REKABENTUK SATU TEKNIK TERA AIR SEPARA RAPUH YANG BERASASKAN 'WAVELET' UNTUK KEGUNAAN PENGESAHAN IMEJ

Oleh

Koh Ling Hock

Disertasi ini dikemukakan kepada
UNIVERSITI SAINS MALAYSIA

Sebagai memenuhi sebahagian daripada syarat keperluan
untuk ijazah dengan kepujian

SARJANA MUDA KEJURUTERAAN (KEJURUTERAAN ELEKTRONIK)

Pusat Pengajian Kejuruteraan
Elektrik dan Elektronik
Universiti Sains Malaysia

Mac 2005

DESIGN OF A WAVELET-BASED SEMI-FRAGILE WATERMARKING TECHNIQUE FOR IMAGE AUTHENTICATION

By

Koh Ling Hock

Dissertation submitted to
UNIVERSITI SAINS MALAYSIA

In partial fulfillment of the requirements
for degree with honors

BACHELOR OF SCIENCE (ELECTRONIC ENGINEERING)

School of Electrical and
Electronic Engineering
Universiti Sains Malaysia

Mac 2005

ABSTRAK

Kemasyhuran jaringan di seluruh pelosok dunia pada awal tahun 1990 telah menunjukkan bahawa ia mempunyai potensi komersial untuk menawarkan sumber multi-media melalui rangkaian digital. Disebabkan peminat-peminat komersial mencuba-cuba untuk menggunakan rangkaian digital bagi menawarkan media digital rekaan mereka demi keuntungan, mereka mempunyai minat yang mendalam untuk melindungi hak-hak milik mereka. Dengan ini, teknik tera air digital telah diusulkan sebagai satu jalan penyelesaian untuk mencapai objektif ini. Teknik tera air digital telah banyak digunakan dalam aplikasi seperti perlindungan harta intelek, hakcipta, pengesahan isi kandungan dan penyembunyian informasi. Sejak tahun 1993, kajian tentang penggunaan teknik tera air digital dalam pengesahan kandungan imej semakin mendapat tempat. Pengesahan kandungan imej dilakukan dengan cubaan membenamkan maklumat ke dalam imej tersebut dengan tujuan untuk mengesan dan mengenal pasti kawasan yang telah diubahsuai. Dalam disertasi ini, satu teknik tera air separa rapuh yang berdasarkan *wavelet* untuk kegunaan pengesahan kandungan imej telah dicadangkan. Kaedah ini adalah sensitif terhadap sebarang pengubahsuaian yang dilakukan ke atas kandungan imej yang telah terbenam. Dalam teknik ini, imej akan dibahagikan kepada blok-blok yang berbentuk segiempat sama tanpa pertindihan antara blok-blok tersebut. Kemudian, purata keamatan bagi piksel-piksel dalam setiap blok dihitung dan seterusnya dibenamkan balik ke dalam imej tersebut sebagai satu tera air separa rapuh untuk melindungi kandungan imej itu. Keputusan dari eksperimen jelas menunjukkan bahawa teknik yang telah dicadangkan adalah terjamin dan berkemampuan untuk mengesan kawasan yang telah diubahsuai.

ABSTRACT

The enormous popularity of the World Wide Web in the early 1990's demonstrated the commercial potential of offering multimedia resources through the digital networks. Since commercial interests seek to use the digital networks to offer digital media for profit, they have a strong interest in protecting their ownership rights. Digital watermarking has been proposed as one way to accomplish this. Digital watermarking serves many purposes, such as intellectual property rights protection, content verification and information hiding. Since 1993, the research of digital watermarking for image content authentication has received more and more attentions. Image content authentication tries to embed information in order to identify and localize any malicious attacks on the image. In this dissertation, a wavelet-based semi-fragile watermarking technique for image content authentication is proposed, which is robust against content-preserving operations, while very fragile to content-changing processing. The image is first divided into equal-sized, non-overlapping blocks. Then, the average of pixel intensities is calculated for each of such blocks and embedded as a semi-fragile watermark back into the image for content protection. The experimental results clearly demonstrate that the proposed algorithm is able to detect malicious attacks on the image, secure and robust to common image processing operations.

Keywords: digital watermarking, image authentication, semi-fragile watermark

ACKNOWLEDGEMENTS

I would like to acknowledge the precious contribution of several people who have, knowingly or not, made the present dissertation possible. My utmost debt of gratitude is to my supervisor, Dr. Khoo Bee Ee, for her continuous and kind support throughout my bachelor's degree. She was always generous with her time and provided me with invaluable guidance in technical and professional matters. She has been particularly important in different writing stages and preparation of presentation, always providing me with advice, imprinted with wisdom earned from years in the field of image processing. I am also thankful to Professor Tang Yuan Liang and Mr. Chen Chun Hung from Chaoyang University of Technology, Taiwan, for their help on practical implications of our field of study and to Mr. Nozomi Ishihara from The University of Electro-Communications, Tokyo, Japan, who provided me with insightful comments on my research. In addition, I want to express my gratitude to Mr. Lim Say Yarn who directed the early stage of my work and who gave me the organizational skills needed for its completion. Special thanks to Mr. Chua Tiong Kin, Ms. Lim Ling Ching, Ms. Neng Shen Hooi and Ms. Ng Wan Yee for their time and guidance, which inspire me in working my dissertation. Finally, I would like to thank my family and all my friends in The University Science of Malaysia Engineering Campus who have supported me through these years.

To all, thank you.

Koh Ling Hock

The University Science of Malaysia Engineering Campus
February 2005

TABLE OF CONTENTS

LIST OF FIGURES

Pages

LIST OF TABLES

Pages

CHAPTER

1 INTRODUCTION

1.1 Motivation

As multimedia becomes an important form of information exchange, a large number of digital products is created and transmitted via the Internet. One of the characteristics of these digital products is that they are easily being created, stored, duplicated, transmitted and modified. This results in a serious security problem, for example, the legal rights of a product could easily be violated and the contents of the multimedia could easily be altered. 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 easily being modified, some works are needed in order to protect the content of these images. As the number of digital images increases rapidly, the direct storage of unique reference patterns becomes impractical. Such a phenomenon leads to the development of digital watermarking techniques for image content authentication. Digital watermarking for image authentication means to insert an image-content-related feature as a watermark into the perceptually less significant locations of the original image to protect the image content from unauthorized alteration while minimize artifacts. In this dissertation, a wavelet-based semi-fragile watermarking technique for image content authentication is proposed, which is robust against content-preserving operations, while very fragile to content-changing processing.

1.2 Objectives of the Proposed Research

There are two main objectives in designing an image content authentication system. First, the relationships among wavelet coefficients for both Discrete Wavelet Transform (DWT) and Stationary Wavelet Transform (SWT) of the image are investigated and compared for designing a more robust and semi-fragile watermark embedding method. From the results, we will choose the most suitable wavelet transformation and find out the most suitable subband for embedding the watermarks.

Second, a technique based on semi-fragile watermarks is designed for image content authentication, considering the characteristics among the wavelet coefficients found previously. The embedded watermark is relatively robust against content-preserving operations, while very fragile to content-changing processing.

1.3 Dissertation Organization

This dissertation is organized as follows. Chapter 2 gives a general overview of the related works in the area of image content authentication. Besides the classification of digital watermarking applications and image content authentication techniques will be delineated, the characteristics of watermarks, the requirements of image authentication scheme, general attacks on image authentication systems and the pros/cons of various proposed techniques will also be included in this chapter as well. The result of the investigation on the relationships among wavelet coefficients between two different types of wavelet transformation is presented in Chapter 3 with the purpose of designing a more robust watermark embedding method. Based on the findings in previous chapter, Chapter 4 describes a proposed semi-fragile watermarking technique for image authentication. Then, the design of Graphical User Interface (GUI) for proposed watermarking technique will be presented in Chapter 5. All experimental results for the proposed semi-fragile watermarking technique for image authentication are delineated in Chapter 6. Finally, Chapter 7 gives some concluding remarks and the future works.

CHAPTER

2 LITERATURE REVIEW

A lot of work has been done in order to develop reliable watermarking systems. Numerous digital medias have been considered for the embedding of information to serve a wide range of applications. In this chapter, a classification of digital watermarking applications is first given. Then, the characteristics of watermarks are discussed. Third, the requirements of image authentication scheme are addressed. Fourth, some general attacks on image authentication systems are considered. Fifth, image content authentication techniques will be classified in order to expose their advantages and disadvantages. Finally, wavelet domain will be discussed briefly.

2.1 Classification of Digital Watermarking Applications

Digital watermarking serves many purposes [Nikolaidis et al, 2001], for example, copyright protection, usage control, content authentication, broadcast monitoring and so forth as shown in Figure 2.1. The classification scheme is based on the kinds of information conveyed by the watermark.

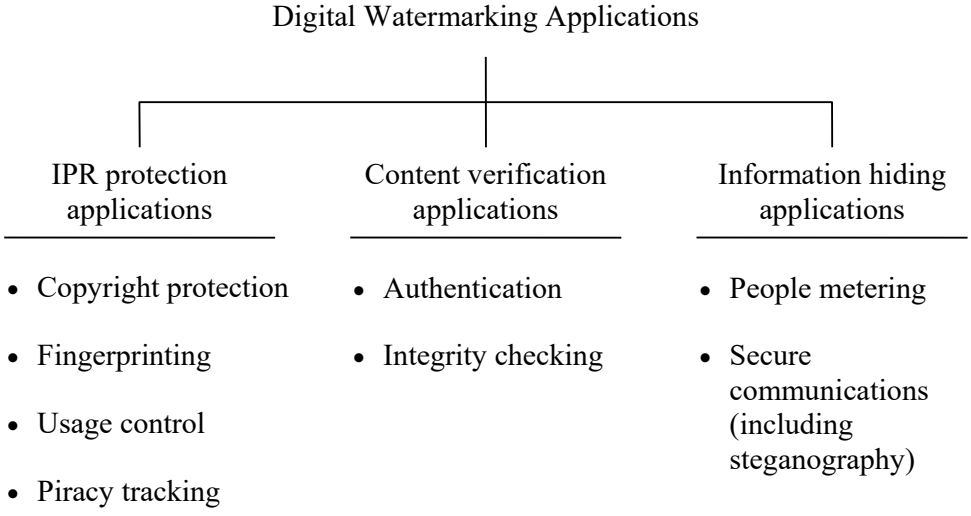

Figure 2.1: A generic classification of digital watermarking applications

2.1.1 Intellectual property rights (IPR) protection applications

Watermarking is used to convey information about content ownership and intellectual property rights [Nikolaidis and Pitas, 1999]. Followings are the typical applications of this class:

i. *Copyright protection*

The copyright owner protects his intellectual property from being manipulated by a user in an illegal way in an effort to remove or destroy the watermark or replace it with his own one.

ii. *Fingerprinting*

The distributor traces illegal copies in order to protect his IPR. The watermark conveys identification information of the user instead of the owner. This information is inserted by the distributor and is different for each copy of the same product, thus characterizing each single transaction.

iii. *Usage control*

The content provider is interested in constraining the level of control that the end user may have on the purchased product. The owner wants to be able to check, for example, how many copies of the content have been made.

iv. *Piracy tracking*

Piracy tracking is done by monitoring stations at the side of the receivers. A watermark is needed to identify the owner and protect his rights during transmission of the corresponding material.

2.1.2 Content verification applications

The watermark indicates whether the multimedia content has undergone any alterations and in certain cases, pinpoints the type and location of alterations. Applications of this category include authentication and integrity checking [Cox et al, 2000].

i. *Authentication*

The copyright owner checks whether the content has been altered, how and to what extent, either by the distributor or by the user.

ii. *Integrity checking*

The user verifies whether the product that he has purchased is authentic or not. Since the embedding is done by the owner, whereas the detection is performed by the buyer, public keys should be employed for security.

2.1.3 Information hiding applications

In this class, watermarks are used as information carriers. The information might be relevant or irrelevant to the product on which they are embedded and may be intended for a specific class of users or a specific use. Applications of this category include people metering and secure communications (including steganography) [Nikolaidis et al, 2001].

i. *People metering*

People metering is done by monitoring stations that decode the watermark which contains information about the identification of the broadcaster and of the broadcast content, as well as the time of broadcast and sometimes the receiver's location.

ii. *Secure communications*

The watermarked media (main channel) conveys side-channel information to the end user. There are three types of side-channel information: public, private and hidden. A public side-channel watermark contains information about the content in which it is embedded, meant to be accessed by any legal buyer or user of the product. A private side-channel watermark contains information that is intended only for specific authorized users. A hidden side-channel watermark is the only important information and the cover media is just the carrier.

2.2 Characteristics of the Watermarks

In the area of image authentication, digital watermarking for image authentication means to insert a watermark into the original image to protect the image content from unauthorized alteration. During authentication, the watermark can be partly recovered from the watermarked image to verify the image content if the correct cryptographically secure key needed for recovery is used. Generally, an excellent digital

watermarking technique must satisfy the following common requirements for watermarking algorithms [Miller et al, 1999]:

i. Imperceptible (fidelity): The embedded watermark should not be noticeable to the human visual system. This means that the perceptual quality of the watermarked image should not be degraded severely compared with the original image. It is the basic requirement of keeping the commercial quality of marked image.

ii. Security: The watermarking algorithm must be public. According to Kerckhoff's assumption [Schneider, 1996], security of the system should reside entirely in the key, not the algorithm.

iii. Blindness (oblivious): The authentication process should not require any information about the original image. In some applications, the original image may not be available, for example, the output image of a digital camera is usually already watermarked.

iv. Robustness: The embedded watermark under the premise that a distorted image quality is acceptable must be robust against attacks to remove it [Tang and Chen, 2004].

2.3 Requirements of Image Authentication Scheme

From our willingness to protect digital data against forgery and tampering, we can extract several requirements that authentication systems must fulfill. Here are the main points to keep in mind in the development and evaluation of watermarking systems. In the context of image protection, an effective authentication scheme should satisfy the following requirements:

i. The original image is not required for watermark detection.

ii. The watermarked image should be visibly indistinguishable from the original.

iii. The authentication process should detect the tampering as well as to localize the tampered areas effectively and efficiently.

iv. It should be difficult for an attacker to use statistical techniques to analyze if a watermark is existent.

2.4 General Attacks on Image Authentication Systems

One must be mindful of potential attacks by malicious parties during the design and evaluation of a marking system. It may be practically impossible to design a system impervious to all forms of attacks, and new methods to defeat marking systems will be invented in time. The possible types of attacks on image authentication system can be divided into four categories [Nikolaidis et al, 2001]:

i. *Removal attacks*: This category includes attacks that aim at removing the watermark without degrading the perceptual quality of the product. For example, collusion attack which tries to combine different watermarked versions of the same image to generate an average image that is very close to the original, thus reducing the watermark strength or totally removing the watermark.

ii. *Presentation attacks*: These attacks aim at manipulating the content in such a way that the detector cannot find the watermark. The intention is essentially the same as in the previous category, but the techniques employed to achieve it are different. Examples of such attacks are rotation, enlargement, and affine transformations in general.

iii. *Interpretation (protocol) attacks*: In this case, the intention of the attacker is to render the watermarking scheme unreliable. This can be done for example by producing a counterfeit original after subtracting a counterfeit watermark from a watermarked image. The attacker can then claim that the watermarked image contains his own watermark and that he has the original product, thus creating an ownership deadlock.

iv. *Legal attacks*: This category implies all the actions that can be taken in a law court in order to damage the credibility of watermarks as proofs of ownership/authenticity in case of disputes. In other words, it does not include manipulations of the watermarked product, but attempts to take advantage of the lack of legal foundation on watermarking as a proof of ownership (i.e. gaps in the legislation on copyright laws), and challenging the credibility of the owner.

Most localized authentication methods rely on some form of block-wise authentication, in which the image is partitioned into non-overlapping blocks and the watermark is inserted into individual blocks. During the image authentication process, each block is authenticated independently. The precise degree of localization depends on the block dimension, therefore, a smaller size provides better localization. Since authentication watermark is fragile, the attacker is not interested in making the authentication watermark unavailable, instead, he/she tries to make a change to the image such that its visual interpretation would be different (forgery) and the system would not detect the change. For block-wise independent watermarking schemes, Holliman and Memon (2000) described a counterfeiting attack called "collage attack": Given a large number of authenticated images, the attacker can approximate a desired image by first searching for authenticated image blocks whose visual intensities are similar to some blocks of the target image, and then combining them to form the target image as a "collage" [Celik et al, 2002]. Figure 2.2 delineates such a process. The system is vulnerable to such an attack generally because the same key is used for authentication of multiple images. To discourage such an attack, one of the solutions suggested by Holliman and Memon is building some relationships among blocks. As a result, the image possesses the block dependency property and the attacker would not be able to form the target image block by block.

2.5 Image Authentication Techniques

Current image authentication techniques can be classified into two approaches according to the information extracted from the image. The first is digital signatures based and the other is digital watermarks based. The difference between them is that the former stores the authentication information in some place other than the image itself, while the latter embeds it into the image directly.

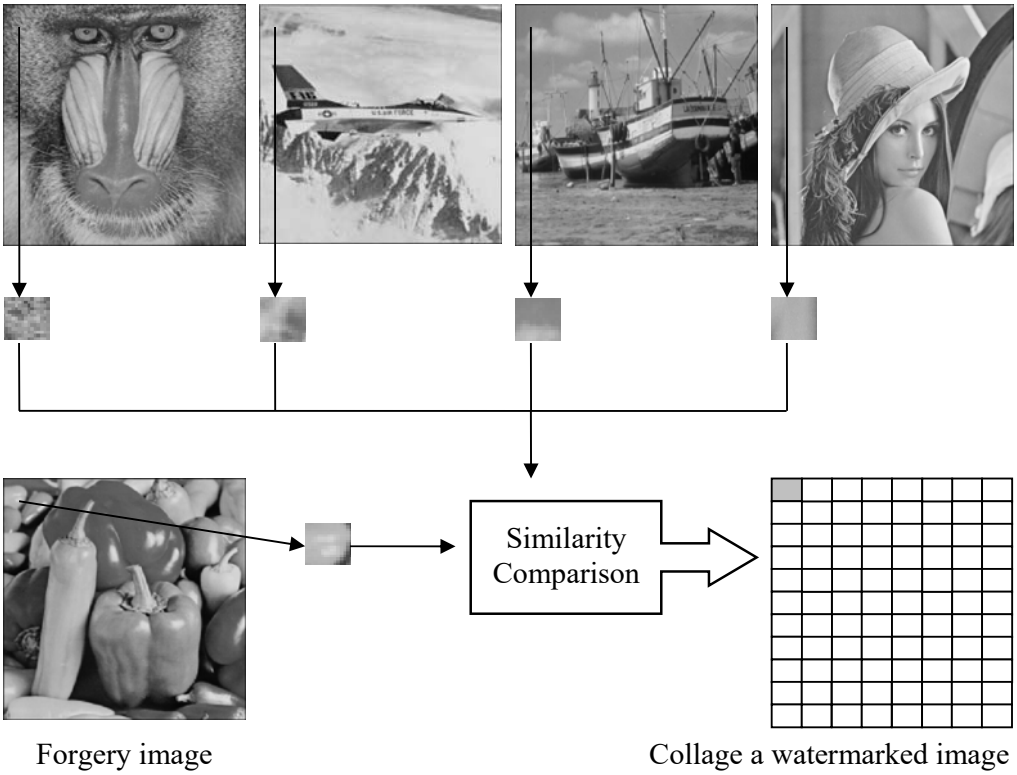

Figure 2.2: Collage attack process.

2.5.1 Authentication with Digital Signatures

Digital signature based schemes store the condensed characteristics of an image in another file for the purpose of authentication. The scenario is as follows. A sender first extracts the features from the original image and then they are encrypted by using a private key to decrypt the information, which is further used to authenticate the received image. Lu & Liao (2003) proposed a signature based technique. They utilized the structure of wavelet transform, which is regarded as a digital signature. The relationship between the parent coefficient and its four child coefficients are computed, decrypted and stored in database for authentication later. Lin and Chang (1998) proposed a signature based scheme, which takes the advantage of the invariance of the relationship between discrete cosine transform (DCT) coefficients. The signature is encrypted with the feature codes and stored in an extra file. When it is necessary to authenticate the image, the signature is decrypted and the image features are extracted, followed by comparison of these two pieces of information to verify the image content.

2.5.2 Authentication with Digital Watermarks

Digital watermarking for image authentication means to insert a watermark into the original image to protect the image content from unauthorized alteration [Lee and Jung, 2001]. The basic idea is to add an image-content-related mark to the host image such that the watermark is unobtrusive and secure but can partly or fully be recovered from the watermarked image later if the correct cryptographically secure key needed for recovery is used. The generic structure of an authentication system consists of two parts: watermark embedding and image authentication. Figure 2.3 shows the watermarking system. The input to the system is the host image and an optional secret key. Depending on the characteristics of the authentication system (exact authentication or inexact authentication), the features to be extracted can be a hash value or invariant image features. To achieve higher fidelity or higher robustness, the host image can be transformed to some frequency domain. The extracted features are then embedded into the original image, which may have been transformed, to produce a watermarked image. The secret key is used to enhance the security of the whole system. The watermark (features) is embedded using such techniques that the watermarked image is perceptually very similar to the original and the embedded watermark is secure.

When a user receives an image, the authentication process is performed to evaluate the authenticity of the image. The system takes the watermarked image, which may have been tampered with, and the secret key as the inputs and performs authenticity verification. The fundamental requirement of the system is the ability to identify any alteration to the image, as well as the tampered locations, if any.

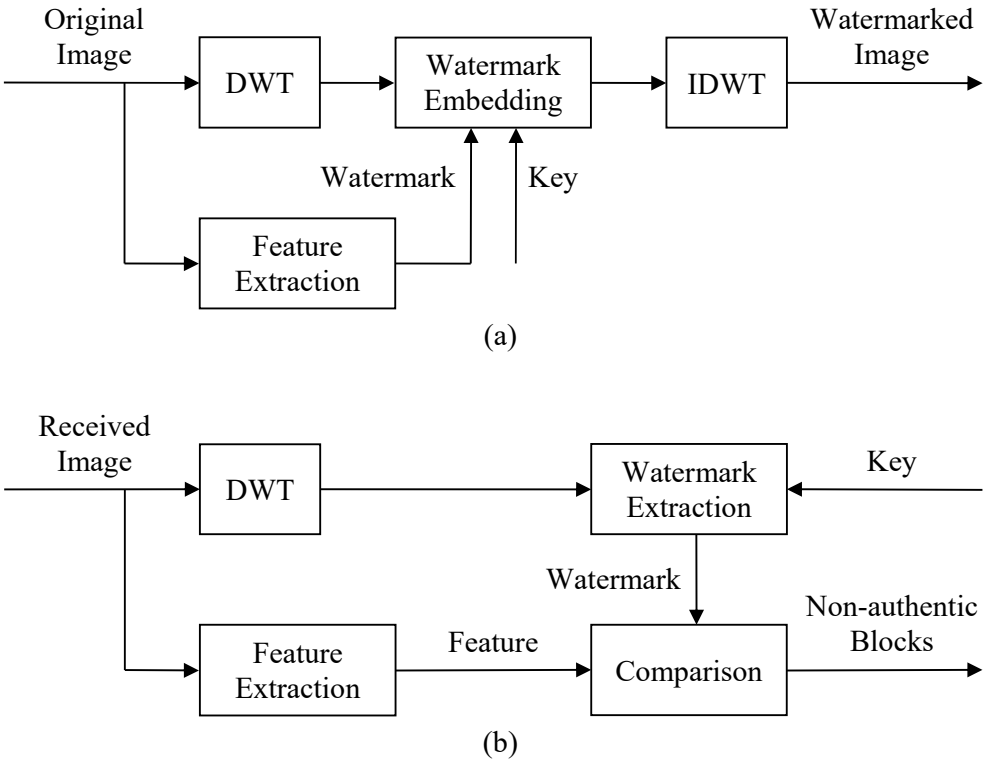

Figure 2.3: A generic structure for image authentication, (a) watermark embedding, (b) authenticity verification.

2.5.3 Private-key/Public-key Watermarking

One of the properties of an ideal watermarking system is that it implements the use of keys to ensure that the approach is not rendered useless the moment that the algorithm becomes known. In the past years, most of the proposed watermarking techniques are based on private-key watermarking schemes [Memon and Wong, 2001]. The private key watermarking techniques use the same secret key for watermark embedding and detection. Therefore, embedding and authentication can be done only by the owner of the key, for example, the owner of the image. The main advantage is that the embedded watermark is difficult to be removed without knowledge of the corresponding secret key. This scenario is illustrated in Figure 2.4.

To supply the public authentication of the image, the watermarking techniques based on public key systems have been proposed and are referred to as public-key watermarking techniques. These systems use a private-key in the embedding process and the corresponding public key is used in the authentication process. The main advantage is that the image owner may embed the watermark using the private key and anyone else can use the public key to verify the authenticity. This scenario is illustrated in Figure 2.5.

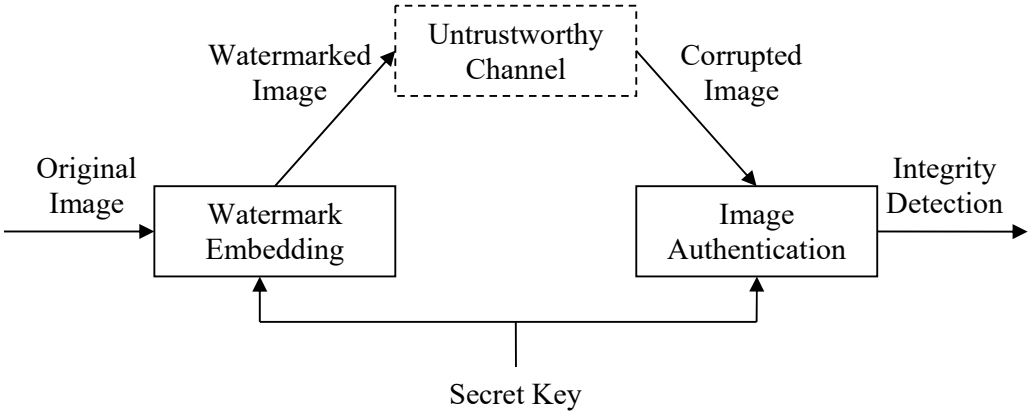

Figure 2.4: The private-key watermarking for image authentication.

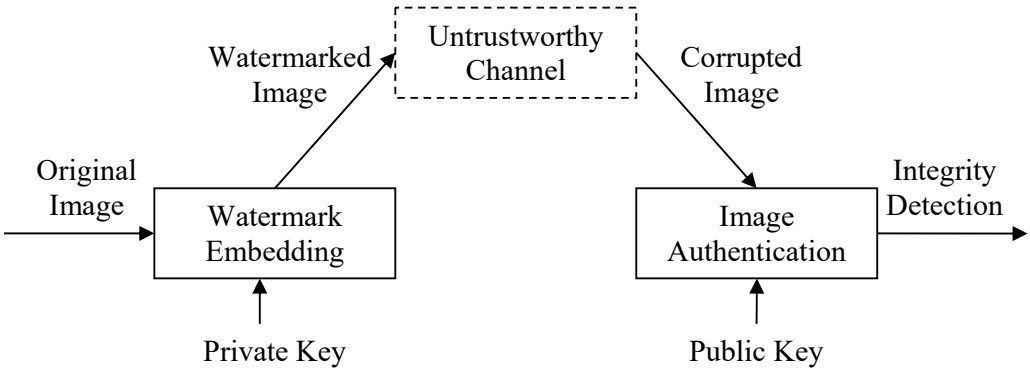

Figure 2.5: The public-key watermarking for image authentication.

2.5.4 Blind/Nonblind Watermarking

For nonblind (or *informed*) authentication systems, the knowledge about the original image is required when extracting the watermark. This kind of authentication needs to access some information about the original image and in general, such a system has better verification performance. However, the owner has to maintain extra space to store the side information derived from the original image. Figure 2.6(a) delineates the concept.

Unfortunately, in most applications the original image may be unavailable; therefore, blind watermarking schemes are preferred [Memon and Wong, 2001]. The blind authentication system verifies the image without specific information related to the original image. Although most former methods are nonblind, it is obvious the trend is towards blind schemes as shown in Figure 2.6(b).

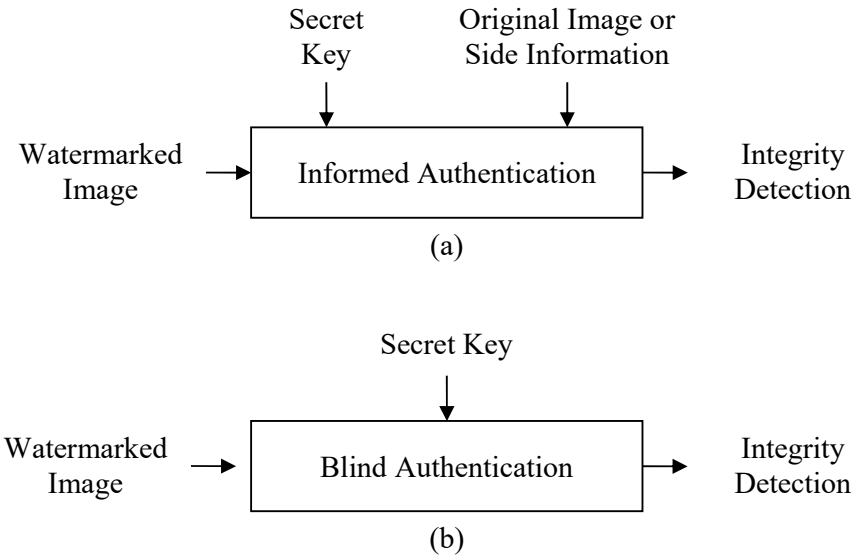

Figure 2.6: (a) Informed authentication, (b) blind authentication.

2.5.5 Spatial/Frequency Domain Watermarking

Depending on the casting domain, watermarking techniques can be divided into spatial and frequency domains. Many early authenticating systems embedded the watermark in the spatial domain of an image. Spatial domain techniques are accomplished by directly modifying pixel intensity of the host image [Lee and Won, 2000; Memon and Wong, 2001]. One of the advantages is their simplicity and efficiency. However, the watermark information will be easily destroyed when various image-processing attacks are applied.

In most multimedia applications, minor data modifications are allowed as long as the content is semantically acceptable. Some authentication schemes [Fridrich et al, 2001; Ishihara and Abe, 2005; Kundur and Hatzinakos, 1999; Wu and Liu, 1998] embed the watermark by altering some frequency bins obtained by transforming the host image into frequency domain, such as DFT (discrete Fourier transformation), DCT (discrete cosine transformation) or DWT (discrete wavelet transformation). The classic and still most popular domain for image processing is that of the DCT. Another possible domain for watermarking embedding is that of the wavelet domain. It is a common belief that the wavelet domain will be the trend in the future. Figure 2.7 delineates the result of a two-scale DWT. Such techniques provide certain degree of robustness against common image processing attacks. Moreover, working in frequency domain may provide better visual fidelity of the images. Nowadays, more and more image authentication schemes are frequency-domain-based.

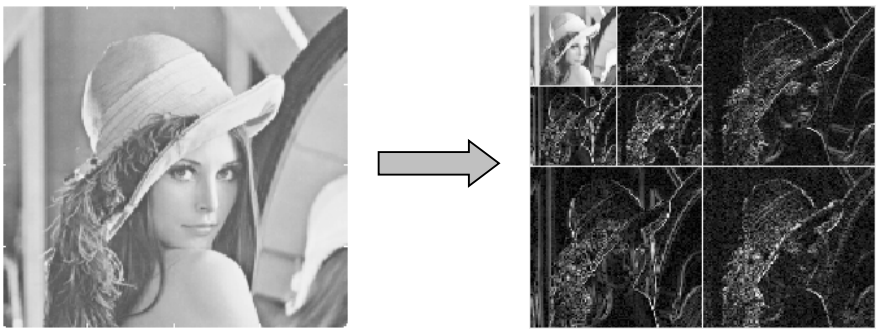

Figure 2.7: The result of two-scale wavelet transformation of the “Lena” image.

2.6 Wavelet Domain

With the standardization process of JPEG2000 and the shift from DCT- to wavelet-based image compression methods, watermarking schemes operating in the wavelet transform domain have become even more interesting. The wavelet transform has a number of advantages [Xia et al, 1998; Lumini and Maio, 2000] over other transforms such as the DCT that can be exploited for both, image compression and watermarking applications. Next, we will study two types of wavelet transform domain for watermarking applications.

2.6.1 Discrete Wavelet Transform

The discrete wavelet transform is the most useful technique for frequency analysis of signals that are localized in time of space. It decomposes signals into basis function that are dilations and translations of a single prototype wavelet function:

$$f(x) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} C_n^m \psi_{m,n}(x) \tag{2.1}$$

Where $\Psi_{m,n}(x)=2^{-m/2}\Psi(2^{-m}x-n)$, are obtained by translates and dilates of the wavelet function $\Psi(x)$. The discrete wavelet transform coefficients C^m_n can be calculated by the inner products ($\Psi_{m,n}(x)$, $f(x)$) which are the estimation of signal components at ($2^{-m}n$, 2^m) in the Time-Frequency place.

Actually, the discrete wavelet transform [Young, 1993] corresponds to multi-resolution approximation expressions. This method permits the analysis of the signal in many frequency bands or at many scales [Mallat, 1989; Daubechies, 1988]. In practice, multi-resolution analysis is carried out using 2 channel filter banks composed of a low-pass (G) and a high-pass (H) filter and each filter bank is then sampled at a half rate (1/2 down sampling) of the previous frequency. By repeating this procedure, it is possible to obtain wavelet transform of any order. The down sampling procedure keeps the scaling parameter constant (n=1/2) throughout successive wavelet transforms so that it benefits for simple computer implementation. In the case of an image, the filtering is

implemented in a separable way by filtering the lines and columns. An example can be illustrated in Figure 2.8.

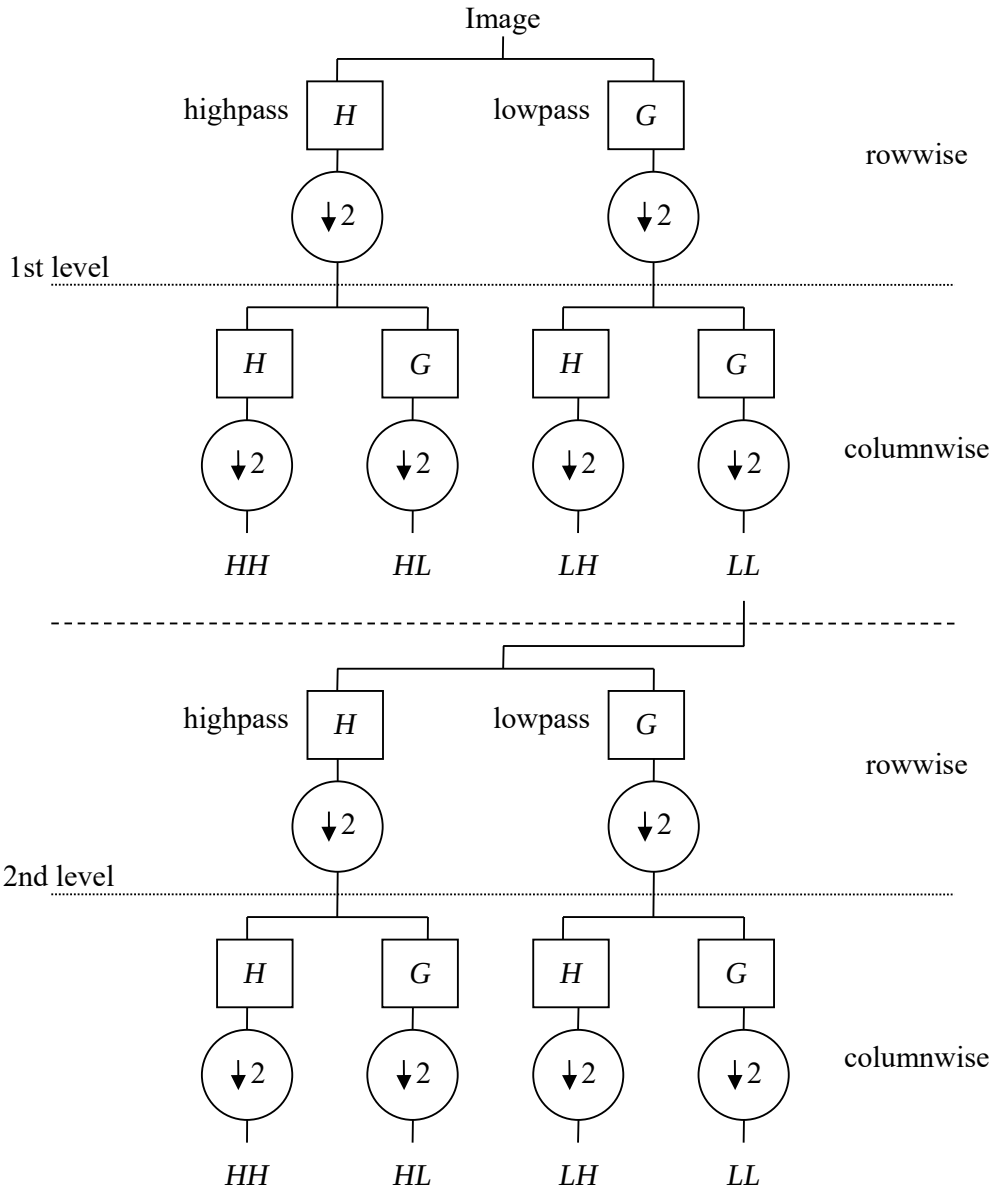

Figure 2.8: Two-level discrete wavelet transform.

According to this procedure, the original image can be transformed into four sub-images, namely:

- LL sub-image: Both horizontal and vertical directions have low-frequencies.
- LH sub-image: The horizontal direction has low-frequencies and the vertical one has high-frequencies.
- HL sub-image: The horizontal direction has high-frequencies and the vertical one has low-frequencies.
- HH sub-image: Both horizontal and vertical directions have high-frequencies.

2.6.2 Stationary Wavelet Transform

The discrete wavelet transform provides the information useful for image analysis and processing. Its fast implementation is usually performed by using multi-resolution analysis. The wavelet coefficients are sampled based on the Nyquist criteria. The representation is accordingly non-redundant and the total number of sample in the representation is equal to the total number of the image pixels. The major inconvenience of this representation is that it does not conserve an essential property in image processing, which is the invariance by translation. Thus pyramidal multi-resolution analysis is not desirable for estimation/detection problems. In order to preserve the invariance by translation, the down sampling operation must be suppressed and the decomposition obtained is then redundant and is called as stationary wavelet transform [Mallat, 1988]. In practice, the structure in cascade of the filter bank does not change, the only operation of down sampling is suppressed.

CHAPTER

3 INVESTIGATION OF THE RELATIONSHIPS AMONG WAVELET COEFFICIENTS FOR BOTH DWT AND SWT

Watermark embedding by quantization of the coefficients in the wavelet domain to embed the watermark has become a popular method. Most of the proposed watermarking techniques quantize individual coefficients to embed the watermark. In this chapter, the value changes on the Discrete Wavelet Transform (DWT) coefficients and Stationary Wavelet Transform (SWT) coefficients before and after common image processing techniques will be investigated and compared for designing a more robust embedding method. From the results, we will choose the most suitable wavelet transformation and find out the most suitable subband for embedding the watermarks.

3.1 Peak Signal-to-Noise Ratio (PSNR)

Signal-to-noise (SNR) measures are estimates of the quality of a reconstructed image compared with an original image [Yuan et al, 2002]. The basic idea is to compute a single number that reflects the quality of the reconstructed image. Reconstructed images with higher metrics are judged better. In fact, traditional SNR measures do not equate with human subjective perception. Several research groups are working on perceptual measures, but for now we will use the signal-to-noise measures because they are easier to compute.

The actual metric we will compute is the peak signal-to-reconstructed image measure which is called PSNR. Assume we are given a source image $f(i, j)$ that contains N by N pixels and a reconstructed image $F(i, j)$ where F is reconstructed by decoding the encoded version of $f(i, j)$. Error metrics are computed on the luminance signal only so the pixel values $f(i, j)$ range between black (0) and white (255). First we compute the mean squared error (MSE) of the reconstructed image as shown in Equation 3.1:

$$MSE = \frac{\sum [f(i,j) - F(i,j)]^2}{N^2} \tag{3.1}$$

The summation is over all pixels. The root mean squared error (RMSE) is the square root of MSE. Then, PSNR in decibels (dB) is computed by using Equation 3.2.

$$PSNR = 20 \log_{10} \frac{255}{\sqrt{MSE}} \tag{3.2}$$

Typical PSNR values range between 20dB and 40dB. They are usually reported to two decimal points (e.g., 25.47dB). The actual value is not meaningful, but the comparison between two values for different reconstructed images gives one measure of quality.

3.2 Relationships among Wavelet Coefficients for DWT and SWT

Wavelet decomposition can be efficiently performed by a pyramidal algorithm [Yuan et al, 2002]. With different combinations of a low-pass filter and a high-pass filter, an image is decomposed into low-low (LL), low-high (LH), high-low (HL) and high-high (HH) bands. To obtain the next coarser scaled wavelet coefficients for DWT, the band LL is further decomposed and subsampled. However, for SWT [Nason and Silverman, 1995], the decomposed image is further decomposed but still has the same length as the original image. Figure 3.1 delineates such a process. This process is repeated several times, which is determined by the requirement of the user.

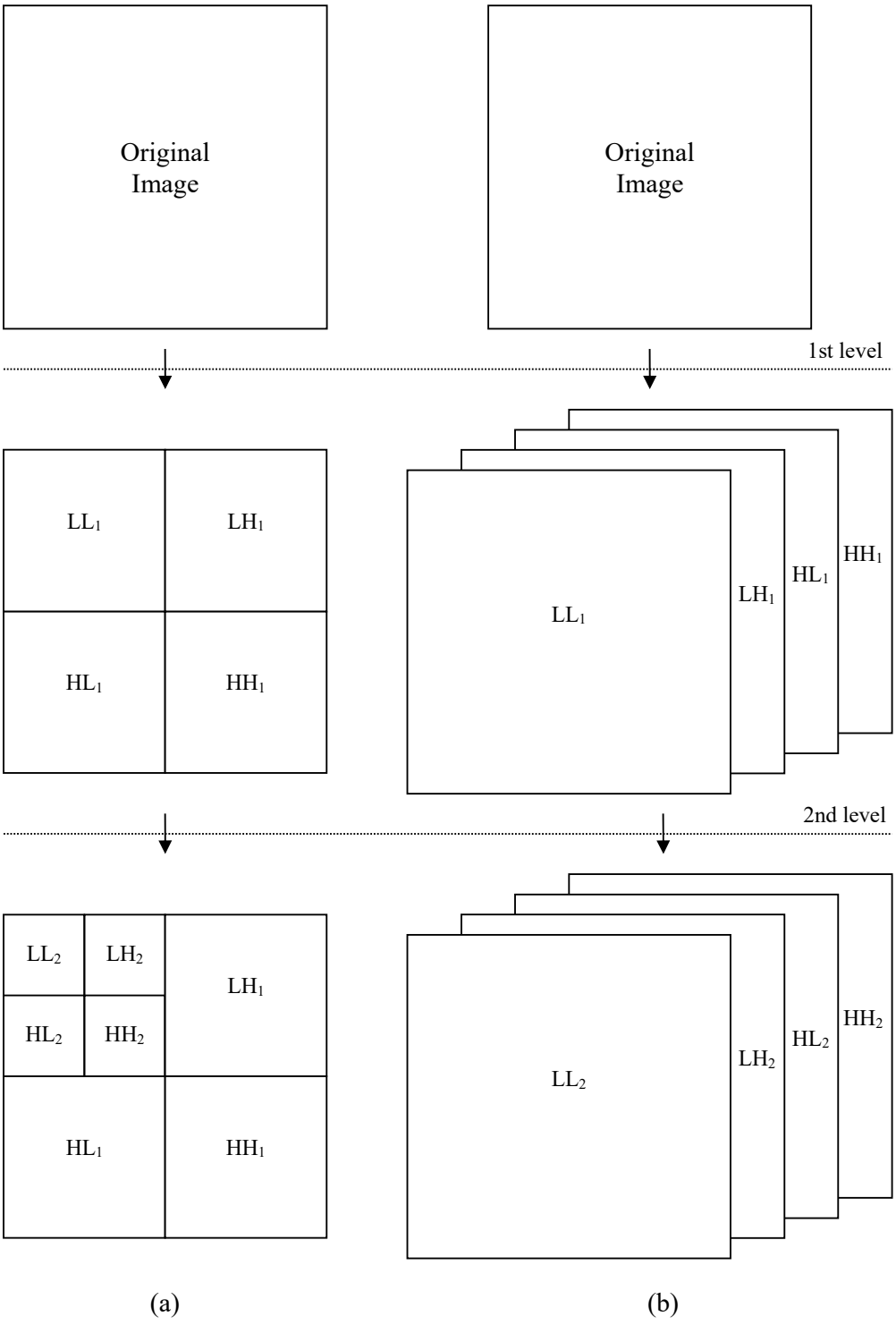

Figure 3.1: Twice wavelet decomposition of (a) DWT and (b) SWT.

Generally, LL band contains coarse wavelet coefficients while HH band contains detailed wavelet coefficients. The watermark embedded in the coarse wavelet coefficients is robust to attack but it will cause degeneration of the image quality. On the contrary, the change in the detailed wavelet coefficients is not detectable to human eyes but it is vulnerable to attacks. Normally, the watermark is embedded in the middle frequency bands for a tradeoff between these two methods. The LH and HL bands are called as middle frequency bands because they include detailed information in one dimension and coarse information in the other dimension. A middle frequency band pair [Tang and Chen, 2004] is defined as a pair of coefficients that are at the same location in LH and HL bands shown in Figure 3.2.

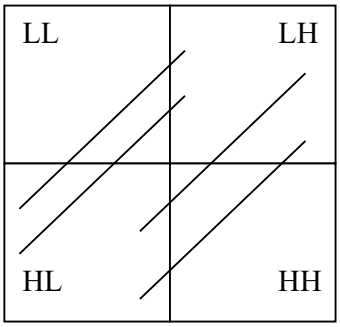

Figure 3.2: Middle frequency band pairs.

3.3 Correlation

The idea of correlation is as one figure changes, we can expect the other to change in a fairly regular way [Edwards, 1976]. Correlation coefficient is also known as Pearson's Correlation Coefficient, represented by the letter *r*, and it is a single number which ranges from -1 (strong negative correlation) to +1 (strong positive correlation). Correlation coefficients, which are close to -1 or +1, indicate a strong correlation. Values close to 0 indicate a weak correlation, with 0 itself indicating no correlation at all. The correlation is defined as:

$$Correlation(x, y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}} \tag{3.3}$$

Yuan et al (2002) indicated that the variation of the sum of the absolute values of the coefficients in LH and HL bands appears to be positively correlated before and after common image processing attacks. Based on Yuan et al findings, Tang and Chen (2004) have come up an important simple conclusion: quantizing the distance of two corresponding coefficients in subbands LH and HL gives a more robust result than quantizing the individual coefficients. In order to go steps further to prove such an argument, we continue the research by conducting experiments on a considerable number of images, as well as applying more image processing techniques, for example, histogram equalization, blurring, sharpening and so-on. The ratio of a coefficient's absolute value and the ratio of a distance's absolute value are used to exhibit the change of the wavelet coefficients. That means, at location (i, j):

$$Ratio(bandname_X(i,j)) = \frac{ABS(bandname_X(i,j))\ \text{after attack}}{ABS(bandname_X(i,j))\ \text{before attack}} \tag{3.4}$$

$$Ratio(LH_X(i,j) - HL_X(i,j)) = \frac{ABS(LH_X(i,j) - HL_X(i,j))\ \text{after attack}}{ABS(LH_X(i,j) - HL_X(i,j))\ \text{before attack}} \tag{3.5}$$

Where *bandname* refers to LH, HL or HH band while *X* refers to level of decomposition.

3.4 Analysis

Several images of size 512×512 are used in our experiment, shown in Figure 3.3.

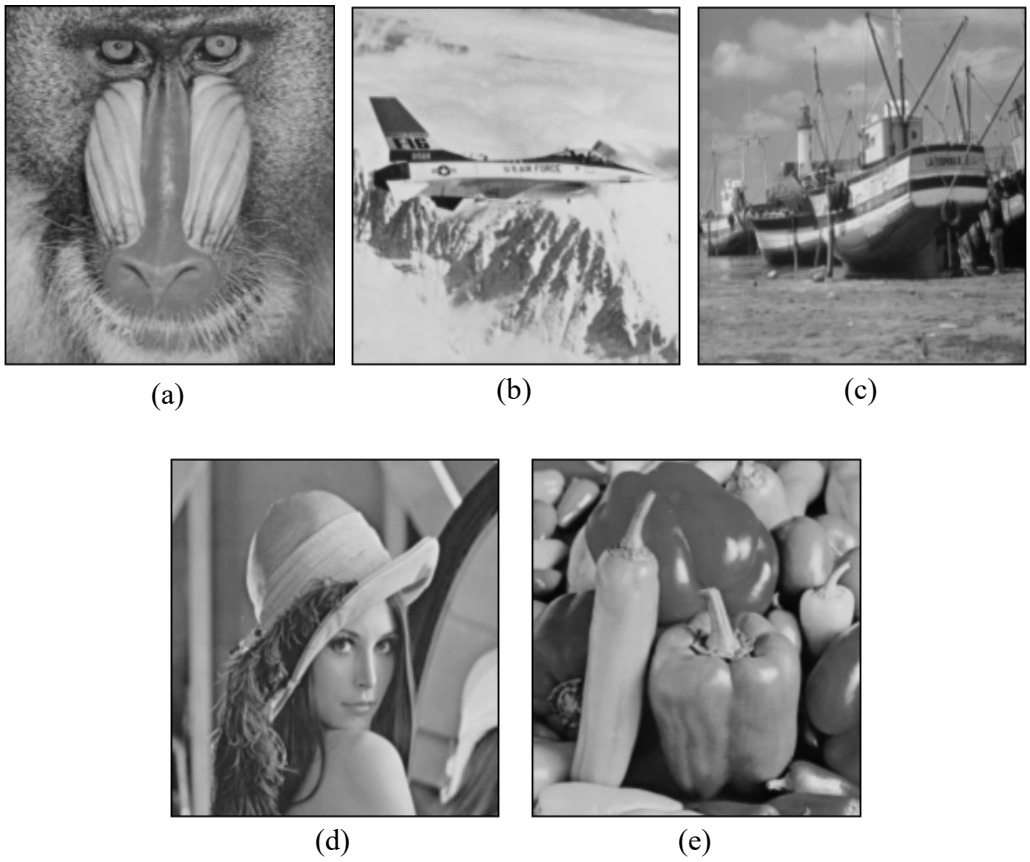

Figure 3.3: Original images: (a) Baboon, (b) F16, (c) Fishingboat, (d) Lena, (e) Peppers.

The image processing attacks include histogram equalization, JPEG compression, JPEG2000 compression, blurring, sharpening, bilinear rotation and a combination of blurring and sharpening.

Experiment 1: PSNR Calculation

In order to exhibit the images' quality after several types of attacks, by applying the Equation 3.2, we have the PSNR values as shown in Table 3.1:

Table 3.1: Comparison of PSNR values for different images and attacks.

Attacks / Image	Histogram Equalization	JPEG Compression	JPEG2000 Compression	Sharpening	Blurring	Sharpening + Blurring	Bilinear Rotation
Baboon	16.32 dB	32.76 dB	27.80 dB	15.89 dB	28.79 dB	27.74 dB	24.33 dB
F16	11.85 dB	36.26 dB	30.46 dB	19.48 dB	31.50 dB	28.99 dB	25.68 dB
Fishingboat	16.56 dB	35.66 dB	29.83 dB	18.91 dB	31.25 dB	29.28 dB	26.56 dB
Lena	18.84 dB	37.06 dB	31.14 dB	20.76 dB	33.31 dB	30.27 dB	27.80 dB
Peppers	20.21 dB	37.61 dB	31.48 dB	21.61 dB	34.00 dB	30.12 dB	26.77 dB

Experiment 2: Correlation Calculation

We decompose the images twice by using DWT and SWT, then LH_2, HL_2, HH_2 and the middle frequency band pairs at level 2 are chosen to check the correlation of the corresponding coefficients between before and after transformation by applying the Equation 3.3. Figure 3.4 shows the comparison between DWT and SWT on correlation values for each type of attacks.