

# Chapter 2

## Watermarking for Image Authentication

Chen Ling and Obaid Ur-Rehman

Digital multimedia has become a ubiquitous part of the modern life with the rapid advancement of network communications systems. Digital multimedia content, in the form of images, audios, and videos, is widely used in electronic commerce, national security, forensics, networked communications, social networking websites, and other fields. Through electronic devices, digital multimedia information can be quickly and conveniently shared over the Internet. With the state-of-the-art powerful signal and image processing techniques and the widely available multimedia editing tools with perfect reconstruction capabilities, digital multimedia content is susceptible to malicious manipulations and alterations. It is difficult for users to establish the authenticity of the multimedia content. As a consequence, the copyright owners may suffer huge economic losses and reduced value of the actual multimedia resources. It might also cause a social disorder if digital multimedia content, such as confidential government documents, judicial evidences, or other significant information, were maliciously tampered. Digital images are one of the most widely used types of multimedia and a basis for video. Therefore, the content protection of multimedia, such as images and videos, is becoming an important problem in the study of information security.

Until now, there are two widely used image authentication techniques, i.e., digital signatures and digital watermarking. Digital signatures based authentication is quite mature and already in wide use. However, its shortcoming is that the signature needs extra bandwidth or establishing a separate secure channel for transmission. As a consequence of using secure hash functions, digital signatures are also susceptible to failed authentication due to the avalanche effect. This might happen due to a change of one or more bits of multimedia data, e.g., due to noise, quantization, or

---

C. Ling (✉)

School of Film and TV Arts & Technology, Shanghai University, 200072 Shanghai, China  
e-mail: lcex@shu.edu.cn

O. Ur-Rehman

Chair for Data Communications Systems, University of Siegen, 57076 Siegen, Germany  
e-mail: obaid.ur-rehman@uni-siegen.de

compression. Digital watermarking can perform content authentication without the aforementioned shortcomings and they are difficult to be removed or tampered.

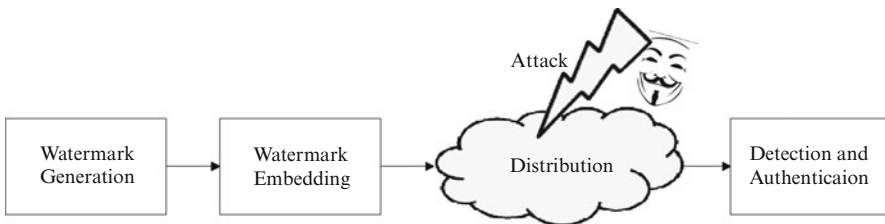
## 2.1 The Basis of Watermarking

In order to cope with the problem of establishing information authenticity, information hiding or watermarking technology has been proposed. Information hiding aims to encode hidden messages in such a way that no one, apart from the sender and the intended recipient, suspects the sole existence of the message. It deals with intelligent ways to embed secret information into unsuspected and rather uninteresting data through clever embedding algorithms. The authenticity of data can be ensured at the receiver based on the embedded secret information. Information hiding makes use of the visual redundancy of the human visual system. It hides digital information into the visual redundancy of other general data; however, the lone existence of this information is invisible to the unintended recipients.

Information hiding can be broadly divided into steganography and digital watermarking. Both steganography and digital watermarking embed data stealthily in noisy signals. Whereas steganography aims for imperceptibility to human senses, digital watermarking considers robustness as its top priority.

Digital watermarking is a covertly embedding technique of digital data with secret information that can be extracted by the recipient. The term “digital watermark” was first coined by Tirkel et al. [1]. The image in which the data to be hidden is embedded, is called the cover image or host. The watermarking process has to be resilient against tampering attacks, keeping the content of a watermark readable in order to be recognizable when extracted by the recipient. Features like robustness and fidelity are the essential features of a watermarking system; however the size of the embedded information has to be considered as well since data becomes less robust as its size increases. Therefore, a trade-off between these features must be considered while developing a watermarking scheme.

Generally, a complete digital watermarking system has three stages: watermark generation, watermark embedding, and watermark extraction for detection and authentication. The general framework is shown in Fig. 2.1.



**Fig. 2.1** General framework of watermarking

In watermark generation stage, the watermark is created such that its content is unique and complex, making it difficult to be extracted or damaged by the attackers. The image feature  $f$  is initially extracted from original cover image  $I_c$ .

$$f = \text{Feature}(I_c) \quad (2.1)$$

Feature( $\cdot$ ) denotes the feature extraction function, where the feature  $f$  uniquely identifies the content of an image. The features can be very simple, such as the grayscale or binary image, edge data of the objects in an image, Wavelet or Fourier coefficients, etc. or they can be much more complicated than these. The image features are then used to generate the watermark,  $wm$ , using a secret key  $K$ .

$$wm = \text{Generation}(f, K) \quad (2.2)$$

where Generation( $\cdot$ ) denotes the watermark generation function.

In cryptography, a secret key is a piece of information (a parameter to the cryptographic algorithm) that determines the functional output of a cryptographic algorithm or a cipher. Without a key, the algorithm would produce no useful results. By changing the key, a significantly different result will be produced. In encryption, a key specifies the particular transformation of plaintext into ciphertext, or vice versa during decryption. The image features generally do not constitute a watermark by itself. Scrambling, random interleaving, and error correcting codes (ECC) are often used for preprocessing to make the secret features more random and useful for watermark generation. The insertion of redundancy (through ECC) in watermarks or in the cover image can improve the extraction process or the reconstruction of watermark after intentional or unintentional distortions. However, ECC can cause collateral effects by increasing the amount of embedded data, which may, in turn, harm the watermark robustness or decrease its data payload, i.e., the capacity of watermark. The most commonly used ECC are: Hamming codes, Bose–Chaudhuri–Hocquenghen (BCH) codes, Reed Solomon (RS) codes, low density parity check (LDPC) codes, and Turbo codes. The watermark is always a pseudo-random sequence or a Gaussian noise sequence.

Embedding of a watermark  $wm$  into the cover image image  $I_c$  is given by:

$$I_{wm} = E(I_c, wm) \quad (2.3)$$

where  $E(\cdot)$  denotes the watermark embedding function and  $I_{wm}$  is the watermarked image. The embedding algorithm should intelligently embed a watermark in the host without destroying the features of the host and making it difficult for an attacker to locate and extract or destroy the embedded watermark.

The transmission of a watermarked image to the receiver is subject to unpredictable distortions and attacks over the communication channel. Some distortions may be unintentional, like channel noise or image compression. Other distortions are intentional, like object removal, replacement, and insertion attacks, etc. Attackers would like to change the original image or remove the watermark to realize their malicious intentions.

At the receiver, the watermarked image  $I'_{wm}$  is received, which might have been tampered and therefore different than  $I_{wm}$ . A watermark detection and extraction algorithm is usually an inverse process of the watermark embedding algorithm:

$$wm' = D(I'_{wm}) \quad \text{or} \quad wm' = D(I'_{wm}, I_c) \quad (2.4)$$

The watermark  $wm'$  is extracted by the extraction function  $D(\cdot)$  from the received watermarked image. There are two different types of watermark extraction methods. The first type is called blind watermark extraction, which does not need the original cover image for watermark extraction. The second type needs the cover image to do the same. From the received watermarked image  $I'_{wm}$ , the receiver generates a watermark  $wm''$  which might not be the same as  $wm$  or  $wm'$  due to noise or tampering.

$$f' = \text{Feature}(I'_{wm}) \quad (2.5)$$

$$wm'' = \text{Generation}(f', K) \quad (2.6)$$

By comparing  $wm'$  and  $wm''$ , it can be determined whether the watermarked image has been tampered or not.

## 2.2 Classification

There is no uniform criterion for the classification of image watermarking schemes [2]. However, a watermarking system has certain generic requirements which must be met when implemented. According to these requirements, watermarking schemes can be broadly classified into seven categories as shown in Fig. 2.2:

- *Perceptibility*: Watermarking schemes can be classified into visible and invisible watermarking based on the perceptibility criterion.
- *Extraction*: Based on the type of extraction method, watermarking schemes are divided into blind watermarking and non-blind watermarking.
- *Platform*: The watermarking schemes are divided into hardware based and software based according to the system platform used for watermarking.
- *Image compression*: According to the image compression methods, the watermarking schemes can be divided into lossy compression based watermarking and lossless compression based watermarking.
- *Embedding domain*: According to the embedding domain classification, there are spatial domain and transform domain watermarking schemes.
- *Robustness*: According to the robustness criterion, the watermarking schemes can be divided into robust, fragile, and semi-fragile watermarking.
- *Lossless*: According to the quality of the watermarked image, there are irreversible watermarking and reversible watermarking.

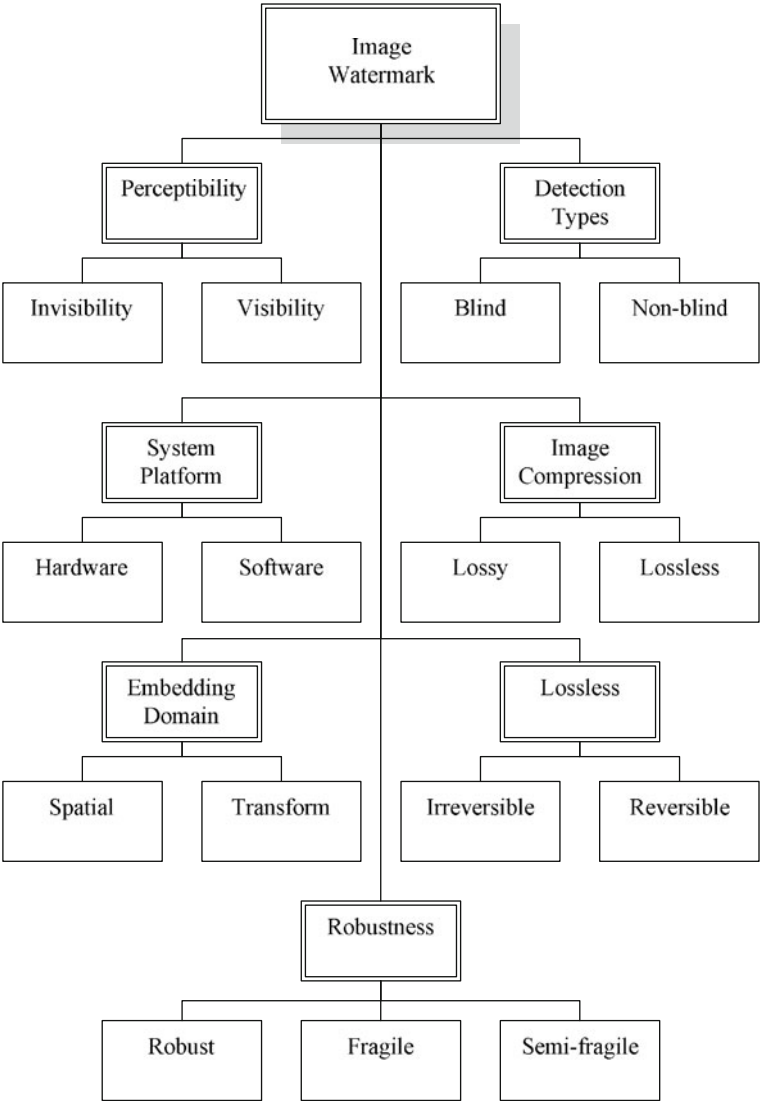


Fig. 2.2 Classification of image watermarking algorithms

2.2.1 *Perceptibility*

Based on the perceptibility criterion, watermarking schemes are generally classified into visible watermarking and invisible watermarking. The visible watermarking technique is a classical approach, as shown in Fig. 2.3a. The word “WATERMARK” is visible in the original Lena image. Everyone can see the embedded information



**Fig. 2.3** Visible watermark (a) and invisible watermark (b)

that the sender wants others to know. This kind of watermarking appeared in Italy during the thirteenth century. It was used to identify the papermaker or the trading company which manufactured the paper. The watermark shown here was created by a wire sewn onto the paper mold. They are still being used in TV station logo and paper currency, etc.

In invisible watermarking, the original cover image and the watermarked image are indistinguishable, as shown in Fig. 2.3b. The shown image has an embedded watermark in the least significant bit (LSB) of the red components. However, this watermark cannot be detected with a human eye. In this chapter, the term “watermarking” refers to the invisible watermarking.

### 2.2.2 *Detection Types*

This classification determines which resources are necessary for an analysis of the extracted watermark from a cover image.

As shown in Eq. 2.4, blind watermarking is a detection type where the original image and watermark data are not available to the receiver. For example, piracy or copy control applications must send different watermarks for each user and the receiver must be able to recognize and interpret these different watermarks.

However, in the non-blind watermarking, the receiver needs either the original image or some information derived from it for the detection process. The same information will also be used in the extraction algorithm. Non-blind watermarking remained a hot research topic in the past few years.

### **2.2.3 System Platform**

According to the system platform, image watermarking can be divided into software-based and hardware-based watermarking. The advantage of software-based watermarking is that it can be easily implemented in the application, not limited by the underlying operating system or hardware. It is also easy to implement any complex algorithm but only limited by the physical computational resources. Generally, software based watermarking is not suitable for real time applications. However, the hardware based watermarking algorithms have relatively powerful processing and computational capabilities. They can meet the requirements of the real-time applications and are especially useful for the limited computing power of small and embedded devices, such as digital cameras.

### **2.2.4 Image Compression**

Images are source coded using compression algorithms to reduce their size for temporal and spatial transmissions. Watermarking can be divided into two types based on the type of image compression schemes: lossless compression based watermarking and lossy compression based watermarking.

Lossless compression is preferred for archival purposes, medical imaging, technical drawings, clip art, comics, etc. Lossy compression methods, especially when used at low bit rates, introduce compression artifacts. Lossy methods are especially suitable for natural images such as photographs in applications where minor loss of fidelity is acceptable to achieve a substantial reduction of the bit rate. There are currently many lossy compression standards in wide use, such as JPEG and JPEG2000.

Watermark can be either directly inserted into the raw image data or integrated during the encoding process after compression. For image compression, the watermarking process can be integrated into image codec or provided as a separate module. The first approach needs to modify the image codec to add watermark embedding and extraction modules. The second one does not change the image codec, but it needs to analyze the encoded stream. It embeds the watermark into the compressed bit stream directly. This is used in the case when the codec cannot be changed and the bit stream can be gained.

### **2.2.5 Embedding Domain**

In the watermark embedding based classification method, two approaches are used for embedding: spatial domain and transform domain. The method used to embed the watermark influences both the robustness against attacks and the detection algorithm, but some methods are very simple and cannot meet the application requirements.

Designing a watermark should consider a trade-off amongst the basic features of robustness, fidelity, and payload.

Spatial domain watermarks insert data in the cover image changing pixels values or image characteristics. The algorithms should carefully weigh the number of changed bits in the pixels against the possibility of the watermark becoming visible. These watermarks have been used for document authentication and tamper detection. The most used spatial domain methods are LSB and spread spectrum (SS).

Transform domain algorithms hide the watermarking data in transform coefficients, thus spreading the data through the frequency spectrum, making it hard to detect. These algorithms are tolerant against many types of signal processing manipulations. The most widely used transforms are discrete cosine transform (DCT) and discrete wavelet transform (DWT).

### 2.2.5.1 LSB

This is the simplest embedding domain watermarking approach, because the LSBs carry less relevant information and the modification of these bits does not cause perceptible changes. Amongst these approaches, some use only the salient points, by encrypting the watermark before embedding it. In this case, the watermark is embedded in the cover image using a key. The key determines which points will be affected and modified by the embedding process.

A watermark extraction algorithm is the inverse of its embedding algorithm. The least significant pixel bits of an image together with the secret key are used in the decoding algorithm to recover the original watermark. Li et al. [3] proposed a multi-block dependency based fragile watermarking scheme to overcome this shortcoming. The images are split into  $8 \times 8$  pixel blocks; a 64-bit watermark is generated for each image block which is then equally partitioned into eight parts. Each part of the watermark is embedded into the LSBs of a different image block which is selected by the corresponding secret key.

Figure 2.4 shows the numerical values of Lena's left eye as a block of  $16 \times 16$  pixel. The cover image and the LSB based watermarked image have only few differences. For example, the value of the first pixel in the cover image is 111, and corresponding pixel has a value of 110 in the watermarked image, which will go unnoticed with a bare human eye.

### 2.2.5.2 SS

Most of the image watermarking methods are based on the ideas known from SS radio communications, namely additive embedding of a pseudo-noise watermark pattern and watermark recovery by correlation [4].

Figure 2.5 illustrates a simple, straightforward example of SS watermarking. The cover image is the red component of the original RGB color image. The binary watermark bits (represented by + 1 for a binary 1 and - 1 for a binary 0) to be embedded



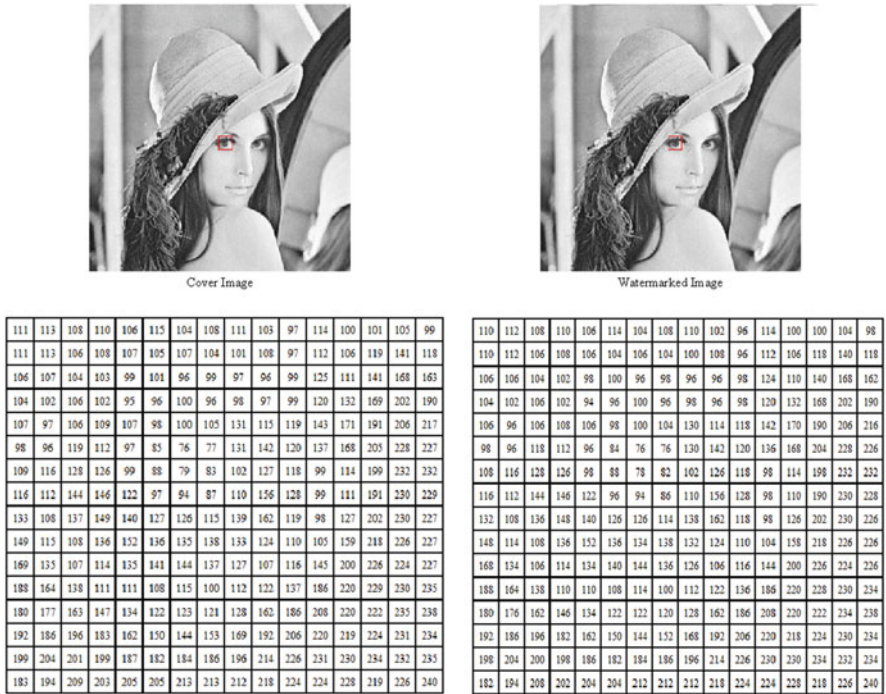
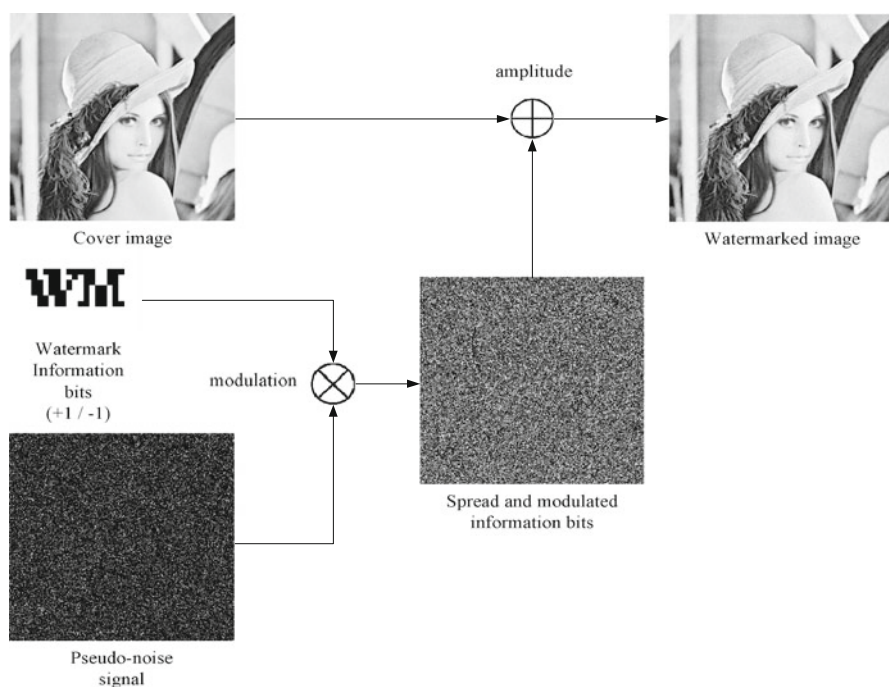


Fig. 2.4 LSB watermarking

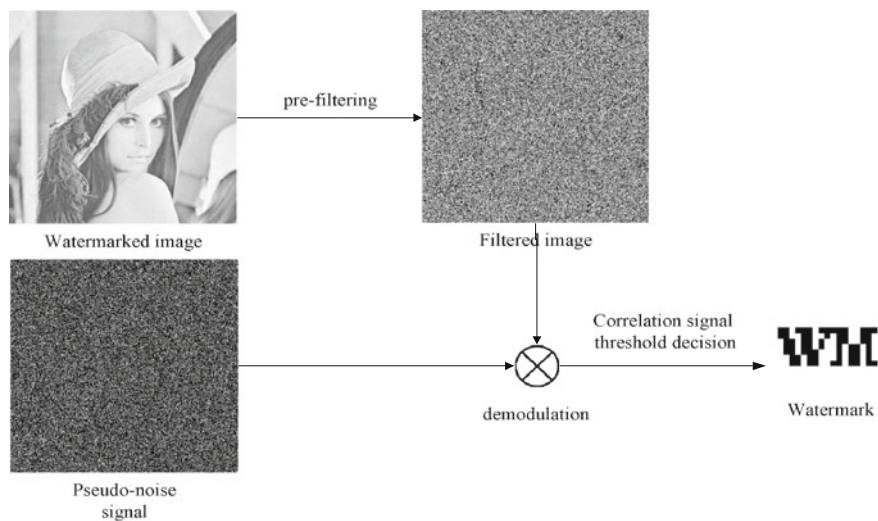
are repeated. The spread information bits are then modulated with a cryptographically secure pseudo noise signal, which is scaled according to the visibility criteria and added to the image pixels.

Figure 2.6 illustrates the corresponding watermark detector based on the principle of a correlation receiver. In order to reduce the cross-talk between an image and a watermark, pre-filtering is done to remove low frequencies from the signal, specifically to remove the local mean. If the original cover image is available to the watermark detector, it is advantageous to replace the filtering by subtraction of the original cover image. The filtered watermarked image is then demodulated using exactly the same pseudo-noise signal previously used in watermark embedding. The samples of a correlation signal and a threshold decision yield the output bits. Thus, the result of the watermark decoding is the same watermark information bits that have been embedded.

The SS watermark can also be embedded in the DCT domain. The 1D watermark vector is rearranged into image structure and by transforming the image into the  $8 \times 8$  DCT domain; the watermark can be added directly to a partially decoded bit stream, making it more robust. However this case is a part of the transform domain algorithm.



**Fig. 2.5** Spread spectrum (SS) watermark embedding



**Fig. 2.6** Spread spectrum (SS) watermark extraction

Robust Image Authentication in the Presence of Noise

Zivic, N. (Ed.)

2015, XV, 187 p. 79 illus., 33 illus. in color., Hardcover

ISBN: 978-3-319-13155-9