order to be efficient are represented. Section 2.5 illustrates several approaches of image authentication that the proposed method will refer to.

2.1 Definition of Authentication

Authentication is the service of ensuring whether a given block of data has *integrity*, (i.e., the associated content has not been modified), and is from the legitimate sender. Authentication is traditionally ensured through mechanisms that involve message authentication codes (MACs) and digital signatures (Stallings, 2000) known as "hard-authenticators." In hard authentication, a MAC (also known as a message digest) or digital signature of the data to protect, called an *authenticator*, is created at the source and transmitted with the data. At the receiver, the authenticator is verified using the received data to deduce if the received information is in fact unmodified and from the alleged sender (Zhao et al., 2004).

When the data represents an image that may travel through a set of diverse distribution chains, then it can be susceptible to content-preserving operations such as compression, trans-coding and other standard format conversions which severely impede the usefulness of hard-authentication mechanisms. Any processing of the image that changes the bit representation, yet still maintains the validity of perceptual content, may be inaccurately categorized as being "inauthentic." Thus, more recently there has been a movement toward schemes that provide "soft-authentication," in which content-preserving processing is distinguished from unlawful content-changing manipulations (Zhao et al., 2004).

One tool-set that has been recently applied to soft-authentication, which will be the partial focus of this work, is called *semi-fragile digital watermarking*. In semi-fragile watermarking, the watermark is embedded such that it is fragile to some pre-defined processing and robust to others. Here, an authenticator which may consist of a MAC or digital signature of salient parts of an image is used to form a *watermark*. This watermark is imperceptibly embedded within the original image (commonly called the *host*). The integration of the authenticator within the image to be secured simplifies the logistical problem of MAC or digital signature data handling during image transmission. Moreover, semi-fragile watermarking can provide information on the degree and location of tampering within an image to make more application-suited decisions on credibility (Kundur & Hatzinakos, 1999) (Zhao et al., 2004).

There are two potential benefits to using watermarks in content authentication. First, watermarks remove any need to store separate, associated metadata, such as cryptographic signatures, this can be important in systems that must deal with legacy issues, such as old file formats that lack fields for the necessary metadata. A second potential advantage of watermarks is more subtle: a watermark undergoes the same transformation as the *Work* in which it is embedded. Unlike an appended signature, the watermark itself changes when the Work is corrupted. By comparing the watermark against a known reference, it might be possible to infer not just that an alteration occurred but what, when, and where changes happened. (Cox et al., 2002)

The image authentication based on digital watermarking can be extended in several commercial applications, such as digital archives, monopolizing of trademark, authenticating of some digital media produced in court as evidence, and any other multimedia that are sold through the Internet or any digital channels. The owner can embed the watermark to authenticate his own multimedia. Then, he can perform the tampering detection procedure when the owner found someone encroaching on or tampering his multimedia.

2.2 Wavelet Analysis

1896

Although the average person probably knows very little about wavelet, their impact on today's technological world is phenomenal. They represent a very powerful mathematical tool commonly used by scientists and engineers, and are currently applied in fields such as signal processing, computer vision and data compression. Several new applications of wavelets are discovered every year and will continue to be in the future.

The first known step toward the development of a unified wavelet theory occurred when a Hungarian mathematician named *Alfred Haar* completed his work on the orthogonal systems of functions. In 1910, he proposed the used of piecewise constant functions to form an orthogonal basis. His system uses a basis function (now referred to as the scaling function ϕ) as a starting point. Then, the mother's

(ψ), daughters', sons', granddaughters', and grandsons' (and so on) functions are obtained by the subsequent scaling and translation of the basis, or father wavelet. *Haar* proved that the obtained set of functions can be used to represent a signal at different levels of detail (Strang & Nguyen, 1996). Furthermore, he demonstrated that a decomposed signal can be reconstructed using the reverse operations. Although it was not called "*wavelets*" back then, the simplest of the wavelet families was nonetheless born, and is now named the *Haar wavelet*.

The real breakthrough in wavelets analysis, however, happened in the late 1980's when a lot of papers now considered classic were published. *Yves Meyer* and *Stephane Mallat* were two important contributors to this newborn field. Investigating the use of wavelets in many different applied, they were amongst the first to develop the concept of multi-resolution analysis for wavelets (Mallat, 1989). This was an important step for the advancement of research on wavelets. As a result, multi-resolution is now an extensively used signal decomposition approach. Malat and Meyer were the first to mention scaling functions of wavelets, which allow researchers and mathematicians to construct their own wavelets using established criteria (Vetterli & Kovacevic, 1995).

Around the same time, a Belgian physicist named *Ingrid Daubechies* employed multi-resolution analysis to create her own family of wavelets. Using construction methods related to filter banks, she introduced (1988) a family of compactly supported orthogonal wavelet systems with arbitrarily high, but fixed regularity. These wavelets offer a number of desirable properties (such as compact support, orthogonality, regularity, and continuity) that make them truly attractive (Strang & Nguyen, 1996). This is why the *Daubechies Wavelets* are now some of the most common ones today.

Daubechies' work was probably the starting point of much focused research on wavelets that has lead to their acceptance as a modern mathematical tool and their wide use in sciences and engineering. Of course, many other researchers have contributed to the advancement of the field in the last decade, and several applications have been found. In particular, wavelet transforms prove to be extremely effective for image coding, and image compression standards, such as JPEG-2000, make use of them. From this, it is clear that wavelets are definitely a tool for the future, and this is why the knowledge of their historical and theoretical bases is of great interest.

2.3 Digital Watermarking

1896

Digital watermarking is a relatively new technology that allows the imperceptible insertion of information into multimedia data. The supplementary information, called *watermark*, is embedded into the over work through its slight modification. This mark is hidden from view during normal use and only becomes visible as a result of a special visualization process. An important point of watermarking techniques is that the embedded mark must carry information about the host in which it is hidden.

In 1988, Komatsu and Tominaga were the first to use the term digital watermarking for their image authentication system (Cox et al., 2002). Although there were several publications in the interval, a cornerstone paper by Cox et al. (1997) was the starting point of more intensified research. Of course, this was not only due to the paper by Cox et al. (1997), but mainly to the organization of the watermarking researchers.

In addition to the exploitation of different host, researchers have been looking at different applications for digital watermarking. The application that attracts the most attention is *copyright protection*. In this context, a watermark is permanently embedded in the work to identify its original owner. In order to be efficient, the embedded mark has to be robust, that is, it has to be detectable as long as the host carries its information, hence, the name of *robust watermarking*. Another use of robust watermarking is for the labeling or fingerprinting of digital media. This application is technically similar to the previous one except that, here, a different mark is embedded in each copy of the same work to allow its tracking.

The watermark is widely used to protect multimedia from illegal usage. Authentication is only one possible application of watermarking and its use on digital work offers other great possibilities. It has been foreseen as a good candidate technology for enhancing multimedia data by the addition of information available to the users for content improvement, copyright protection, authentication, and so forth.

Many of the first papers published on digital watermarking are about its use for copyright and ownership protection related functions. Thereby, most of the bases and theories associated with the technology are laid out in relation of this particular application. The paper published by Cox et al. (1997) constitutes an important step towards the installation of watermarking as a technology in its own right. Presenting a watermarking approach for the copyright protection of digital content, Cox et al. capture the most important concepts of robust watermarking. They demonstrate that their technique is robust to common signal processing procedures and geometric transformations, and is able to deal with simple collusive attacks, thus ensuring good copyright protection of images. They conclude by stating, without implementing, that watermarking systems should take explicit advantages of the characteristics of the human visual system, HVS.

This paper directly led to another cornerstone paper. Cox et al. (1999) examine the similarities and differences between watermarking and traditional communications. They stress the importance of the used of characteristics of the HVS in the embedding process; both for maximizing the robustness, and for minimizing the perceptual distortion introduced. They argue that an appropriate distortion model for watermarking applications includes a significant correlation between the distortion vectors (watermarks) and content vectors they are applied to.

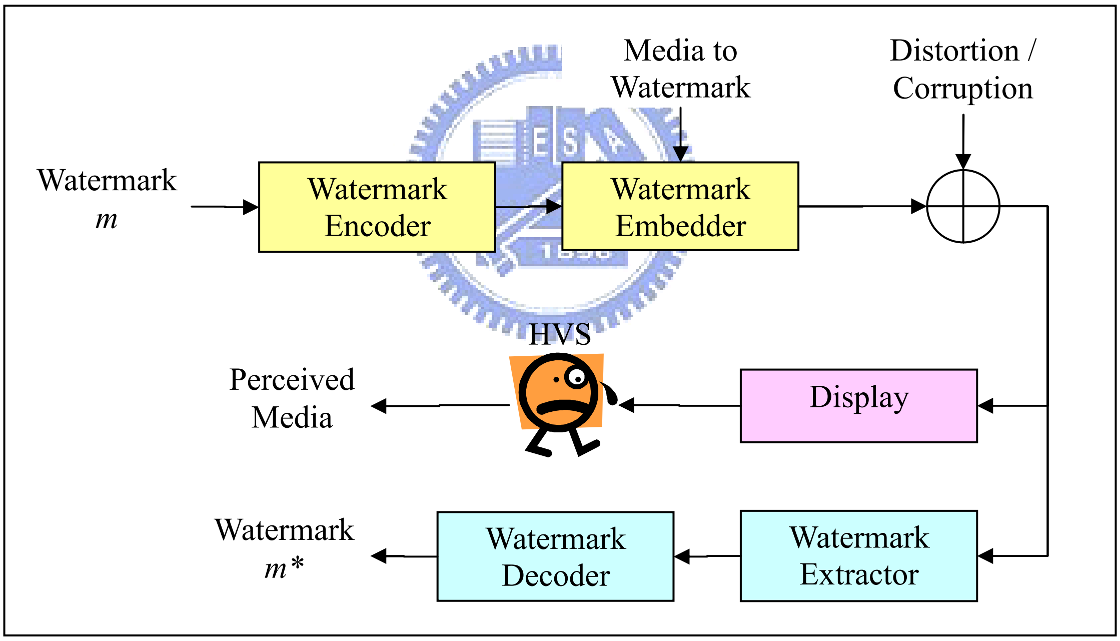


Figure 2.1: Watermarking scheme with perceptual model.

Figure 2.1 represents the underlying principles of watermarking as it takes into consideration the perception of the marked content by a potential user in parallel with the decoding procedure. The authors conclude by explaining the design of a blind optimal threshold-based detector. This discards the need to access the original image in the detection process, thus opening the field of watermarking to a wider range of applications.

2.4 Requirements of Authentication Schemes

Authenticating images and multimedia content in general differs from the traditional problem of authentication in cryptography. The goal in image authentication is to authenticate the content and not the specific representation of the image. As a result, a main requirement of such authentication systems is that minor modifications such as lossy compression which do not alter the content of the data preserve the authenticity of the data, whereas modifications which do modify the content render the data inauthentic. This requirement is difficult to formalize as the notion of content is difficult to specify precisely. Furthermore, as images can be considered as points in a continuous space, there is not a sharp boundary between authentic and inauthentic data since a sharp boundary implies that there are authentic and inauthentic images which are similar to each other.

More realistically is the diagram shown in Figure 2.2, where the region of surely authentic images is separated from the surely inauthentic images by a fuzzy region where the authenticity of the images is difficult to ascertain. β_a and β_m indicate radii in the case these regions are spheres in the underlying space. (Wu, 2002) In this figure, these regions are illustrated as spheres in some suitable metric space to facilitate characterization, although, in general, they can have more complicated shapes. Anther way to represent this is by assigning to each image a "degree of authenticity," a number between zero and one with zero meaning surely inauthentic and one meaning surely authentic. Thus, there are three answers when authenticating an image: authentic, inauthentic and do not know.

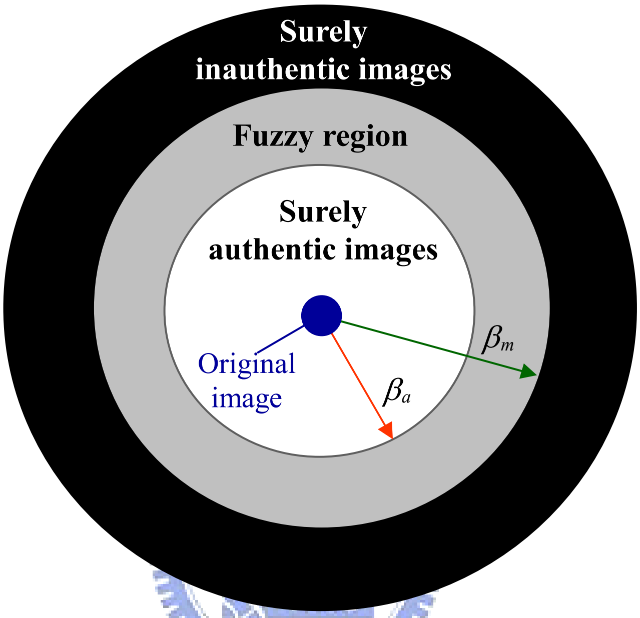


Figure 2.2: The authenticity of the image. (Wu, 2002)

From our willingness to protect digital data against forgery and tampering, and also based on semi-fragile techniques already proposed, we can extract several requirements that authentication systems must fulfill. We argue that traditional authentication approaches for data are not well suited for images, sound, and video; to be practically useful a tamper-proofing technique must not only detect the presence of modifications in a signal but should also provide information helpful to characterize the distortions. As Kundur & Hatzinakos (1999) proposed, a tamper proofing method must be able to do the following:

- Indicate with high probability that some form of tampering has or has not occurred;
- Provide a measure of the relative degree of distortion of the signal;
- Characterize the type of distortion, such as filtering, compression or replacement, without access to the original host signal or any other signal-dependent information; it should be possible to detect changes due to compression or random bit errors and make application- dependent decisions concerning whether or not the signal still has credibility;
- Validate the signal and authenticate the source without requiring the maintenance and synchronization of additional data separate from the signal.

There has been a recent trend toward addressing the problems of tamper proofing and authentication using a digital watermarking approach. The attraction of such an approach is that no additional data are required for signal verification. In addition, the verification information is discretely watermarked which adds an additional level of security against attacks to modify both the signal and the verification data.

Here are the main points to keep in mind in the development and evaluation of certification systems. In the context of image protection, an effective authentication scheme should be able to do the following:

- Determine whether an image has been altered or not;
- Find the location in the image where the alterations, if any, are made;
- Integrate authentication data within the host image rather than storing the data in a separate file;
- Be robust to acceptable manipulations such as lossy compression or to

other content-preserving manipulations defined by the original owner of the work to protect;

- Include security features preventing the forgery of manipulation of the reference mark. In essence, this means that the reference key used for authentication must be securely stored. In addition, the embedding protocol must depend on the secret key in order to enhance the security of the authentication scheme.

Some authors have also added the *recovery* capability as a prerequisite of image authentication systems (Lin & Chang, 2001). This means that, after the detection process, it should be possible to find out the original content of the tampered areas, and also, that the recovered data shown, be of the same quality as the original. This concept is interesting, and it has also lead to the development of *erasable watermarking systems*. An erasable watermark can be removed from its associated cover work to obtain an exact copy of the original unwatermarked work. It is however, impossible to design an erasable watermarking scheme that can be uniquely applied on all the work of a specific family of digital contents. For example, it is impossible to use the same erasable watermarking scheme on all digital images (Cox et al., 2002). Erasable watermarking schemes are still highly prototypic and this is why, in the present thesis, we have strictly been concerned by the detection and localization of alterations, and have not attempted the subjects of reconstructing tampered regions or deleting embedded marks. Nevertheless, the use of digital watermarking for image authentication clearly presents some advantages.

The advantages of watermarking approaches for content authentication are twofold. First, the direct embedding of a mark in the host data removes the need to store a separate authentication signature. Second, as the watermark undergoes the same alterations as the host, the mark is modified by the host's corruption (Cox et al., 2002). Using a reference pattern in the embedding and decoding procedures allows for the identification and delimitation of tampered regions. In addition, some basic requirements of digital watermarking are helpful in the authentication context. The fact that the embedded mark must stay invisible allows the watermarked data to be as close as possible to the original data, therefore, preserving the original content.

2.5 Approaches to Authentication

The raising of interest for content authentication has accelerated the development of fragile watermarking systems. As for other watermarking types, the fragile watermarking techniques proposed can be divided in two general categories in terms of the embedding process: the ones acting directly in the spatial domain and the others, working in different transform domains. Besides fragile watermarking techniques, semi-fragile watermarking and digital signature techniques are also used in image authentication methods in recent years. Each has pros and cons that we highlight here.

2.5.1 Watermarking in Spatial Domain

Fragile watermarking techniques that embed hidden information in the spatial domain, such as Queluz & Lamy (2000), Tefas & Pitas (2000), Tirkel, Osborne, & Schyndel (1996) and Yeung & Mintzer (1997), are definitely more straightforward, and less computationally expensive than the ones using transforms. Therefore, this kind of embedding is probably more suitable for real-time implementation.

Yeung & Mintzer (1997) propose one of the first watermarking methods for high-quality color and grey-scale image verification and authentication. A watermark image is embedded in to the source image in the spatial domain by the modification of the pixel values. The stamped image produced is visibly identical to the original one. A verification *key*, stored and known only to authorized parties, is also produced and is used in the verification process in order to extract the image inserted in the host. The extraction procedure can detect and localize spatial alterations done on previously watermarked images. The technique therefore provides a way of ensuring data integrity, adds to the security of the digital content, and allows the recipients of an image to verity the image as well as to display the ownership information of that image. The embedding process is however, fragile to unintentional image distortions introduced by basic image processing operations (e.g. compression) done for storage purposes.

Another spatial embedding watermarking method is that proposed by Tefas & Pitas (2000). In addition to allowing the identification of modified regions in tampered images, it is able to reject small distortions introduced by high quality image compression (for which Yeung & Mintzer (1997) is fragile). A pseudorandom watermark is embedded on randomly selected pixels using a neighbor-dependant function. In the detection process, the pixels surrounding the marked ones are used to create a mapping of false detections. The identification of changes in small details of the image is based on mathematical morphology; altered pixels are linked together in order to indicate tampered areas. Finally, the decision about the image's authenticity is made by comparing the ratio of correctly detected watermark with a predefined threshold.

However, both techniques (Yeung & Mintzer, 1997) (Tefas & Pitas, 2000) suffer from the following major drawback of spatial domain watermarking: the difficulty of the frequency localization of modifications. In fact, because the marks are inserted in certain particular pixels, it is often impossible to localize frequency alterations applied to the entire image. The reason why the localization of frequency alterations is important is twofold: one, it is a step towards *telltaling*, the characterization of the specific process used for the alteration of the content; and two, it provides a measure of the relative degree of image distortion.

In addition to the impossibility of identifying frequency tampering, image authentication systems based on the embedding of watermarks in the spatial domain have the drawback of being more susceptible to malicious attacks. In fact, search and collage attacks are a threat to spatial-based-watermarking and particularly block-based-watermarking authentication approaches (Cox et al, 2002). In a search attack, the aggressor, who has access to the watermark decoder, creates altered versions of the work and processes them through the decoder by brute force, until one is declared authentic. Since the mark is embedded directly in the pixel intensity values, it is possible, although lengthy, to extract a pattern from the multiple watermarked image and then use it to create *authentic* images. On the other hand, collage attacks are much more possible and easy to realize.

In summary, spatial-based authentication watermarking methods show speed advantage that can be favored for real-time implementations. This is why such techniques have often been extended to the authentication of video data (Bartollini et al., 2001). However, for all the reasons mentioned above, more compliant techniques must be developed for still image authentication.

2.5.2 Watermarking in Frequency Domain

The techniques using frequency domain are, of course, a little bit more complex and computationally expensive than the spatial domain ones. Yet, they offer a higher degree of robustness against common image processing operations (Cox et al., 1997). Once could wonder why robustness is important for fragile watermarking systems. This is simply because it is highly preferable that basic image processing operations – ones that are typically used for storage of watermarked images – do not alter the embedded marks.

Some authors have proposed taking advantage of the knowledge of current image compression standards to develop semi-fragile watermarking techniques in the discrete cosine transform (DCT) domain (Lin & Chang, 2001) (Wu & Liu, 1998).

Lin & Chang (2001) introduce an authentication scheme that accepts JPEG lossy compression performed on the watermarked image up to a pre-determined lowest quality factor while rejecting crop and replacement processes. Their authentication procedure is based on JPEG invariant properties of DCT coefficients. Their technique also allows for the recovery of original visual information after tampering. To achieve these goals, two binary sequences are created. The authentication bits (Φ), used to determine if any tampering has occurred, are computed from the relationship between two DCT coefficients of the same position in two separate (8 by 8) image blocks. This value is used since it is invariant to JPEG compression at a given quality factor. On the other hand, the recovery bits (Ψ), used to reconstruct the approximation of the original blocks of pixels after tampering, are obtained by the reduction, compression and encoding of the original (*unmarked* and *uncompressed*) image. The two are then embedded independently by the quantization of DCT coefficients using secret *block-selection* functions in relation with JPEG quantization tables. Selecting quantization levels greater than JPEG ones guarantees that the embedded marks stay unaltered up to a lowest compression quality threshold. In the decoding step, the private authentication process first reconstructs the authentication bits, and then, reconstructs altered regions, if needed. Finally, the capacity of the system to endure JPEG compression with quality factor great than 50, and to reconstructed altered regions, is explicitly demonstrated.

Although DCT approaches show some potential, it is the wavelet domain that attracts the most attention among all the transform domains used as it has been shown to yield the highest degree of robustness to simple image processing operations (Tsai & Hung, 2004). Furthermore, as DCT techniques (Lin & Chang, 2001) are mainly block-based, they are also highly susceptible to collage attacks. In terms of decomposition the main advantage of wavelets over Fourier and DCT analysis is that they allow for combined spatial and frequency resolutions. Wavelet transform allows for the decomposition of the signal in narrow levels of detail, while keeping the basis signal space limited. This is certainly of great importance when dealing with real signals, especially when spatial localization is to be considered.

Moreover, as stated earlier, the availability of numerous mother wavelets gives flexibility to the analysis and allows it to be truly adaptive to a particular application. It is also possible to develop new basis functions to fulfill specific requirements.

Finally, the use of the wavelet domain – as opposed to spatial or DCT domains – to embed the watermark provides simultaneous spatial localization and a frequency spread of the watermark in the host image (Kundur & Hatzinakos, 1999). All these gains certainly explain why wavelet transform attracts so much attention for a wide range of image processing application, including digital watermarking for image authentication (Kundur & Hatzinakos, 1999) (Lu & Liao, 2001) (Yu et al., 2001) (Zhao et al., 2004) and the image compression standard, JPEG-2000.

Kundur & Hatzinakos (1999) present a fragile watermarking technique for the tamper proofing of still images, as show in Figure 2.3 and Figure 2.4. They propose to embed a mark in the discrete wavelet domain by the quantization of the image's corresponding wavelet coefficients. The first operation is the decomposition of the image by the computation of its discrete wavelet transform (DWT). The authors make use of the *Haar* wavelet exclusively, and propose an algorithm in which the changes in the wavelet coefficients guarantee integer changes in the spatial domain. Once the image is decomposed in L levels of detail, a watermark can be inserted. First, an author identification key is produced by the generation of a pseudo-random binary sequence (zeros and ones) of length N_w. This sequence is kept secret and known only by the original owner of the work. Then, a quantization map is created based on a user-defined quantization step Δ.

Figure 2.5 shows the rounding of specific DWT coefficients to even or odd quantization step values that embeds the zeros and ones of the watermarks. The selection of embedding locations is pseudo-random and well spread spatially and throughout each resolution level to be able to assess changes to all image components. The location information is stored in the coefficient selection key (*ckey*). In addition, an image-dependant quantization key (*qkey*) is introduced to improve security against forgery, and monitor specific changes to the image. The last step of the embedding process is the construction of the *tamper-proofed* image by the computation of the inverse discrete wavelet transform (IDWT).

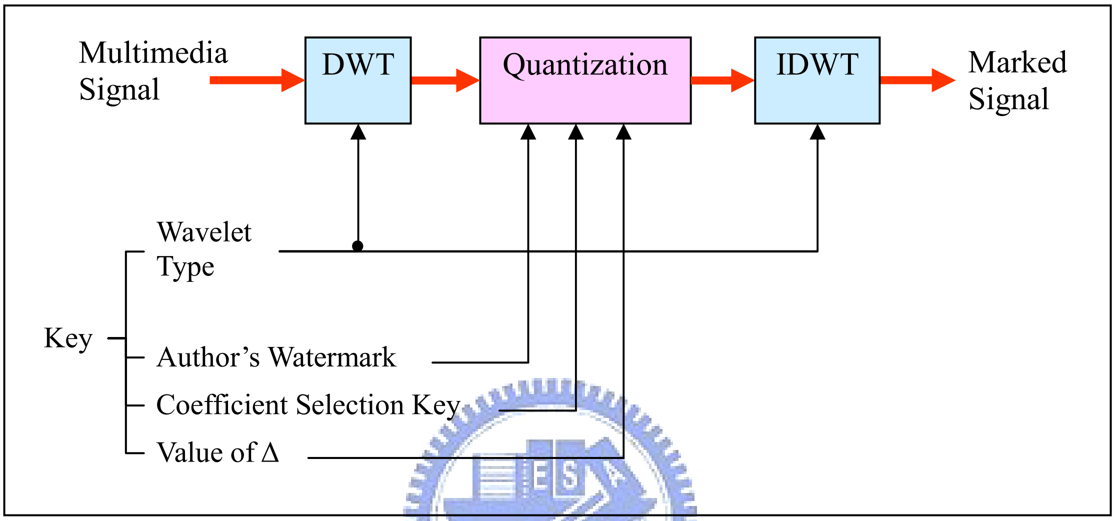


Figure 2.3: The embedding process of Kundur & Hatzinakos's proposed image authentication method.

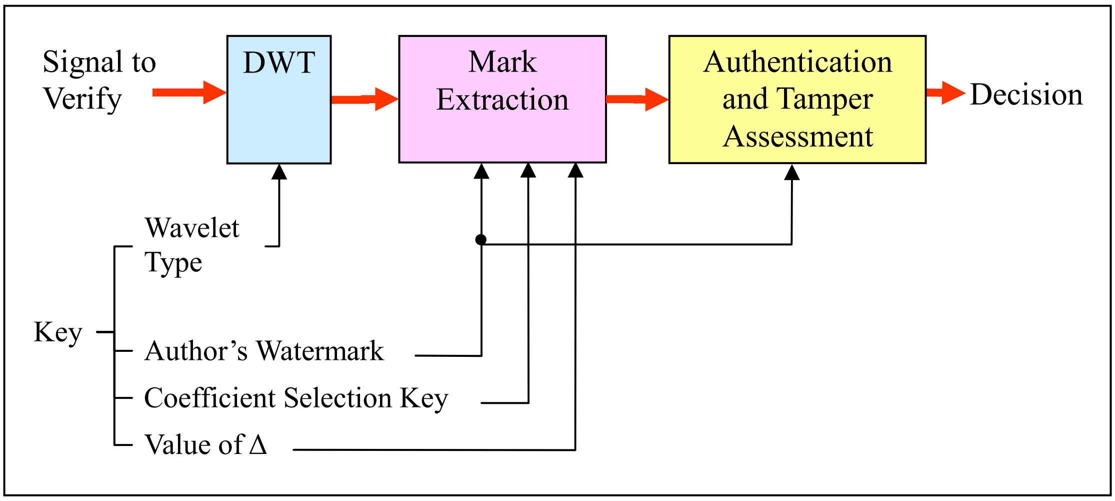


Figure 2.4: The tamper assessment process of Kundur & Hatzinakos's proposed image authentication method.

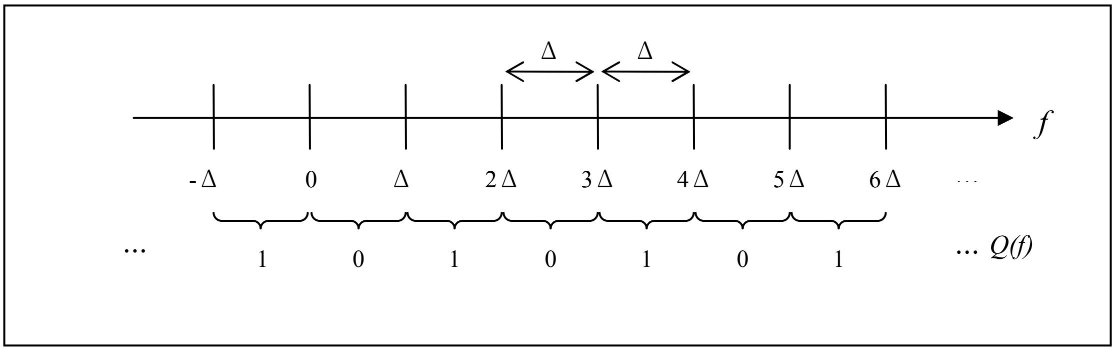


Figure 2.5: The quantization function used by Kundur & Hatzinakos.

In the decoding process, the DWT is performed on the *possibly* tampered image and locations of original watermark embedding are selected using *ckey*. Then, the embedded mark is blindly extracted by the computation of the mark extracted with the originally embedded one. The approach permits tamper detection in localized spatial and frequency region, therefore making possible the identification of specific modified frequencies in an image. To assess the extent of tampering (the difference between the embedded mark ω and the extracted one $\tilde{\omega}$), a temper assessment function, TAF, is computed with the following:

$$TAF(\omega,\tilde{\omega}) = \frac{1}{N_\omega}\sum_{i=1}^{N_\omega}\omega(i)\oplus\tilde{\omega}(i) \qquad (3.1)$$

Comparing the TAF with a predefined threshold T, allows the user to make application-dependant decisions concerning the credibility of the received data. Examining how a known embedded watermark has been changed gives the possibility to investigate how a work has been corrupted. This type of watermarking is referred to as a *telltale* watermarking. Thus, the users are allowed to make context-dependant decisions on the validity of the images received. However, the total capacity of the system, given by the mark's length N_ω, is not specified. In addition, no strategy is propose to deal with a combination of malicious tampering and incidental distortion for the choice of Δ or T.

The same line of thought, Yu et al. (2001) developed a digital images authentication procedure that allows for the detection of malicious tampering while staying robust to incidental distortion introduced by compression.

As in Kundur & Hatzinakos's (1999) proposed, they embed a binary watermark in the wavelet transform domain. Once again, the insertion is done by the even or odd quantization of selected wavelet coefficients. Quantization-based watermarking is the simplest protocol because it requires the least storage of information. It is however, very sensitive to image modification. For this reason, the authors propose to make the embedded watermark more robust by quantizing the mean value of weighted magnitudes of wavelet coefficients. The quantization of regions of wavelet coefficients is performed using a predetermined function *Q*.

The same function is used in the blind detection process as well, to privately retrieve the mark by *reversed quantization*, that is, determining the parity (in terms of quantization level) of the mean value of the wavelet packet coefficients. In order to distinguish malicious tampering from incidental distortion, the amount of modification on wavelet coefficients introduced by incidental *versus* malicious tampering is modeled as *Gaussian* distributions with small *versus* large variance. The probability of watermark error due to incidental alterations is smaller than malicious tampering because they produce a comparatively smaller variance difference with the embedded marks.

To state the validity of possibly tampered images, a tamper response function (TRF) is defined for each decomposition level. It compares original quantization values $x_l(i,j)$ with wavelet coefficients $x_l(i^*,j^*)$ of the possibly tampered image, as shown below:

$$TRF(x_l(i^*,j^*),x(i,j))=\frac{\max\{|i^*-i|,|j^*-j|\}}{(\sum_{k=1}^{(Density(x_l(i^*,j^*)+1)}k^2)} \qquad (3.2)$$

The TRF allows for the estimation of tampering depth. Furthermore, the computation of the *Chess-Board* distance among altered coefficients permits the mapping of the tamper response. This serves as the basis for the decision rules to measure the malevolence of attacks. The integration of the tamper response at each scale of the wavelet decomposition allows for the discrimination of malicious tampering from incidental ones. This grants a certain degree of robustness to the system as the method is able to blindly authenticate JPEG compressed images. In spite of this, the authors do not explicate the degree to which the image can be compressed, and never explain how the quantization parameters are chosen.

The main flaw with the two techniques described above is that they both involve post-processing operations to determine the validity of the content. As shown in Table 2.1, the comparison with watermark embedding Domains for Image Authentication was listed. In (Kundur & Hatzinakos, 1999), the user has to set a threshold below which a mark can be considered authentic, while in (Yu et al., 2001), the tampering distribution has to be examined. Furthermore, in (Yu et al., 2001), the users might have to determine the tampering manually at each scale if the tampered area is too small, or if there are many small unconnected tampered regions. In fact, both systems in themselves are not robust to JPEG compression, and only the detection processes allow this specific operation to go unnoticed. In conclusion, truly robust automated image authentication techniques in the wavelet-domain have yet to be developed.

Table 2.1: The comparison with watermarking in spatial and frequency domain for image authentication.

		Advantages	Disadvantages
Spatial Domain		⇨ Fast. ⇨ Good spatial localization of tampering.	⇨ Sensitive to attacks. ⇨ No localization of frequency tampering. ⇨ Most techniques are sensitive to compression.
Frequency Domain	DCT Domain	⇨ Offers robustness to JPEG compression. ⇨ Adequate spatial localization of tampering.	⇨ Sensitive to block-based attacks. ⇨ Localization of frequency tampering is not straightforward.
	Wavelet Domain	⇨ Combines frequency and spatial localization of tampering. ⇨ Highly secure.	⇨ Post-processing operations needed to assess the malevolence of tampering.

3 Proposed Technique of Image Authentication

The last two techniques presented in Subsection 2.5.2 protected digital images from malicious tampering and unauthorized processing, while allowing the compression of images with small compression ratios. In this chapter, the proposed technique of image authentication will be detailed as follow. Section 3.1 first introduces the drawbacks of Kundur & Hatzinakos proposed method that our method can improve well. Section 3.2 gives an overview of our proposed technique, the general concept of the authentication scheme, and the system point of view of our image authentication system decision process. Finally, Section 3.3 and Section 3.4 illustrate the detail algorithms about our embedding process and decoding process for well image authentication and tampering detection.

3.1 The Improvement of Telltale Tamper Proofing Method

In the watermark-based image authentication approaches detection of tampering is based on the fragility of a hidden watermark. Subsection 2.5.2 has introduced Kundur & Hatzinakos's method of telltale tampering proofing and authentication. They make use of the *Haar* wavelet transform, in which the coefficients at each resolution level l are rational numbers of the form $r/2^l$ where $r \in \mathrm{Z}$. For this reason, their approach uses $\delta 2^l$ as the size of a quantization interval, where δ is a pre-specified positive integer, $l = 1,\ldots,L$, and L is the number of scales used in the wavelet transform. However, the following lists some points that were not considered in Kundur & Hatzinakos's algorithm:

- In their quantization-based method, a watermark value is encoded by modulating a selected wavelet coefficient into a quantized interval. Basically, the quantity they used for modulation, which is monotonically increased from high resolution to low resolution, violates the capacity constraint of the human visual system (HVS). (Warson et al., 1997)
- They defined a tamper assessment function (TAF), which is the ratio of the number of tampered coefficients to the total number of coefficients in a specific subband, in order to measure the degree of tampering. They also pointed out if the TAF values decrease monotonically from high resolution to low resolution, then it is very likely that the manipulation is JPEG compression. However, they did not address the situation in which an instance of malicious tampering and an incidental manipulation are imposed simultaneously. (Yu et al., 2001)
- Their quantization-based method encodes a watermark so that the hidden watermark is more/less sensitive to modifications at high/low frequency in the wavelet domain. In this context, over-sensitivity may occur at the small-to- medium scale while under-sensitivity may only happen at the medium-to- large scale. With this understanding, one could make application- dependent decisions on whether an image is credible or not when encountering some modifications.
- The problem associated with their approach is that the results of tampering detection are very unstable. It is well known that the perturbation applied to a wavelet coefficient may make the extracted mark different from or still the same as the embedded one. In other words, the extracted result may be

completely unpredictable. (Lu & Liao, 2003)

- A major drawback is that the method cannot resist incidental modifications and consequently cannot distinguish incidental modifications from malicious tampering.

In this thesis proposed method, it takes the limitations of the human visual system (HVS) into consideration by fixing the quantization step size in each wavelet scale. On the other hand, since any modification applied to an image will change its wavelet coefficients, it is reasonable to expect that their corresponding watermark symbols will be changed, too. By comparing the extracted watermark values with the original hidden ones, the maliciously attacked area can be located. Although the fragility of the watermark proposed in Kundur & Hatzinakos is able to reveal malicious tampering, that watermark is not robust enough to tolerate incidental modifications. Therefore, we address this problem as well by using fine tuning at encoder and fuzzy tampering detection at decoder to draw an unknown fuzzy region between malicious tampering and incidental modifications.

3.2 Overview of the Proposed Technique

The proposed method is described in the context of watermarking still images, but it also works for general multimedia signals. We make use of the DWT domain opposed to spatial or DCT domains to embed the watermark because it provides both a simultaneous spatial localization and a frequency spread of the watermark within the host image. We argue that characterizing the modifications in terms of localized space-frequency distortions in more effective and practical for tampering detection than attempting to parameterize the distortion.

The fundamental advantage of the proposed method lies not only in its ability to detect, with high probability, the spatial and frequency components of the image which are untampered and still credible but also in its ability to distinguish incidental modifications and malicious tampering.

The problem we address is that of the image tampering detection for authentication. As shown in Figure 3.1, the general concept of image authenticated and tampering detection problem can be stated as follows.

Consider the existence of an original image I. After an image authenticated method can get an authentic image I^*. Given an image $\tilde{I}$, which is a possibly modified version of I^*, determine to a high degree of probability, whether $\tilde{I} = I$ without explicit knowledge of the original image I or the authentic image I^*. Thus, if it can be shown that $\tilde{I}$ is equal to I^* almost for certain, then the image $\tilde{I}$ is considered to be *credible*.

Although the general concept of image authentication and tampering detection process is easy to think, we should take this process specifically into a system point of view to put this idea into practice. Figure 3.2 illustrates the image authentication system decision process. At the first step, one has to deal with is to do quality check of the image with human eyes to identify whether an incoming image is credible or not. When a user receives an image, the authenticity of it could be rapidly determined by the pre-attentive perceptibility. That is, if the quality of a received image is too poor to be acceptable (including a highly compressed image), then it is considered not acceptable; otherwise, it is sent to an image authentication system for further verification (at the second step).

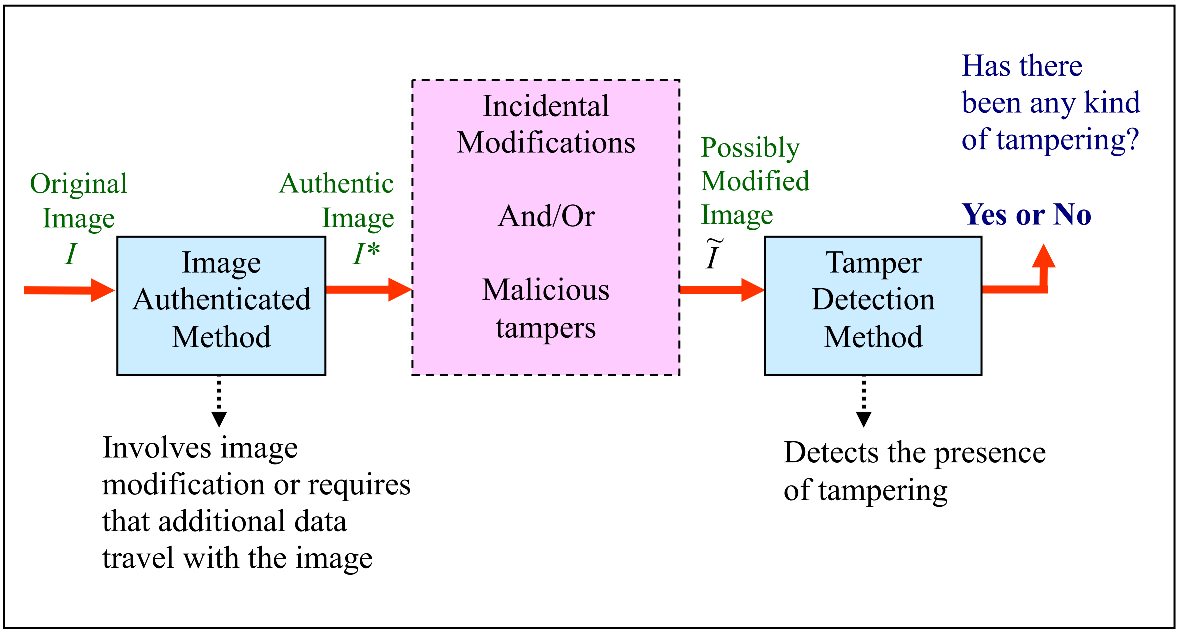


Figure 3.1: The general concept of image authentication and tampering detection.

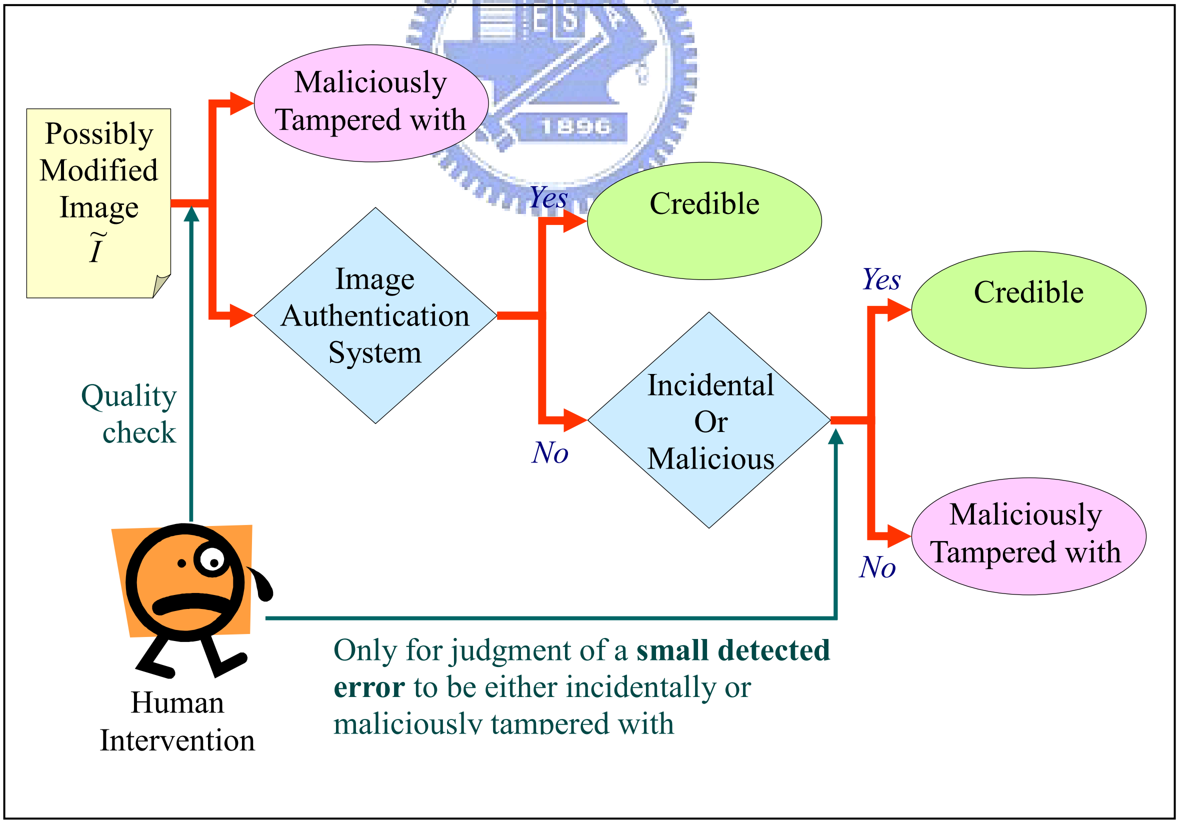


Figure 3.2: The system point of view of image authentication system decision process.

After the verification process, errors might be either detected or not found. If there is no error detected, then the received image is definitely credible; otherwise, it might have been maliciously tampered with or incidentally modified depending on the degree of detected errors. When entering into the third step, it requires human intervention. If the value of *CoSIM* is smaller than a pre-determined threshold, then the received image is not credible. Otherwise, the received image is either incidentally manipulated or maliciously modified. However, sometimes the above mentioned situations are very confusing. Therefore, the human intervention should be introduced to distinguish between these two cases. The assumption is that a meaningful tampering should have the affected pixels aggregate together instead of spreading over the whole image.

3.3 Embedding Process

The starting assumption of the proposed approach is that any modification to an image leads to changes in the corresponding wavelet coefficients and the embedded watermark (Yu et al., 2000). As explained, small modifications in the wavelet coefficients do not change the image significantly, while minor changes in the image alter the coefficients locally, but noticeably. This characteristic is a good premise for watermark invisibility and fragility. In fact, this is the first reason why we have chosen the wavelet domain for our embedding procedure. As shown in Figure 3.3, the main steps of the technique developed are presented here, along with the specific advantages of wavelet coefficients relationships between each other.

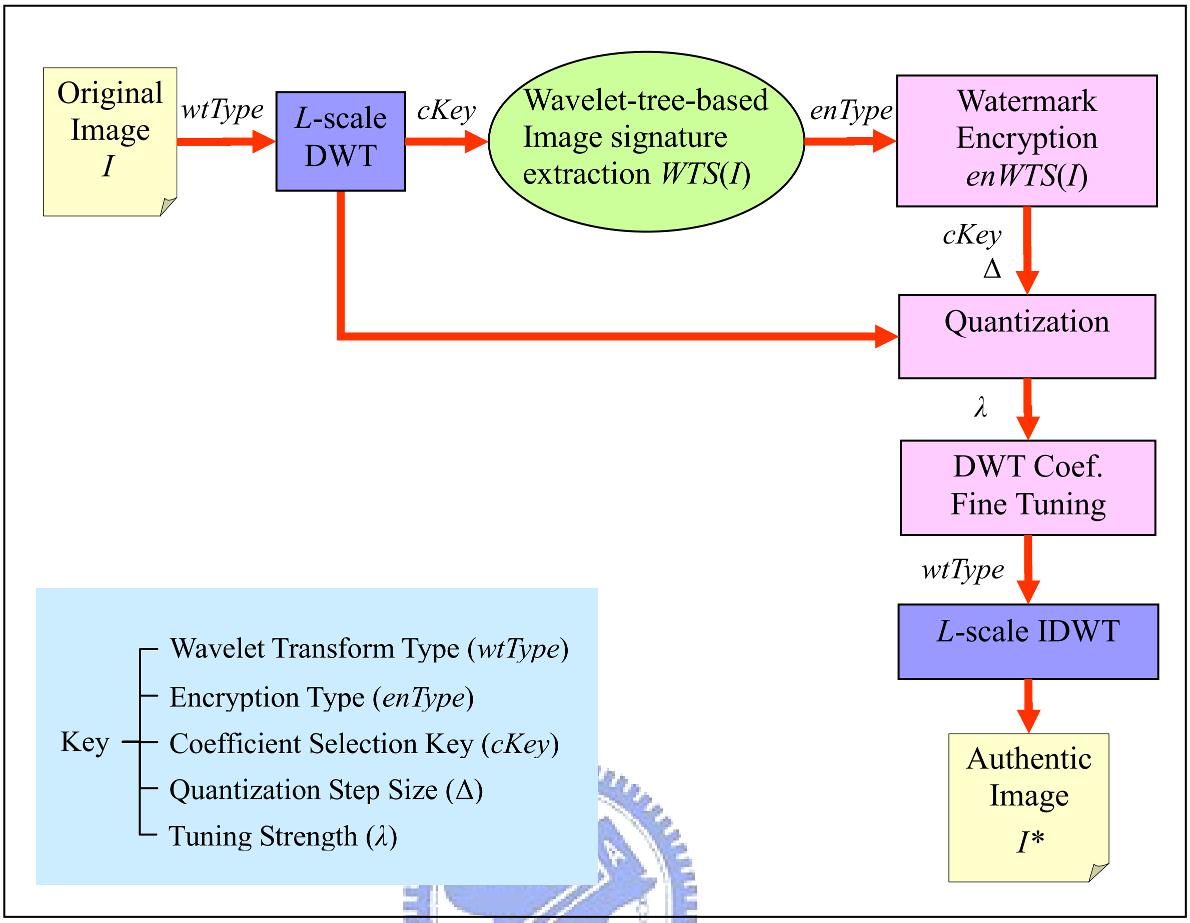
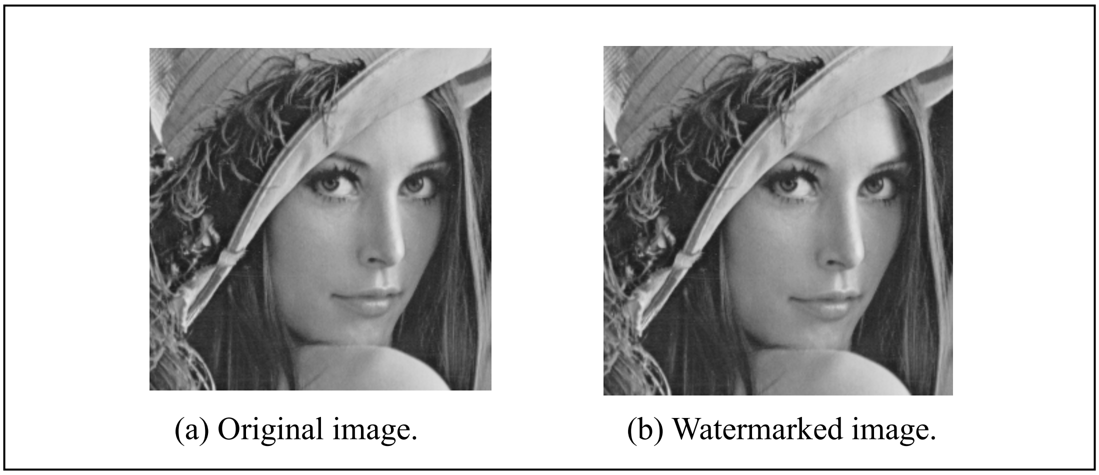

Original Image I
wtType
L-scale DWT
cKey
Wavelet-tree-based Image signature extraction WTS(I)
enType
Watermark Encryption enWTS(I)
cKey
Δ
Quantization
λ
DWT Coef. Fine Tuning
wtType
L-scale IDWT
Authentic Image I*
Key
Wavelet Transform Type (wtType)
Encryption Type (enType)
Coefficient Selection Key (cKey)
Quantization Step Size (Δ)
Tuning Strength (λ)

Figure 3.3: The proposed watermark embedding process.


(a) Original image.

(b) Watermarked image.

Figure 3.4: An example of the original image and the watermarked image. The watermarked image is visually identical to the original unmarked image (*PSNR*=38.47).

1. *L*-scale Discrete Wavelet transform (DWT) of the original image (*I*) is performed. Here we need to determine which wavelet transform algorithm, such as *Haar*, *Daubechies*, *Coiflets*, etc., is used and to decide the scale of wavelet transform by the Wavelet Transform Type (*wtType*). In order to enhance the security of this scheme, we can also use different filters in each wavelet decomposition scale or different wavelet decomposition schemes instead of traditional pyramid scheme. Those would avoid the invader to conjecture the key of *wtType* and gives higher variation and security in the extraction of wavelet tree-based binary image signature (*WTS*) illustrated follows.
2. The wavelet tree-based binary image signature (*WTS*(*I*)) is then extracted through calculating the wavelet tree-based relationships between parent coefficient and its four child coefficients. The Coefficient Selection Key (*cKey*) is needed to allocate the coefficient location (orientation and position) of the image to extract the image signature. This procedure is specifically described in Subsection 3.3.1. In both steps of *WTS* extraction and quantization, the same *cKey* can be used or different *cKey* are used for higher security.
3. The extracted wavelet tree-based binary image signature (*WTS*(*I*)) is then encrypted by the chosen encryption algorithm (Encryption Type, *enType*) to form the encrypted wavelet-tree-based image signature (*enWTS*(*I*)).
4. According to *cKey* and the selected Quantization Step Size (Δ), *enWTS*(*I*) is watermarked into the *L*-scale DWT image by the quantization technique. This quantization technique is detailed in Subsection 3.3.2. After the step of extracting *WTS*(*I*) mentioned in point two, this image-dependent

watermark, obtained from the parent and child's wavelet coefficients relationship, was fixed. Hence, even though the quantization procedure might change the parent-child relationship in watermark embedding process, it doesn't make the watermark detection errors since the original watermark *WTS*(*I*) is necessary in tampering detection procedure.

5. After the above-mentioned embedding step, the image authenticated process was almost finished. However, by reason of increasing the authentic elasticity to distinguish incidental modification and malicious tampering, it needs to tune the DWT coefficients value appropriately to resist incidental small modification. The Subsection 3.3.3 describes the detail of DWT coefficients fine tuning procedure.
6. Finally, the image, through Inverse Discrete Wavelet Transform (IDWT) to go back to spatial domain, changes into an authentic image (*I**). As shown in Figure 3.4, this authentic image (*I**) produced is visually identical to the original unmarked image (*I*).

3.3.1 Wavelet Tree Based Binary Image Signature Extraction

The wavelet-tree-based coefficients context is used in binary image signature extraction. The relationship between parent wavelet coefficient and its child wavelet coefficients can be illustrated in Figure 3.5.

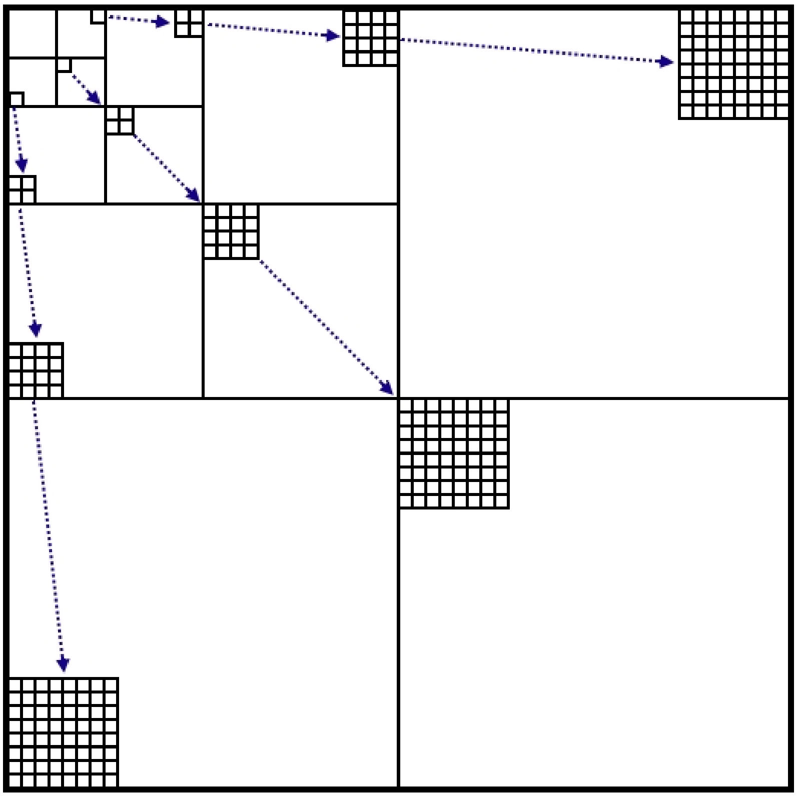


Figure 3.5: The wavelet coefficients relationships between parent coefficient and its four child coefficients.

1896

Let $f_{l,k}(x, y)$ represent a wavelet coefficient, at scale *l*, orientation *k*, and position (*x, y*), in the orthogonally down-sampled wavelet transform domain of an image *I*. Suppose a *L*-scale wavelet transform is performed, then $0 \le l \le L$. It is well known that a large/small scale represents a coarser/finer resolution of an image, i.e., the low/high frequency part. The orientation *k* may be in a horizontal, vertical, or diagonal direction.

The inter-scale relationships of wavelet coefficients can then be converted into the relationships between the parent node $f_{l+1,k}(x, y)$ and its four child nodes $f_{l,k}(2x+i, 2y+j)$ with

$$\begin{cases} |f_{l+1,k}(x,y)| \geq |f_{l,k}(2x+i, 2y+j)| \\ |f_{l+1,k}(x,y)| < |f_{l,k}(2x+i, 2y+j)| \end{cases} \tag{1}$$

where $0 \leq l < L$, $0 \leq i, j \leq 1$, $0 \leq x < N$ and $0 \leq y < M$ ($N \times M$ is the image size). In order to design a reliable scheme for image authentication, we define this signature method as wavelet-tree-based image signature, *WTS*.

According to the inter-scale relationship existing among wavelet coefficients, there are two possible relationship types of a *WTS*:

1) The magnitude of a parent node *p* is larger than that of its child node *c*.
2) The magnitude of a parent node *p* is smaller than that of its child node *c*.

Therefore, we can define the *WTS* as for each $f_{l,k}(x,y)$ where *l=1,2,…,L*, and *k=h,v,d*:

$$\begin{cases} WTS = 1, & \text{if } |p| \geq |c| \\ WTS = 0, & \text{if } |p| < |c| \end{cases} \tag{2}$$

The *WTS* can be obtained by observing the inter-scale relations of wavelet coefficients of an image. The basic concept of *WTS* relies on the following (Lu & Liao, 2003):

1) The inter-scale relationship should be difficult to be destroyed after content-preserving manipulations.
2) This inter-scale relationship should be difficult to be preserved after content changing manipulations.

Because these inter-scale relationships result from the structure of an image *I*, we define them as the wavelet-tree-based image signature of *I*, and call it *WTS*(*I*).

3.3.2 Quantization

Watson et al. (1997) investigated the sensitivity of the human eye and then proposed a wavelet-based human visual system (HVS). According to the HVS, the wavelet coefficients can be modified without causing visual artifacts. In order for a watermarked image to satisfy the transparency requirement, the quantization interval will be defined as the maximally allowable modification quantity based on the HVS.

The basic concept of this thesis proposed is that if the modification quantity of a wavelet coefficient does not exceed its corresponding masking threshold, then this modification will not raise visual awareness. Otherwise, we can say the modification is a malicious one. (Yu et al., 2001)

For an arbitrary wavelet transform, the detail coefficients $\{ f_{l,k}(x,y) \}$ are real numbers. We perform quantization on the wavelet coefficients in the following manner. Every real number is assigned a binary number, as shown in Figure 3.6.

We denote the quantization function by $Q(\cdot)$ which maps the real number set to $\{0,1\}$. Specifically

$$Q(f)=\begin{cases} 0, & \text{if } \left\lfloor \dfrac{f_{l,k}(x,y)}{\Delta} \right\rfloor \text{is even} \\ 1, & \text{if } \left\lfloor \dfrac{f_{l,k}(x,y)}{\Delta} \right\rfloor \text{is odd} \end{cases} \qquad (3)$$

where Δ is positive real number called the quantization step size, $\lfloor \cdot \rfloor$ is the *floor* operator. The following assignment rule is used to embed the image signature *enWTS*(*I*) into the selected coefficient *cKey*(*i*). We denote the coefficient selected by *cKey*(*i*) as $f_{l,k}(x,y)$.

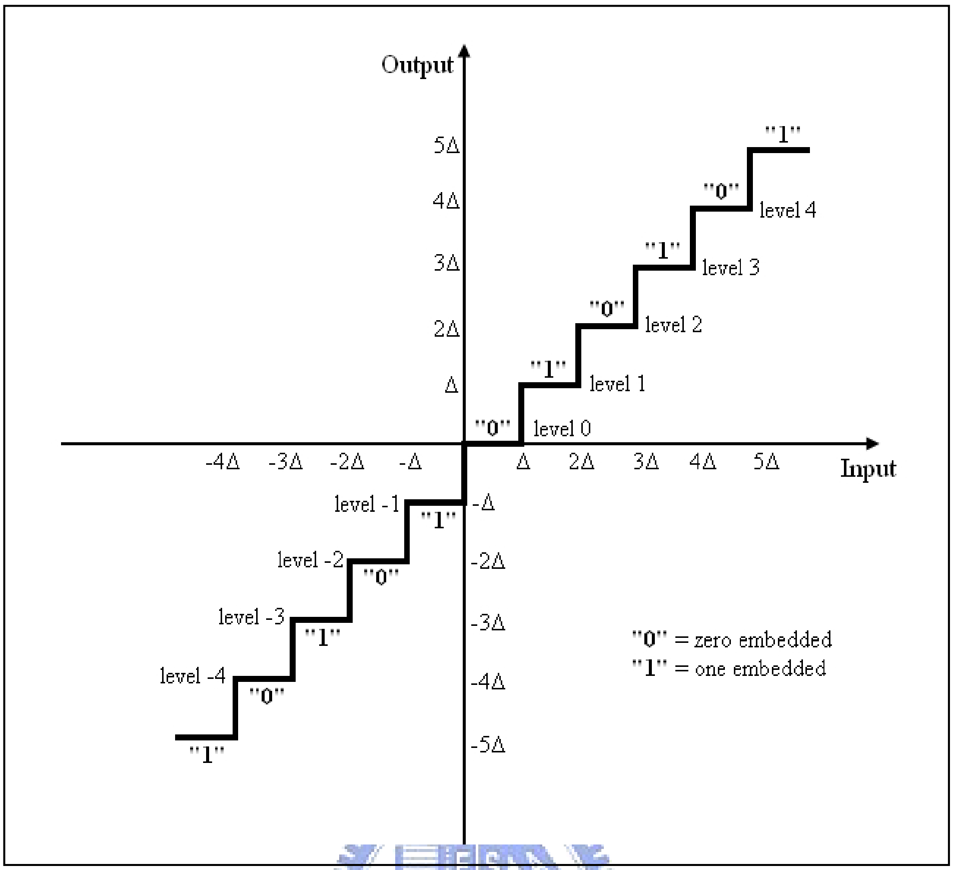


Figure 3.6: The Input/Output relationship in the quantization process.

1896

1. If $Q(f_{l,k}(x.y)) = enWTS(i)$, then no change in the coefficient is necessary.
2. Otherwise, change $f_{l,k}(x,y)$ so that to force $Q(f_{l,k}(x.y)) = enWTS(i)$, using the following assignment:

$$f_{l,k}(x,y) := \begin{cases} f_{l,k}(x,y) + \Delta & \text{if } f_{l,k}(x,y) \le 0 \\ f_{l,k}(x,y) - \Delta & \text{if } f_{l,k}(x,y) > 0 \end{cases} \tag{4}$$

where Δ is the same quantization step size as in Figure 4.6, and (3), and := is the assignment operator.

The nature of the assignment in (4) bas been experimentally found to change the image with the least visual degradation for a given magnitude of Δ. The quantization step size Δ is user defined and is set to establish an appropriate sensitivity to changes in the image. A smaller value of Δ will make the quantization process finer image quality and hence makes minor changes in the image easier to detect.

It is assumed that the specific wavelet transform used is unknown to make forgery difficult. If the wavelet transform were known, it would be possible for a faker to apply it to any arbitrary image and quantize the coefficients using the knowledge of $Q(\cdot)$ in the same way in which it appears in the original watermarked image so that the forgery appears authentic. Therefore, the use of an image-dependent image signature (e.g. *WTS*) to quantize the image is a way to overcome this handicap.

3.3.3 Fine Tuning of Wavelet Coefficients

After quantizing the wavelet coefficients to force the image signature *WTS*, the image authenticated process was almost finished. However, by reason of increasing the authentic elasticity to distinguish incidental modification and malicious tampering and getting better tampering detection result, it needs to tune the DWT coefficients value appropriately to resist incidental small modification.

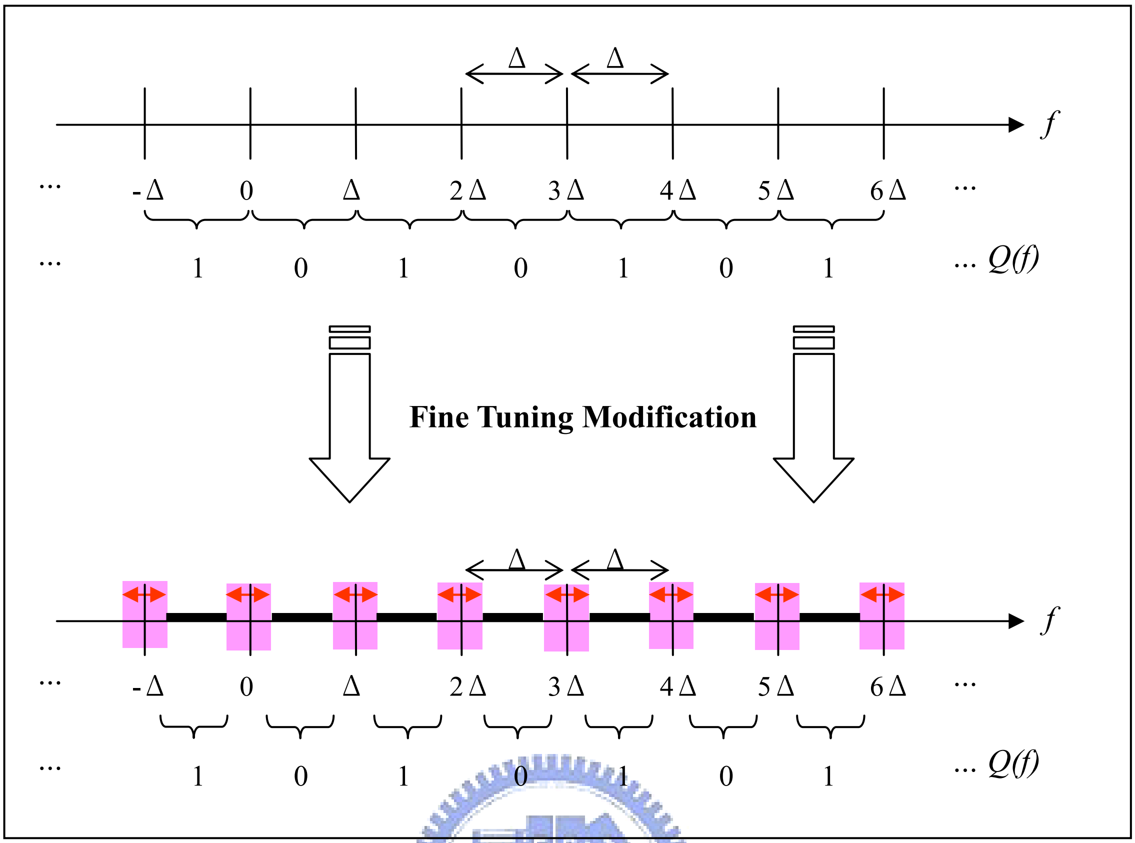


Figure 3.7: The Input/Output relationship after fine tuning modification in the quantization process.

As shown in Figure 3.7, the tuning operation centralizes the value of coefficient to approach the center of each quantization level. In other words, the value of coefficient which situates is close to the edge of each quantization level would be avoided. And then, the output number of quantization function will become discrete. This operation would later make the tampering detection procedure more elastically. The fine tuning operation can be specified as:

$$r = f_{l,k}(x,y) - \left\lfloor \frac{f_{l,k}(x,y)}{\Delta} \right\rfloor \times \Delta \tag{5}$$

$$f^*_{l,k}(x,y) = \begin{cases} f_{l,k}(x,y) + \lambda\Delta & \text{if } r < \lambda\Delta \\ f_{l,k}(x,y) - \lambda\Delta & \text{if } r > (1-\lambda)\Delta \end{cases} \tag{6}$$

where r is the wavelet coefficient's remainder after quantizing step, λ is the tuning strength, usually between 0 to 0.5. The larger the tuning strength, the more centralized the wavelet coefficient close to the center of its quantization level.

3.4 Detecting Process

At the other end of the communication channel or after the image has been stored, the watermarked content needs to be authenticated. Therefore, in order to extract the embedded mark, the tampering detection process focuses on the authenticity of the received image and localizes the tampering if needed. The first few steps of the decoding procedure are identical to the embedding ones. The detailed account of tampering detection procedure is given below, and is revealed in Figure 3.8.

1. L-scale DWT of the possibly modified image ($\tilde{I}$) is performed. The wavelet transform algorithm and the scale of wavelet transform are the same as the embedding one and be recorded in *wtType*.
2. The wavelet tree-based binary image signature ($WTS(\tilde{I})$) is then extracted. The Coefficient Selection Key (*cKey*) is needed to allocate the coefficient location (orientation and position) of the image to extract the image signature.

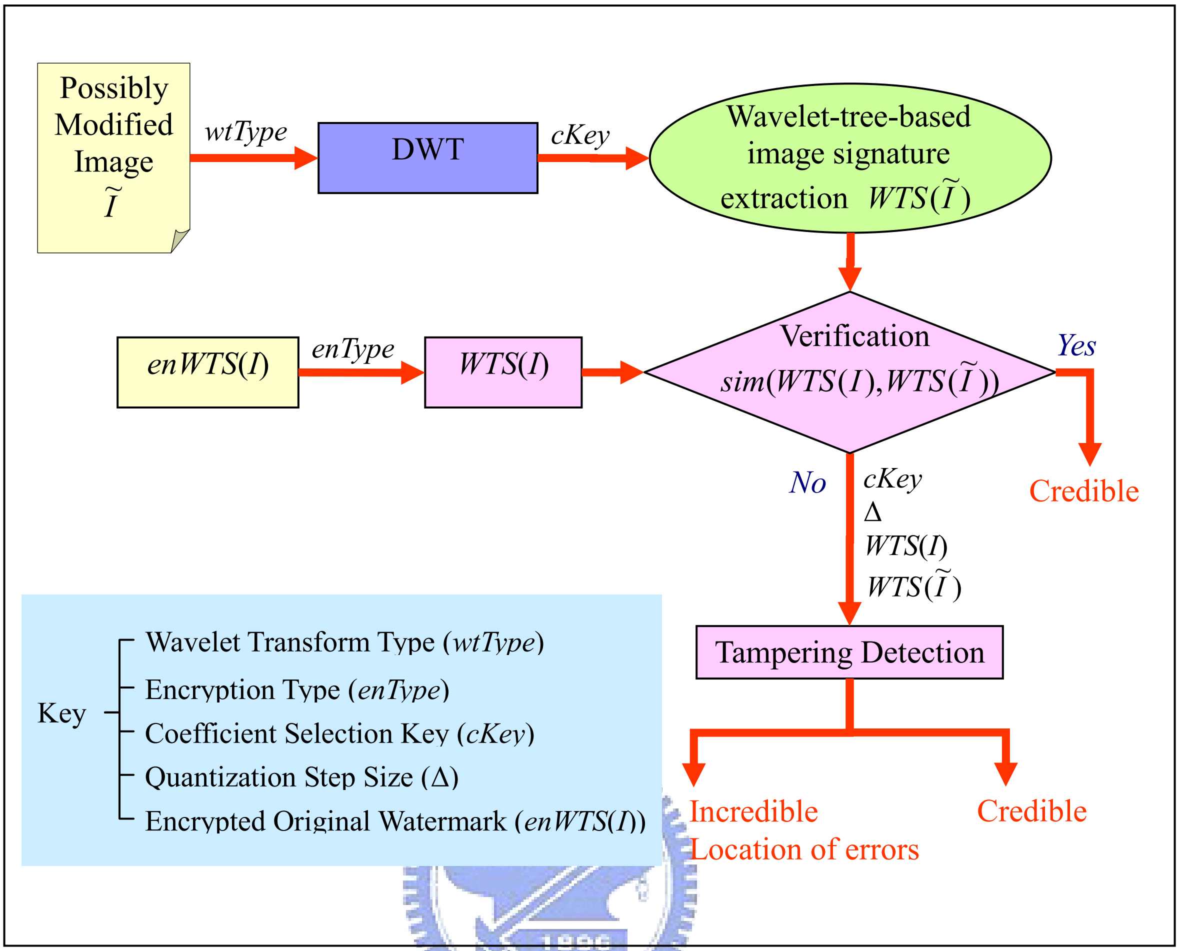

Possibly Modified Image
$\tilde{I}$
wtType
DWT
cKey
Wavelet-tree-based image signature extraction $WTS(\tilde{I})$
enWTS(I)
enType
WTS(I)
Verification
$sim(WTS(I),WTS(\tilde{I}))$
Yes
Credible
No
cKey
Δ
WTS(I)
$WTS(\tilde{I})$
Tampering Detection
Incredible
Location of errors
Credible
Key
Wavelet Transform Type (wtType)
Encryption Type (enType)
Coefficient Selection Key (cKey)
Quantization Step Size (Δ)
Encrypted Original Watermark (enWTS(I))

Figure 3.8: The proposed tampering detection process for image authentication.

3. The encrypted original watermark (*enWTS*(*I*)) is decrypted to obtain the original watermark (*WTS*(*I*)). Here needs the knowledge of encryption type (*enType*) to get the decryption key, with the result that the concern of security is protected.
4. The Co-similarity (*CoSIM*) between *WTS*(*I*) and $WTS(\tilde{I})$ is calculated. If the result of *CoSIM* is upper than the threshold that the user pre-set, we call the image $\tilde{I}$ is credible; otherwise, the image $\tilde{I}$ may suffer some tamper and needs further detection. The step 2 to step 4 is details in

Subsection 3.4.1.

5. If the number of *CoSIM* is lower than the threshold in step 4, it would go to step 5 to carry out the tampering detection procedure and locate the tempering position. This step is described in Subsection 3.4.2.

3.4.1 Wavelet Tree Based Binary Image Signature Verification

To extract the possibly modified image's ($\tilde{I}$) wavelet-tree-based digital signature ($WTS(\tilde{I})$) is the first step of image tamper detection. The extraction technique is similar to the quantization step in watermark embedding process. After the L-scale wavelet transform, the detail coefficients $\{\tilde{f}_{l,k}(x,y)\}$ of $\tilde{I}$ are real numbers. So that $WTS(\tilde{I})$ which maps the real number set to $\{0,1\}$ is extracted in the following operation. For each $\tilde{f}_{l,k}(x,y)$:

1896

$$\begin{cases} WT\tilde{S}_{l,k}(x,y)=0 \quad \text{if } \left\lfloor \dfrac{\tilde{f}_{l,k}(x,y)}{\Delta} \right\rfloor \text{ is even} \\ WT\tilde{S}_{l,k}(x,y)=1 \quad \text{if } \left\lfloor \dfrac{\tilde{f}_{l,k}(x,y)}{\Delta} \right\rfloor \text{ is odd} \end{cases} \tag{7}$$

where $l=1,2,...,L$, and $k=h,v,d$ in accordance with *ckey*. Δ is pre-defined quantization step size, $\lfloor \cdot \rfloor$ is the *floor* operator.

Afterward, the co-similarity (*CoSIM*), between $WTS(I)$ and $WTS(\tilde{I})$, needs to be calculated. One can say the parent and child inter-scale relationship of a pair $\langle p,c \rangle$ in original image *I* is still unchanged in the possibly modified image $\tilde{I}$ if their signature symbols (in a set of $\{0,1\}$) are the same. That is, the relation

$$sym\langle p, c\rangle = sym\langle \tilde{p}, \tilde{c}\rangle \tag{8}$$

hold, where the pair $\langle \tilde{p}, \tilde{c}\rangle$ in $\tilde{I}$ is the corresponding pair of $\langle p, c\rangle$ in I. Finally, the calculate the co-similarity (*CoSIM*), which is defined as

$$CoSIM\left(WTS(I), WTS(\tilde{I})\right) = \frac{N^{+} - N^{-}}{\left|WTS(I)\right|} \tag{9}$$

where N^{+} means the number of pairs satisfying formula (8) and N^{-} means the number of pairs violating this formula. $\left|WTS(I)\right|$ is used to denote the number of parent-child pairs in $WTS(I)$. From formula (9), we know that *CoSIM* will well fall into the interval [-1, 1]. In other words, the *CoSIM* represents ratio of how many parent-child pairs are preserved to satisfy their inter-scale relationships. A larger *CoSIM* means the suspect image $\tilde{I}$ is reliable; otherwise, it means $\tilde{I}$ has been maliciously tampered with. In addition, the location of a tampering region can be easily detected from those tree-based pairs whose signature symbols have been updated.

3.4.2 Tampering Detection Elasticity – Implement of Fuzzy Region

In Section 2.4, the requirements of authentication schemes were introduced. As shown in Figure 2.1, Wu (2002) introduced that the original image is surrounded by a set of images which are surely to be authentic and separated from the set of surely inauthentic images by a fuzzy region where the authenticity of the image is uncertain.

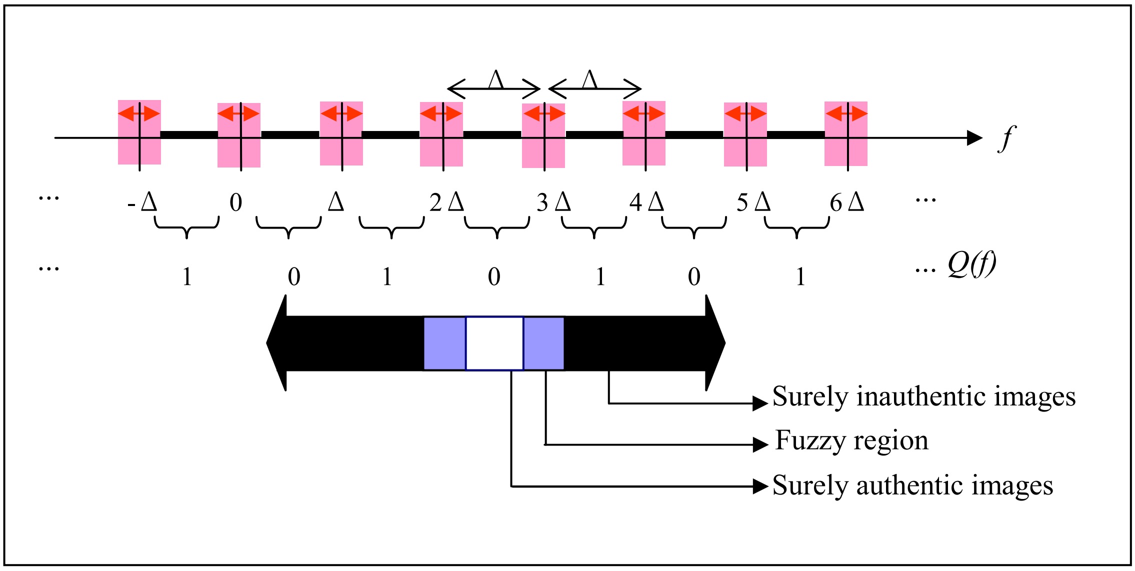


Figure 3.9: Implementation of the fuzzy region to resist incidental modification.

For this reason, the proposed tampering detection technique tries to put the idea of fuzzy region into practice to get better display of location the tampering area. As mentioned in embedding process, the fine tuning technique is just used in increasing the authentic elasticity to distinguish incidental modification and malicious tampering. Figure 3.9 is a diagram which shows the fine tuning technique's capability in implementing the fuzzy region. For each $\tilde{f}_{l,k}(x,y)$,

$$d = \tilde{f}_{l,k}(x,y) - \left\lfloor \frac{\tilde{f}_{l,k}(x,y)}{\Delta} \right\rfloor \times \Delta \tag{10}$$

$$fz = \begin{cases} 0, & \text{if } \rho\Delta \le d \le (1-\rho)\Delta \\ 1, & \text{if } d < \rho\Delta \text{ or } d > (1-\rho)\Delta \end{cases} \tag{11}$$

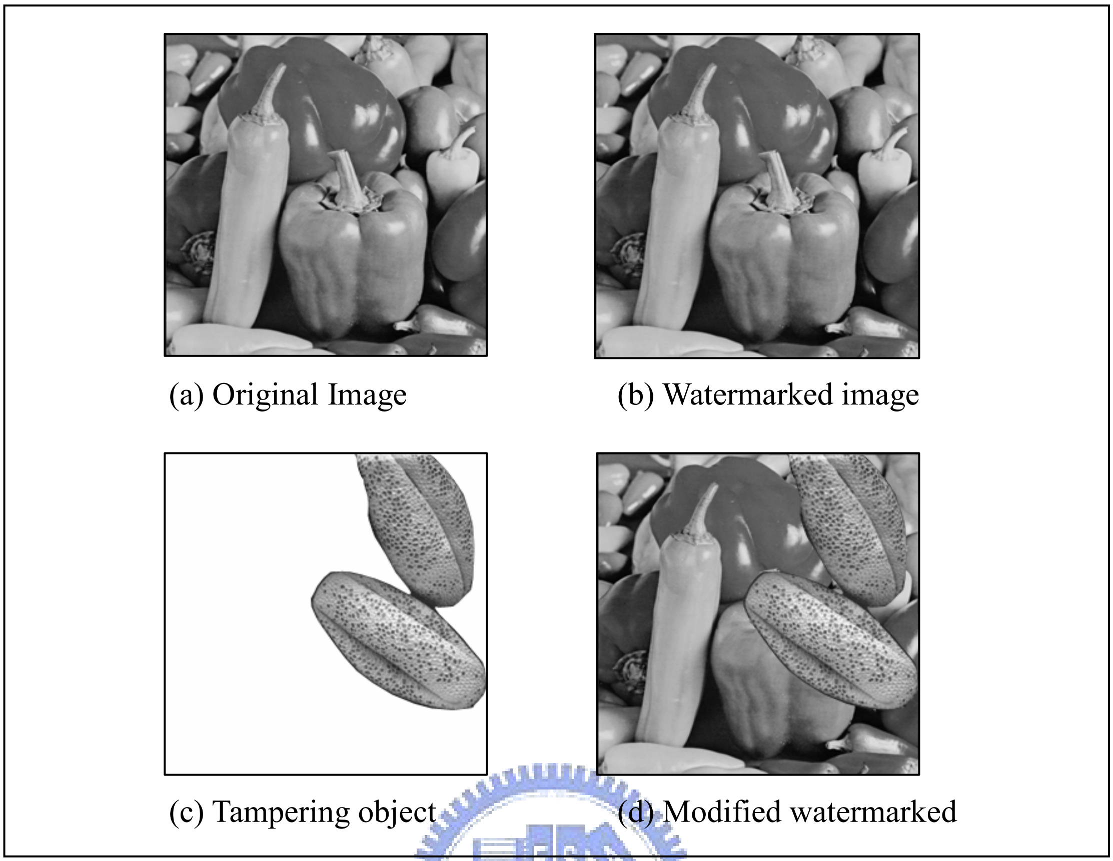
where d is the remainder after the quantization calculation of $\tilde{I}$, fz is fuzzy count of $\tilde{I}$, and ρ is fuzzy strength which is between 0 to 0.5. The larger the fuzzy strength, the larger fuzzy region is held.

4 Experimental Results

To demonstrate the power of the proposed method of image authentication system, Section 4.1 first introduce the experimental setup, and then Section 4.2 presents the parameters settings by way of parametric inference in accordance with a set of test images. The detection results obtained under various incidental distortions is presented in Section 4.3. In Section 4.4, the experimental results are obtained by applying both malicious tampering and incidental manipulation. A set of test images processed be combining different incidental and malicious manipulations was used to estimate the area that was maliciously tampered with. Section 4.5 illustrates the system prototype in accordance with the best parametric inference in Section 4.2. A comparison of the performance of the conventional quantization-based approach and this thesis proposed approach will be made in Section 4.6.

4.1 Experimental Setup

The images used in the experiment were of size 256×256 with256 gray levels and five scale of discrete wavelet transformation. Figure 4.1 is an example showing how a watermarked image is tampered with, including the original image, the watermarked image, the altered area, and the final altered image. The PSNR of the watermarked image shown in Figure 4.1(b) was 38.99dB. Two flower seeds (Figure 4.1(c)) were added as shown in Figure 4.1(b) and formed an image that had been tampered with, as shown in Figure 4.1(d). This set of data was used to test the performance of the proposed approach in the subsequent experiments.



(a) Original Image

(b) Watermarked image

(c) Tampering object

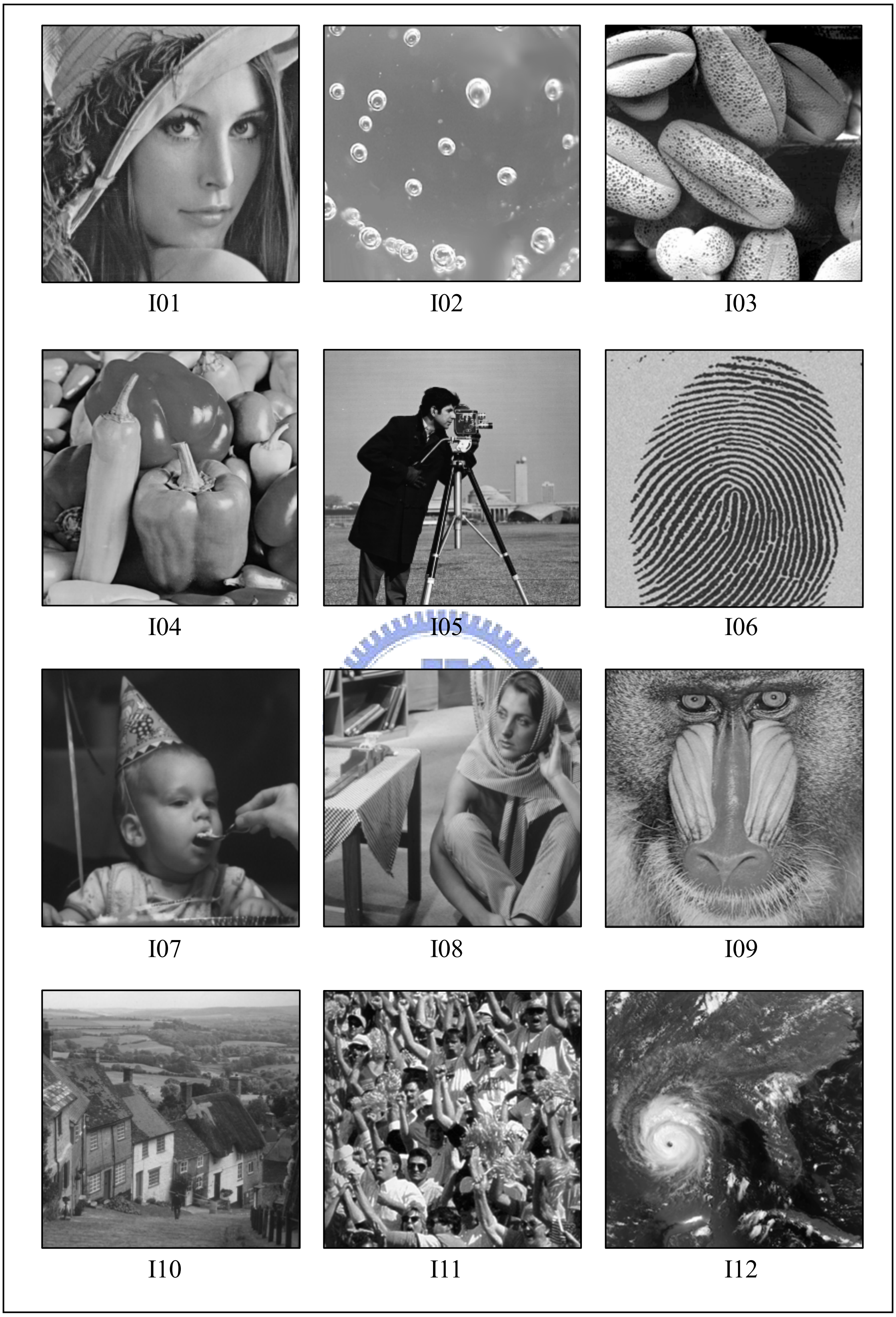
(d) Modified watermarked

Figure 4.1: An example showing malicious tampering by object replacement.

1896

Figure 4.2 shows a set of test images that was used in the experiments. These test image include a low level of detail (Figure 4.2(I01)~(I04)), a medium level of detail (Figure 4.2(I05)~(I08)), and a relatively large amount of detail (Figure 4.2(I09)~(I12)). The complexity of the image composition is an important point to observe the representations of the experimental result.

The set of incidental attacks used in the experiments included JPEG compression, blurring, and sharpening. The mask sizes used in the blurring operation were 3×3, 5×5, and 7×7, respectively. The quality factors adopted for JPEG compression were from 0% to 100%, and the factors used in the sharpening operation were from 10% to 90%. In the experiments, the watermark sequence was embedded in accordance with *ckey* at each scale of a wavelet-transformed image.



I01 I02 I03

I04 I05 I06

I07 I08 I09

I10 I11 I12

Figure 4.2: A set of test images.

4.2 The Parametric Inferences

In both watermark embedding and decoding process, some parameters, such as quantization step size and tuning strength, are set by client side. Choosing a set of appropriate parameters, which will highly affect the experimental results, is very important in both image authentication process and tampering detection task. For this reason, this section will discuss the setting of parameters in accordance with testing the set of test images.

Figure 4.3 and Figure 4.4 present the average *PSNR* and *CoSIM* in accordance with changing tuning strength (λ) from 0 to 0.5. The statistics in Figure 4.3 are as quantization step size (Δ) = 8 and quantization selection key (*ckey*) =*d* as well as Figure 4.4 are as λ=0.3 and *ckey*=*d*. The two figures show that tuning strength does not influence the value of *CoSIM*, but quantization step size suffers by comparison. Having an observation to the relationship between *PSNR* and *CoSIM*, the tuning strength among 0.3 to 0.4 and the quantization step size among 5 to 20, which *PSNR* is above 30dB, have better experimental results. Besides, in Figure 4.3, it is reasonable that the better image quality is gotten as tuning strength lies near the quantization center in between 0.3 to 0.4 since the distribution of wavelet coefficients in each quantization level approximates to normal distribution.

Figure 4.5 shows the average of *CoSIM* suffered from JPEG compression under different Δ and λ settings. The experimental result displays the higher Δ and λ, the better resistance to the JPEG compression. To inspect the settings of Δ and λ, it gets finer result in higher setting of Δ than in higher setting of λ.

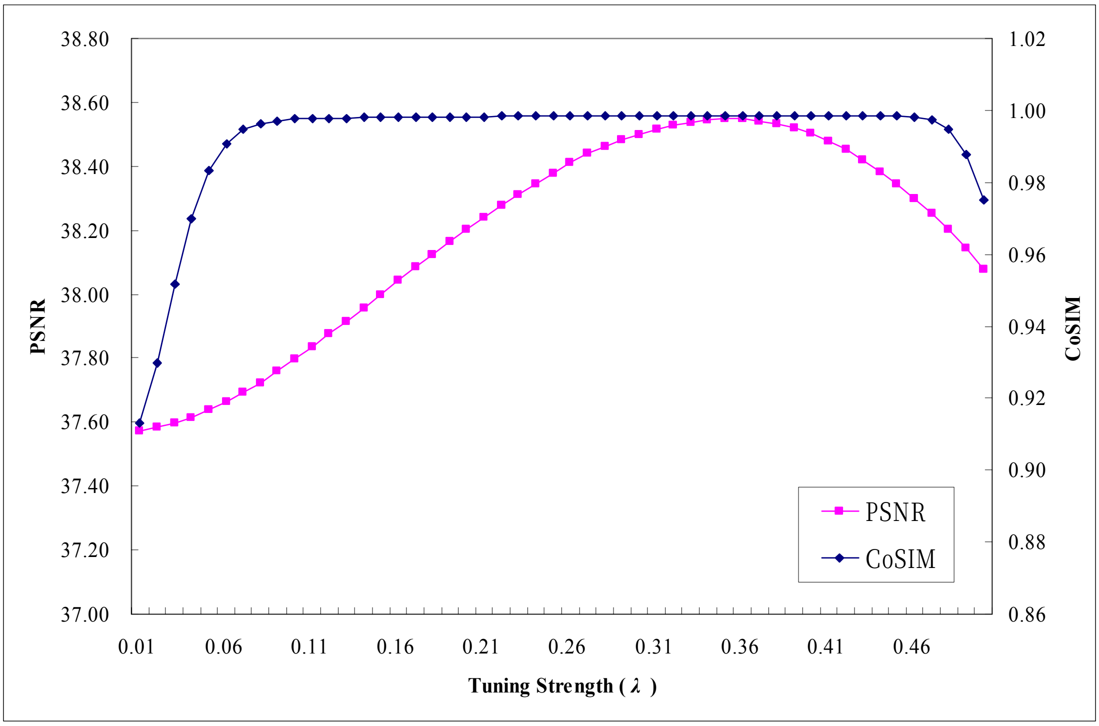


Figure 4.3: The relationship between *PSNR* and *CoSIM* by changing tuning strength.

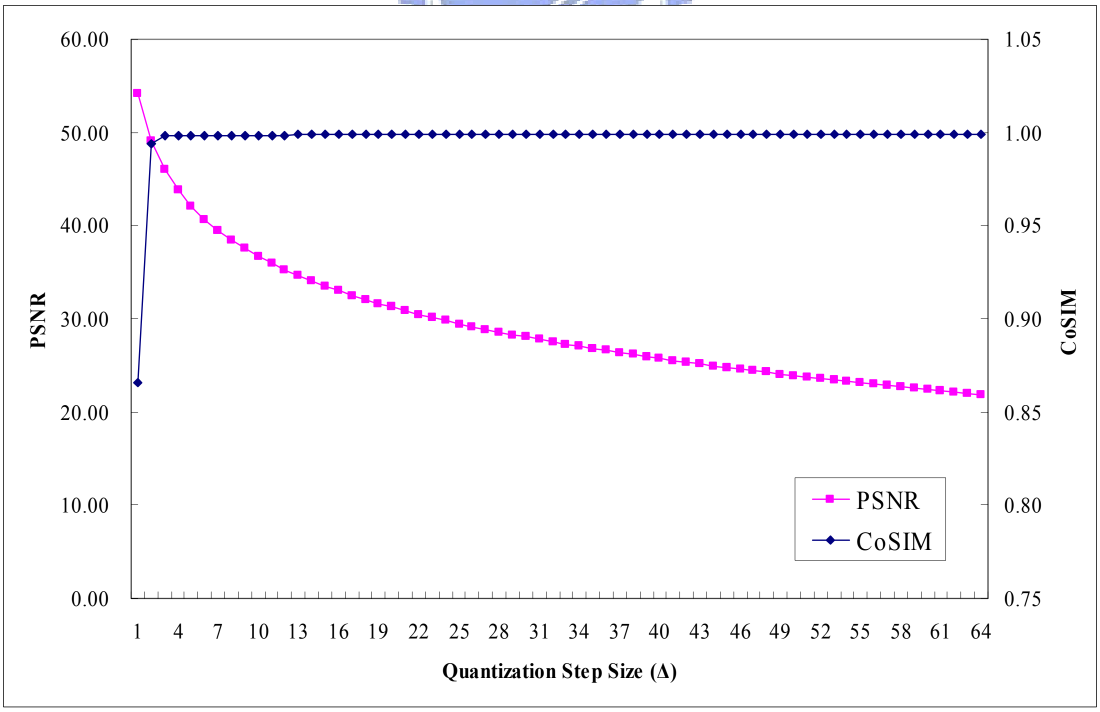


Figure 4.4: The relationship between *PSNR* and *CoSIM* by changing Quantization Step Size.

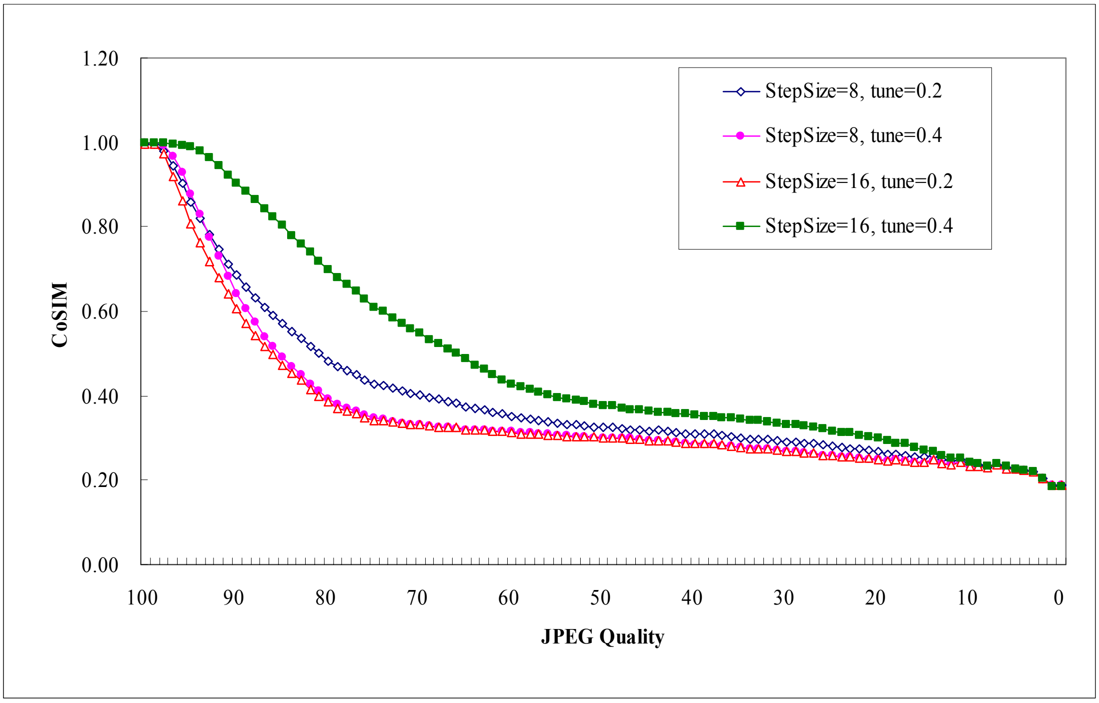


Figure 4.5: The value of *CoSIM* in different JPEG compression quality factor. (*ckey=h,v,d*)

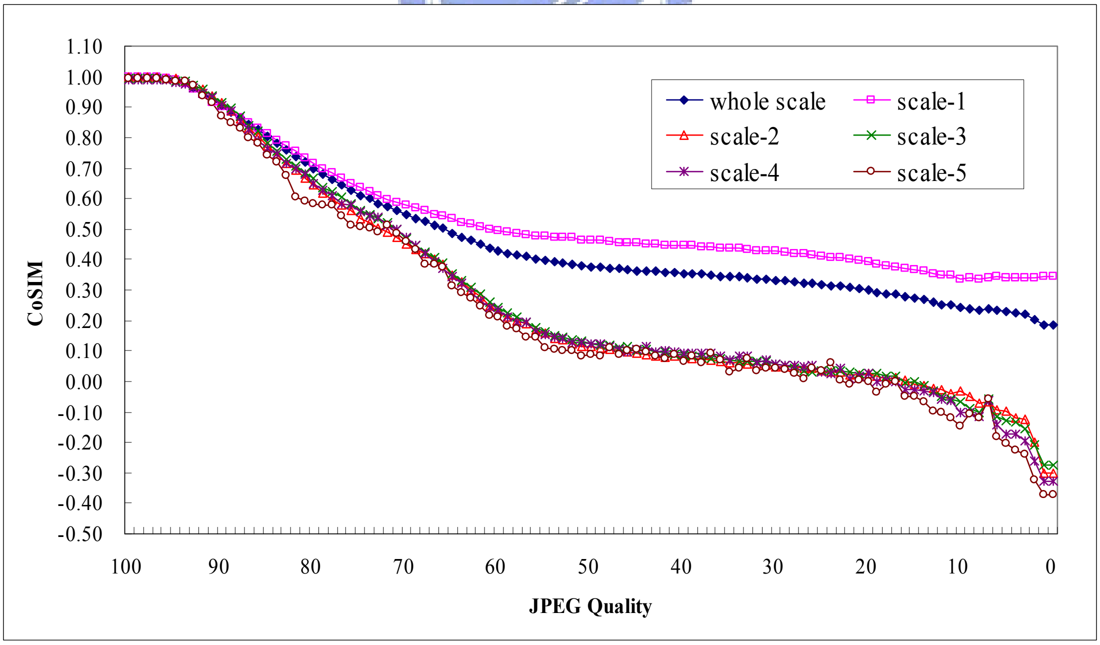


Figure 4.6: The value of *CoSIM* with each scale of wavelet transform in different JPEG compression quality factor. (*ckey=h,v,d*, Δ=16, λ=0.4)

In Figure 4.6, the details of *CoSIM* under different JPEG compression quality factor with each scale of wavelet transformation (where Δ=16 and λ=0.4) are illustrated. It can get the inferences that $CoSIM(WTS(I), WTS(\tilde{I}))$ is heavily induced by $CoSIM(WTS_1(I), WTS_1(\tilde{I}))$. Furthermore, thinking over what degree of JPEG compression is so called incidental modification can be an argument and is subjectively under personal perception. However, on purpose of giving an objective measurement for our proposed method, 80% quality of JPEG compression is set to be defined as incidental modification where *CoSIM* is probably about 0.8.

4.3 The Resistance on Incidental Distortions

In this section, the experiments will check whether the proposed approach could tolerate a number of incidental operations with different degree of alteration. The incidental operations that were applied to the set of test images included JPEG compression, blurring, and sharpening.

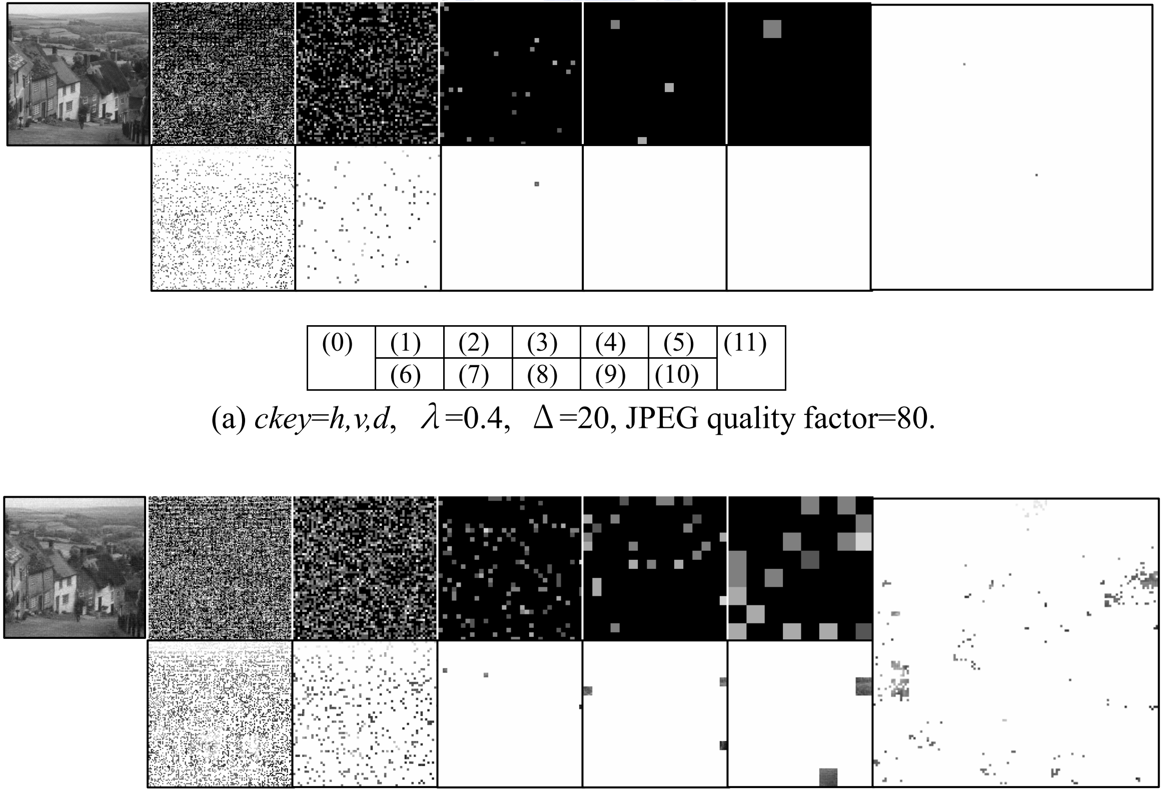
Table 4.1 to Table 4.3 list the results obtained in this experiment. A ✓ symbol indicates that the proposed method treats the operation as an incidental distortion, as a ✗ symbol indicates that the proposed method mistakenly considered the operation to be a malicious one.

Figure 4.7 to Figure 4.9 are the tampering detection image. For example, Figure 4.7(a)-(0) is an image that was modified by performing 80% quality factor of JPEG compression. The detected watermark errors and its strength as scales 1 to 5 are shown in Figure 4.7(a)-(1) to 4.7(a)-(5), respectively. It can be seen that the watermark errors in scale 5 caused by the JPEG compression are much fewer than those caused in scale 1 to 4. The detected watermark errors were then converted into the probability of been maliciously tampered with as shown in Figure 4.7(a)-(6) to 4.7(a)-(10). After performing information fusion, the final detected altered areas were those shown in Figure 4.7(a)-(11). It is apparent that the maliciously modified regions were detected correctly (In this example, no maliciously tempered with).

From the tables and figures, it is obvious that the proposed method could successfully pass almost all the JPEG-compressed images down to quality factor 40%. As for the sharpening operation, the proposed method could successfully tolerate most of the sharpened images up to a 50% sharpening factor. However, in the case of the blurring operation, the proposed method did not work well. The short experimental result on blurring distortion is referred to the inherent defect of the quantization operation, and is reasonable beforehand.

Table 4.1: Tampering detection with JPEG compressing for a set of incidentally manipulated test images. (where ✓ is though as incidental modification; ✗ is though as malicious tampering)

Image Operation	Response											
	I01	I02	I03	I04	I05	I06	I07	I08	I09	I10	I11	I12
JPEG (Q=60%)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
JPEG (Q=55%)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
JPEG (Q=50%)	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓
JPEG (Q=45%)	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓
JPEG (Q=40%)	✓	✓	✗	✓	✓	✗	✗	✓	✓	✓	✓	✓
JPEG (Q=35%)	✗	✓	✗	✗	✗	✗	✗	✗	✗	✓	✓	✗
JPEG (Q=30%)	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
JPEG (Q=25%)	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
JPEG (Q=20%)	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗



(0)	(1)	(2)	(3)	(4)	(5)	(11)
	(6)	(7)	(8)	(9)	(10)	

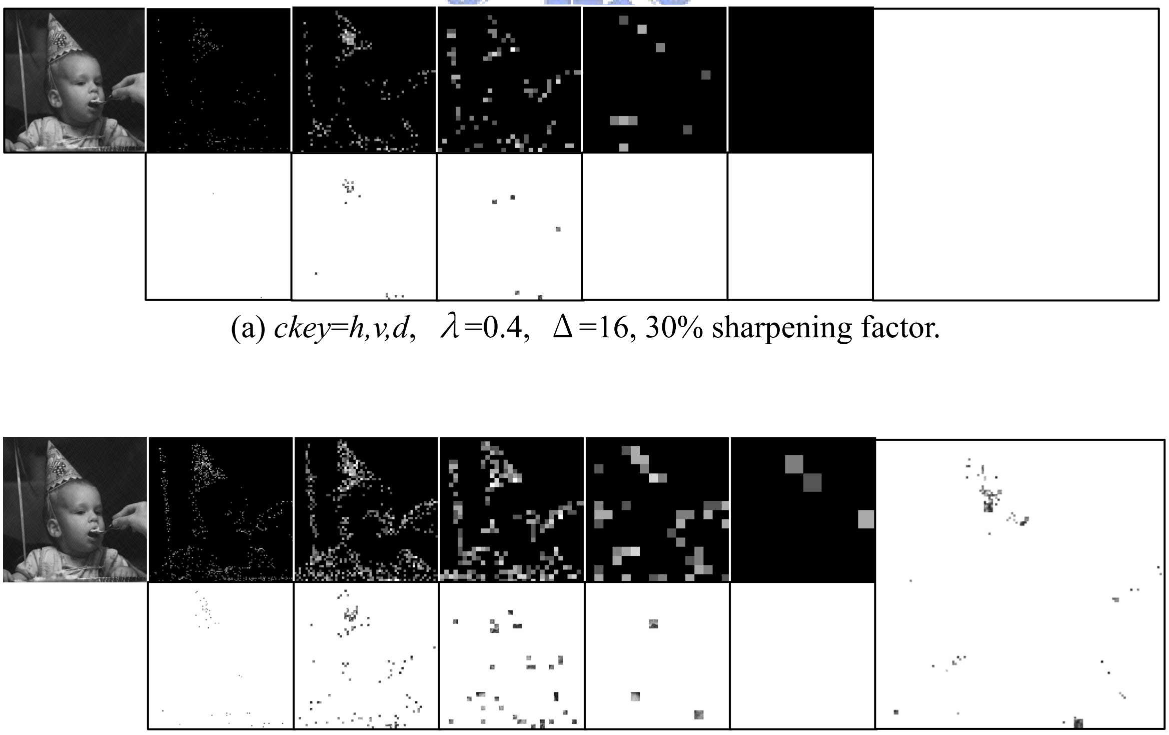
(a) *ckey=h,v,d*, λ=0.4, Δ=20, JPEG quality factor=80.

(b) *ckey= h,v,d*, λ=0.4, Δ=20, JPEG quality factor=50.

Figure 4.7: Tampering detection with different quality factor of JPEG compression.

Table 4.2: Tampering detection with sharpening the whole image for a set of incidentally manipulated test images. (where ✓ is though as incidental modification; × is though as malicious tampering)

Image Operation	Response											
	I01	I02	I03	I04	I05	I06	I07	I08	I09	I10	I11	I12
Sharpen (a=10%)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sharpen (a=20%)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sharpen (a=30%)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sharpen (a=40%)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sharpen (a=50%)	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	×	✓
Sharpen (a=60%)	✓	✓	×	✓	×	×	✓	✓	✓	✓	×	×
Sharpen (a=70%)	×	✓	×	✓	×	×	✓	✓	×	✓	×	×
Sharpen (a=80%)	×	×	×	×	×	×	✓	×	×	×	×	×
Sharpen (a=90%)	×	×	×	×	×	×	×	×	×	×	×	×



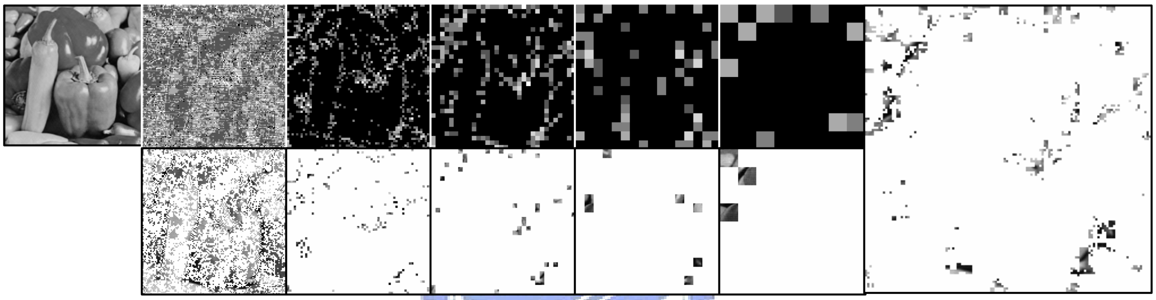
(a) *ckey=h,v,d*, λ=0.4, Δ=16, 30% sharpening factor.

(b) *ckey=h,v,d*, λ=0.4, Δ=16, 50% sharpening factor.

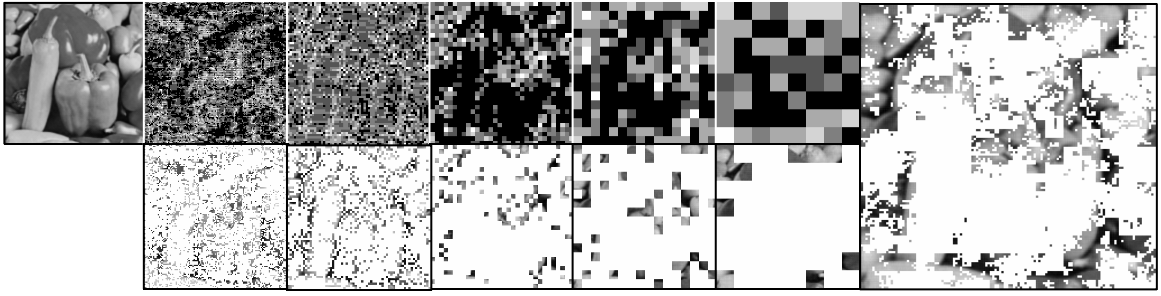
Figure 4.8: Tampering detection with sharpening the whole image.

Table 4.3: Tampering detection with blurring the whole image for a set of incidentally manipulated test images. (where ✓ is though as incidental modification; ✗ is though as malicious tampering)

Image Operation	Response											
	I01	I02	I03	I04	I05	I06	I07	I08	I09	I10	I11	I12
Blurring (3×3)	✓	✓	✗	✓	✗	✗	✓	✓	✗	✓	✗	✗
Blurring (5×5)	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Blurring (7×7)	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗



(a) *ckey=h,v,d*, λ=0.4, Δ=20, 3×3 blurring mask.



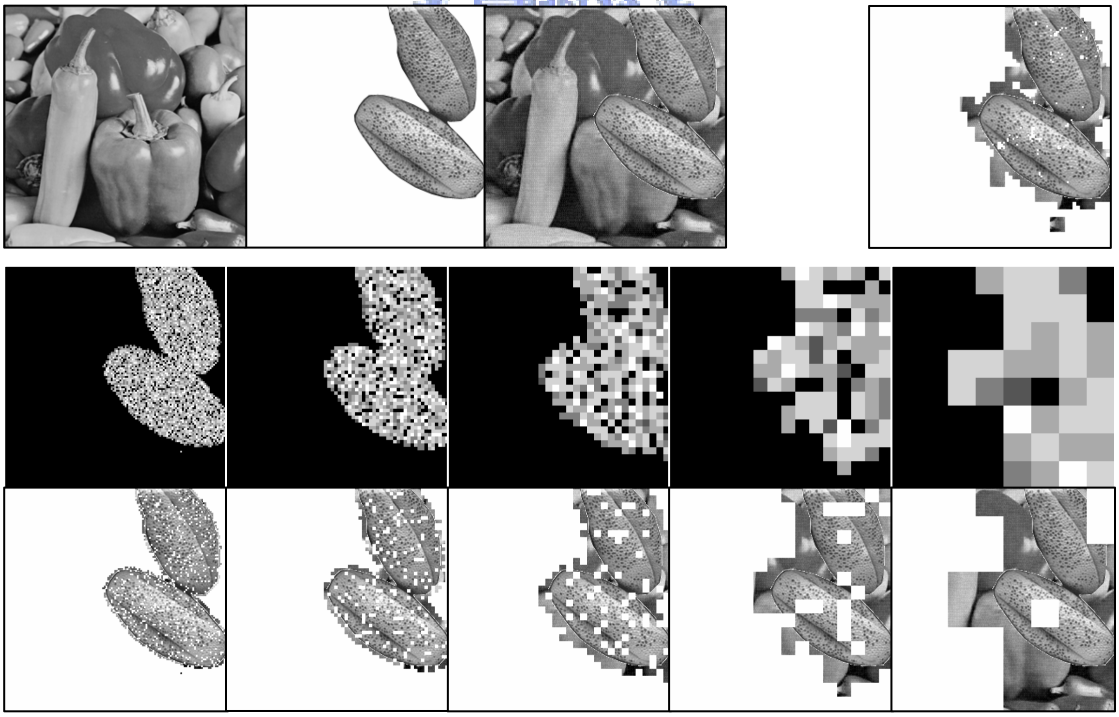
(b) *ckey=h,v,d*, λ=0.4, Δ=16, 5×5 blurring mask.

Figure 4.9: Tampering detection with blurring the whole image.

1896

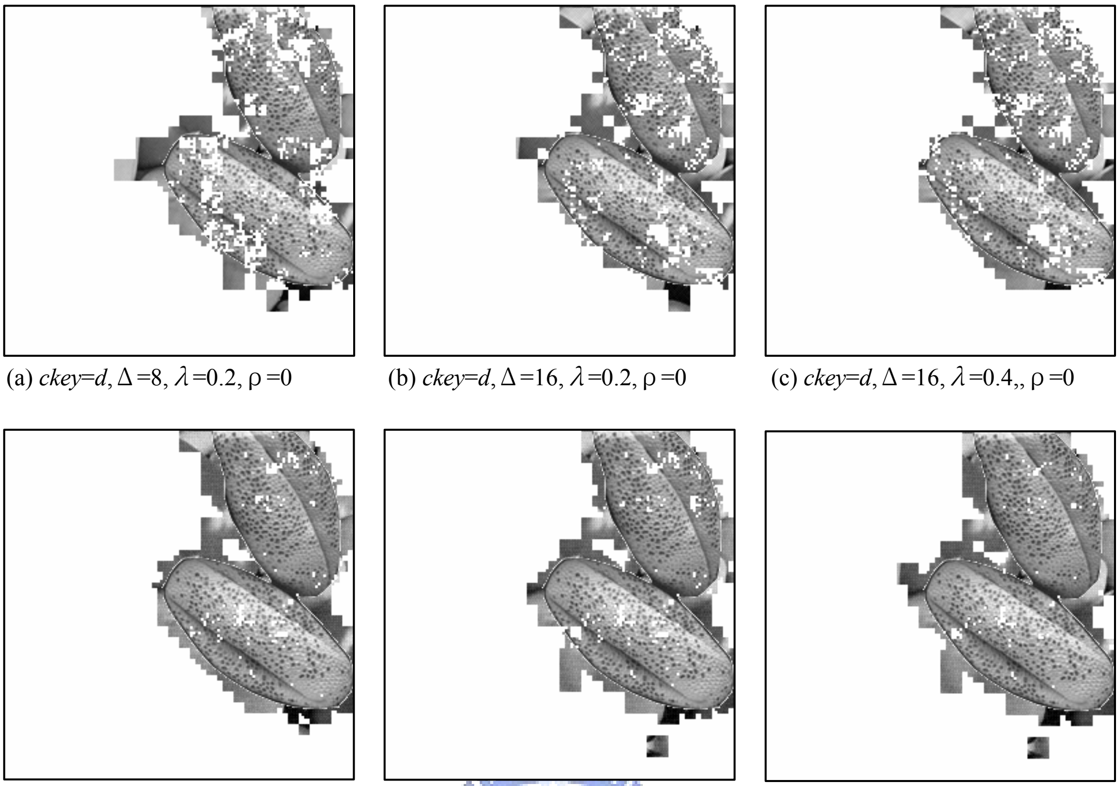
4.4 Tampering Detection of Malicious Distortion

Figure 4.10 is a pepper image that was modified by two seeds replacement. The replacement procedure shown in Figure 4.10(a) to 4.10(c) was mentioned in Section 4.1. The other compositions in Figure 4.10-(1) to 4.10-(10) are the same as Figure 4.7(a)-(1) to 4.7(a)-(11). It is obvious that the coefficients having the sparse type all had lower probability of having been maliciously tampered with at each scale. On the other hand, the areas that corresponded to the regions that were maliciously tampered with all had higher probability of having been maliciously tampered with. The final detected altered areas were represented in Figure 4.10-(11). It is apparent that the maliciously modified regions were detected correctly.



(a)	(b)	(c)		(11)
(1)	(2)	(3)	(4)	(5)
(6)	(7)	(8)	(9)	(10)

Figure 4.10: Malicious distortion of object replacement and its tampering detection result (*ckey=h,v,d*, λ=0.4, ρ=0.3, Δ=16).

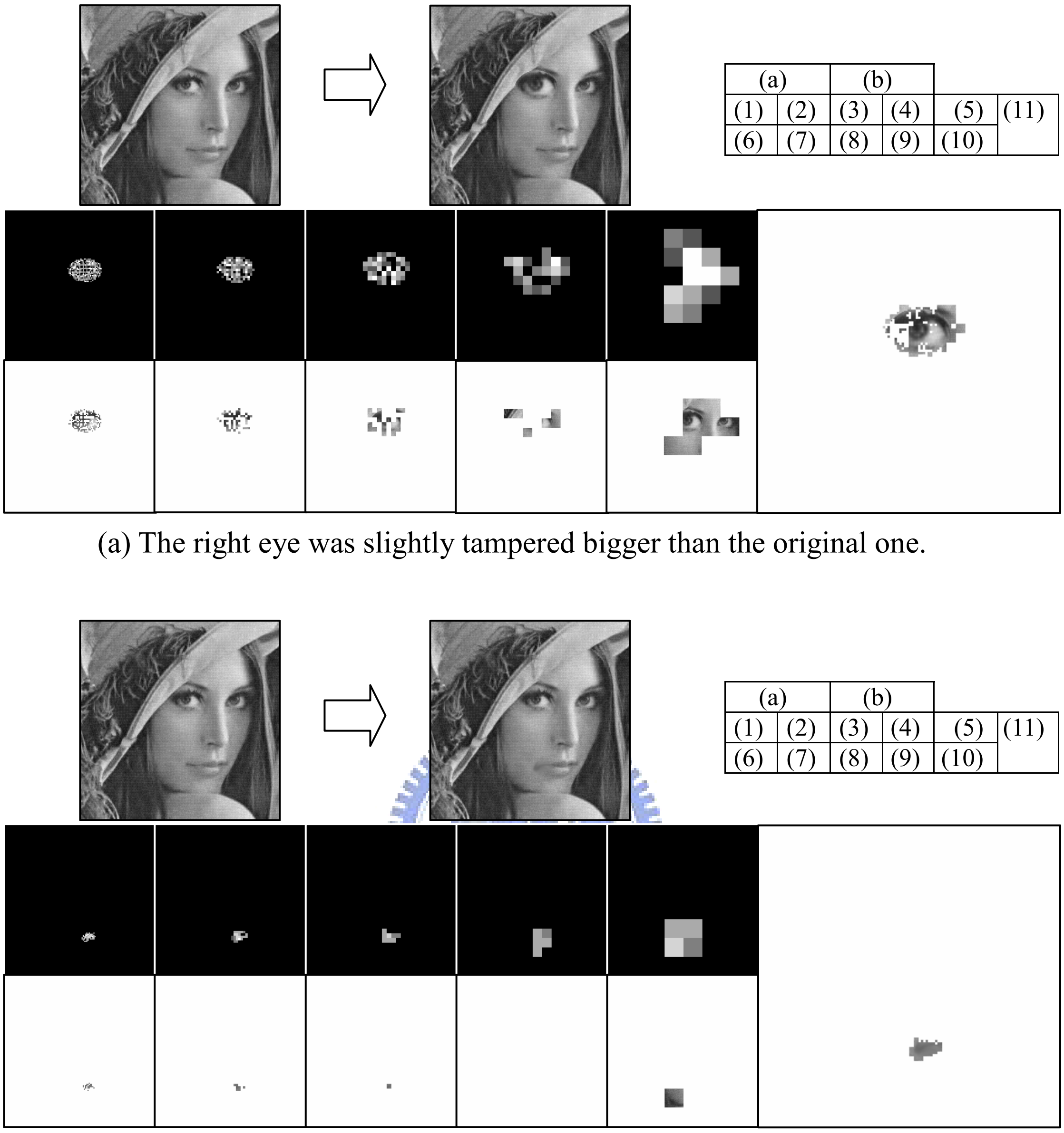


(a) *ckey=d*, Δ=8, λ=0.2, ρ=0 (b) *ckey=d*, Δ=16, λ=0.2, ρ=0 (c) *ckey=d*, Δ=16, λ=0.4,, ρ=0

(a) *ckey=h,v,d*, Δ=16, λ=0.2, ρ=0 (b) *ckey=h,v,d*, Δ=16, λ=0.4, ρ=0 (c) *ckey=h,vd*, Δ=16, λ=0.4, ρ=0.3

Figure 4.11: Malicious distortion of object replacement and their tampering detection results with different settings of parameters.

Figure 4.11 shows the detection results obtained by the proposed method using different settings of parameters. It obviously shows that given larger quantization step size (Δ), tuning strength(λ), or fuzzy strength(ρ) would get better detection result. On the other hand, selecting much more coefficients (*ckey*) to embedding watermarks is also a good choice to enhance the representation of the detected watermark errors. An important condition has to notice is that, in order to obtaining a good tampering detecting result, it must have the image quality dropping off at the expense in watermark embedding process. For this reason, how image quality and detection result are acceptable is depending on the applications.



(a) The right eye was slightly tampered bigger than the original one.

(b) Droop the right corner of the mouth.

Figure 4.12: Sensitivity test of against small image modifications.
(*ckey*=*h,v,d*, Δ=16, λ=0.4, ρ=0)

An experiment to test the sensitivity of the proposed algorithms to small image modifications was performed. The test image used was the Lena (I01) image as shown in Figure 4.2. Figure 4.12(a) shows the right eye was slightly tampered bigger than the original one, and Figure 4.12(b) shows to droop the right corner of the mouth. It is obvious that the proposed method was able to detect it correctly.

4.5 The Resistance on Complex Modification

In this section, some experimental results obtained by applying malicious tampering and an incidental manipulation simultaneously were represented. The objective of these experiments was to check whether the proposed approach could successfully tolerate an incidental manipulation while detecting a malicious attack.

Figure 4.13 is a pepper image that was modified by performing JPEG compression of QF=70, and then followed by two seeds replacement. The replacement procedure was mentioned in Section 4.1 and was shown in Figure 4.1. The compositions in Figure 4.13 are the same as Figure 4.7(a)-(1) to 4.7(a)-(11). It is apparent that the maliciously modified regions were detected correctly.

Figure 4.14 shows another 6 detection results obtained using the proposed algorithms with different quality factor of JPEG compression. In the whole set of experiments, the resolution of the wavelet transform was taken up to 5 scales. From Figure 4.14, it is apparent that our approach did work quite well in most cases when tolerating incidental manipulation like JPEG compression to QF=50.

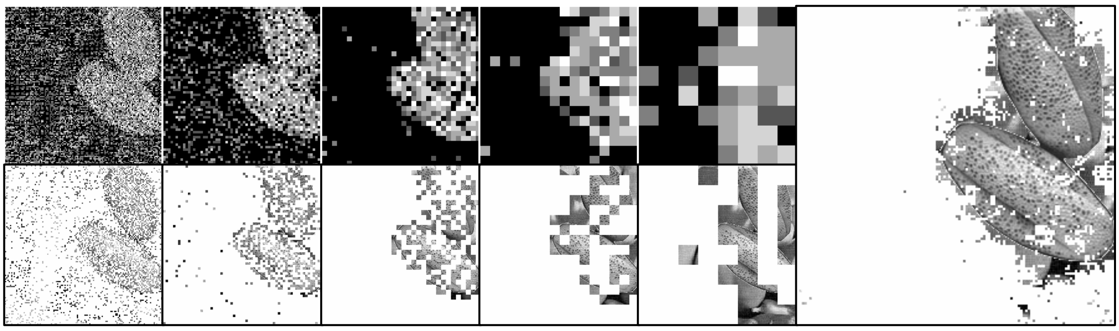
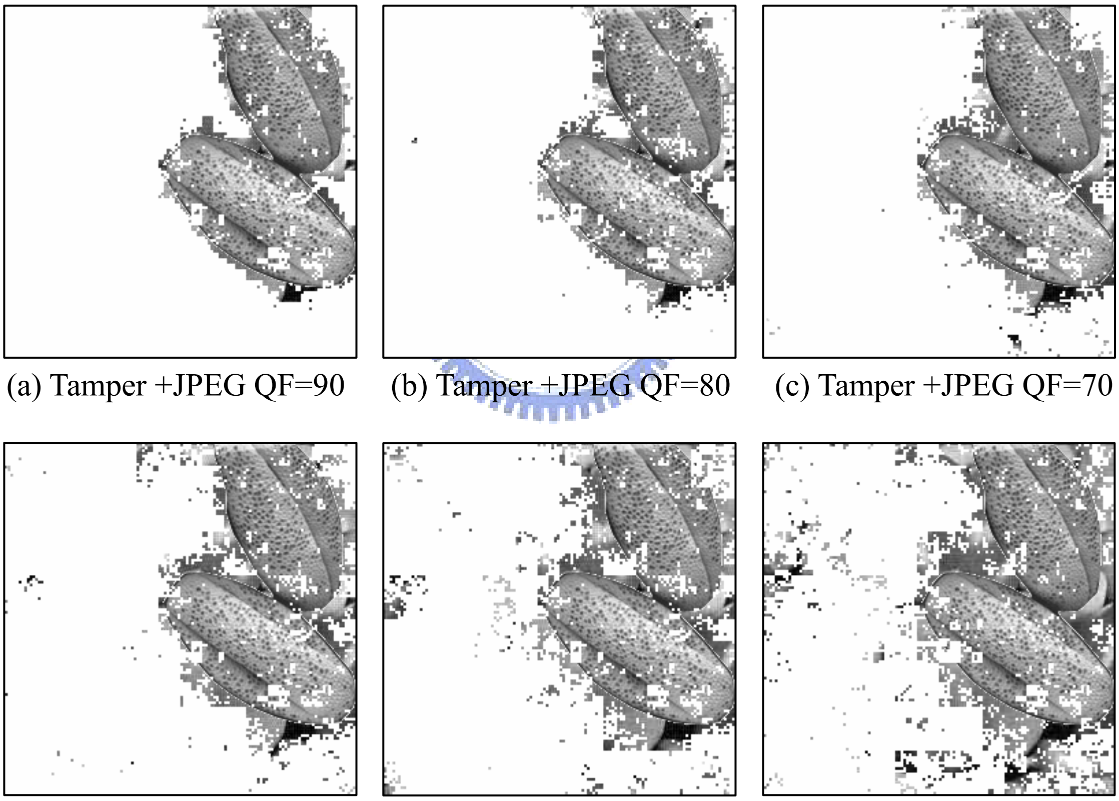


Figure 4.13: Complex distortion of object replacement and JPEG QF=70, and its tampering detection result with each wavelet scale.
(*ckey=h,v,d*, Δ=20, λ=0.4, ρ=0)



(a) Tamper +JPEG QF=90 (b) Tamper +JPEG QF=80 (c) Tamper +JPEG QF=70

(d) Tamper +JPEG QF=60 (e) Tamper +JPEG QF=50 (f) Tamper +JPEG QF=40

Figure 4.14: Complex distortion of object replacement and different JPEG quality factor, and their tampering detection results.
(*ckey=h,v,d*, Δ=20, λ=0.4, ρ=0)

4.6 System Prototype Implementation

The environment of the system is established on c program and Microsoft .Net framework. The input image file can be in JPEG format or BMP format.

In Encoder, users should give image size and set the parameters of tuning strength and quantization step size. The default settings of these parameters are image size=256╳256, tuning strength=0.2, and quantization step size=16. The screen shot of the encoder is shown in Figure 4.15.As shown in Figure 4.16, in the encoding process, system gave the user a watermarked image in JPEG format and its watermark which was already embedded in image in .txt format.

Later, if user can not authenticate the image in human eyes, user can put the possibly modified image and its watermark that is gotten in encoding procedure into tampering detection process. Figure 4.17 and Figure 4.18 represents the screen shot of decoder and the results after tampering detection process. The 5-level images of detected watermark errors, and the image of final detection result would be displayed. User can then authenticate the image in human eyes again according to these tampering detection images.

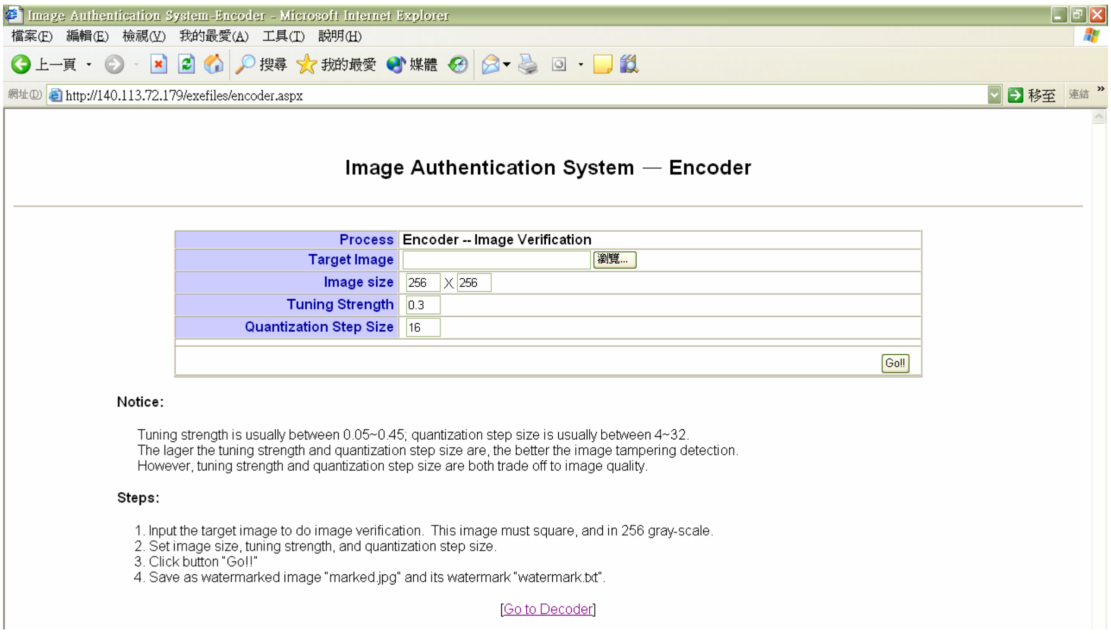


Image Authentication System-Encoder - Microsoft Internet Explorer
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
上一頁 搜尋 我的最愛 媒體
網址(D) http://140.113.72.179/exefiles/encoder.aspx
移至 連結

Image Authentication System — Encoder

Process	Encoder -- Image Verification
Target Image	瀏覽...
Image size	256 X 256
Tuning Strength	0.3
Quantization Step Size	16
	Go!!

Notice:

Tuning strength is usually between 0.05~0.45; quantization step size is usually between 4~32.
The lager the tuning strength and quantization step size are, the better the image tampering detection.
However, tuning strength and quantization step size are both trade off to image quality.

Steps:

1. Input the target image to do image verification. This image must square, and in 256 gray-scale.
2. Set image size, tuning strength, and quantization step size.
3. Click button "Go!!"
4. Save as watermarked image "marked.jpg" and its watermark "watermark.txt".

[Go to Decoder]

Figure 4.15: Screen shot of the system prototype – Encoder.

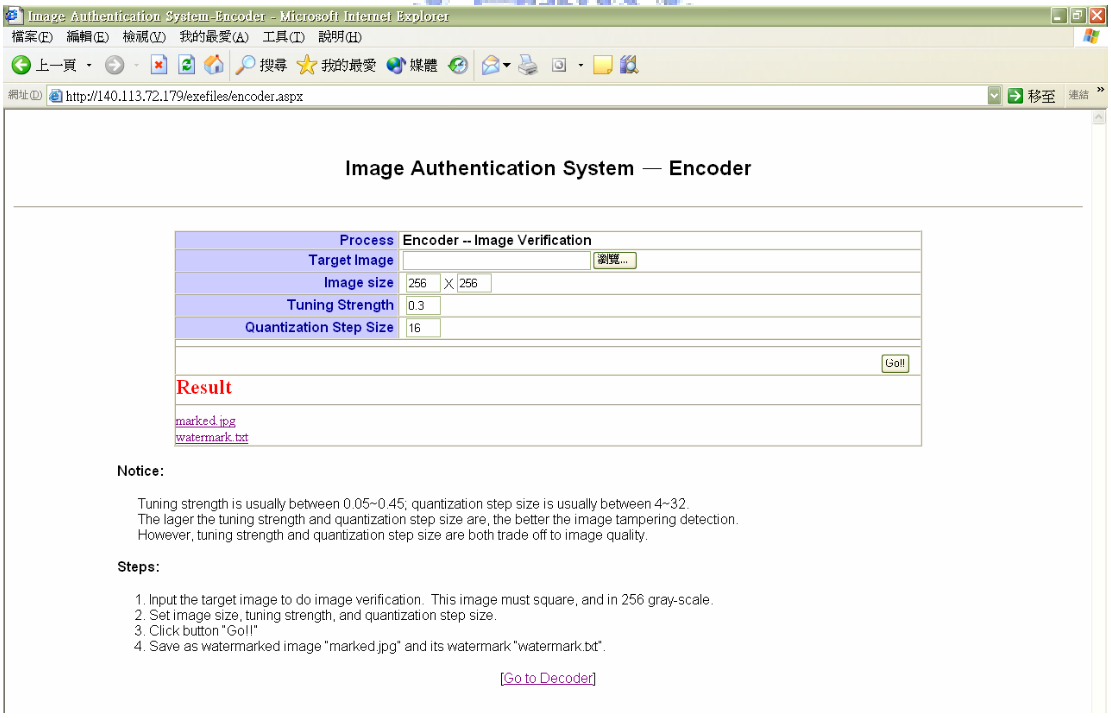


Image Authentication System-Encoder - Microsoft Internet Explorer
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
上一頁 搜尋 我的最愛 媒體
網址(D) http://140.113.72.179/exefiles/encoder.aspx
移至 連結

Image Authentication System — Encoder

Process	Encoder -- Image Verification
Target Image	瀏覽...
Image size	256 X 256
Tuning Strength	0.3
Quantization Step Size	16
	Go!!

Result

marked.jpg
watermark.txt

Notice:

Tuning strength is usually between 0.05~0.45; quantization step size is usually between 4~32.
The lager the tuning strength and quantization step size are, the better the image tampering detection.
However, tuning strength and quantization step size are both trade off to image quality.

Steps:

1. Input the target image to do image verification. This image must square, and in 256 gray-scale.
2. Set image size, tuning strength, and quantization step size.
3. Click button "Go!!"
4. Save as watermarked image "marked.jpg" and its watermark "watermark.txt".

[Go to Decoder]

Figure 4.16: System replied the watermarked image and its watermark to user after encoding process.

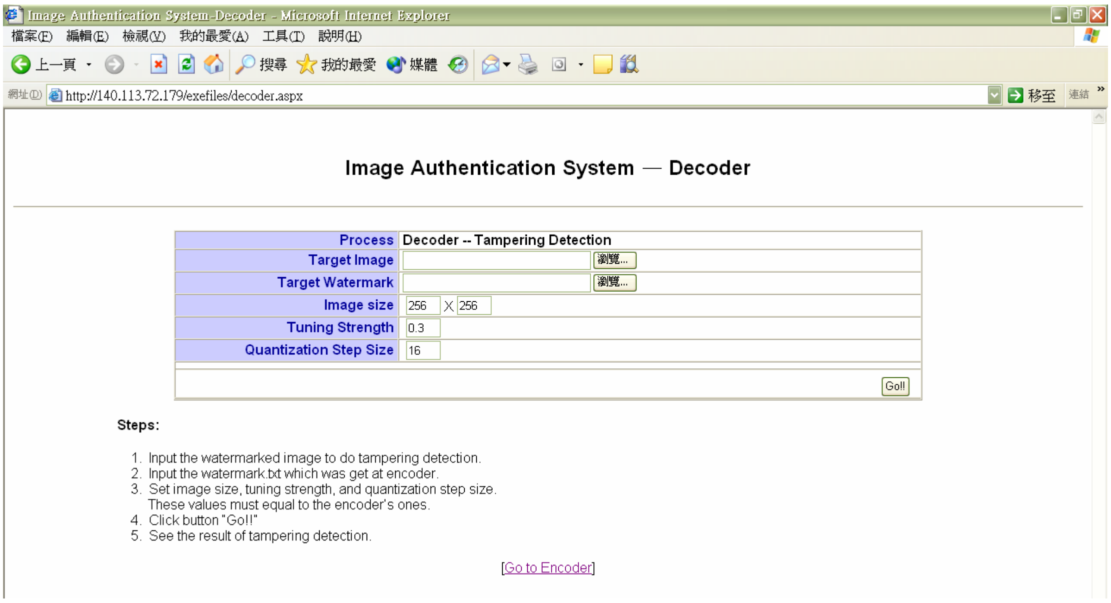


Image Authentication System-Decoder - Microsoft Internet Explorer
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
上一頁 搜尋 我的最愛 媒體
網址(D) http://140.113.72.179/exefiles/decoder.aspx
移至 連結
瀏覽...
Go!!

Image Authentication System — Decoder

Process	Decoder -- Tampering Detection
Target Image	
Target Watermark	
Image size	256 X 256
Tuning Strength	0.3
Quantization Step Size	16

Steps:

1. Input the watermarked image to do tampering detection.
2. Input the watermark.txt which was get at encoder.
3. Set image size, tuning strength, and quantization step size. These values must equal to the encoder's ones.
4. Click button "Go!!"
5. See the result of tampering detection.

[Go to Encoder]

Figure 4.17: Screen shot of the system prototype – Decoder.

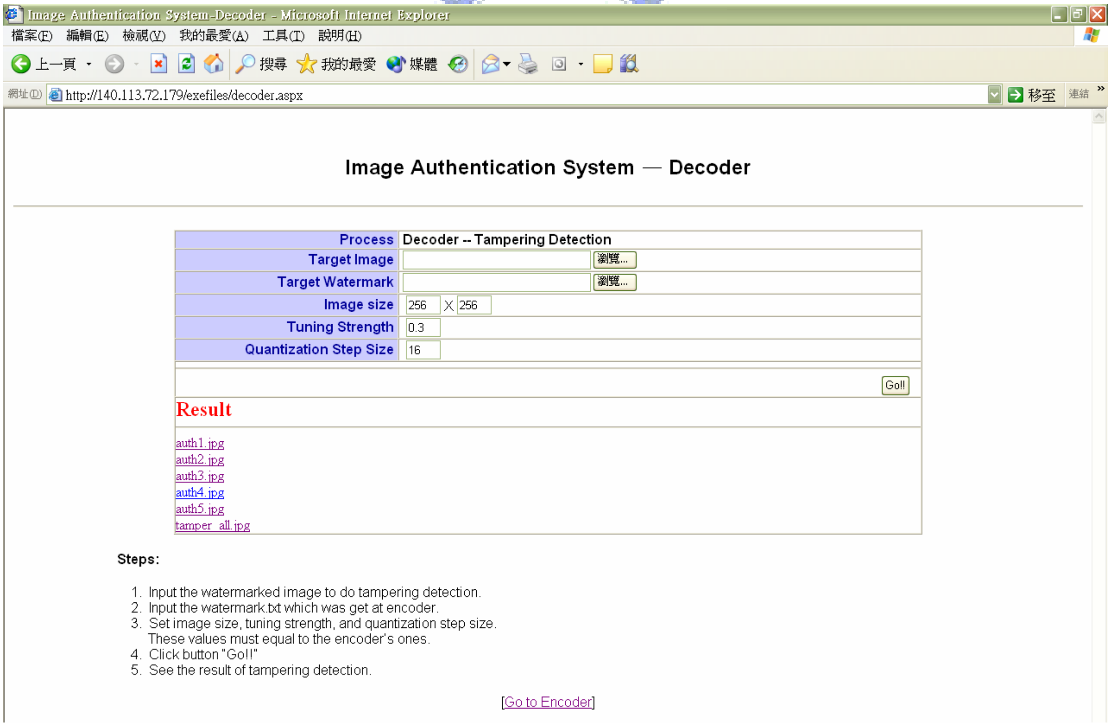


Image Authentication System-Decoder - Microsoft Internet Explorer
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
上一頁 搜尋 我的最愛 媒體
網址(D) http://140.113.72.179/exefiles/decoder.aspx
移至 連結
瀏覽...
Go!!

Image Authentication System — Decoder

Process	Decoder -- Tampering Detection
Target Image	
Target Watermark	
Image size	256 X 256
Tuning Strength	0.3
Quantization Step Size	16

Result

auth1.jpg
auth2.jpg
auth3.jpg
auth4.jpg
auth5.jpg
tamper_all.jpg

Steps:

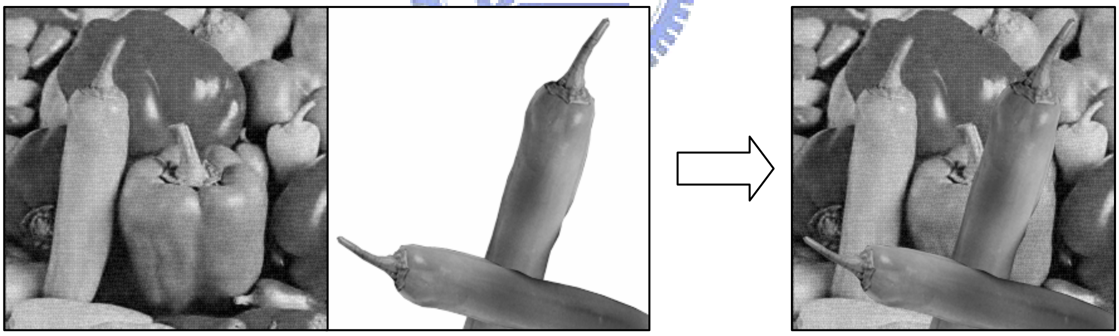
1. Input the watermarked image to do tampering detection.
2. Input the watermark.txt which was get at encoder.
3. Set image size, tuning strength, and quantization step size. These values must equal to the encoder's ones.
4. Click button "Go!!"
5. See the result of tampering detection.

[Go to Encoder]

Figure 4.18: The 5-level images of detected watermark errors showed after tampering detection process.

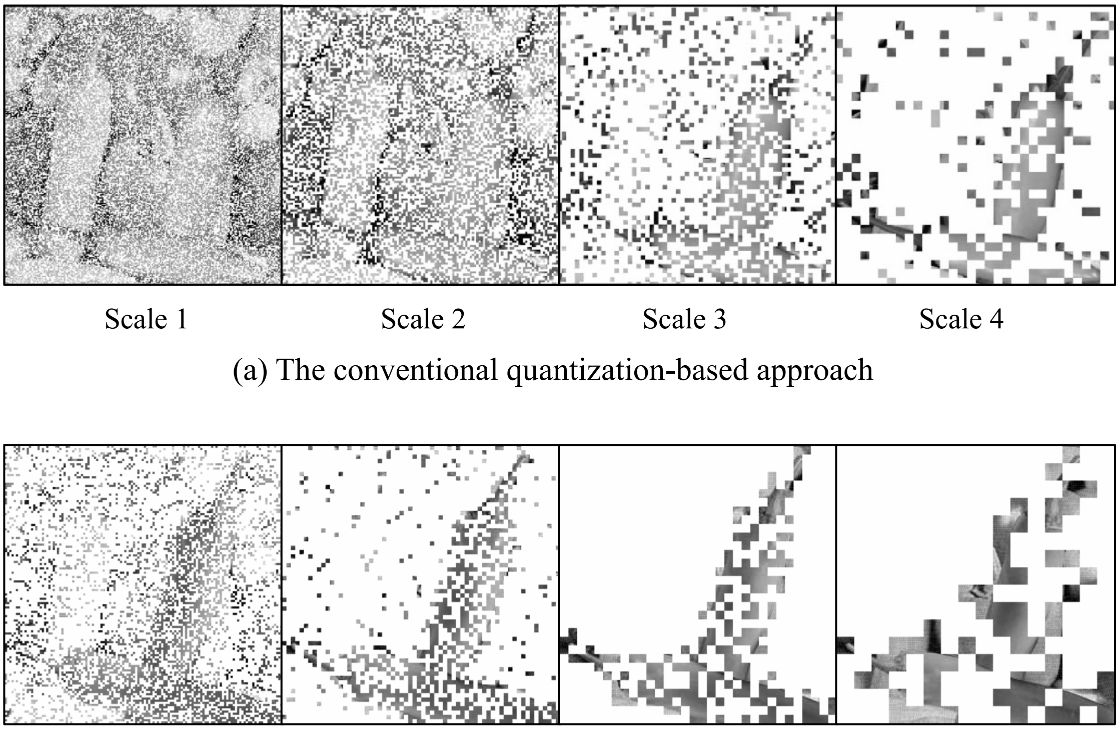
4.7 Comparison with the Conventional Quantization-based Approach

In this Section, our proposed approach is comparison with the conventional approach (Kundur & Hatzinakos, 1999). The maliciously attacked image shown in Figure 4.19(c), subjected to JPEG compression with a quality factor = 60, was used as the test image. The watermark errors (at scales 1 to 4) obtained by applying the conventional quantization-based approach and the proposed approach are shown in Figure 4.20(a) and 4.20(b), respectively. It is obvious that the results obtained by applying the proposed approach are better than those obtained by applying the conventional approach.



(a) Watermarked image (b) Tampering object (c) Tampered image

Figure 4.19: The maliciously attacked image.



Scale 1 Scale 2 Scale 3 Scale 4

(a) The conventional quantization-based approach

Scale 1 Scale 2 Scale 3 Scale 4

(b) The proposed approach

Figure 4.20: Comparison of detected watermark errors obtained using the conventional quantization-based approach and our proposed approach. (ckey=h,v,d, Δ=20, λ=0.4, ρ=0)

5 Conclusion

The protection of visual content is becoming an important issue as the use of digital images increases. In the context, this thesis has studied the utilization of watermarking for content protection of digital images. In summary the first part of the thesis gives an overview of wavelet transform, digital watermarking, and image authentication methods. The emphasis is, however, mainly put on the development of a wavelet transform based digital watermarking for image authentication and tampering detection.

The main contribution of the thesis is the development of a novel, semi-fragile watermarking-technique for image authentication. This thesis elaborates a secure watermarking technique for which the specific domain of embedding known only by the creator. In addition to the contribution to watermarking for image authentication, the overall image authentication and tampering detection methods are categorized, and use this classification to summarize previously proposed techniques.

This thesis clearly improves existing image authentication techniques and introduces new concepts for the use of watermarking. Nonetheless, there are still many aspects that can be further investigated, and regarding which the overall image authentication technology can be improved. This discussion can be divided into three part of algorithm, display of the detected watermark errors, and system implementation.

In the aspect of image authentication algorithm, the optimization of the embedding process for particular types of images may augment the embedding capacity – the amount of information carried by a host. So that the overall performance of the proposed method may be further enhanced by the use of color images, since their capacity to accept an invisible mark is greater that one of gray-scale images, due to the presence of chrominance information, in addition to the luminance. On the other hand, in order to increase the sensitivity of small modification, using different image-dependence watermark extracting technique, such as features, histogram, and so on, is also a point to thought to get better tampering detection result. Additional work, such as using data hiding technique covering up the image information to reach the recovered image after tampering detecting process; or using different watermarking embedding method, instead of quantization method, to enhance the robust of watermark slightly that can better resist on some attacks in spatial domain such as blurring. The augmentation of the mark's resistance to JPEG-2000 compressions by the extraction of compression-invariant images' characteristics in the wavelets' domain and their use in the embedding process is another idea about image authentication.

According to the experiment results of incidental modification in Section 4.2, we can obviously observe that different modification would make different kind of watermark detected errors in each scale of wavelet transform. For example, when an image suffers from the modification of JPEG compression, it is usually having serious watermark detected errors in low scale of wavelet transform, but scarce errors in high scale. In addition, if the watermark detected errors are usually made along the object contour, it might be inferred that the image are modified as blurring or sharpening distortion. In this context, if we could use some technologies, such as vector analysis or neural network, to correlate the relationship between each scale of wavelet transform, the better representation of the tampering position and the precise contour of the tampering object would be well displayed.

Semi-fragile watermarking techniques provide an effective means of protecting the content of digital media. Furthermore, the use on digital images allows for the detection and localization of unauthorized tampering, while permitting the efficient storage of visual information. These combined characteristics are of primary importance in applications, such as courtroom evidence or medical imaging for which the information contained in images is of utmost importance, while the number of images available requires proficient storage techniques. In this context, the use of semi-fragile watermarking techniques clearly augments the value of digital images. For all these reasons, the development of certification systems for digital data will become an increasingly important issue in the future. Thus, the introduction of our wavelet transform based digital watermarking technique for image authentication and tampering detection and its system prototype are the small steps in the advancement of overall digital security.

References

Adams, W. C. JR., & Giesler, C. E. (1978). Quantizing characteristics for signals having Laplacian amplitude probability density function. IEEE Transactions on Communications, com-26(8), 1295-1297.

Arnold, M, Schmucker, M, & Wolthusen, S. D. (2003). Techniques and applications of digital watermarking and content protection. MA, USA: Artech House.

Bartollini, F., Tefas, A., Barni, M., & Pias, I. (2001). Image authentication techniques for surveillance applications. Proceedings of the IEEE 89(10), 1403-1418.

Celik, M. U., Sharma, G., Saber, E., & Tekalp, A. M. (2002). Hierarchical watermarking for secure image authentication with localization. IEEE Transactions on Image Processing 11(6), 585-595.

Cope, B., & Freeman, R. (2001). Digital rights management and content development. Australia: RMIT University.

Cox, I. J., Kilian, J., Leighton, T., & Shamoon, T. (1997). Secure spread spectrum watermarking for images, audio, and video. Transactions on Image Processing 6(12), 1673-1687.

Cox, I. J., Miller, M. L., & Bloom, J. A. (2002). Digital watermarking. CA, USA: Academic Press.

Cox, I. J., Miller, M. L., & Bloom, J. A. (1999). Watermarking as communications with side information. Proceeding of the IEEE 87(7), 1127-1141.

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. Communications on Pure Applied Mathematics XLI(41), 909-996.

Gladney, H. M., & Bennett, J. L. (2003). What do we mean by authentic? What's the real McCoy? D-Lib Magazine 9(7/8). Retrieved December 4, 2003 from the World Wide Web: http://www.dlib.org/dlib/july03/gladney/07gladney.html

Kundur, D., & Hatzinakos, D. (1999). Digital watermarking for telltale tamper proofing and authentication. Proceedings of the IEEE 87(7), 1167-1180.

Lin, C. -Y., & Chang, S.- F. (2003, Fourth Quarter). Robust digital signature for multimedia authentication: A summary. IEEE Circuits and Systems Magazine, 23-26.

Lin, C. –Y., & Chang, S. –F. (2001). A robust image authentication method distinguishing JPEG compression from malicious manipulation. IEEE Transactions on Circuits and Systems of Video Technology 11(2), 153-168.

Lu, C. –S., & Liao, H. –Y. M. (2003). Structural digital signature for image authentication: An incidental distortion resistant scheme. IEEE Transactions on Multimedia 5(2), 161-173.

Lu, C. –S., & Liao, H. –Y. M. (2001). Multipurpose watermarking for image authentication and protection. IEEE Transactions on image processing 10(10), 1579-1592.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. IEEE Transactions on pattern Analysis and Machine Intelligence 11(7), 674-693.

Pan, J. –S., Huang, H. –C., & Jain, L. C. (2004). Intelligent watermarking techniques. MA, USA: World Scientific.

Queluz, M. P., & Lamy, P. (2000). Spatial watermark for image verification. In SPIE Conference on Security and Watermarking of Multimedia Contents II, vol. 3971, 120-130.

Rosenblatt, B., Trippe, B., & Mooney S. (2002). Digital rights management: Business and technology. NY, USA: M&T Books.

Stallings, W. (2000). Network security essentials: Applications and standards. NJ, USA: Prentice Hall.

Strang, G. & Nguyen, T. (1996). Wavelets and filter banks. Wellesley -Cambridge Press.

Tefas, M. A., & Pitas, I. (2000). “Image authentication based on chaotic mixing.” In IEEE International Symposium on Circuits and Systems (ISCAS’2000), vol. I, 216-219.

Trkel, A. Z., Osborne, C. F., & van Schyndel R. G. (1996). Image watermarking: A spread spectrum technique. In IEEE 4th International Symposium on Spread Spectrum Techniques and Applications, vol. II, 785-789.

Tsai, M.J., & Hung, H.Y. (2004). DCT and DWT-based Image Watermarking by Using Subsampling. In Proceeding of the 6th International Workshop on Multimedia Network Systems and Applications (MNSA'2004) in conjunction with The 24th International Conference on Distributed Computing Systems (ICDCS-2004), March 23-26, 2004, Tokyo, Japan, 184-189.

Vetterli, M. and Kovacevic, J. (1995). Wavelets and Subband Coding. USA, NJ: Prentice Hall.

Gonzalez, R. C., & Woods, R. E. (2002). Digital image processing. USA, NJ: Prentice Hall.

Wang, Y., Doherty, J. F., & van Dyck, R. E. (2002). A wavelet-based watermarking algorithm for ownership verification of digital images. IEEE Transactions on image processing 11(2), 77-88.

Watson, A. B., Yang, G. Y., Solomon, J. A., & Willasenor, J. (1997). Visibility of wavelet quantization noise. IEEE Transactions on Image Processing 6(8), 1164-1175.

Wu, C. W., (2002). On the design of content-based multimedia authentication systems. IEEE Transactions on Multimedia 4(3), 385-393.

Wu, G., Yang, E. –H., & Sun W. (2003). Optimization strategies for quantization watermarking with application to image authentication. ICASSP, v-672-675.

Wu, M., & Liu, B. (1998). Watermarking for image authentication. In IEEE International Conference on Image Processing (ICIP'1998), vol. II, 437-441.

Yeung, M. M., & Mintzer, F. (1997). An invisible watermarking techniques for image verification. In IEEE International Conference on Image Processing (ICIP'1997), vol. III, 548-551, Santa Barbara, California, October 1997.

Yu, G. –J., Lu, C. –S., & Liao, H. –Y. M. (2001). Mean-quantization-based fragile watermarking for image authentication. Optical Engineering 40(7), 1396-1408.

Yu, G. –J., Lu, C. –S., Liao, H. –Y. M., & Sheu J. –P. (2000). Mean quantization blind watermarking for image authentication. In IEEE International Conference on Image Processing (ICIP'2000), vol. III, 706-709, Vancouver, BC, Canada, 2000.

Zhao, Y., Campisi, P., & Kundur, D. (2004). Dual domain watermarking for authentication and compression of cultural heritage images. IEEE Transactions on Image Processing 13(3), 430-448.