Multipurpose Watermarking for Image Authentication and Protection

Chun-Shien Lu and Hong-Yuan Mark Liao

Institute of Information Science, Academia Sinica,
Taipei, Taiwan.
Tel: +886-2-27883799
Fax: +886-2-27824814
E-mail: {lcs, liao}@iis.sinica.edu.tw

Abstract

We propose a novel multipurpose watermarking scheme, in which robust and fragile watermarks are simultaneously embedded, for copyright protection and content authentication. By quantizing a host image's wavelet coefficients as masking threshold units (MTUs), two complementary watermarks are embedded using cocktail watermarking and they can be blindly extracted without access to the host image. For the purpose of image protection, the new scheme guarantees that, no matter what kind of attack is encountered, at least one watermark can survive well. On the other hand, for the purpose of image authentication, our approach can locate the part of the image that has been changed and tolerate some incidental processes that have been executed. Experimental results show that the performance of our multipurpose watermarking scheme is indeed superb in terms of robustness and fragility.

Keywords: Robust watermarking, Copyright protection, Fragile watermarking, Authentication.

1 Introduction

Copyright marking is a relatively new technique used for hiding multimedia information [24]. Its application is broad, including ownership protection [3, 7, 17, 18, 19, 25, 30], content authentication [8, 9, 13, 15, 32, 34, 35], side information conveyance [22], and so on. For ownership protection, robustness is one of the major points of concern [18, 19]. Watermarks embedded for this purpose are called robust watermarks. For content authentication, the embedded watermark should be fragile so that changes or modifications of an image will be reflected in the hidden watermark. This type of watermark is called a fragile watermark. In side information conveyance, a watermark is required to convey more information than a robust watermark does. As a consequence, less redundancy can be employed in this type of watermark [22]. Usually, people call this kind of watermark a captioning watermark. Most of the existing watermarking schemes are designed for either ownership protection or content authentication. If there are multiple purposes, then multiple watermarks must be embedded. Because watermarks of different sorts play different roles, as Mintzer and Braudaway [22] noted, the order for hidden watermarks is important. They suggested that ownership watermarks should be embedded first, captioning watermarks should be embedded next, and fragile watermarks should be embedded last. Wu and Liu [35] also presented a similar concept for combining their own image authentication scheme with an existing ownership protection scheme for "double watermarking." In other words, if multiple watermarks having different missions are to be embedded, then one has to worry about the order of hiding. However, an effective mechanism which can embed multiple watermarks simultaneously is always preferable.

It is well known that an effective watermarking scheme has to satisfy a set of typical requirements, including transparency (perceptual invisibility), robustness, oblivious detection, universality, non-invertibility, and so on. In this paper, our purpose is to develop an oblivious yet highly robust watermarking scheme which can achieve the goal of image authentication and protection simultaneously. As to the robustness requirement, we have proposed the concept of cocktail watermarking [18, 19], which can resist different kinds of attacks. However, the first version of the cocktail watermarking algorithm was not oblivious, which may leave open the possibility that the original sources were stolen when they were transmitted over the network. Some previous works [1, 7, 10, 12, 14, 21, 31] have achieved the oblivious detection requirement but at the expense of robustness especially under stronger attacks or repeated (combined) attacks. For instance, Kutter *et al.* [14] predicted an original DCT coefficient based on the distorted DCT coefficients in a local region. Barni *et al.* [1] skipped the largest N DCT coefficients and tried to decorrelate the low-frequency part of a host image and that of an extracted watermark. To eliminate cross-talk between the video signal and the watermark signal, Hartung *et al.* [10] applied high-pass filtering to the attacked watermarked video. The authors in [8, 12] directly used the information of a distorted image as if it came from the original image. Su and Kuo [31] constructed a pseudo host image from their multi-threshold wavelet codec ($MTWC$) based on the assumption that the largest coefficients are not easily attacked. Recently, Lu and Liao [21] used the generalized Gaussian to model a host image. A set of parameters (secret keys) which can be used to reconstruct the host image is stored. Then, the host image is reconstructed from the stored keys according to the relative positions with respect to the corresponding watermarked image. In sum, the correlation values detected using most of the above mentioned methods [1, 7, 10, 12, 14, 31] are, basically, relatively low under strong attacks.

As to image authentication, the previous techniques [9, 32, 34, 36] focused on detecting whether an image was tampered with or not. However, they did not clearly specify how and where the image was changed. Kundur and Hatzinakos [13] proposed a telltale tamper-proofing method to determine the extent of tampering using a statistics-based tamper assessment function. The quantization process they designed is more/less sensitive to modifications at high/low frequency in the wavelet domain. In their scheme, over-sensitivity may occur at the small-to-middle scale while under-sensitivity may only happen at the middle-to-large scale. Under the circumstances, a user can make application-dependent decisions about whether an image, which is $JPEG$ compressed, still has credibility. However, their approach violates the nature of the human visual system [33]; thus, their system is confused when an image is $JPEG$ compressed first and then maliciously tampered. Another disadvantage associated with Kundur and Hatzinakos's approach [13] is that their tampering detection results are very unstable. As we can see from their quantization process, the value of an extracted watermark is binary, i.e., 0 or 1, depending on which quantization interval the tampered coefficient falls into. Perturbation of a wavelet coefficient to the left or to the right by a certain quantity will make the extracted mark different from the embedded one. Basically, the above mentioned mechanism is suitable for the detection of modifications. However, if the perturbation exceeds one quantization interval, then the extracted watermark value can be either the same as or different from the embedded one. In other words, some severe modifications beyond the capability of the human visual system will not be identified. Hence, the watermark value may be determined accidentally, and by the same token, not every affected pixel is guaranteed to be detected. In fact, what they wanted to get was a localized assessment of the degree of distortion experienced by a group of coefficients in the wavelet domain through a statistical measure.

In order to make the designed image authentication system survive $JPEG$ compression, Lin and Chang [15] proposed preserving the invariance between the DCT coefficients before and after quantization such that a digital signature can be formed. However, it was not mentioned whether their method can survive a compression attack like EZW or other incidental manipulations. They also mentioned that the authenticator is sufficient to accept those images that are compressed using $JPEG$ up to a certain compression ratio or quality factor. To resist other non-malicious processing, their authenticator can adapt to different situations by adjusting thresholds. For feature-based authentication systems, Bhattacharjee and Kutter [2] proposed generating a digital signature by encrypting the feature points of an image, which are relatively less affected by lossy compression. Authentication is then accomplished by comparing the positions of the feature points with those decrypted from the previously encrypted feature points. Again, it is not clear whether this approach can resist $JPEG$ compression with middle-to-low quality factors because the feature points are liable to shift under $JPEG$ compression with middle-to-high ratios. Recently, Dittmann *et al.* [6] presented a content-based digital signature approach for image/video authentication using edge characteristics. Their content feature is similar to [2], but different extraction techniques are used. Unfortunately, the above mentioned digital signature-based methods can only be used for image authentication but not for copyright protection since the original image is not watermarked. More complete reviews of image protection and image authentication can be found in [24, 30] and [13, 16], respectively.

In this paper, we propose a multipurpose watermarking scheme which can simultaneously achieve copyright protection and content authentication. The proposed scheme can fulfill the above mentioned purposes by hiding several multipurpose watermarks at the same time. The validity of our method is based on simultaneous detection of the robust watermark and the fragile watermark. As a consequence, the order of hiding [22] is no longer an important issue. We propose to quantize the selected wavelet coefficients into masking threshold units. Then, the watermarks can be embedded by modulating the quantization result into either a right or a left masking threshold unit using cocktail watermarking [18, 19]. In the meantime, the original quantization result can be recorded as the hidden watermark because it is the closest neighbor to the modulated quantization. Hence, the hidden watermark carries the information of the host image, which can be used to recover the host image with indistinguishable perceptual degradation. Basically, this information is beneficial to the detection of robust watermark and fragile watermark.

The major contribution of this work is twofold. First, a new oblivious watermark detection technique which is associated with our previously developed cocktail watermarking scheme is proposed.

Since the good characteristics of cocktail watermarking are still maintained, the new oblivious scheme guarantees high robustness for copyright protection. Second, the extent of modification can be estimated by comparing the hidden watermark with the extracted one. Under the circumstances, any malicious tampering can be detected while some incidental manipulations can be tolerated. Here, if the amount of modification exceeds a threshold based on the human visual system (HVS), then these modification will be regarded as malicious.

The remainder of this paper is organized as follows. In Sec. 2, the non-oblivious cocktail watermarking scheme is briefly reviewed. Then, multiple watermark hiding for image protection and authentication is described in detail in Sec. 3. Analysis of our method with respect to fragile watermarking is conducted in Sec. 4. Finally, simulation results and conclusions are given in Secs. 5 and 6, respectively.

2 Review of Cocktail Watermarking

In [18, 19], we proposed a novel image protection scheme called "cocktail watermarking." We analyzed and pointed out the inadequacy of the available modulation techniques commonly used in ordinary spread spectrum watermarking methods and visual model-based ones. To resolve this inadequacy, two watermarks which play complementary roles are simultaneously embedded into a host image using a complementary modulation strategy which includes positive modulation (PM) and negative modulation (NM). The first watermark is inserted based on a positive modulation rule employed to increasingly modulate the transformed coefficients of a host image. In addition, the second watermark is embedded based on a negative modulation rule which is used to decreasingly modulate the transformed coefficients of a host image. Based on analysis on the behaviors of attacks, we have confirmed that the new watermarking scheme guarantees that, no matter what kind of attack is encountered, at least one watermark can survive well. We also conduct a statistical analysis to derive the lower bound of the worst likelihood that the better watermark (out of the two) can be extracted. With this "high" lower bound, it is ensured that a "better" extracted watermark will always be obtained for noise-like watermark hiding [18] as well as bipolar watermark hiding [19] under the constraint that the original image is required in the detection process.

In the cocktail watermarking scheme [18, 19], there are three major ways to achieve robustness. They are: (1) bipolar watermarking (the designated watermark); (2) complementary modulation (the hiding rule); and (3) use of a wavelet-based human visual system [33] to control the hiding strength.

Our theoretical analysis and experimental results have shown that cocktail watermarking can really achieve the requirement of high robustness. In cocktail watermarking [18, 19, 20], the complementary modulation rules used in the embedding process is summarized as follows.

Positive modulation:

$$T^m(x_p, y_p) = \begin{cases} T(x_p, y_p) + J(x_p, y_p) \cdot n_{bottom} \cdot w, & T(x_p, y_p) \geq 0; \\ T(x_p, y_p) + J(x_p, y_p) \cdot n_{top} \cdot w, & T(x_p, y_p) < 0. \end{cases} \tag{1}$$

Negative modulation:

$$T^m(x_n, y_n) = \begin{cases} T(x_n, y_n) + J(x_n, y_n) \cdot n_{top} \cdot w, & T(x_n, y_n) \geq 0; \\ T(x_n, y_n) + J(x_n, y_n) \cdot n_{bottom} \cdot w. & T(x_n, y_n) < 0. \end{cases} \tag{2}$$

$T(.,.)$ here can be a coefficient of any transformation [18, 19, 20]. $J(.,.)$ represents the JND values obtained from a visual model, and n_{top}/n_{bottom} represents the value retrieved from the top/bottom of the sorted watermark sequence n. w is an image-dependent weight used to control the maximum possible modification that will lead to the least image quality degradation. A watermark embedded using negative modulation is called a "negatively modulated watermark," and a watermark embedded using positive modulation is called a "positively modulated watermark."

3 The Proposed Multipurpose Watermarking Algorithm

This section will elaborate on the proposed approach in detail. Our scheme has been developed for gray-scale images and color images but is also feasible for other types of media. The wavelet-transformed domain is adopted due to its excellent multiscale and precise localization properties. Furthermore, the availability of public masking thresholds [33] is another important reason why this domain was chosen. The lowest wavelet subband used in this work is constrained to be 16×16. Conventional noise-like watermarking [3, 18, 25] or bipolar watermarking [10, 12, 19] cannot be used in the current scheme. We shall describe in the subsequent sections what type of a watermark should be designed.

3.1 Basic Concept

In order to satisfy copyright protection and content authentication requirements simultaneously, a hidden watermark should be designed in a form that can carry the approximate information of a host image. In this manner, two purposes are accomplishable: (1) For fragile watermarking, the amount of change of a hidden watermark can be correctly calculated without accessing the host image. (2) For robust watermarking, the polarity of change of a hidden watermark can be determined obliviously. The amount of change and the polarity of change corresponding to a hidden watermark are helpful in calculating the detector responses of a robust watermark and a fragile watermark. To record the host image's information in the hiding process, the selected wavelet coefficients should be modulated using the cocktail watermarking strategy [18, 19]. At this stage, the hiding places are randomly divided into two groups: one group is for positive modulation, and the other one is for negative modulation. Because the JND values of a wavelet-based human visual system are publicly available [33], they are utilized to transform the host image's wavelet coefficients into another set of data using a quantization process. This new set of data can be regarded as the watermark values, which will be used in the hiding process. In the watermark detection process, the watermark values will be blindly extracted through a quantization operation. Finally, the host image can be recovered using a strategy which will be described in Sec. 3.3. It is noteworthy that based on the extracted watermarks and the hidden watermarks, the amount of modification can be determined and then used for fragile watermarking. On the other hand, if the recovered host image is also taken into consideration, then the polarity of modification can be easily detected and then used for robust watermarking. Note that this multipurpose watermarking scheme is performed by embedding watermarks only *once* without considering their hiding order. The details of this process will be given in Sec. 3.2. In general, the proposed multipurpose watermarking scheme can be applied to a variety of applications covering different media. For a specific application, a suitable watermark detection process should be determined by the user. The details of the watermark detection scheme will be presented in Sec. 3.4 to explain why our watermarks can be either robust or fragile.

3.2 Hiding: Quantization of Wavelet Coefficients as Masking Threshold Units (MTUs)

In this section, we will describe how to embed watermarks and record the host image's information. Conventionally, watermarks are embedded in transformed coefficients which are larger in magnitude. Let $w_{s,o}(x, y)$ be a selected wavelet coefficient with scale s, orientation o, and position (x, y). We modulate $w_{s,o}(x, y)$ to derive $w^m_{s,o}(x, y)$ using the i-th watermark value, $k_H(i)$. The relation between

(x, y) and i is a mapping function map, which can be defined as follows [18, 19]:

$$map(x,y) \quad = \quad \begin{cases} i, & \text{for positive modulation;} \\ -i, & \text{for negative modulation.} \end{cases} \tag{3}$$

Basically, the mapping results are stored for watermark detection.

The main goal in using k_H to modulate $w_{s,o}(x, y)$ is to quantize the selected wavelet coefficients into masking threshold units (MTUs), which are publicly available [33]. Let $JND_{s,o}(x, y)$ be the masking threshold [33] corresponding to $w_{s,o}(x, y)$; we can calculate an integer quantization value, q, as

$$q(|map(x,y)|) = \lfloor \frac{w_{s,o}(x,y)}{JND_{s,o}(x,y)} \rfloor, \tag{4}$$

where $\lfloor\cdot\rfloor$ denotes the $floor$ operation; as a result, $q(|map(x, y)|) \in \mathcal{Z}$ and $|q(|map(x, y)|)| \geq 1$ because $|w_{s,o}(x, y)| \geq |JND_{s,o}(x, y)|$. Using $q(|map(x, y)|)$, we can define the interval of the $q(|map(x, y)|)$-th MTU, $\eta^{(q(|map(x,y)|))}$, as follows:

$$\eta^{(q(|map(x,y)|))} \quad = \quad \begin{cases} [q(|map(x,y)|) \cdot JND_{s,o}(x,y) \quad (q(|map(x,y)|)+1) \cdot JND_{s,o}(x,y)), \\ \qquad\qquad\qquad\qquad \text{if } q(|map(x,y)|) \geq 1; \\ ((q(|map(x,y)|)+1) \cdot JND_{s,o}(x,y) \quad q(|map(x,y)|) \cdot JND_{s,o}(x,y)], \\ \qquad\qquad\qquad\qquad \text{if } q(|map(x,y)|) \leq -1. \end{cases} \tag{5}$$

In order to obtain a transparent watermarked image, the modulated quantity should not exceed $JND_{s,o}(x, y)$. According to our complementary modulation strategy [18, 19], a watermarking procedure tends to move $w_{s,o}(x, y)$ located at the $q(|map(x, y)|)$-th MTU to its neighboring unit. Using cocktail watermarking, we can embed two watermarks, respectively, based on a negative modulation (NM) rule and a positive modulation (PM) rule as follows [18, 19].

Negative modulation:

$$w^m_{s,o}(x,y) \quad = \quad \begin{cases} q(-map(x,y)) \cdot JND_{s,o}(x,y) - 1, & \text{if } w_{s,o}(x,y) > JND_{s,o}(x,y); \\ q(-map(x,y)) \cdot JND_{s,o}(x,y) + 1, & \text{if } w_{s,o}(x,y) < -JND_{s,o}(x,y). \end{cases} \tag{6}$$

Positive modulation:

$$w^m_{s,o}(x,y) = Q(map(x,y)) \cdot JND_{s,o}(x,y), \tag{7}$$

where

$$Q(map(x,y)) = \lceil \frac{w_{s,o}(x,y)}{JND_{s,o}(x,y)} \rceil \tag{8}$$

and

$$|Q(map(x,y))| = |q(map(x,y))| + 1.$$

$\lceil \cdot \rceil$ denotes the *ceiling* operation.

The modulated wavelet coefficient, $w^m_{s,o}(x,y)$, either falls into the $(q(-map(x,y))-1)$-th MTU, $\eta^{(q(-map(x,y))-1)}$ after the negative modulation rule is applied or falls into the $Q(map(x,y))$-th MTU, $\eta^{(Q(map(x,y)))}$ after the positive modulation rule is applied. In other words, the original and the modulated wavelet coefficients are located at different but contiguous MTUs, no matter what type of modulation rule is applied. From the modulated wavelet coefficients shown in Eqs. (6) and (7), one can calculate the modulated quantization index, q^m, as

$$|q^m(|map(x,y)|)| = \begin{cases} |q(-map(x,y))| - 1, & \text{for negative modulation;} \\ |q(map(x,y))| + 1, & \text{for positive modulation.} \end{cases} \tag{9}$$

The integer value $q^m(|map(x,y)|)$ is regarded as an embedded watermark value, $k_H(i)$, using either negative modulation or positive modulation based on the sign of $map(x,y)$. If we want to take $k_H(i)$ into consideration, the watermark hiding rules in Eqs. (6) and (7) should be rewritten as follows:

Negative modulation:

$$w^m_{s,o}(x,y) = \begin{cases} (k_H(i)+1) \cdot JND_{s,o}(x,y) - 1, & \text{if } w_{s,o}(x,y) > JND_{s,o}(x,y); \\ (k_H(i)-1) \cdot JND_{s,o}(x,y) + 1, & \text{if } w_{s,o}(x,y) < -JND_{s,o}(x,y). \end{cases} \tag{10}$$

Positive modulation:

$$w^m_{s,o}(x,y) = |k_H(i)| \cdot JND_{s,o}(x,y). \tag{11}$$

The hidden watermark k_H can be used to evaluate the robustness and the fragility of the extracted watermark without accessing the original image. As a result, the original image will never be used again; thus, we can call the proposed multipurpose watermarking scheme an oblivious one.

3.3 Host Image Recovery

Using the hidden watermark k_H, we can approximately reconstruct a host image with negligible degradation. In what follows, we shall show how this can be done. Let the i-th watermark value be $k_H(i)$; it is equal to the quantization index, $q^m(|map(x,y)|)$, as indicated in Eq. (9). The recovered quantization value, $q^r(|map(x,y)|)$, can be derived from Eq. (9) as follows:

$$|q^r(|map(x,y)|)| = \begin{cases} |q^m(-map(x,y))| + 1 = |k_H(-map(x,y))| + 1, & \text{for NM;} \\ |q^m(map(x,y))| - 1 = |k_H(map(x,y))| - 1, & \text{for PM.} \end{cases} \tag{12}$$

The difference, Δ, between a recovered wavelet coefficient and its corresponding original wavelet coefficient is bounded by $JND_{s,o}(x,y)$. That is,

$$\Delta = |q^r(|map(x,y)|) \cdot JND_{s,o}(x,y) - w_{s,o}(x,y)| < JND_{s,o}(x,y), \tag{13}$$

where $w_{s,o}(x,y)$ is a selected wavelet coefficient for hiding. Since our scheme has been designed based on the characteristics of the human visual system, the recovered host image should be perceptually indistinguishable from the original image.

3.4 Watermark Detection

In this section, we shall describe how an embedded watermark can be detected. Let $w^a_{s,o}(x,y)$ be a modulated wavelet coefficient which has experienced attacks; the positively/negatively modulated watermark value can be extracted without accessing the original image using a quantization process:

$$k^e_H(|map(x,y)|) = \lfloor \frac{w^a_{s,o}(x,y)}{JND_{s,o}(x,y)} \rfloor, \tag{14}$$

which depends on the sign of $map(x,y)$ (defined in Eq. (3)). By comparing the hidden watermark (k_H) and the extracted one (k^e_H), the purpose of fragile watermarking can be achieved. On the other hand, by comparing the hidden watermark, the extracted watermark and the host image's information (q^r), the goal of robust watermarking can be achieved. Note that the detector response (robustness or fragility) can be separately calculated in our scheme. In what follows, we shall describe in detail how this can be done.

3.4.1 Detection of Robust Watermarks

Robust watermarks are expected to resist any attack of any strength (gentle or severe). If the signs of $(k_H(i) - q^r(i))$ and $(k^e_H(i) - q^r(i))$ are the same, i.e., the majority of the transformed coefficients in the modulation and attacking processes are updated toward the same polarity, then they contribute positively to the detector response. A higher detector response provides stronger evidence that k^e_H is a genuine watermark. The detector response of robust watermarking (called "robust detector response") is defined as

$$\rho_{robust}(k_H, k^e_H) = \frac{\sum_{i=1}^{N_w} sign(k_H(i) - q^r(i)) \cdot sign(k^e_H(i) - q^r(i))}{N_w}, \tag{15}$$

where N_w is the watermark length and

$$sign(u) \quad = \quad \begin{cases} 1, & u \geq 0; \\ -1, & u < 0. \end{cases} \tag{16}$$

For robust watermarking, two detector responses are obtained with respect to the two complementary watermarks. The larger one is chosen as the final detector response; meanwhile, the type of incoming attack can be understood [18, 19, 20].

3.4.2 Detection of Fragile Watermarks based on the Characteristics of the Human Visual System

Fragile watermarks are different from robust watermarks and are supposed to be sensitive to tampering. Based on the standard of the human visual system, an image pixel is considered to have been tampered with if the difference between a hidden watermark value and its corresponding extracted watermark value is larger than t ($t \geq 1$) masking units. When t is set to be 1, this means that if the amount of modification exceeds the tolerance of the human visual system, then this modification will be considered to be malicious. However, images may be unavoidably manipulated by some incidental processes, such as compression. Under these circumstances, we cannot think of these incidental processes as malicious ones. In other words, a fragile watermarking scheme should be robust to incidental distortions. As we have noted with respect to cocktail watermarking [18, 19], incidental modification like compression tends to decrease the magnitudes of the transformed coefficients. On the other hand, incidental modification like sharpening tends to increase the magnitudes of the transformed coefficients. In what follows, we shall discuss the safe range into which a fragile watermark should fall when incidental distortions are encountered. Suppose a wavelet coefficient x was originally located at the $(j+1)-th$ masking unit, is moved to the $j-th$ masking unit and, thus, becomes x^M after NM. x^M is considered to not have been tampered with as long as the tampered coefficient, x^T, falls in the range between the $(j-t)-th$ and the $(j+1)-th$ masking units. Under these circumstances, the number of masking units (corresponding to NM) in the left interval and the right interval of x^T is t and 1, respectively. In other words, the untampered range is *asymmetric* with respect to x^T. This situation also applies similarly to PM. The tampered region and the robust region corresponding to negative modulation and positive modulation are illustrated in Fig. 1. Basically, our watermarking strategy makes the authentication process more robust (less fragile) to incidental distortions. If x^T is obtained by applying a compression/enhancing process, then x^M still has a good chance of being credible because the left/right interval of x^M is longer. On the other hand, fragility is determined from the other (shorter) interval (only *one* masking unit). Hence, tampering detection of a negatively modulated watermark is defined as

$$T^{neg}(i) = \begin{cases} 1, & |k_H(i)| > |k_H^e(i)| \wedge |k_H(i) - k_H^e(i)| > t; \\ 1, & |k_H(i)| \leq |k_H^e(i)| \wedge |k_H(i) - k_H^e(i)| > 1; \\ 0, & otherwise, \end{cases} \quad (17)$$

where $\wedge$ is an "*and*" operation. On the other hand, tampering detection of a positively modulated watermark is defined as

$$T^{pos}(i) \quad = \quad \begin{cases} 1, & |k_H(i)| < |k_H^e(i)| \wedge |k_H(i) - k_H^e(i)| > t; \\ 1, & |k_H(i)| \geq |k_H^e(i)| \wedge |k_H(i) - k_H^e(i)| > 1; \\ 0, & otherwise. \end{cases} \tag{18}$$

In sum, the global detector response of fragile watermarking (called "fragile detector response") is defined as

$$\rho_{fragile}^{neg}(k_H, k_H^e) = \frac{\sum_{i=1}^{N_w} T^{neg}(i)}{N_w}$$

and

$$\rho_{fragile}^{pos}(k_H, k_H^e) = \frac{\sum_{i=1}^{N_w} T^{pos}(i)}{N_w},$$

respectively, for negative modulation and positive modulation. Note that different t values enable our authentication scheme to adapt to various distortions. The fragility of an incidental process can be determined by

$$Min(\rho_{fragile}^{neg}(k_H, k_H^e), \rho_{fragile}^{pos}(k_H, k_H^e)). \tag{19}$$

Robustness and the perception-based fragility will be carefully analyzed in Sec. 4.

3.4.3 Detection of Fragile Watermarks based on Tendency of Attacks

As we have described in Sec. 3.4.2, incidental tampering is said to have occurred if the detector response of a fragile watermark (Eq. 19) is smaller than a preset threshold. However, the threshold is sometimes difficult to determine. In this section, a criterion is provided to judge the fragility based on the assumption that incidental manipulation tends to behave consistently while malicious one dose not. The consistency of attacking behavior can be defined as $BR_{fragile}$, which is expressed as

$$BR_{fragile} = \frac{MAX(\rho_{fragile}^{pos}(\cdot,\cdot), \rho_{fragile}^{neg}(\cdot,\cdot))}{MIN(\rho_{fragile}^{pos}(\cdot,\cdot), \rho_{fragile}^{neg}(\cdot,\cdot))}, \tag{20}$$

where $MAX(\cdot,\cdot)$ and $MIN(\cdot,\cdot)$ are the maximum and the minimum operations, respectively. Incidental processing will have the tendency to have a large $BR_{fragile}$ value. The threshold used for deciding the existence of non-malicious tampering is easier to derive than the one chosen in Eq. 19.

3.4.4 Detection of Fragile Watermarks based on Invariance Property

Tampering can also be detected by checking some invariance properties. It has been found that perception-based fragility can resist compression ($JPEG$ or $SPIHT$) up to the middle compression ratio. Previous feature-based image authentication methods [2, 6] suffered from the problem of shifting feature points when the compression ratios ranged from middle to high. Lin and Chang [15] proposed a solution which is able to resist $JPEG$ compression with any ratios. In this work, we shall adopt their invariance property to check the degree of similarity between watermarks. Because two complementary watermarks are embedded in our multipurpose watermarking scheme, the invariance property is checked based on the two watermarks. It is expected that the relationship between the two hidden watermarks will be maintained after incidental manipulations. Let k_{neg} and k_{pos} be the two watermarks hidden by means of negative modulation and positive modulation, respectively, and let k^e_{neg} and k^e_{pos} be the two extracted watermarks. We define the invariance property between a pair of watermark values located in the same position as follows:

- if $k_{neg}(i) - k_{pos}(i) > 0$ then $k^e_{neg}(i) - k^e_{pos}(i) \geq 0$,
- if $k_{neg}(i) - k_{pos}(i) < 0$ then $k^e_{neg}(i) - k^e_{pos}(i) \leq 0$,
- if $k_{neg}(i) - k_{pos}(i) = 0$ then $k^e_{neg}(i) - k^e_{pos}(i) = 0$.

If any one of the above three conditions is satisfied, then we can say that there no tampering has occurred.

3.5 Normalization of the Hidden Watermark k_H

The hidden watermark k_H is designed to carry the information of a host image and is, therefore, dependent on the host image. Any randomly selected watermark k^r may be highly correlated with the hidden watermark k_H, and this will cause a severe false positive problem. Hence, k_H should be normalized to $N(0,1)$ as Cox *et al.* did in [3]. This procedure will make k_H and k^r statistically independent. Let (m, σ) be the mean and the standard deviation of the hidden watermark k_H. The normalized k_H is denoted as k_G, where

$$k_G(i) = \frac{k_H(i) - m}{\sigma}. \tag{21}$$

To compute the false positive and false negative probability, the Gaussian distributed watermark k_G is used. The pair (m, σ) is regarded as an image-dependent watermark (IDW) key and is jointly used with k_G to generate k_H (using Eq. (21)) for watermark hiding (Sec. 3.2), host image recovery (Sec. 3.3) and watermark detection (Sec. 3.4).

4 Analysis

In this section, we will analyze the robustness and the fragility of the proposed multipurpose watermarking scheme used for integrity verification. Suppose a wavelet coefficient $w_{s,o}(x,y)$ was originally located at the $(j+1)-th$ masking unit, is moved to the $j-th$ masking unit after NM and becomes $w^m_{s,o}(x,y)$. The tampering effect can be modeled as follows:

$$w^a_{s,o}(x,y) = w^m_{s,o}(x,y) + n_{s,o}(x,y), \tag{22}$$

where $w^a_{s,o}(x,y)$ is the tampered coefficient and the amount of modification $n_{s,o}(x,y)$ is assumed to be Gaussian distributed with zero mean and variance σ. Under negative modulation, $w^a_{s,o}(x,y)$ is thought of as unchanged if $w^a_{s,o}(x,y)$ falls into the robust range, which is t masking units to the left of $w^m_{s,o}(x,y)$ and 1 masking unit to the right of $w^m_{s,o}(x,y)$. Therefore, the probability that a coefficient will still be credible after attacks is defined as

$$p^{neg}_{rr} = P\{-t \cdot JND_{s,o}(x,y) < n_{s,o}(x,y) < 1 \cdot JND_{s,o}(x,y)\}. \tag{23}$$

p^{neg}_{rr} can be rewritten as

$$\begin{aligned} p^{neg}_{rr} &= P\{-t \cdot JND_{s,o}(x,y) < n_{s,o}(x,y) < 0\} + P\{0 < n_{s,o}(x,y) < 1 \cdot JND_{s,o}(x,y)\} \\ &= erf(\frac{t \cdot JND_{s,o}(x,y)}{2\sigma}) + erf(\frac{JND_{s,o}(x,y)}{2\sigma}) \\ &= P^{neg}_l + P^{neg}_r, \end{aligned} \tag{24}$$

where $erf(\cdot)$ is the error function, defined as

$$erf(\epsilon) = \frac{2}{\sqrt{\pi}} \int_0^\epsilon e^{-u^2} du.$$

For positive modulation, a similar result can be derived, where

$$\begin{aligned} p^{pos}_{rr} &= P\{-JND_{s,o}(x,y) < n_{s,o}(x,y) < t \cdot JND_{s,o}(x,y)\} \\ &= P\{-JND_{s,o}(x,y) < n_{s,o}(x,y) < 0\} + P\{0 < n_{s,o}(x,y) < t \cdot JND_{s,o}(x,y)\} \\ &= erf(\frac{JND_{s,o}(x,y)}{2\sigma}) + erf(\frac{t \cdot JND_{s,o}(x,y)}{2\sigma}) \\ &= P^{pos}_l + P^{pos}_r. \end{aligned} \tag{25}$$

According to Eqs. (24) and (25), we know that

$$P^{neg}_l = P^{pos}_r \ \geq \ P^{neg}_r = P^{pos}_l.$$

P_l^{neg} and P_r^{pos} reflect the degree of robustness against incidental distortions, such as compression and sharpening. On the other hand, the degree of sensitivity in response to malicious tampering is determined by P_r^{neg} and P_l^{pos}.

The three parameters, JND, t, and σ, are closely related to p_{rr}^{neg} and p_{rr}^{pos}. First, the larger $JND_{s,o}(x,y)$ is, the more robust (less fragile) the watermark is. This is because either p_{rr}^{neg} or p_{rr}^{pos} is large. If t and σ are fixed at all wavelet-transformed scales, then our scheme is more sensitive to distortions at lower frequencies in terms of fragility. Secondly, t controls the tradeoff between robustness and fragility in our fragile watermarking scheme, as described in Sec. 3.4.2. In other words, the larger t is, the larger P_l^{neg} and P_r^{pos} are. This means that under NM/PM, tampering on the left/right interval of $x_{s,o}^m(x,y)$ is more robust than tampering on the right/left interval of $x_{s,o}^m(x,y)$. This again confirms our assertion with regard to perception-based fragile watermark detection given in Sec. 3.4.2. Thirdly, it should be noted that the smaller σ is, the larger P_l^{neg} and P_r^{pos} are. This implies that like distortions with smaller σ are easy to overcome because they are similar to modifications like compression with small-to-middle ratios. For manipulation like content replacement, σ is often larger and the manipulation is expected to be detected whenever t is not very large.

5 Experimental Results

A series of experiments was conducted to demonstrate the robustness and the fragility of the proposed multipurpose watermarking scheme. In addition to gray-scale images, color images were also considered. Among the existing color systems, $YCbCr$ was chosen for two reasons: (1) it has been adopted in many compression standards; (2) masking thresholds are available [33]. For color image watermarking, the watermarks were embedded and detected in the Y channel because humans are more sensitive to this channel. The flowchart of our method is illustrated in Fig. 2.

5.1 Results of Fragile Watermarking

The degree of fragility was verified using the gray-scale "MonaLisa" image, size 256 × 256, as shown in Fig. 3(a). The length of a watermark depends on both the host image and the wavelet-based visual model. Here, its length was dynamically determined to 3442. Using cocktail watermarking [18, 19], 6884 wavelet coefficients were modulated. The PSNR of the watermarked image shown in Fig. 3(b) was 43.37 dB. It is noted that no perceptual distortion could be observed on a computer screen at a distance of 32 *in* [33]. As expected, no tampering results were detected from the unmodified image

(Fig. 3(b)). Next, the watermarked MonaLisa image was slightly modified at the position of her face by means of texturing, as shown in Fig. 3(c). We wanted to see whether our fragile watermarks were sensitive to this type of malicious modification. Figs. 3(d)$\sim$(f) show when $t = 1$, the tampering detection results at different scales. Figs. 3(g)$\sim$(i) show another set of results when $t = 10$. It is found that in Fig. 3 that the altered regions were almost located. It is worth noticing that for different t values, the difference between k_H and k_H^e only slightly reduced even when t has been changed from 1 to 10. This implies that our multipurpose watermarking scheme is indeed fragile enough because the change of t would not affect fragility significantly. Furthermore, a color beach image, size 512×512 (shown in Fig. 4), was also used to demonstrate the fragility of our approach. The watermarks were embedded in the illumination channel, and the PSNR was 42.8 dB (Fig. 4(b)). An umbrella was placed on the watermarked image to change the image, as shown in Fig. 4(c). Figs. 4(d)$\sim$(i) show the tampering detection results at different scales with respect to $t = 1$ and $t = 10$, respectively. Again, we can see that all the altered regions were successfully detected.

In addition to malicious tampering, some incidental modifications caused by compression were used to check the robustness of our fragile watermarking scheme. The perception-based fragility and the invariance-based fragility were compared with respect to $JPEG$ and $SPIHT$ compression, respectively. The results of these comparisons are summarized in Table 1 and Table 2. From the two tables, it is obvious that the watermark embedded using negative modulation (NM) was more robust to compression than was that embedded using positive modulation (PM) by comparing their fragile detector response. The reason why this is true was given in our previous work [18, 19]. Basically, cocktail watermarking is designed to resist different kinds of attacks. Besides, a threshold (e.g., 0.15) can be used [13] to judge the robustness of a fragile watermarking scheme. This threshold may be application-dependent and is sometimes hard to determine. However, if the fragile detector response with respect to incidental modification could be controlled to be as small as possible, then it will be helpful to the selection of a threshold. As we can see in Table 1 and Table 2, the fragile detector responses with respect to NM are much smaller than those of [13] and are almost comparable with those of INV (invariance). This means that the robustness of incidental manipulation could be achieved to some extent while preserving fragility of malicious tampering. On the other hand, if the INV between watermark values is utilized, our approach can be extremely robust to compression.

In addition to compression, there are also some incidental manipulations [6] needed to be handled for fragile watermarking; for example, rescaling, histogram equalization, bright/contrast change, and noise addition. The criterion mentioned in Sec. 3.4.3 was used to measure the robustness of our fragile watermarking scheme when incidental manipulations were encountered. The cocktail watermarked MonaLisa image (Fig. 3(b)) was modified by SPIHT with compression ratio (64 : 1), JPEG compression with quality factor (20%), rescaling, histogram equalization, contrast enhancement, and Gaussian noise addition, respectively, as shown in Figs. 5(a)∼(f). The behavior ratio of fragility ($BR_{fragile}$) with respect to t ($1 \leq t \leq 10$) is depicted in Fig. 5(g). As we have described previously, there was no significant fragility loss even t was increased from 1 to 10. It can be observed from Fig. 5(g) that all curves turned flat when t was increased. The above mentioned experimental results indicate that a larger t will be beneficial to robustness but will not seriously affect fragility. On the other hand, a larger behavior ratio of fragility ($BR_{fragile}$) resulted from a larger t reflects that the behaviors of an attack can be captured by NM or PM. These phenomena confirmed what we have discussed in Sec. 3.4.3. The experimental results shown in Fig. 5(g) can be summarized as follows. The value of $BR_{fragile}$ is always larger than or equal to 4 as t increases under the $SPIHT$ compression, contrast adjusting, and $JPEG$ compression. All of these manipulations can thus be considered as incidental. On the other hand, our approach fails to tolerate Gaussian noise adding because $BR_{fragile}$ is too small. For the cases of histogram equalization and rescaling, our approach sometimes works but sometimes doesn't.

5.2 Results of Robust Watermarking

In this section, we shall discuss the experimental results with regard to robust watermarking. The "sailboat" image, size 256 × 256, as shown in Fig. 6(a), was used to evaluate the robustness of our scheme. The length of every single hidden watermark was 5928, and a total of 11856 wavelet coefficients were modulated. The PSNR of the watermarked image shown in Fig. 6(b) was 40.33 dB. Under these circumstances, no perceptual distortion was observed on a computer screen at a distance of 32 *in* [33]. 23 different attacks, including blurring, median filtering, rescaling, histogram equalization, jitter attack, changing the brightness/contrast, the negative film effect, segmentation, Gaussian noise adding, mosaicing, sharpening, texturizing, shading, the ripple effect, netdotting, uniform noise adding, the twirl effect, $SPIHT$ compression, $JPEG$ compression, StirMark, dithering, pixel spreading, and cropping were selected to test the robustness of our watermarking scheme. Some of the 23 attacked watermarked images are shown in Figs. 6(c)∼(h). Fig. 7 shows the robust watermark detection results obtained under the 23 attacks. For each pair of detected watermarks, one watermark could be destroyed (with lower response) while the other survived well (with higher detector response). The lowest detector response as shown in Fig. 7 was 0.32 (the 9-th attack), which corresponds to the

Gaussian noise attack. We used the worst result to verify the uniqueness requirement, i.e., to show the false positive probability. Fig. 8 shows the detector responses with respect to 10000 random marks (including the hidden one, i.e., the 5000-th mark). It is obvious that the response with respect to the hidden one is a recognizable spike. Basically, more accuracy is needed in selecting a reasonable threshold to determine the existence of an extracted watermark. For this reason, analysis of false positive and false negative probability is indispensable [19].

In effect, we have found that the current oblivious cocktail watermarking scheme is as good as the previously proposed non-oblivious one [18, 19] in terms of robustness.

6 Conclusion

A multipurpose watermarking scheme which can be applied to achieve both authentication and protection of multimedia data has been presented in this paper. Watermarks are embedded once in the hiding process and can be blindly extracted for different applications in the detection process. The proposed scheme has three special features: (1) The approximation information of a host image is kept in the hiding process by utilizing masking thresholds defined based on the human visual system [33]. (2) Oblivious and robust watermarking is achieved. (3) An asymmetric robust range is adopted for fragile watermarking to achieve malicious tampering detection and non-malicious tampering tolerance. Using this multipurpose watermarking scheme, we not only want to verify data integrity but also want to confirm the rightful ownership. Experimental results have demonstrated that our watermarking scheme is extremely effective for content authentication and copyright protection. In addition to images (gray-scale and color), this method has been extended to audio watermarking. To the best of our knowledge, this is the first method that combines both robust watermarking and fragile watermarking into one approach.

Future work will focus on eliminating the need of storing and retrieving the mapping file and the hidden watermarks (considered as the secret keys) for watermark detection. It is believed that secret key detection will encumber the automation and portability of watermarking. We shall put more time on studying the public key issue [27]. This is because public key detection admits of reading the watermarks for everybody and of removing them only by an authorized person.

Acknowledgment: The authors thank Dr. Martin Kutter for providing the color images shown in Fig. 4.

References

[1] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, "Copyright Protection of Digital Images by Embedded Unperceivable Marks", *Image and Vision Computing*, Vol. 16, pp. 897-906, 1998.

[2] S. Bhattacharjee and M. Kutter, "Compression Tolerant Image Authentication", *IEEE Inter. Conf. on Image Processing*, USA, 1998.

[3] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure Spread Spectrum WaterMarking for Multimedia", *IEEE Trans. Image Processing*, Vol. 6, pp. 1673-1687, 1997.

[4] S. Craver, N. Memon, B.-L. Yeo, and M. M. Yeung, "Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, Attacks, and Implications", *IEEE Journal on Selected Areas in Communications*, Vol. 16, pp. 573-586, 1998.

[5] J. F. Delaigle, C. De Vleeschouwer, and B. Macq, "Watermarking Algorithms based on a Human Visual Model", *Signal Processing*, Vol. 66, pp. 319-336, 1998.

[6] J. Dittmann, A. Steinmetz, and R. Steinmetz, "Content-based Digital Signature for Motion Pictures Authentication and Content-Fragile Watermarking", *IEEE Inter. Conf. Multimedia Computing and Systems*, Vol. II, Italy, 1999.

[7] J. Fridrich, "Combining Low-frequency and Spread Spectrum Watermarking", *Proc. SPIE Int. Symposium on Optical Science, Engineering, and Instrumentation*, 1998.

[8] J. Fridrich, "Methods for Detecting Changes in Digital Images", *Proc. IEEE Int. Workshop on Intell. Signal Processing and Communication Systems*, 1998.

[9] G. L. Friedman, "The Trustworthy Digital Camera: Restoring Credibility to the Photographic Image", *IEEE Trans. Consumer Electronics*, Vol. 39, pp. 905-910, 1993.

[10] F. Hartung and B. Girod, "Watermarking of uncompressed and compressed Video", *Signal Processing*, Vol. 66, pp. 283-302, 1998.

[11] F. Hartung, J. K. Su, and B. Girod, "Spread Spectrum Watermarking: Malicious Attacks and Counterattacks", *Proc. SPIE: Security and Watermarking of Multimedia Contents*, Vol. 3657, 1999.

[12] D. Kundur and D. Hatzinakos, "Digital Watermarking Using Multiresolution Wavelet Decomposition", *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Vol. 5, pp. 2969-2972, 1998.

[13] D. Kundur and D. Hatzinakos, "Digital Watermarking for TellTale Tamper Proofing and Authentication", *Procceedings of the IEEE*, Vol. 87, pp. 1167-1180, 1999.

[14] M. Kutter, F. Jordan, and F. Bossen, "Digital Signature of Color Images using Amplitude Modulation", *Journal of Electronic Imaging*, Vol. 7, pp. 326-332, 1998.

[15] C.-Y. Lin and S.-F. Chang, "A Robust Image Authentication Method Surviving JPEG Lossy Compression", *SPIE Storage and Retrieval of Image/Video Database*, Vol. 3312, San Jose, 1998.

[16] C.-Y. Lin and S.-F. Chang, "Issues and Solutions for Authenticating MPEG Video", *SPIE Inter. Conf. on Security and Watermarking of Multimedia Contents*, Vol. 3657, San Jose, 1999.

[17] C. S. Lu, S. K. Huang, C. J. Sze, and H. Y. Mark Liao, "A New Watermarking Technique for Multimedia Protection", to appear in *Multimedia Image and Video Processing*, eds. L. Guan, S. Y. Kung, and J. Larsen, CRC Press Inc, 2000.

[18] C. S. Lu, H. Y. Mark Liao, S. K. Huang, and C. J. Sze, "Cocktail Watermarking on Images", *Proc. 3rd International Workshop on Information Hiding*, Dresden, Germany, LNCS 1768, pp. 331-345, Sept. 29-Oct. 1, 1999. (patent pending)

[19] C. S. Lu, H. Y. Mark Liao, S. K. Huang, and C. J. Sze, "Highly Robust Image Watermarking Using Complementary Modulations", *Proc. 2nd International Information Security Workshop*, Malaysia, LNCS 1729, pp. 136-153, Nov. 6-7, 1999.

[20] C. S. Lu, Y. V. Chen, H. Y. Mark Liao, and C. S. Fu, "Complementary Watermarks Hiding for Robust Protection of Images Using DCT", *Proc. Inter. Symposium on Signal Processing and Intelligent System: special session on Image Processing and Pattern Identification*, China, pp. 293-298, Nov. 26-28, 1999 (Invited Paper).

[21] C. S. Lu and H. Y. Mark Liao, "Oblivious Watermarking Using generalized Gaussian", *5th Joint Conf. on Information Sciences (JCIS), Vol. II: 3rd Inter. Conf. on Computer Vision, Pattern Recognition and Image Processing*, pp. 260-263, Atlantic City, USA, Feb. 28-Mar. 2, 2000.

[22] F. Mintzer and G. W. Braudaway, "If one watermark is good, are more better?", *Inter. Conf. on Acoustic, Speech, and Signal Processing*, pp. 2067-2070, 1999.

[23] F. Petitcolas, R. J. Anderson, and M. G. Kuhn, “Attacks on Copyright Marking Systems”, *Second Workshop on Information Hiding*, USA, pp. 218-238, 1998.

[24] F. Petitcolas, R. J. Anderson, and M. G. Kuhn, “Information Hiding: A Survey”, *Proceedings of the IEEE: special issue on Protection of Multimedia Content*, Vol. 87, pp. 1062-1078, 1999.

[25] C. I. Podilchuk and W. Zeng, “Image-Adaptive Watermarking Using Visual Models”, *IEEE Journal on Selected Areas in Communications*, Vol. 16, pp. 525-539, 1998.

[26] J. J. K. Ruanaidh and T. Pun, “Rotation, Scale, and Translation Invariant Spread Spectrum Digital Image Watermarking”, *Signal Processing*, Vol. 66, pp. 303-318, 1998.

[27] R. van Schyndel, A. Z. Tirkel, and I. D. Svalbe, “Key Independent Watermark Detection”, *IEEE Inter. Conf. on Multimedia Computing and Systems*, Italy, Vol. I, pp. 580-585, 1999.

[28] M. D. Swanson, B. Zhu, and A. H. Tewfik, “Multiresolution Scene-Based Video Watermarking Using Perceptual Models”, *IEEE Journal on Selected Areas in Communications*, Vol. 16, pp. 540-550, 1998.

[29] M. D. Swanson, B. Zhu, A. H. Tewfik, and L. Boney, “Robust Audio Watermarking Using Perceptual Masking”, *Signal Processing*, Vol. 66, pp. 337-356, 1998.

[30] M. D. Swanson, M. Kobayashi and A. H. Tewfik, “Multimedia Data-Embedding and Watermarking Technologies”, *Proc. of the IEEE*, Vol. 86, pp. 1064-1087, 1998.

[31] P. C. Su, C.-C. Jay Kuo, and H. J. Wang, “Blind Digital Watermarking for Cartoon and Map Images”, *SPIE International Symposium Electronic Imaging*, 1999.

[32] S. Walton, “Image Authentication for A Slippery New Age”, *Dr. Dobb’s Journal*, Vol. 20, pp. 18-26, 1995.

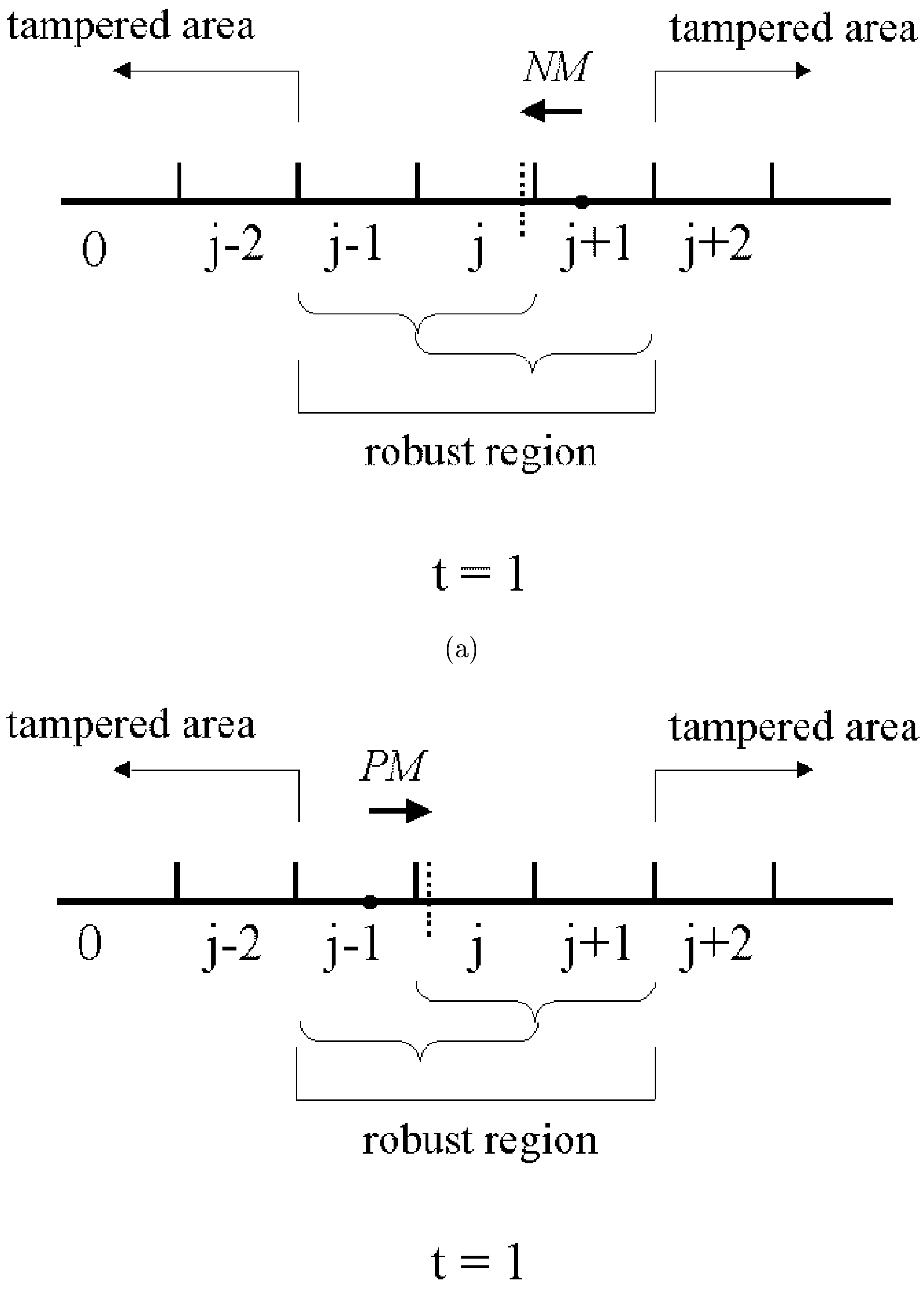
[33] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, “Visibility of Wavelet Quantization Noise”, *IEEE Trans. Image Processing*, Vol. 6, pp. 1164-1175, 1997.

[34] R. B. Wolfgang and E. J. Delp, “Fragile Watermarking Using the VW2D Watermark”, *Proc. SPIE/IS&T Inter. Conf. Security and Watermarking of multimedia Contents*, Vol. 3657, pp. 40-51, 1999.

[35] M. Wu and B. Liu, “Watermarking for Image Authentication”, *IEEE Inter. Conf. on Image Processing*, 1998.

[36] M. M. Yeung and F. Mintzer, “An Invisible Watermarking Technique for Image Verification”, *IEEE Conf. Image Processing*, Vol. 2, pp. 680-683, 1997.

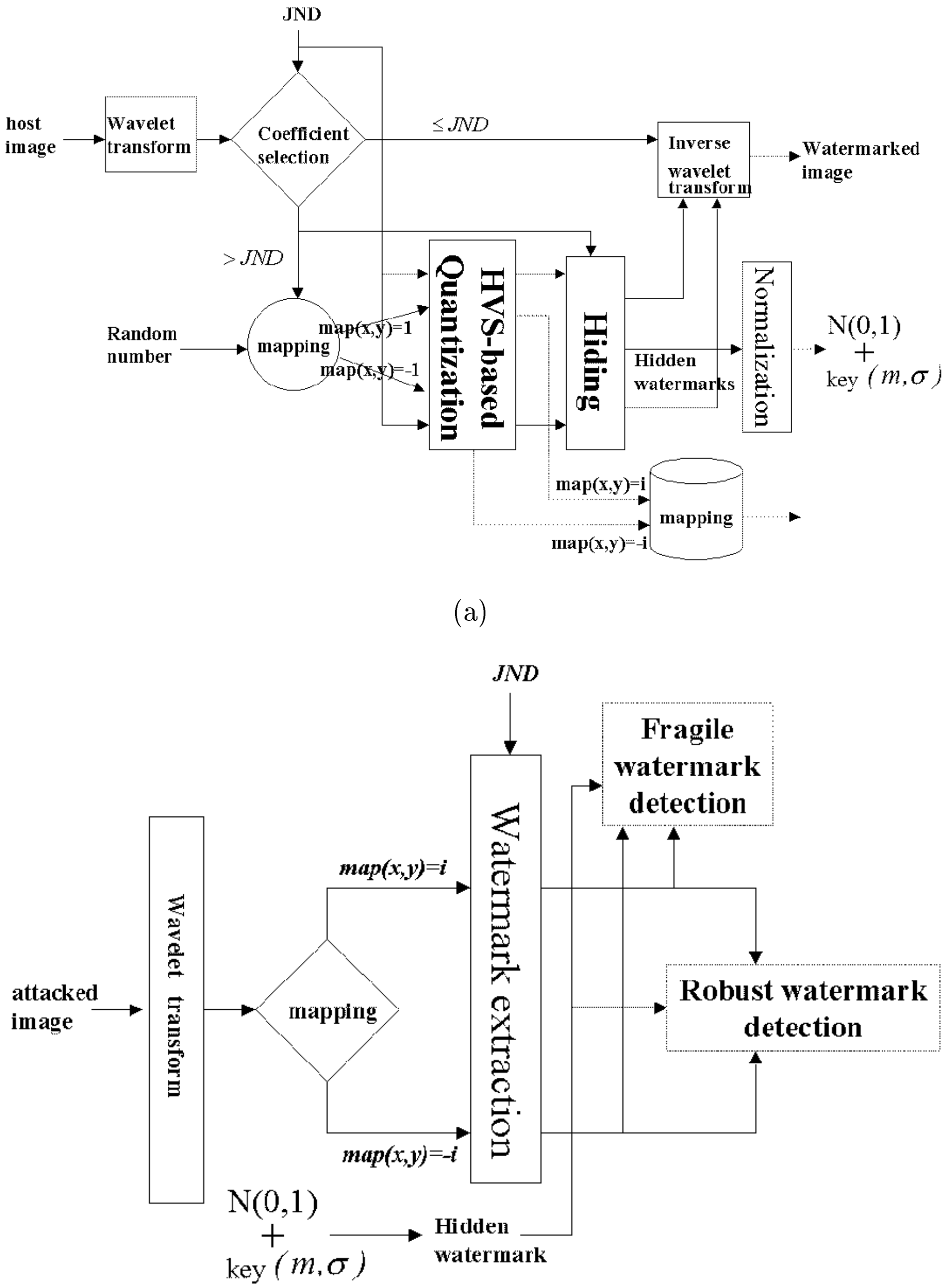
[37] B. Zhu, M. D. Swanson, and A. H. Tewfik, “Transparent Robust Authentication and Distortion Measurement Technique for Images”, *The 7th IEEE Digital Signal Processing Workshop*, pp. 45-48, 1996.



(a)

(b)

Figure 1: Illustration of the tampered region and the robust region for (a) negative modulation (NM) and (b) positive modulation (PM) when t=1. The arrow with the label NM or PM indicates the direction of alternation in the hiding process. Note that the robust region (indicated with {}) is asymmetric with respect to j for $t > 1$.



(a)

(b)

Figure 2: The flowchart of our multipurpose watermarking scheme: (a) watermark hiding; (b) watermark detection.

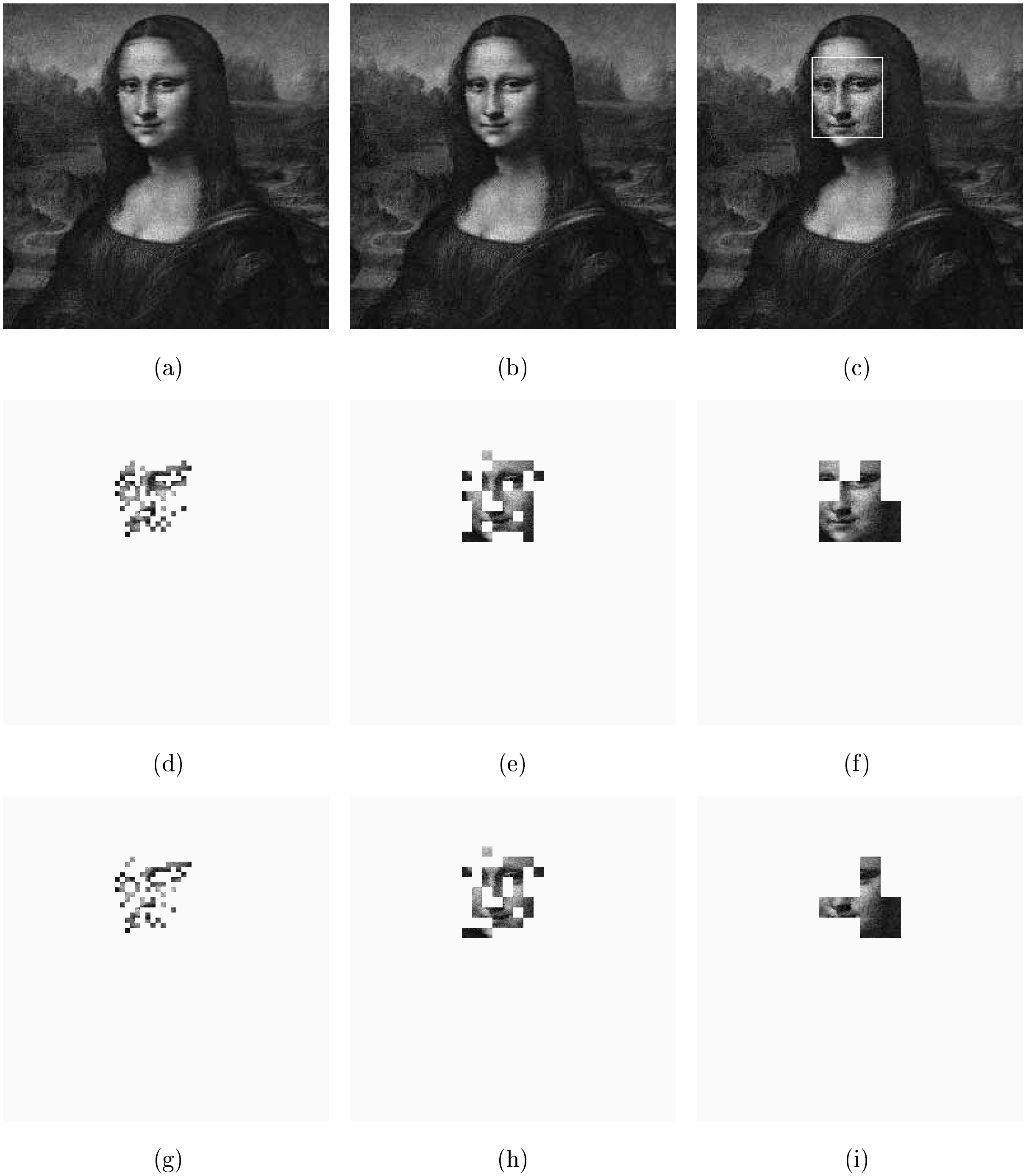


Figure 3: Malicious tampering detection: (a) host image; (b) watermarked image; (c) image after malicious tampering; (d)$\sim$(f) the tampering detection results at the $2^2 \sim 2^4$ scales with respect to $t = 1$; (g)$\sim$(i) the tampering detection results at the $2^2 \sim 2^4$ scales with respect to $t = 10$.

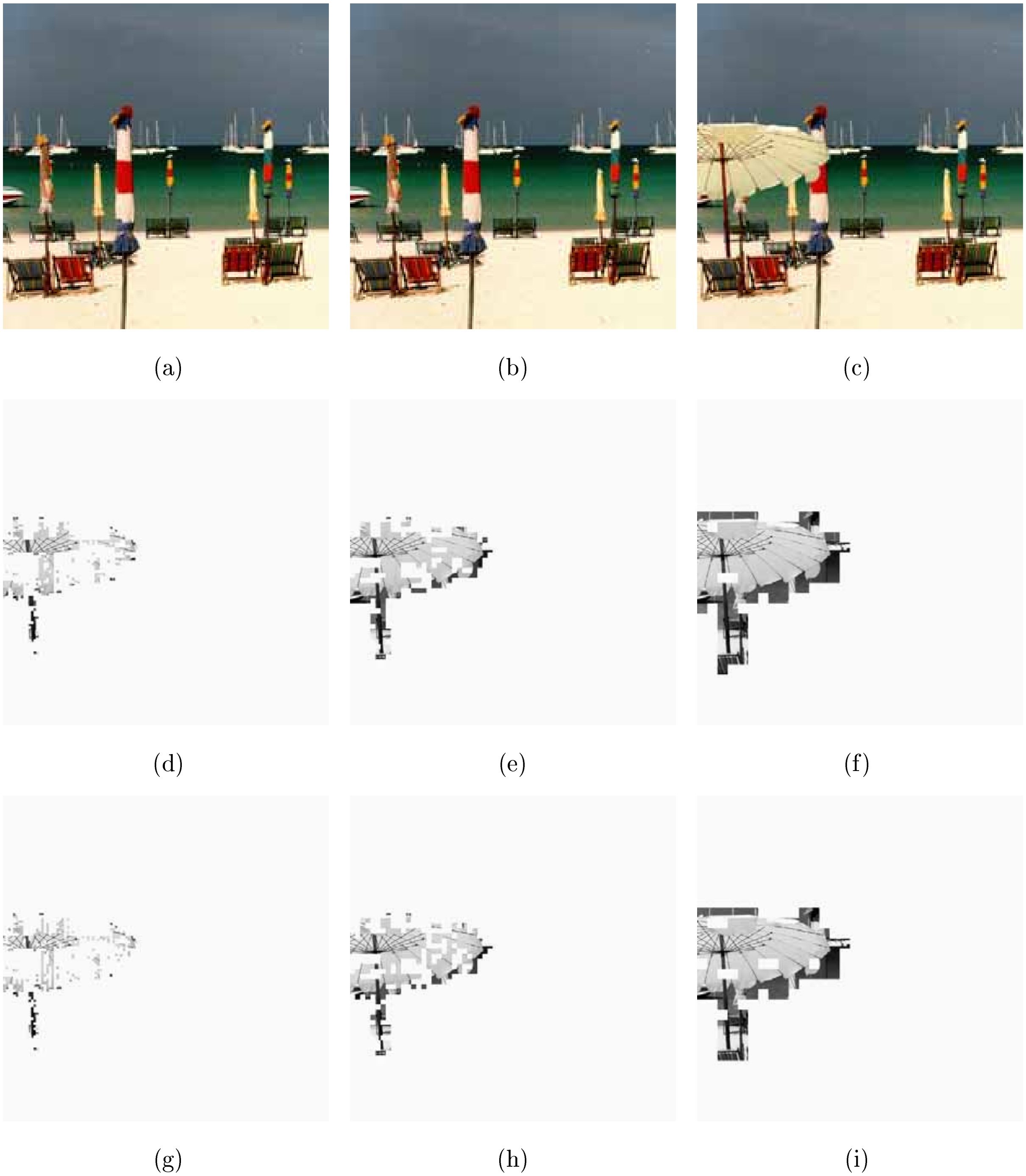


Figure 4: Tampering detection of object placement: (a) host image; (b) watermarked image; (c) image after object placing; (d)∼(f) the tampering detection results at the $2^2 \sim 2^4$ scales with respect to $t = 1$; (g)∼(i) the tampering detection results at the $2^2 \sim 2^4$ scales with respect to $t = 10$.

Table 1: **Tampering degree evaluation under *JPEG* compression.**

Compression Ratio (Quality Factor %)	*Degree of tampering*						
	t=1		*t=2*		*t=3*		*INV*
	NM	*PM*	*NM*	*PM*	*NM*	*PM*	
6.07(70%)	0.037	0.129	0.035	0.128	0.035	0.128	0.039
7.54(60%)	0.024	0.200	0.058	0.199	0.057	0.199	0.043
8.93(50%)	0.076	0.264	0.068	0.261	0.067	0.261	0.049
10.84(40%)	0.104	0.336	0.091	0.332	0.088	0.331	0.049
13.70(30%)	0.142	0.416	0.115	0.407	0.111	0.407	0.049
19.57(20%)	0.191	0.560	0.145	0.541	0.135	0.534	0.054
32.09(10%)	0.274	0.717	0.216	0.684	0.183	0.670	0.073

Table 2: **Tampering degree evaluation under *SPIHT* compression.**

Compression Ratio	*Degree of tampering*						
	t=1		*t=2*		*t=3*		*INV*
	NM	*PM*	*NM*	*PM*	*NM*	*PM*	
4	0.001	0.023	0.000	0.023	0.000	0.023	0.017
8	0.013	0.086	0.010	0.086	0.009	0.086	0.022
16	0.079	0.379	0.050	0.374	0.046	0.373	0.025
32	0.132	0.671	0.066	0.665	0.047	0.665	0.023
64	0.242	0.845	0.145	0.833	0.094	0.827	0.030

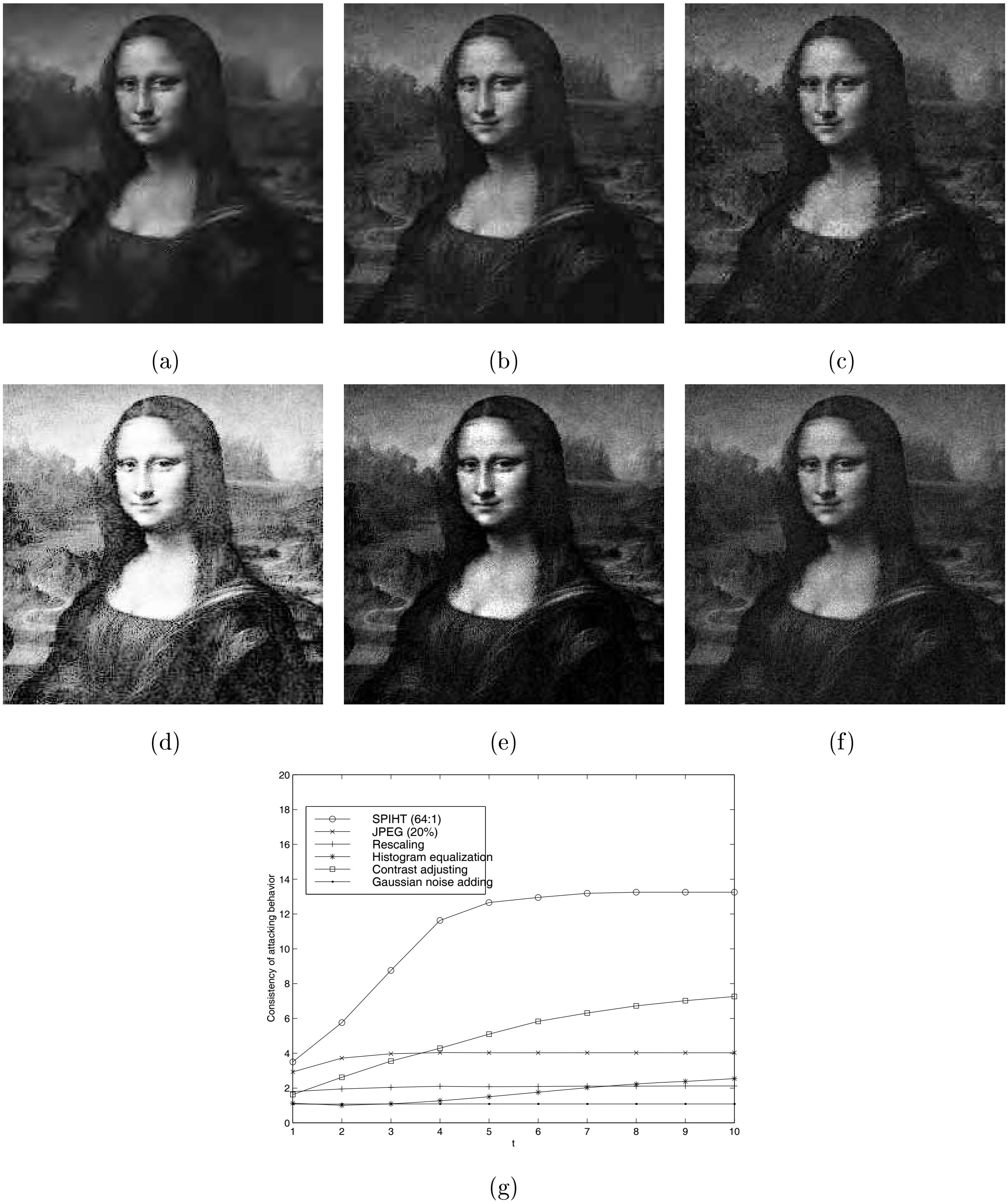
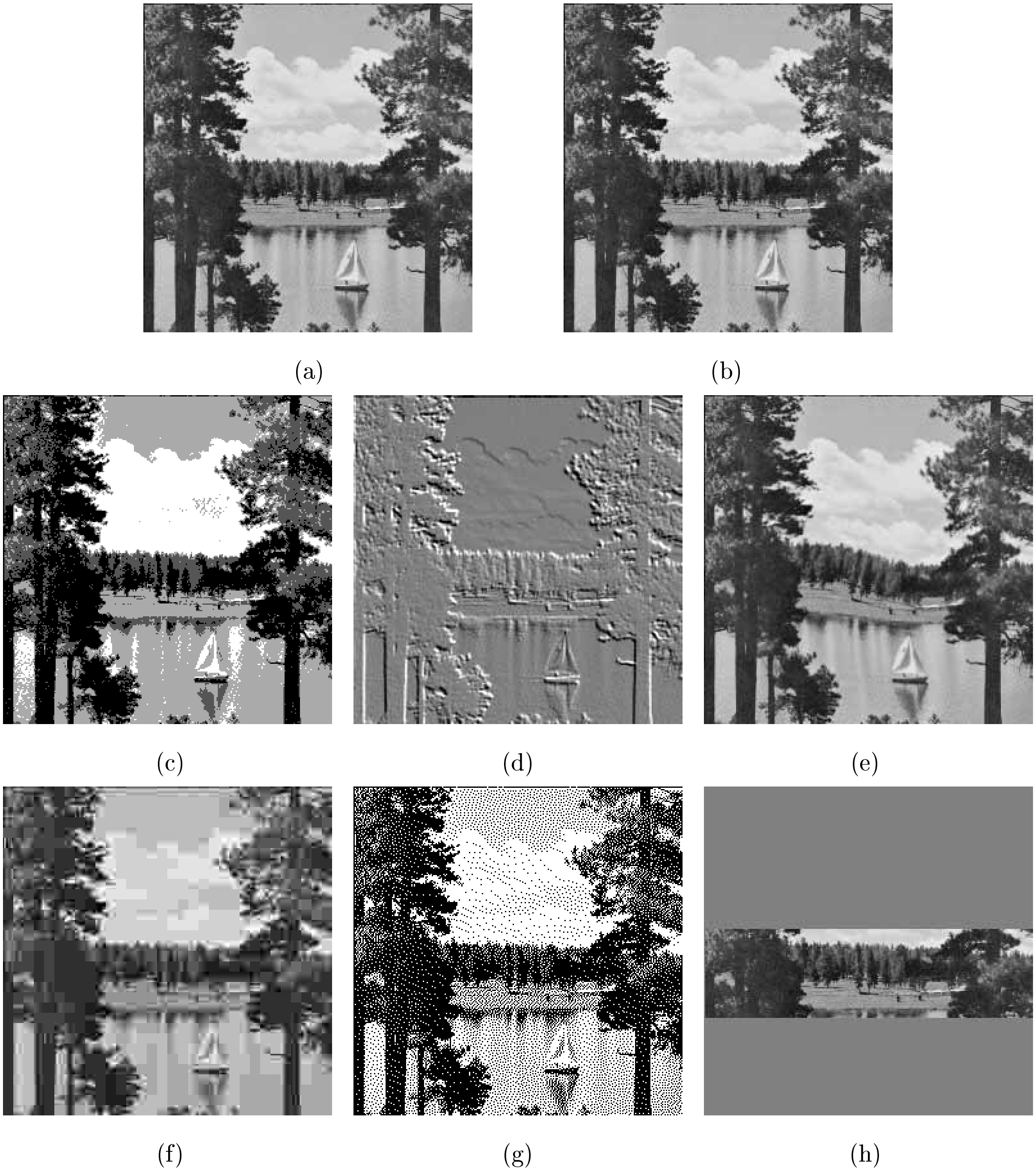


Figure 5: Fragile watermarks facing incidental tampering: (a) SPIHT with compression ratio 64 : 1; (b) JPEG with quality factor 20% (compression ratio $\approx 20:1$); (c) rescaled; (d) histogram equalized; (e) contrast adjusted; (f) Gaussian noise added; (g) the $BR_{fragile}$ values obtained at different t ($1 \leq t \leq 10$) with respect to six distinct incidental manipulations.



(a) (b)

(c) (d) (e)

(f) (g) (h)

Figure 6: Robust watermarking: (a) host image (b) watermarked image; (c)∼(h) attacked image corresponding to segmentation, shading, the twirl effect, $JPEG$ compression, dithering, and cropping.

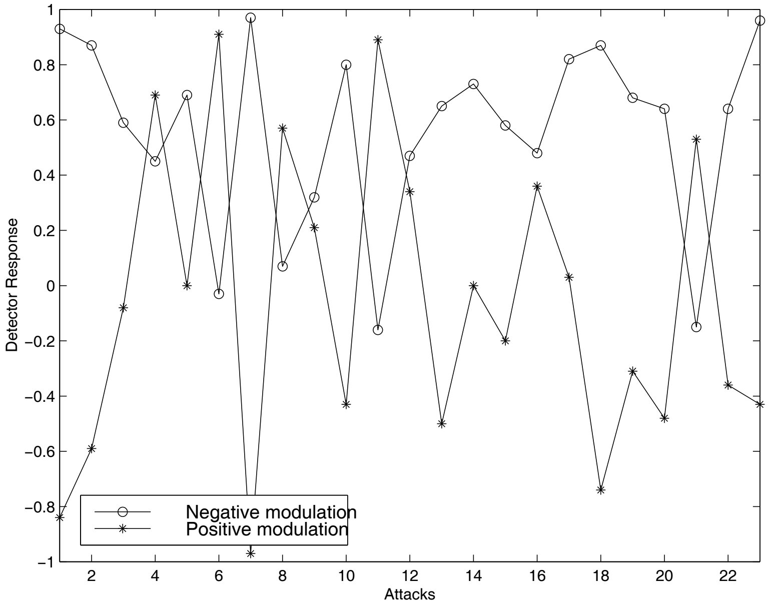
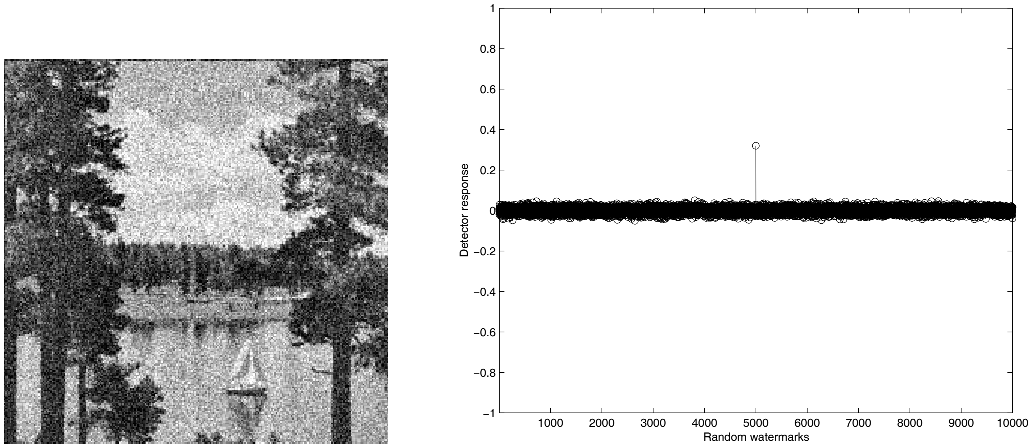


Figure 7: Detector response for robust watermarks.



(a)

(b)

Figure 8: Uniqueness verification of robust watermarking under a Gaussian noise adding attack: (a) attacked image after the Gaussian noise was added; (b) the detector responses of the extracted watermark with respect to 10000 random marks (including the hidden one, the 5000-th mark).