Mean Quantization-based Fragile Watermarking for Image Authentication

Gwo-Jong Yu*, Chun-Shien Lu**, and Hong-Yuan Mark Liao**,†
* Department of Computer Science and Information Engineering
National Central University, Chung-Li, Taiwan
** Institute of Information Science, Academia Sinica, Taipei, Taiwan
{yugj, lcs, liao}@iis.sinica.edu.tw

Abstract

The existing digital image editing tools have made the authentication of digital images an important issue. The objective of this paper is to propose an image authentication scheme, which is able to detect malicious tampering while tolerating some incidental distortions. By modeling the magnitude changes caused by incidental distortion and malicious tampering as Gaussian distributions with small and large variances, respectively, we propose to embed a watermark by using a mean quantization technique in the wavelet domain. The proposed scheme is superior to the conventional quantization-based approaches in terms of the credibility of authentication. Statistical analysis is conducted to show that the probabilities of watermark errors caused by malicious tampering and incidental distortion will be, respectively, maximized and minimized when our new scheme is applied. Experimental results demonstrate that the credibility of our method is superior to that of the conventional quantization-based methods under malicious attack followed by an incidental modification, such as JPEG compression, sharpening or blurring.

Keywords: Fragile watermarking, image authentication, fragility, robustness, wavelet transform, human visual system.

†To whom all correspondence should be addressed.

1 Introduction

The invention of the Internet provides a brilliant way of transmitting digital media. When digital media contain important information, their credibility must be ensured. As a consequence, a reliable media authentication system is indispensable when digital media are transmitted over a network. In order to save bandwidth and storage space, digital media are usually transmitted or stored in a compressed format. In addition, media like images may be processed by blurring or sharpening for specific purposes. Under these circumstances, an image authentication system should be able to tolerate some commonly used incidental modifications, such as JPEG compression, sharpening, and/or blurring. In this paper, we shall focus our discussion on image authentication.

In the literature, image authentication methods can be roughly classified as being either digital signature-based or watermark-based. The digital signature-based approach [1, 2, 3, 4] does not modify the content of an image. Instead, it extracts either global features or relational features from media for authentication purposes. For example, Bhattacharjee and Kutter [1] used the positions of a set of feature points as a digital signature. By examining the existence of feature points, images can be authenticated. Lin and Chang [4] computed the invariant relations between the coefficients of two randomly selected DCT blocks and then used them as a digital signature. Their method is able to resist JPEG compression with compression ratios (CR) up to 20:1. The major limitation of a digital signature-based method is that it can only be used for the purpose of verification, not copyright protection.

The watermark-based image authentication approaches, on the other hand, detect potential tampering based on the fragility of a hidden watermark [5, 6, 7, 8, 9, 10, 11, 12]. In Kundur and Hatzinakos' [5] quantization-based method, a watermark value is encoded by modulating a selected wavelet coefficient into a quantized interval. Basically, the quantity they used for modulation, which is monotonously increased from high resolution to low resolution, violates the capacity constraint of the human visual system [13]. They defined a tamper assessment function (TAF), which is the ratio of the number of tampered coefficients to the total number of coefficients in a specific subband, in order to measure the degree of tampering. They also point out if the TAF values decrease monotonously from high reso-

lution to low resolution, then it is very likely that the manipulation is JPEG compression. However, they did not address the situation in which an instance of malicious tampering and an incidental manipulation are imposed simultaneously. Recently, Lu et al. [6, 7] proposed a multipurpose watermarking technique for image/audio authentication and protection. They combined a complementary modulation [14, 15] strategy and an image/audio dependent quantization mechanism to hide watermarks. In addition, they proposed several detection techniques to perform authentication and protection simultaneously.

In [2], Dittmann *et al.* mentioned that incidental distortions, such as JPEG compression, blurring or sharpening should not be treated as malicious tampering. They also mentioned that if a watermarked image is tampered with maliciously, then the portions where the watermark errors emerge should be the manipulated areas. Their argument is only partially true because incidental operations which are not malicious also cause watermark errors. Under these circumstances, one cannot judge whether a modification is malicious or not simply by looking at watermark errors. Therefore, the objective of this paper is to increase the credibility of the embedded watermark by maximizing or minimizing, respectively, the probabilities of watermark errors caused by malicious tampering and incidental distortion. Under these circumstances, an instance of malicious tampering can be easily distinguished from an incidental modification. In general, the probability of watermark error caused by an incidental distortion can be reduced by either enlarging the quantization interval or reducing the quantity of modifications on coefficients. However, it is well known that the maximum quantization interval should be bounded by the human visual system [13] so that visual quality can be maintained. As a consequence, the only methodology that we can adopt here is to increase the robustness by decreasing the variance of coefficients. Owing to the fact that the variance of the sub-block mean is smaller than that of an individual sample, we know that a watermark value encoded by quantizing the mean of a set of coefficients is more robust than one encoded by quantizing a single coefficient.

Under a reasonable assumption that the quantity of modifications caused by an incidental distortion is smaller than that caused by a malicious distortion, the modifications caused by an incidental distortion or an instance of malicious tampering can be, respectively, modeled as a Gaussian distribution with smaller or larger variance. In a good image authentication

system, it is expected that the embedded watermark should be robust enough to tolerate incidental distortion and fragile enough to detect malicious tampering. However, it is also well known that robustness and fragility are two factors that compete against each other. Therefore, we need to seek a tradeoff between them that can lead to the best outcome. In order to achieve the above mentioned goal, a mechanism which can be used to encode a watermark such that the probabilities of watermark errors caused by malicious tampering and incidental distortion are, respectively, maximized and minimized, is indispensable.

In this paper, we propose a mean quantization-based fragile watermarking approach which can be used to judge the credibility of a suspect image. The mean quantization approach embeds a watermark by taking the mean value of a set of wavelet coefficients. Through theoretical analysis of the probabilities of watermark errors caused by malicious tampering and incidental distortion, the best number of coefficients needed to embed a watermark at each scale can be computed such that the tradeoff between robustness and fragility can be optimized. Since the probability of watermark errors caused by incidental distortion at each scale is different, the detection responses at all scales should be integrated so as to obtain a global estimation of the maliciously attacked area. Then, we can use some decision rules to judge whether a suspect image has been tampered with or not.

The remainder of this paper is organized as follows. The mean quantization-based fragile watermarking approach will be described in Sec. 2. An information fusion technique which can be used to integrate the detection results at multiple scales will be addressed in Sec. 3. Experimental results and conclusions will be given in Sec. 4 and Sec. 5, respectively.

2 Mean Quantization: A New Mechanism to Achieve better Authentication

In this section, we shall describe the proposed mean quantization-based fragile watermarking approach. In order to protect the original source, our watermark extraction process will be designed in a blind detection manner. Blind detection means the original source is not required for watermark extraction. Among the existing blind watermarking schemes,

the quantization-based watermarking approach is the simplest one to achieve the above mentioned goal. This is because in a quantization-based approach, a watermark is encoded and decoded by the same quantization operation. In the following, we shall first introduce the conventional quantization-based approach and point out its disadvantages. Then, a mean quantization-based approach will be proposed to eliminate these disadvantages. Finally, we will propose a systematic way to determine an optimal number of coefficients for mean quantization.

2.1 Disadvantages of the Conventional Quantization-Based Scheme

The quantization-based fragile watermarking approach [5] divides a real number axis in the wavelet domain into intervals with equal size at each scale and assigns watermark symbols to each interval periodically. Assuming that x is a wavelet coefficient, and that q is the size of a quantization interval, the watermark symbol, which is either 0 or 1, is determined by a quantization function Q, where

$$Q(x,q) = \begin{cases} 0, & \text{if } tq \le x < (t+1)q \text{ for } t = 0, \pm 2, \pm 4, \ldots \\ 1, & \text{if } tq \le x < (t+1)q \text{ for } t = \pm 1, \pm 3, \pm 5, \ldots. \end{cases} \tag{1}$$

Let w denote the target watermark value which is to be encoded for a wavelet coefficient x. The encoding rule is as follows: If $Q(x,q) = w$, then no modification is necessary for x; otherwise, x is updated to x^* by

$$x^* = \begin{cases} x + q, & \text{if } x \le 0 \\ x - q, & \text{if } x > 0. \end{cases} \tag{2}$$

Kundur and Hatzinakos' approach [5] uses $\delta 2^l$ as the size of a quantization interval, where δ is a pre-specified positive integer, $l = 1, \ldots, L$, and L is the number of scales used in wavelet transform. In their approach, the size of a quantization interval increases monotonically from high frequency to low frequency. However, this kind of design violates the characteristics of the human visual system [13]. In our design, we take the limitations of the human visual system into consideration. On the other hand, since any modification applied to an image will change its wavelet coefficients, it is reasonable to expect that their corresponding watermark symbols will be changed, too. By comparing the extracted watermark values with

the original hidden ones, the maliciously attacked area can be located. Although the fragility of the watermark proposed in [5] is able to reveal malicious tampering, their watermark is not robust enough to tolerate incidental distortions. Therefore, we shall seriously address this problem as well.

2.2 The Proposed Scheme

Watson *et al.* [13] investigated the sensitivity of the human eye and then proposed a wavelet-based human visual system (HVS). According to HVS, the wavelet coefficients can be modified without causing visual artifacts. In order for a watermarked image to satisfy the transparency requirement, the quantization interval will be defined as the maximally allowable modification quantity based on their HVS [13]. Our basic concept is that if the modification quantity of a wavelet coefficient does not exceed its corresponding masking threshold, then this modification will not raise visual awareness. Otherwise, we can say the modification is a malicious one.

Statistically, the mean value of a set of samples has variance smaller than that of a single sample. We expect that if the watermark is embedded by modulating the mean value rather than a single coefficient, the probability of watermark errors will be smaller. This is because the mean value is more difficult to move beyond the quantization interval where it is originally located. For a specific subband, let the size of a quantization interval be denoted as q, and let a set of n wavelet coefficients be denoted as x_i, $i = 1, \ldots, n$. The mean value of x_i's can be computed as follows:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{3}$$

For the purpose of robustness, a watermark value should be encoded by moving its mean $\bar{x}$ to the middle of a corresponding quantization interval such that the modulated $\bar{x}$ can not be easily moved away from the current interval. The mean quantization-based fragile watermarking approach operates as follows. Let w be the target watermark symbol to be encoded, and let $\bar{r}$ be the quantization noise defined as

$$\bar{r} = \bar{x} - \left\lfloor \frac{\bar{x}}{q} \right\rfloor \cdot q, \tag{4}$$

where $\lfloor \cdot \rfloor$ is the *floor* operator. To encode w, the amount of update, $\bar{u}$, added to the mean coefficient $\bar{x}$, can be determined as follows:

$$\bar{u} = \begin{cases} -\bar{r} + 0.5q, & \text{if } Q(\bar{x}, q) = w \\ -\bar{r} + 1.5q, & \text{if } Q(\bar{x}, q) \neq w \text{ and } \bar{r} > 0.5q \\ -\bar{r} - 0.5q, & \text{if } Q(\bar{x}, q) \neq w \text{ and } \bar{r} \leq 0.5q. \end{cases} \tag{5}$$

As a consequence, the new mean coefficient becomes $\bar{x}^* = \bar{x} + \bar{u}$. In Eq. (5), $0.5q$ and $1.5q$ are used to shift a mean coefficient $\bar{x}$ to the middle of a quantization interval, such that $\bar{x}^*$ is relatively difficult to move away from the current interval. However, updating the mean coefficient implies that all the constituent coefficients need to be updated accordingly by

$$x_i^* = x_i + \bar{u}, \ (1 \leq i \leq n), \tag{6}$$

where x_i^* is an updated wavelet coefficient.

Let a modified wavelet coefficient $\hat{x}$ be modeled as

$$\hat{x} = x^* + \Delta, \tag{7}$$

where x^* is the wavelet coefficient defined in Eq. (6) and Δ represents the amount of update caused by tampering. In the case of incidental modification, Δ_I can be modeled as a Gaussian distribution with a smaller variance, that is

$$\Delta_I \sim N(0, \sigma_I^2), \tag{8}$$

where σ_I denotes the variance of the modification quantities due to an incidental distortion. On the other hand, for an instance of malicious tampering, Δ_M can be modeled as a Gaussian distribution with a larger variance, that is,

$$\Delta_M \sim N(0, \sigma_M^2), \tag{9}$$

where σ_M denotes the variance of modification quantities caused by malicious tampering. Usually, it is assumed that the variance of modification quantities caused by an incidental distortion is smaller than that caused by an instance of malicious tampering, i.e., $\sigma_I < \sigma_M$. Lin and Chang [4] have provided some reference values for σ_I and σ_M in the spatial domain.

Fig. 1 illustrates the statistical distributions of updates of wavelet coefficients corresponding to incidental and malicious modifications. In Fig. 1, each quantization interval has a corresponding binary watermark symbol, 0 or 1. The watermark symbol associated with coefficient x changes when the amount of tampering Δ is greater than $|0.5q|$. Based on the above design, a wavelet coefficient is put at the middle of a quantization interval in order to reduce the probability of watermark errors caused by tampering. Since the same watermark symbol appears periodically, the watermark symbol may not be changed even for $\Delta > |0.5q|$. For example, if $\Delta = 2.0q$, the tampered coefficient $\hat{x}$ will fall into the interval $[(2t+2)q, (2t+3)q]$ with the watermark symbol "0", which is the same as the original watermark symbol carried by x. This is the common drawback of a quantization-based watermarking approach. However, since the variance of modification quantities caused by an instance of malicious tampering is larger than that caused by incidental distortion, we can expect that an incidentally distorted coefficient has greater possibility of falling into the interval $[-0.5q, 0.5q]$. Thus, we have the hypothesis that the probability of watermark errors caused by an incidental distortion is smaller than that caused by an instance of malicious tampering. In addition, we will conduct an analysis in the next paragraph to prove that our scheme can improve the common drawback of the conventional quantization-based approach.

In order to ensure that an authentication system is incidental-distortion-tolerant, the credibility of a fragile watermark should be increased so that an incidental modification won't be misunderstood as a malicious one. Because the sum of more than two random variables with Gaussian distribution is still a Gaussian distribution but with smaller variance, we have

$$\bar{\Delta} \sim N(0, \frac{1}{n}\sigma^2) \tag{10}$$

when $\Delta \sim N(0, \sigma^2)$. Thus, when mean quantization is applied, the distribution of modification quantities caused by malicious or incidental distortions will become

$$\bar{\Delta}_I \sim N(0, \frac{1}{n}\sigma_I^2), \tag{11}$$

or

$$\bar{\Delta}_M \sim N(0, \frac{1}{n}\sigma_M^2), \tag{12}$$

respectively. Eqs. (11) and (12) indicate that the proposed mean quantization-based approach can reduce the variance of modification quantities caused by incidental and malicious

distortions, respectively. From Eqs. (11) and (12), it is obvious that when the number of coefficients, n, used to encode a watermark value is increased, the probability of watermark errors will be decreased. In order to increase the credibility of a fragile watermark for image authentication, the watermark errors caused by an instance of malicious tampering should be maximized, and those caused by an incidental distortion should be minimized. Under these circumstances, if the above mentioned n is too small, then the embedded watermark will be too fragile to tolerate incidental manipulation. On the other hand, if n is too large, the embedded watermark will be too robust to detect malicious tampering. Therefore, the number n used to encode a watermark value is a key factor in balancing the tradeoff between robustness and fragility. We will conduct an analysis with regard to this tradeoff in Sec. 2.3.

2.3 Choosing an Optimal n for Mean Quantization

In this section, we shall provide a formal proof to show that the proposed mean quantization-based fragile watermarking scheme is superior to the conventional quantization-based approach [5]. In [5], Kundur and Hatzinakos assumed that the distributions of modification quantities caused by an instance of malicious tampering and an incidental distortion are both Gaussian distributions. They also mentioned that the major difference between the two distributions is that the variance of the distribution caused by an instance of malicious tampering is larger than that caused by an incidental distortion. Since the operation of a mean quantization will make the variance of all distributions smaller, in this section, we shall devise a systematic way to determine an optimal number of coefficients that should be adopted in the mean quantization process.

Given a distribution of tampering $N(0, \sigma^2)$, and a quantization interval size q, the probability of watermark errors computed using a quantization-based approach is

$$E = 2\sum_{j=0}^{\infty} \int_{(2j+\frac{1}{2})q}^{(2j+\frac{3}{2})q} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x}{\sigma})^2} dx \tag{13}$$

$$= 2\sum_{j=0}^{\infty} \lim_{r\to\infty} \sum_{k=0}^{r} \frac{1}{\sqrt{2\pi}\sigma} \frac{q}{r} e^{-\frac{1}{2}(\frac{(2j+\frac{1}{2}+\frac{k+\frac{1}{2}}{r})\cdot q}{\sigma})^2}. \tag{14}$$

Since Eq. (13) is not in a discrete format, we use the form shown in Eq. (14) instead to compute the probability of watermark errors with respect to σ and q because σ and q are

two important factors which will influence the results. Fig. 2 shows the relations between the variance of tampering σ, the size of a quantization interval q, and the probability of watermark errors E. The X-axis and Y-axis in Fig. 2 represent $\frac{\sigma}{q}$ and E, respectively. However, owing to the fact that the maximum q is bounded by the characteristics of the human visual system [13], the probability of watermark errors cannot be arbitrarily reduced. On the other hand, for a fixed q, a larger σ value will lead to a larger E value. If the variance σ can be reduced, then the probability of watermark errors caused by a malicious distortion or an incidental distortion will be reduced.

From Fig. 2, we know that the probability of watermark errors is a function of $\frac{\sigma}{q}$. Therefore, we can represent the probability by means of $f(t)$, where $t = \frac{\sigma}{q}$ denotes the ratio between σ and q. In general, the range of t can be divided into three zones. A robust zone means the value of $f(t)$ is very close to 0. On the other hand, a fragile zone means the value of $f(t)$ is close to 0.5. There is a transition zone in between, which we call a semi-fragile zone. The value of $f(t)$ changes from 0 to 0.5 within the transition zone. Therefore, there are two critical points that need to be determined. One is the point at which the value of $f(t)$ changes from zero to non-zero. The other is the point where $f(t)$ starts to saturate at 0.5. We call these points t_1 and t_2, respectively. Furthermore, since the semi-fragile zone is an ambiguous zone, we would like to make it as small as possible. As a consequence, the values of t_1 and t_2 can be determined by solving the following constraint optimization problem:

$$g(t_1, t_2) = \alpha \cdot |f(t_1)| + \beta \cdot |f(t_2) - 0.5| + \gamma \cdot \frac{t_2}{t_1}, \tag{15}$$

where $g(\cdot)$ is a cost function to be minimized. The first term and the second term on the right hand side of Eq. (15) are the constraints that force the values of $f(t_1)$ and $f(t_2)$ to be as close as possible to 0 and 0.5, respectively. As to the $\frac{t_2}{t_1}$ term, it is used to keep the size of the transition zone as small as possible. In our experiments, we set the values of the leading coefficients α, β, and γ to be 1000, 1000, and 1, respectively. Based on the above setting, t_1 and t_2 can be determined. They are 0.15 and 1.15, respectively.

Let the distribution of an instance of malicious tampering and an incidental distortion be denoted as $N(0, \sigma_I^2)$ and $N(0, \sigma_M^2)$, respectively. From Lin and Chang's [4] previous experience, we know that σ_M is larger than σ_I, and they have a relation $\sigma_M = c\sigma_I$ with

$c > 1$. Let n denote the number of coefficients used in calculating a mean coefficient (Eq. (3)); the new distributions of modification quantities caused by a malicious tampering and an incidental distortion become $N(0, (\sigma_I^*)^2)$ and $N(0, (\sigma_M^*)^2)$, respectively, where $\sigma_I^* = \frac{1}{\sqrt{n}}\sigma_I$ and $\sigma_M^* = \frac{1}{\sqrt{n}}\sigma_M = \frac{c}{\sqrt{n}}\sigma_I$. Let the size of a quantization interval, q, be determined according to the human visual system [13]. This means that q is fixed with respect to the human visual system. The question is how to determine the best n such that the probability of watermark errors caused by an instance of malicious tampering will be maximized and that caused by an incidental distortion will be minimized. If the relation $\frac{\sigma_M^*}{q} \geq t_2$ holds (as described in the previous paragraph), then the probability of watermark errors caused by a malicious tampering will definitely be maximized. Therefore, we have

$$\frac{\sigma_M^*}{q} \geq t_2 \Rightarrow \frac{c\sigma_I}{\sqrt{n}q} \geq t_2 \Rightarrow \frac{c\sigma_I}{t_2 q} \geq \sqrt{n}. \tag{16}$$

Similarly, if the relation $\frac{\sigma_I^*}{q} \leq t_1$ holds, then the probability of watermark errors caused by an incidental distortion will be minimized. That is,

$$\frac{\sigma_I^*}{q} \leq t_1 \Rightarrow \frac{\sigma_I}{\sqrt{n}q} \leq t_1 \Rightarrow \frac{\sigma_I}{t_1 q} \leq \sqrt{n}. \tag{17}$$

Combining Eqs.(16) and (17), we obtain

$$\frac{\sigma_I}{t_1 q} \leq \sqrt{n} \leq \frac{c\sigma_I}{t_2 q}. \tag{18}$$

It is obvious that the minimum n that satisfies Eq. (18) is an n_1 which makes $\frac{\sigma_I}{t_1 q} = \sqrt{n_1}$. Therefore, we have

$$n_1 = (\frac{\sigma_I}{t_1 q})^2. \tag{19}$$

n_1 will lead to the minimum probability of watermark errors caused by an incidental distortion. On the other hand, the maximum n that will satisfy Eq. (18) is an n_2 which makes $\sqrt{n_2} = \frac{c\sigma_I}{t_2 q}$. Thus, we have

$$n_2 = (\frac{c\sigma_I}{t_2 q})^2. \tag{20}$$

n_2 will lead to the maximum probability of watermark errors caused by an instance of malicious tampering. In order to find the best n that will bypass an incidental distortion while detecting an instance of malicious tampering, we should select an n which is bounded by n_1 and n_2, i.e., $n \in [n_1, n_2]$.

In what follows, we shall conduct a theoretical analysis to determine an ideal n. From Fig. 2, we know that the probability of watermark errors is a function of $\frac{\sigma}{q}$. Since q is a constant when a specific human visual model [13] is adopted, t is proportional to the value of σ. Let the probabilities of watermark errors caused by an incidental distortion and a malicious tampering be $f(\hat{t}_1)$ and $f(\hat{t}_2)$, respectively, where $\hat{t}_1 = \frac{\sigma_I}{q}$ and $\hat{t}_2 = \frac{\sigma_M}{q}$. Because $\sigma_M = c\sigma_I$, we have

$$\hat{t}_2 = \frac{\sigma_M}{q} = \frac{c\sigma_I}{q} = c\hat{t}_1. \tag{21}$$

When a mean quantization operation covering n coefficients is applied, σ_I and σ_M will be updated to $\frac{\sigma_I}{\sqrt{n}}$ and $\frac{\sigma_M}{\sqrt{n}}$, respectively. In order to obtain the best mean quantization result, the difference between the watermark error caused by an instance of malicious tampering and an incidental distortion should be maximized. That is, $f(\hat{t}_2) - f(\hat{t}_1)$ should be maximized. The physical meaning of maximizing $f(\hat{t}_2) - f(\hat{t}_1)$ is to make the watermark errors caused by an instance of malicious tampering as large as possible and those caused by an incidental one as small as possible. Using the optimization scheme, one can decide on an optimal value of n such that $f(\hat{t}_2) - f(\hat{t}_1)$ is maximized. The simplest way to calculate the ideal n is to compute the values of $f(\frac{c\sigma_I}{\sqrt{n}q}) - f(\frac{\sigma_I}{\sqrt{n}q})$ using various integers $n \in [n_1, n_2]$. The integer that leads to the largest outcome is the ideal n.

To obtain an ideal n in the interval $[n_1, n_2]$, Eq. (18) should hold. However, if $c \leq \frac{t_2}{t_1}$, then Eq. (18) doesn't hold. Under these circumstances, the relation between n_1 and n_2 becomes

$$n_2 = (\frac{c\sigma_I}{t_2 q})^2 \leq (\frac{t_2}{t_1}\frac{\sigma_I}{t_2 q})^2 = (\frac{\sigma_I}{t_1 q})^2 = n_1. \tag{22}$$

Given two values n_A and n_B, if $n_1 > n_A > n_B$, then

$$\frac{\sigma_I}{q\sqrt{n_1}} < \frac{\sigma_I}{q\sqrt{n_A}} < \frac{\sigma_I}{q\sqrt{n_B}} \Rightarrow f(\frac{\sigma_I}{q\sqrt{n_1}}) < f(\frac{\sigma_I}{q\sqrt{n_A}}) < f(\frac{\sigma_I}{q\sqrt{n_B}}). \tag{23}$$

Therefore, if we choose an n which is as close to n_1 as possible, its corresponding probability of watermark error will be smaller. Similarly, given two values n_C and n_D, if $n_C > n_D > n_2$, then

$$\frac{c\sigma_I}{q\sqrt{n_C}} < \frac{c\sigma_I}{q\sqrt{n_D}} < \frac{c\sigma_I}{q\sqrt{n_2}} \Rightarrow f(\frac{c\sigma_I}{q\sqrt{n_C}}) < f(c\frac{\sigma_I}{q\sqrt{n_D}}) < f(\frac{c\sigma_I}{q\sqrt{n_2}}). \tag{24}$$

Under these circumstances, if we choose an n which is as close to n_2 as possible, then its corresponding probabilities of watermark error will be larger. The optimization problem

now is that of finding an ideal $n \in [n_2, n_1]$ such that $f(\frac{c\sigma_I}{\sqrt{n^* q}}) - f(\frac{\sigma_I}{\sqrt{n^* q}})$ is maximized. Once the ideal n is determined, the watermark errors E corresponding to a malicious tampering and an incidental distortion are $f(\frac{c\sigma_I}{\sqrt{nq}})$ and $f(\frac{\sigma_I}{\sqrt{nq}})$, respectively. Since making a comparison between the mean-quantization approach and the conventional quantization-based approach requires real data, we did this in an experiment.

3 Tampered Area Estimation using Information Fusion

For image authentication, the wavelet-based fragile watermarking method proposed in [5] only shows the tampering detection results at multiple scales. In this section, we will present an information fusion technique which can be used to integrate the results obtained at multiple scales. In addition, the proposed technique has the merit of suppressing sparse watermark errors spread out over the subimages at multiple scales.

In the following, we first define the function T as

$$T(z) = \begin{cases} 1, & \text{if } Q(z,q) \neq w \text{ (watermark error)} \\ 0, & \text{if } Q(z,q) = w \text{ (no watermark error)} . \end{cases} \tag{25}$$

The T function is used to indicate whether or not a watermark error occurs on a wavelet coefficient z, where q is the size of a quantization interval and w is the target watermark symbol. To quantitatively calculate the degree of tampering, the chessboard distance [16] is used, where

$$d_{chess}((i_1, j_1), (i_2, j_2)) = \max\{|i_1 - i_2|, |j_1 - j_2|\}. \tag{26}$$

This metric can measure the spatial distance between two coefficients which have been tampered with and are located at (i_1, j_1) and (i_2, j_2). Next, the density of a coefficient $x_l(i, j)$ at scale l is defined as

$$D(x_l(i,j)) = \begin{cases} \min\limits_{x_l(i',j') \in R_l} \{d_{chess}((i,j),(i',j'))\}, & \text{if } T(x_l(i,j)) = 1 \\ 0, & \text{if } T(x_l(i,j)) = 0, \end{cases} \tag{27}$$

where $R_l = \{x_l(i^*, j^*) | T(x_l(i^*, j^*)) = 1 \text{ and } (i^*, j^*) \neq (i, j)\}$. The altered coefficient with $D(\cdot) = 1$ will form a *dense* region while the coefficient with $D(\cdot) > 1$ will form a *sparse* region. Let the probability of watermark errors for an altered coefficient x and all its neighboring

eight neighbors be denoted as $p(x)$; the probability that x is dense will be $(1.0 - (1.0 - p(x))^8)$. The relation between the probability of forming a dense region and the probability of watermark errors is shown in Fig. 3. From Fig. 3, we observe that if the probability of watermark errors is greater than 0.25, then the probability that this coefficient is dense is 90%. To make the probability that a coefficient is dense as small as 10%, the probability of watermark errors should be as small as 0.02. For these reasons, an ideal n should be chosen so as to minimize the probability of watermark errors caused by incidental distortion and maximize the probability of watermark errors caused by malicious tampering. If the above mentioned concept can be realized, then the watermark errors caused by a malicious or incidental distortion should have high probability of being dense or sparse, respectively. Thus, we can locate the area that has been maliciously tampered with by grouping those areas with dense responses.

Let N_l^{total}, N_l^{tamper}, N_l^{dense}, and N_l^{sparse} denote the total number of coefficients, the total number of altered coefficients, the number of altered coefficients which are dense and the number of altered coefficients which are sparse at scale l, respectively, where $N_l^{tamper} = N_l^{dense} + N_l^{sparse}$. Furthermore, let the *tampering ratio* at scale l be defined as

$$TR_l = N_l^{tamper}/N_l^{total}. \tag{28}$$

This ratio is used to measure the degree of tampering. During the process of information fusion, the following rules are applied to judge whether a modification is malicious or incidental:

Rule 1: If $TR_l = 0$ at every scale, then the target image was neither maliciously tampered with nor incidentally distorted.

Rule 2: If $TR_l = 0$ for some scale l, then the target image only encountered incidental distortions.

Rule 3: Assume l^* represents the scale where $TR_{l^*} = \min_l\{TR_l\}$. If $TR_{l^*} > 0$ and $N_{l^*}^{dense} < \alpha \times N_{l^*}^{tamper} (0.5 \leq \alpha \leq 1.0)$, then the target image only encountered incidental distortions.

Rule 4: If $N_l^{dense} = N_l^{tamper}$ at every scale l, then the target image was only maliciously tampered with.

Rule 5: If none of the above rules fits, then the target image was both maliciously tampered with and incidentally distorted.

The above mentioned rules should be applied sequentially. For a target image, if one of the first 3 rules is matched, then the image is considered to be free of tampering. Otherwise, the image is regarded as having been tampered with and we have to further compute the size of the affected area.

In order to retain the responses caused by malicious tampering (dense responses) and remove those caused by incidental modifications (sparse responses) at scale l, the watermark errors are transformed into a Tamper Response Map (TRM) as follows:

$$TRM_l(i,j) = \sum_{i^*,j^*} TRF(x_l(i^*,j^*), x_l(i,j)), \tag{29}$$

where $TRF(\cdot)$ is a *tamper response function* (TRF). The TRF of a wavelet coefficient, $x_l(i^*,j^*)$, is defined as

$$TRF(x_l(i^*,j^*), x_l(i,j)) = \begin{cases} \frac{d_{chess}((i^*,j^*),(i,j))}{A_l(i^*,j^*)}, & \text{if } d_{chess}((i^*,j^*),(i,j)) \leq (D(x_l(i^*,j^*))+1) \\ 0, & \text{otherwise,} \end{cases} \tag{30}$$

where $A_l(i^*,j^*) = \sum_{k=1}^{D(x_l(i^*,j^*)+1)} k^2$ is a normalization factor and $D(x_l(i^*,j^*))$ is the density of the coefficient $x_l(i^*,j^*)$ as defined in Eq. (27). The TRF is used to point out where the dense watermark errors are located. In order to distinguish the degree of importance of watermark errors at each scale, a weighting factor associated with each scale l is defined as

$$WGT_l = (N_l^{dense}/N_l^{tamper})^2. \tag{31}$$

The tamper response maps at all scales are then weighted by their corresponding WGT's and then integrated to form the final tamper response map, i.e.,

$$TRM^{final}(i,j) = \sum_l WGT_l \times TRM_l(i,j). \tag{32}$$

Let N_l^T and p_l^T denote the number of coefficients and the probability of watermark errors, respectively, in an area that has been maliciously tampered with at scale l. Because the

watermark errors in such an area are dense, we have the relation $N_l^{dense} = N_l^T \times p_l^T$. Thus, an estimation of the ratio of a *tampered area* (TA) with respect to the entire image is defined as

$$\begin{aligned} TA &= \frac{N_{l^*}^T}{N_{l^*}^{total}} = \frac{N_{l^*}^{dense}}{p_{l^*}^T \cdot N_{l^*}^{total}} = \frac{1}{p_{l^*}^T} \cdot \frac{N_{l^*}^{denes}}{N_{l^*}^{total}} \\ &= \beta \times (N_{l^*}^{dense}/N_{l^*}^{total}), \end{aligned} \tag{33}$$

where $l^* = arg\min_l\{TR_l\}$ and $\beta = \frac{1}{p_{l^*}^T}$. From Fig. 3, it is clear that when the probability of watermark errors is close to 0.5, the probability that the corresponding coefficient is dense is large. Let $p_{l^*}^T = 0.5$; then β is 2.0. Let the values of $TRM^{final}(i,j)$ be sorted into $TRM^{final}(i_k, j_k)$ in descending order; then the set of pixels, $\{TRM^{final}(i_{k^*}, j_{k^*}) | 1 \leq k^* \leq (TA \times N_I)\}$, will be marked as a set of maliciously attacked areas, where $k, k^* = 1, \ldots, N_I$ and N_I is the size of the image.

Basically, the above mentioned decision rules can be used to detect most of the areas that have been maliciously tampered with. However, when such an area is very small, it is difficult to distinguish it from an area that has encountered incidental distortion. This is because an instance of malicious tampering and an incidental distortion both generate watermark errors of the sparse type. On the other hand, if the probability of watermark errors caused by an incidental distortion is very small (zero is the ideal case), then one can claim that the detected watermark errors were completely obtained from an area that was maliciously tampered with.

4 Experimental Results

To demonstrate the power of our image authentication system, we will first introduce the experimental setup in Sec. 4.1 and give the detection results obtained under various incidental distortions in Sec. 4.2. In Sec. 4.3, we will present some experimental results obtained by applying both malicious tampering and incidental manipulation. A set of test images processed by combining different incidental and malicious manipulations was used to estimate the area that was maliciously tampered with. A comparison based on the performance of the conventional quantization-based approach and our approach will be made in Sec. 4.4.

4.1 Experimental Setup

The images used in the experiment were of size 512×512 with 256 graylevels. Fig. 4 is an example showing how a watermarked image is tampered with, including the original image, the watermarked image, the altered area, and the final altered image, respectively. The PSNR of the watermarked image shown in Fig. 4(b) was 35.91 dB. Two peppers (Fig. 4(c)) were added as shown in Fig. 4(b) and formed an image that had been tampered with, as shown in Fig. 4(d). This set of data was used to test the performance of our approach in the subsequent experiments.

The set of incidental attacks used in the experiments included JPEG compression, blurring, and sharpening. The mask sizes used in the blurring operation were 3×3, 5×5 and 7×7, respectively. The quality factors adopted for JPEG compression were from 10% to 90%, and the parameters used in the sharpening operation were from 10% to 50%. In the experiments, the watermark sequence was embedded in the LH subband at each scale of a wavelet transformed image. As to the determination of the best n at every scale of a wavelet transform, this can be calculated by scanning the interval $[n_1, n_2]$ for large c ($c > 7.67$) or by scanning the interval $[n_2, n_1]$ for small c ($1 < c \leq 7.67$), where n_1 and n_2 are computed using Eqs. (19) and (20), respectively. We use $\mathbf{n} = (n_1, n_2, \ldots, n_s)$ to represent the number of coefficients used at every scale in the mean quantization process, where n_i is the number of coefficients used to derive a mean at scale i and s is the total number of scales used. From Eqs. (19) and (20), the best set of n could be theoretically determined as (9, 16, 16, 11) when the total number of scales was chosen to be 4. On the other hand, for the purpose of easy implementation, $3 \times 3, 4 \times 4, 4 \times 4$ and 4×3 were used as the block sizes at scales 1, 2, 3, and 4, respectively.

4.2 Detection Results Obtained by Applying Incidental Distortions Only

In this section, we shall check whether our approach could tolerate a number of incidental operations with different degrees of alteration. Fig. 5 shows a set of test images which was used in the experiments. The incidental operations which were applied to the set of

test images included JPEG compression, blurring, and sharpening. Table 5 lists the results obtained in this experiment. A "$\surd$" symbol shown in a box indicates that our system considered the operation to be an incidental one. On the other hand, an "$\times$" symbol indicates that our system mistakenly considered the operation to be a malicious one. From the table, it is obvious that our system could successfully bypass almost all the JPEG compressed images up to quality factor of 20%. As for the sharpening operation, our system could successfully tolerate most of the sharpened images up to a 50% sharpening factor. However, in the case of the blurring operation, our system only worked well when the window size was 3×3.

4.3 Detection Results Obtained by Applying Malicious Tampering and Incidental Manipulation Simultaneously

In this section, we shall give some experimental results obtained by applying malicious tampering and an incidental manipulation simultaneously. The objective of these experiments was to check whether our approach could successfully tolerate an incidental manipulation while detecting a malicious attack. Fig. 6(a) is a pepper image that was tampered with by performing 60% (quality factor) JPEG compression, followed by two-pepper replacement. The detected watermark errors at scales 1 to 4 are shown in Figs. 6(b)-(e), respectively. It can be seen that the watermark errors caused by the JPEG compression are much fewer than those caused by malicious tampering. The detected watermark errors were then converted into the tamper response maps shown in Figs. 6(g)-(j). It is obvious that the coefficients having the sparse type all had weak responses in the tamper response map at each scale. On the other hand, the areas that corresponded to the regions that were maliciously tampered with all had strong responses in the tamper response maps. After performing information fusion, the final detected altered areas were those shown in Fig. 6(f). It is apparent that the maliciously modified regions were detected correctly.

Fig. 7 shows another 21 detection results obtained using the proposed mean quantization-based fragile watermarking technique. The symbols "T," "B," "J," and "S" denote malicious tampering, blurring, JPEG compression and sharpening, respectively. The number following each symbol is the parameter used in an incidental distortion. For example, "T+B 3×3"

shown in Fig. 7(b) means an image was maliciously tampered with and then blurred with a mask of size 3×3. In the whole set of experiments, the resolution of the wavelet transform was taken up to 4 scales. The optimal number of coefficients used to perform mean quantization at each scale was 9, 16, 16, 12, respectively. That is, $\mathbf{n} = (9, 16, 16, 12)$. From Fig. 7, it is apparent that our approach did work well in most cases, especially in tolerating incidental manipulation like JPEG. Fig. 7(l) indicates that when the quality factor reached 20%, the detection result was still good. In the case of a combined attack including 7×7 blurring (Fig. 7(d)), the result was bad. But when the window size were 3×3 and 5×5, the detection results were good. In the case of a combined attack involving sharpening, the results were good when the sharpening factor was smaller than 50%. When the sharpening factor reached or exceeded 70%, the detected results were completely wrong.

4.4 Comparison with the Conventional Quantization-based Approach

In this section we shall compare our approach with the conventional approach. The maliciously attacked image shown in Fig. 4(d) subjected to JPEG compression with a quality factor 60% was used as the test image. The watermark errors (at scales 1 to 4) obtained by applying the conventional quantization-based approach [5] and the proposed mean quantization-based approach with $\mathbf{n} = (9, 16, 16, 12)$ are shown in Figs. 8(a) and 8(b), respectively. It is obvious that the results obtained by applying our approach are better than those obtained by applying the conventional approach.

5 Conclusion

In this paper, a mean quantization-based fragile watermarking approach has been proposed for image authentication. Our system is able to maximize the probability of watermark errors caused by an instance of malicious tampering and minimize the probability of watermark errors caused by an incidental distortion. In addition, an information fusion procedure which can integrate detection responses at each scale in the wavelet domain has been presented

which can be used to estimate the area that has been maliciously tampered with.

Our future work will proceed in two directions. First, the capability of our image authentication system in distinguishing malicious tampering and incidental distortion will be further improved so that incidental distortion with large variance of modification, such as histogram equalization, can also be tolerated. Secondly, we will extend the mean quantization-based watermarking approach to multipurpose watermarking [6], so that an embedded watermark can be used in multiple applications.

References

[1] S. Bhattacharjee and M. Kutter, "Compression tolerant image authentication," in *IEEE Inter. Conf. on Image Processing*, vol. 1, (Chicago, USA), pp. 4–7, October 1998.

[2] J. Dittmann, A. Steinmetz, and R. Steinmetz, "Content-based digital signature for motion pictures authentication and content-fragile watermarking," in *IEEE Inter. Conf. Multimedia Computing and Systems*, vol. II, (Italy), 1999.

[3] G. L. Friedman, "The trustworthy digital camera: Restoring credibility to the photographic image," *IEEE Trans. on Consumer Electronics*, vol. 39, pp. 905–910, October 1993.

[4] C.-Y. Lin and S.-F. Chang, "A robust image authentication method surving JPEG lossy compression," in *SPIE Inter. Conf. on Storage and Retrieval of Image/Video Database*, vol. 3312, (San Jose, USA), January 1998. EI'98.

[5] D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," *Proceedings of the IEEE*, vol. 87, pp. 1167–1180, July 1999.

[6] C. S. Lu, H. Y. M. Liao, and L. H. Chen, "Multipurpose audio watermarking," in *15th Inter. Conf. on Pattern Recognition*, (Spain), 2000.

[7] C. S. Lu, H. Y. M. Liao, and C. J. Sze, "Combined watermarking for image authentication and protection," in *Proceeding of the 1st IEEE Inter. Conf. on Multimedia and Expo*, (USA), 2000.

[8] R. B. Wolfgang and E. J. Delp, "Fragile watermarking using the VM2D watermark," in *SPIE Inter. Conf. on Security and Watermarking of Multimedia Contents*, vol. 3657, (San Jose, CA), pp. 204–213, January 1999. EI'99.

[9] P. W. Wong, "A public key watermark for image verification and authentication," in *IEEE Inter. Conf. on Image Processing*, (Chicago, USA), October 1998.

[10] M. Wu and B. Liu, "Watermarking for image authentication," in *IEEE Inter. Conf. on Image Processing*, (Chicago, USA), October 1998.

[11] M. Yeung and F. Mintzer, "An invisible watermarking technique for image verification," in *IEEE Inter. Conf. on Image Processing*, (Santa Barbara, USA), October 1997.

[12] B. Zhu, M. D. Swanson, and A. H. Tewfik, "Transparent robust authentication and distoration measurement technique for images," in *The 7th IEEE Digital Signal Processing Workshop*, pp. 45–48, September 1996.

[13] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Trans. on Image Processing*, vol. 6, pp. 1164–1175, August 1997.

[14] C. S. Lu, H. Y. M. Liao, S. K. Huang, and C. J. Sze, "Cocktail watermarking on images," in *Proceeding of the 3rd Inter. Workshop on Information Hiding*, LNCS 1768, (Dresden, Germany), pp. 333–347, 1999.

[15] C. S. Lu, H. Y. M. Liao, S. K. Huang, and C. J. Sze, "Highly robust image watermarking using complementary modulations," in *Proceeding of the 2nd Inter. Information Security Workshop*, LNCS 1729, (Malaysia), pp. 136–153, 1999.

[16] R. Jain, R. Kasturi, and B. G. Schunck, *Machine Vision*. MacGraw-Hill, Inc., 1995.

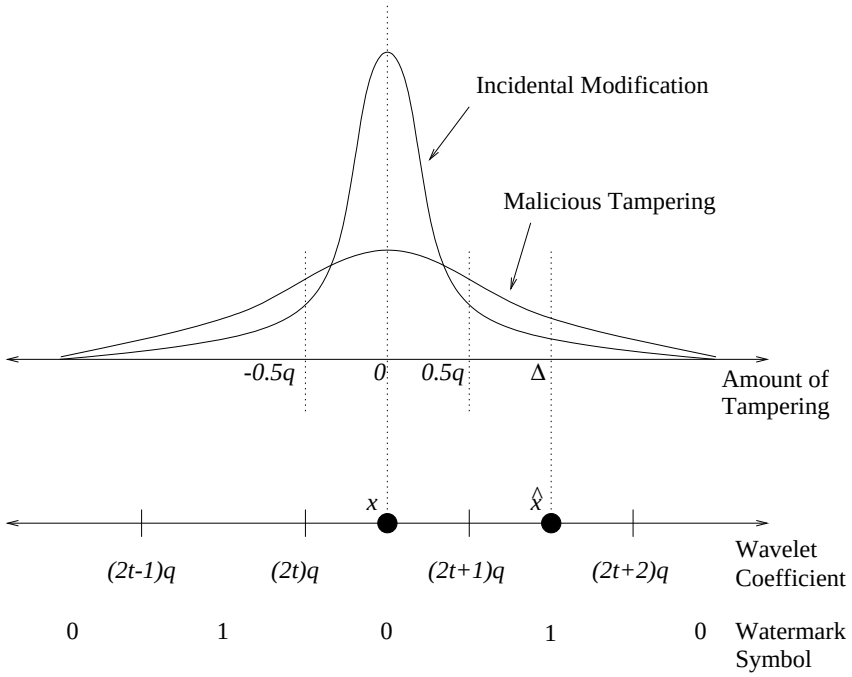


Figure 1: The statistical distribution of incidental modification and malicious tampering on wavelet coefficients (top) and an illustration of quantization-based watermarking (bottom).

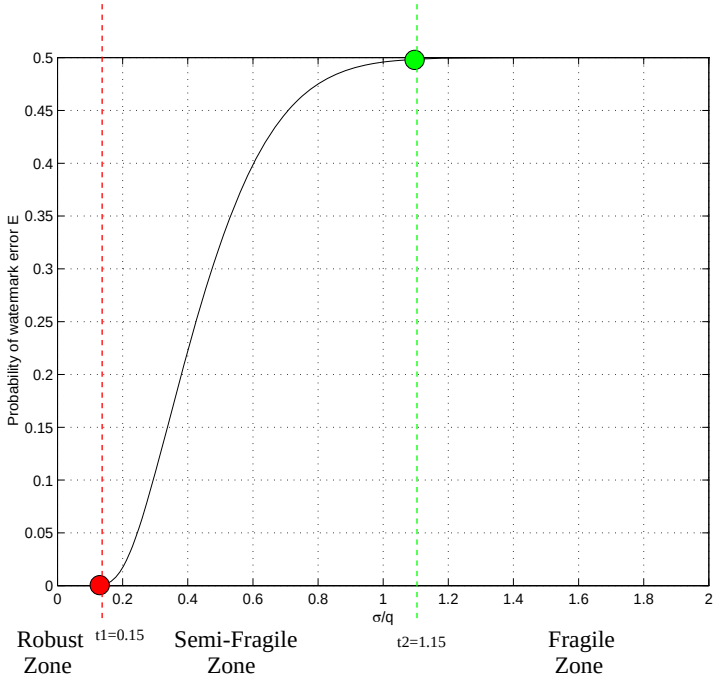


Figure 2: The relation between the variance of tampering σ, the quantization interval's size q and the probability of watermark errors.

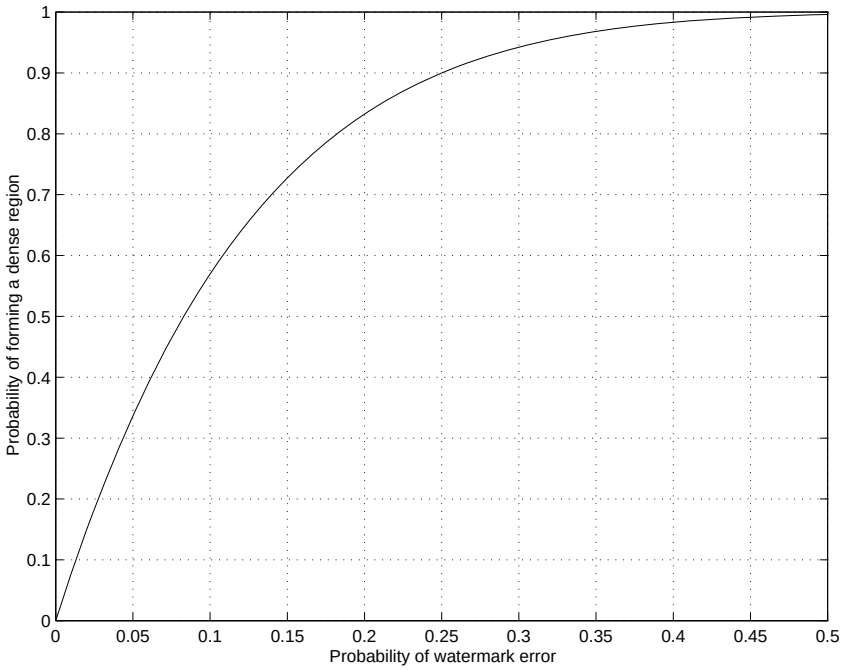


Figure 3: The relation between the probability of watermark errors and the probability of forming a dense region.

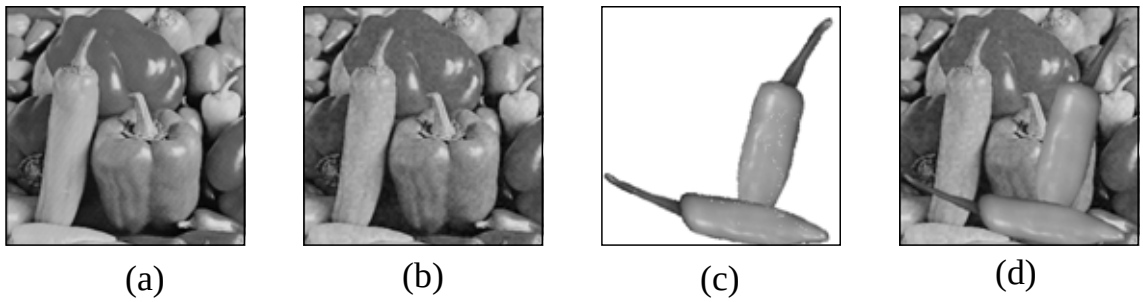
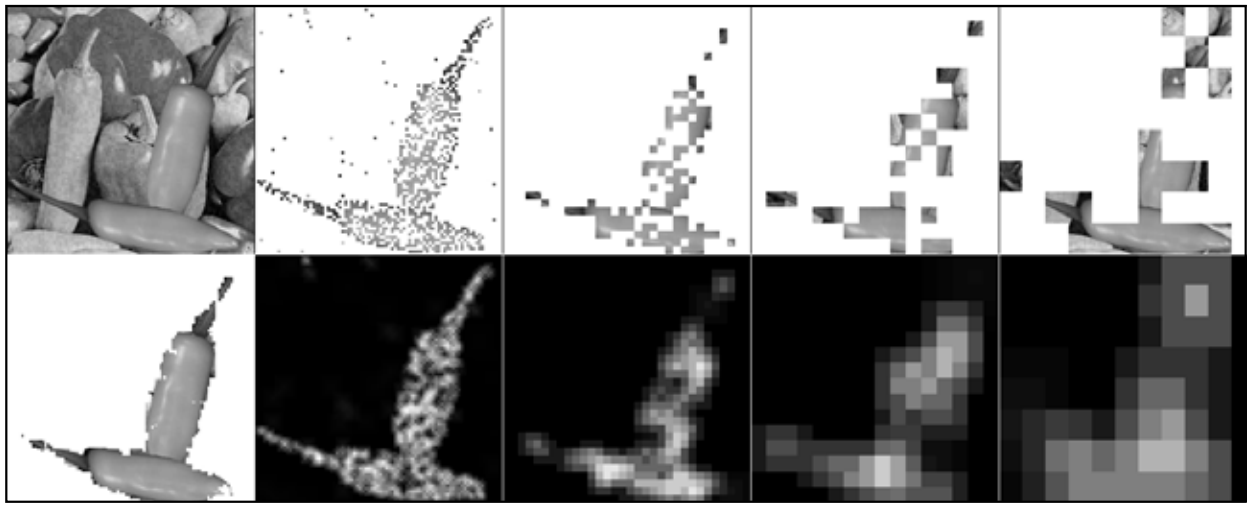


Figure 4: An example showing malicious tampering by means of object replacement: (a) original image; (b) watermarked image; (c) objects used for tampering; (d) modified watermarked image.

Figure 5: A set of test images.

Table 1: Tampering detection for a set of incidentally manipulated test images. A " $\surd$ " symbol indicates that our system treats the operation as an incidental distortion while an " $\times$ " symbol indicates that the operation is malicious tampering.

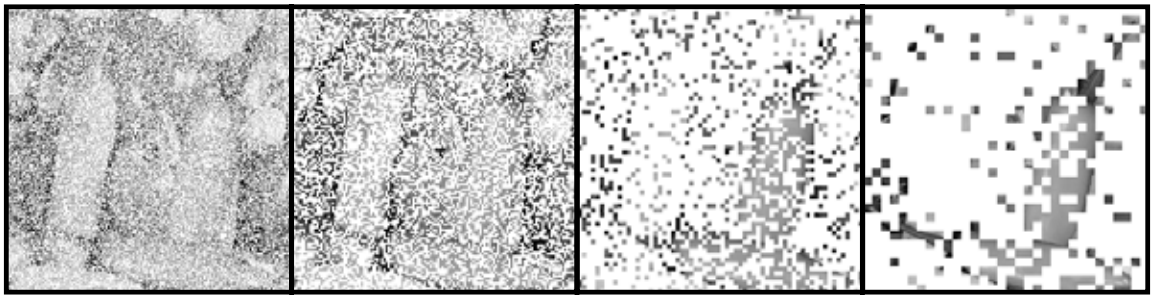
Image	Image Label											
Operation	A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12
Blur (3×3)	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\times$	$\surd$	$\times$	$\surd$	$\surd$	$\surd$	$\surd$
Blur (5×5)	$\times$	$\times$	$\times$	$\surd$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\surd$	$\surd$
Blur (7×7)	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
Sharpen (F=10%)	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
Sharpen (F=20%)	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
Sharpen (F=30%)	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
Sharpen (F=40%)	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\times$	$\surd$	$\surd$	$\surd$	$\surd$
Sharpen (F=50%)	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\times$	$\surd$	$\times$	$\surd$	$\times$	$\times$	$\surd$
Sharpen (F=60%)	$\surd$	$\times$	$\surd$	$\times$	$\times$	$\times$	$\times$	$\times$	$\surd$	$\times$	$\times$	$\surd$
Sharpen (F=70%)	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\surd$	$\times$	$\times$	$\surd$
Sharpen (F=80%)	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\surd$
Sharpen (F=90%)	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
JPEG (QF=90%)	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
JPEG (QF=80%)	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
JPEG (QF=70%)	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
JPEG (QF=60%)	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
JPEG (QF=50%)	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
JPEG (QF=40%)	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
JPEG (QF=30%)	$\surd$	$\surd$	$\surd$	$\surd$	$\times$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
JPEG (QF=20%)	$\surd$	$\surd$	$\surd$	$\surd$	$\times$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
JPEG (QF=10%)	$\times$	$\times$	$\times$	$\surd$	$\times$	$\surd$	$\surd$	$\times$	$\times$	$\times$	$\times$	$\times$



(a)	(b)	(c)	(d)	(e)
(f)	(g)	(h)	(i)	(j)

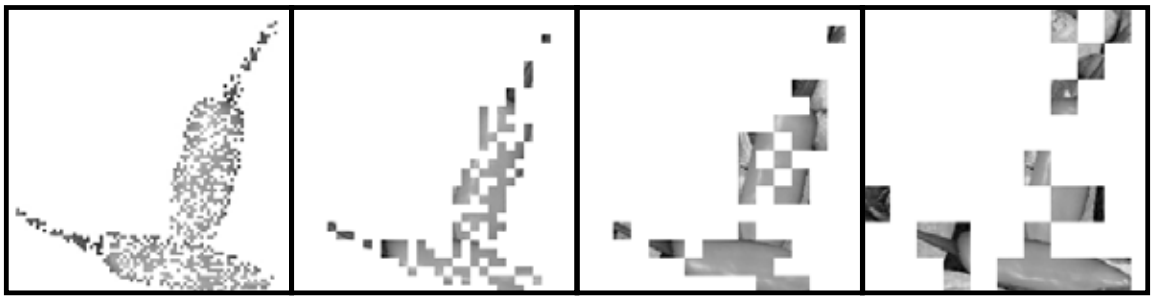
Figure 6: Tampering with object placement and JPEG compression: (a) is a tampered image with two objects added; (b)-(e) are the detected watermark errors from scales 1 to 4, respectively; (g)-(j) are the tamper response maps derived from scales 1 to 4, respectively; (f) is the final result after performing information fusion.

Figure 7: A set of detection results obtained by applying our mean quantization-based method. (a) is the detection result when the attack is object placement only; (b)-(d) show the detection results when the attack is object placement followed by blurring with mask sizes of 3×3, 5×5, and 7×7, respectively; (e)-(m) show the detected results when the attack is object placement followed by JPEG compression with a quality factor ranging from 90% to 10%; (n)-(v) show the detection results when the attack is object placement followed by sharpening with a sharpening factor ranging from 10% to 90%, respectively.



Scale 1 Scale 2 Scale 3 Scale 4

(a) Kundur and Hatzinakos' approach.



Scale 1 Scale 2 Scale 3 Scale 4

(b) Mean quantization approach with **n** =(9, 16, 16, 12).

Figure 8: Comparison of detected watermark errors obtained using the conventional quantization-based approach and the mean quantization-based approach with $\mathbf{n} = (9, 16, 16, 12)$.