Structural Digital Signature for Image Authentication: An Incidental Distortion Resistant Scheme[*]

Chun-Shien Lu and Hong-Yuan Mark Liao

Institute of Information Science,
Academia Sinica, Taipei, Taiwan.
E-mail: {lcs, liao}@iis.sinica.edu.tw

Abstract

The existing digital data verification methods are able to detect regions that have been tampered with, but are too fragile to resist incidental manipulations. This paper proposes a new digital signature scheme which makes use of an image's contents (in the wavelet transform domain) to construct a structural *digital signature (SDS) for image authentication. The characteristic of the SDS is that it can tolerate content-preserving modifications while detecting content-changing modifications. Many incidental manipulations, which were detected as malicious modifications in the previous digital signature verification or fragile watermarking schemes, can be bypassed in the proposed scheme. Performance analysis is conducted and experimental results show that the new scheme is indeed superb for image authentication.*

keywords: Digital signature, Wavelet transform, Authentication, Fragility, Robustness.

[*] The preliminary version of this paper will be published in [14] (http://smart.iis.sinica.edu.tw/∼lcs).

1 Introduction

Because of the easy-to-copy nature of digitized media, it is very easy for one to tamper with digital data without leaving any clues. Under these circumstances, integrity verification has become an important issue in the digital world. Conventionally, the methods used for media verification can be classified into two kinds: digital signature-based [2, 3, 5, 7, 8] and watermark-based [4, 6, 9, 12, 17, 18, 19, 20, 21, 23]. A digital signature is a set of features extracted from a media, and these features are stored as a file, which will be used later for authentication. A very important characteristic of a digital signature is that it sufficiently represents the content of the original media. Watermarking, on the other hand, is a media authentication/protection technique that embeds invisible (or inaudible) information into a media. For content authentication, the embedded watermark can be extracted and used for verification purposes. The major difference between a watermark and a digital signature is that the embedding process of the former requires the content of a media to change. However, both the watermark-based approach and the digital signature-based approach are expected to be sensitive to any malicious modification applied to the media. For an incidental modification such as *JPEG* compression or blurring, a good authentication system should be able to tolerate it. Unfortunately, most of the existing media authentication systems, though they can detect malicious tampering successfully, are vulnerable to incidental modifications. The main reason for the above mentioned problem is that the existing methods do not consider carefully the tradeoff between robustness and fragility. In the whole course of this study, we shall focus our discussion on the image authentication system.

The underlying techniques used to implement the digital signature-based or watermark-based approaches can be roughly classified into quantization-based [6, 12, 22], feature point-based [2, 3], and relation-based [7, 8]. As to a quantization-based approach, Kundur and Hatzinakos [6] designed a quantization technique to encode a watermark so that the hidden watermark is more/less sensitive to modifications at high/low frequency in the wavelet domain. Usually, over-sensitivity may occur at the small-to-medium scale while under-sensitivity may only happen at the medium-to-large scale. With this understanding, one could make application-dependent decisions on whether an image is credible or not when encountering some modifications. The major problem associated with [6] is that the tampering detection results are very unstable. It is well known that the perturbation applied to a wavelet coefficient may make the extracted mark different from or still the same as the embedded one. In other words, the extracted result may be completely unpredictable. Another drawback of [6] is that the method cannot resist incidental modifications. Recently, we have proposed a multipurpose watermarking scheme [12, 13] for image/audio authentication and protection. Our method combines a media data-dependent quantization technique and a complementary watermark hiding strategy [10, 11] to conceal watermarks. We have also proposed several detection methods to optimize the tradeoff between robustness and fragility.

As to feature point-based authentication systems, Bhattacharjee and Kutter [2] proposed to generate a digital signature by encrypting the feature points' positions in an image. Authentication is then accomplished by comparing the positions of the feature points extracted from a questionable image with those decrypted from the previously encrypted digital signature. It is not certain that this approach can resist $JPEG$ compression with middle-to-high compression ratios because the feature points are liable to be shifted. Recently, Dittmann *et al.* [3] presented a content-based digital signature approach for image/video authentication using edge characteristics. Their content features are similar to [2], but different extraction techniques are used.

A typical relation-based technique for developing an image authentication system has been reported by Lin and Chang [7, 8]. In order to make the designed image authentication system tolerate $JPEG$ compression, Lin and Chang [7, 8] dedicated themselves to exploring the operation in a $JPEG$-based system. They proposed to extract a digital signature by using the invariant relation existing between any two DCT coefficients, which are at the same position of two different 8×8 blocks. They found that the invariance properties could always be preserved before and after $JPEG$ compression. However, they didn't mention clearly whether their method could survive other incidental manipulations. Although they used the invariance property to achieve their goal, the extracted relation is random by nature. In other words, the merit of the image structure, which is a very important feature, was not utilized.

In this paper, we will develop a new digital signature-based image authentication scheme which is completely different from the existing methods. In the proposed method, commonly adopted features such as the position of feature points or the relationship of any two random coefficients are not used at all. On the contrary, we propose to use the "*structure*" of an image as a digital signature. In the proposed scheme, the structure of an image's contents is composed of a number of parent-child pairs in the wavelet domain. We build up a structural digital signature and check to see if it is robust under content-preserving manipulations and fragile under content-changing manipulations. Performance analysis on the proposed new image authentication system has been conducted and the experimental results have proven the powerfulness of the system.

The remainder of this paper is organized as follows. In Sec. 2, we will present the proposed structural digital signature-based image authentication scheme. This will include the construction and verification of a structural digital signature. An analysis on the performance of our proposed scheme will be conducted in Sec. 3. We will discuss the false positive and false negative problems when incidental distortions and/or malicious tampering are encountered. In addition, we will analyze the effect that occurs when the size of a structural digital signature changes. Based on the analysis, a systematic way can be derived to determine the best size for use. In Sec. 4, a series of experiments will be conducted and their results will be reported. Concluding remarks will be given in Sec. 5.

2 Structural Digital Signature (SDS)

Our digital signature scheme is based on the wavelet transform due to its excellent multiscale and precise localization properties. Basically, the multiscale representation of an image is by nature highly suitable for designing a structural digital signature. In Sec. 2.1, we will introduce how to define a structural digital signature based on the interscale relation of wavelet coefficients. The rules for instructing how to label an SDS will be described in Sec. 2.2. The metric and the procedure used to authenticate an incoming unknown image will be detailed in Sec. 2.3. Analysis issues about the size and the complexity of an SDS will be elaborated on in Secs. 2.4 and 2.5, respectively.

2.1 Defining SDS based on Interscale Relation of Wavelet Coefficients

Let $w_{s,o}(x, y)$ represent a wavelet coefficient (at scale s, orientation o, and position (x, y)) in the orthogonally downsampled wavelet transform domain of an image $\mathbf{I}$. Suppose a J-scale wavelet transform is performed, then $0 \leq s < J$. It is well known that a large/small scale represents a coarser/finer resolution of an image, i.e., the low/high frequency part. The orientation o may be in a horizontal, vertical, or diagonal direction. The interscale relationships of wavelet coefficients can then be converted into the relationships between the parent node $w_{s+1,o}(x, y)$ and its four child nodes $w_{s,o}(2x+i, 2y+j)$ with

$$|w_{s+1,o}(x, y)| \geq |w_{s,o}(2x+i, 2y+j)|, \tag{1}$$

or

$$|w_{s+1,o}(x, y)| < |w_{s,o}(2x+i, 2y+j)|, \tag{2}$$

where $0 \leq s < J$, $0 \leq i, j \leq 1$, and $1 \leq x \leq N$ and $1 \leq y \leq M$ ($N \times M$ is the image size). Combining Eqs. (1) and (2), the above two relations can be rewritten as

$$||w_{s+1,o}(x, y)| - |w_{s,o}(2x+i, 2y+j)|| \geq 0. \tag{3}$$

In order to design a reliable scheme for image authentication, we propose a new signature method called structural digital signature (SDS). The new signature can be obtained by observing the interscale relations of wavelet coefficients of an image. The basic concept of the new scheme relies on the following: (i) the interscale relationship should be difficult to be destroyed after content-preserving manipulations; and (ii) this interscale relationship should be difficult to be preserved after content-changing manipulations. Because these interscale relationships result from the structure of an image (say $\mathbf{I}$), we define them as the structural digital signature of $\mathbf{I}$ and call it $SDS(\mathbf{I})$.

The structural digital signature of an image consists of a set of parent-child pairs which satisfy

$$||w_{s+1,o}(x,y)| - |w_{s,o}(2x+i, 2y+j)|| \geq \sigma \ (\sigma > 0). \tag{4}$$

The above constraint is stricter than the original interscale relationship of wavelet coefficients shown in Eq. (3). The size of σ will determine the number of parent-child pairs recorded in an $SDS(\mathbf{I})$. The smaller the σ is, the larger the amount of elements in an SDS. We do not intend to keep all the parent-child pairs as elements of an SDS because some of the elements may not be significant enough. The significance of a parent-child pair is completely dependent on their magnitude difference. The larger the difference, the more significant the parent-child pair is. A parent-child pair whose magnitude difference is small is equivalent to having a "small" quantization interval in the quantization-based approaches [6, 12, 22]. Therefore, it will be very sensitive to modifications including some minor incidental ones. In order to design a robust image authentication scheme, we only consider those parent-child pairs whose magnitude differences are large as the elements of a structural digital signature. In order to appropriately detect malicious tampering while tolerating an incidental modification, we use the size of a structural digital signature to control the tradeoff between fragility and robustness. In general, the construction of a structural digital signature is very easy because there is no feature point selection involved [2, 3].

Once the parent-child pairs are selected by the constraint defined in Eq. (4), each pair is assigned a symbol that represents what kind of relationship this pair carries. These symbols will be formally defined in Sec. 2.2. The above mentioned symbols and their locations in the wavelet domain will be encrypted by a public key algorithm such as the famous RSA method [15]. Finally, the encrypted information will be stored and used for image authentication later.

2.2 Labeling an SDS

According to the interscale relationship existing among wavelet coefficients, there are four possible relationship types of an SDS. Assume the magnitude of a parent node p is larger than that of its child node c (i.e., $|p| > |c|$), then the four possible relationships of the pair, $< p, c >$, are: (i) $p \geq 0, c \geq 0$; (ii) $p \geq 0, c \leq 0$; (iii) $p \leq 0, c \geq 0$; (iv) $p \leq 0, c \leq 0$. Consider the case when $|p| > |c|$ and c is small. In order to make $< p, c >$ still credible when incidental modifications are encountered, the value of c is not important. Therefore, the relations (i) and (ii) can be merged to form a signature symbol I under the condition that $p \geq 0$ and c don't care. On the other hand, the relations (iii) and (iv) can be merged to form another signature symbol II, under the condition that $p \leq 0$ and c don't care. In other words, we intend to keep the sign of the larger element unchanged while disregarding the smaller one under the constraint that their original interscale relationship is still preserved. Similarly, signature symbol III (under the condition that $c \geq 0$ and p don't care) and IV (under the condition that $c \leq 0$ and p don't care) can be defined under the constraint $|p| < |c|$. For those pairs that are not recorded in an SDS are all labeled by the fifth signature symbol V. Hence, we represent the signature symbol of a parent-child pair as $sym(< p, c >)$, which can be one of the above defined symbol types. In the following section, we shall describe how the verification process is executed.

2.3 Verification

In the verification process, if one would like to verify an unknown image $\tilde{\mathbf{I}}$, it is first wavelet transformed and then its structural digital signature $SDS(\tilde{\mathbf{I}})$ that should be constructed. The encrypted structural digital signature of the original image $\mathbf{I}$ is retrieved and then decrypted to obtain its corresponding $SDS(\mathbf{I})$. One can say the interscale relationship of a pair $< p, c >$ in $\mathbf{I}$ is still unchanged in $\tilde{\mathbf{I}}$ if their signature symbols are the same. That is, the relation

$$sym(< p, c >) = sym(< \tilde{p}, \tilde{c} >) \tag{5}$$

holds, where the pair $< \tilde{p}, \tilde{c} >$ in $\tilde{\mathbf{I}}$ is the corresponding pair of $< p, c >$ in $\mathbf{I}$. Finally, we calculate the completeness of the SDS ($CoSDS$) in $\tilde{\mathbf{I}}$, which is defined as the similarity degree, Sim, between $SDS(\mathbf{I})$ and $SDS(\tilde{\mathbf{I}})$:

$$CoSDS(\tilde{\mathbf{I}}) = Sim(SDS(\mathbf{I}), SDS(\tilde{\mathbf{I}})) = \frac{N^+ - N^-}{|SDS(\mathbf{I})|}, \tag{6}$$

where N^+ represents the number of pairs satisfying Eq. (5) and N^- represents the number of pairs violating Eq. (5). $|SDS(\mathbf{I})|$ is used to denote the number of parent-child pairs in $SDS(\mathbf{I})$. From

Eq. (6), we know that $CoSDS(\tilde{\mathbf{I}})$ will fall into the interval $[-1 \;\; 1]$. In other words, the completeness of SDS represents the ratio of how many parent-child pairs are preserved to satisfy their interscale relationships. A larger $CoSDS$ means the suspect image $\tilde{\mathbf{I}}$ is reliable; otherwise, it means $\tilde{\mathbf{I}}$ has been maliciously tampered with. In addition, the location of a tampering region can be easily detected from those parent-child pairs whose signature symbols have been updated.

2.4 How the Size of an $|SDS|$ influences the Compromise between Robustness and Fragility

In this subsection, we shall discuss how the constituent parent-child pairs of an SDS influence a compromise between robustness and fragility. Let the magnitudes of the differences of parent-child pairs in a structural digital signature be arranged in a decreasing order. It is known that the elements (parent-child pairs) with larger magnitudes are not vulnerable to attacks while those with smaller magnitudes tend to be easily attacked. Therefore, one can use the larger elements to indicate robustness and use the smaller elements to reflect fragility. Under the circumstances, when the size of a structural digital signature becomes large, the elements with smaller magnitudes tend to be changed so that the robustness property is more or less affected. On the other hand, the modification of the smaller elements will reflect accurately the degree of fragility. So, if $|SDS|$ is small enough such that elements are all with larger magnitudes, then the fragility property may disappear. In Sec. 3, we will give a systematic way to determine σ (which also determines the $|SDS|$) by a statistical analysis of the distributions on an SDS and the behavior of an attack.

2.5 Complexity Analysis on an SDS

In this section, the complexity of a structural digital signature will be analyzed. Let the number of parent-child pairs in an SDS be n. The first part of an SDS we should store is the child locations of the n parent-child pairs. The reason why the child locations are examined instead of the parent locations is that they are easily tracked. For example, if a child node's location is (x, y), then its parent's location is $(\lfloor \frac{x}{2} \rfloor, \lfloor \frac{y}{2} \rfloor)$. On the contrary, if a parent node's location is (x, y), there are four possible locations for a child. They are $(2x+i, 2y+j)$ where $0 \leq i, j \leq 1$. For the n parent-child pairs, $2 \times n$ bytes are required to store their locations because each location needs two bytes. In addition, each parent-child pair in an SDS has four possible interscale relationships. Since each interscale relationship needs two bits to express it, a total of $\frac{n}{4}$ bytes is required to store all the interscale relationships.

In fact, the storage can be further reduced if the locations of child nodes are stored based on their pre-defined ordering. Under the circumstances, the number of occurrences of every signature symbol is counted. For the first four types of symbols, we store the number of parent-child pairs and then the locations of these pairs. In this way, the memory used for storing the signature symbols will be reduced from $\frac{n}{4}$ bytes to 4 bytes. That is, a total of $(2n + 4)$ bytes is required to store a structural digital signature before encryption.

3 Performance Analysis

Usually, a watermark-based or digital signature-based authentication method must be justified by the false positive (false alarm) and false negative (miss detection) probability analyses like those that have been done in [6, 7, 11]. For an image authentication system, a false positive probability means an image is detected to be maliciously tampered but in fact it is not. On the other hand, a false negative probability means an image is actually modified by a malicious tampering but some tampered areas are not detected. A practical signature system should ensure that both the false positive and false negative probabilities are reasonably small. The analysis on the false positive and the false negative probabilities will be elaborated in Secs. 3.1 and 3.2, respectively. The relationship between the predetermined threshold σ and the strength of attacks will be discussed in Sec. 3.3. The security issues will be discussed in Sec. 3.4

3.1 False Positive due to Incidental Manipulations

An incidental modification like the $JPEG$ compression is a kind of "attack" that we would like to bypass. If an incidental attack is detected, it will cause a false positive type error. Let $\mathbf{I}$ be an image, A be any incidental manipulation, and ψ be a wavelet function. A distorted image, $\mathbf{I}^A$, can be derived by $\mathbf{I} * A$, where $*$ is a convolution operator. Since the authentication process is conducted in the wavelet domain, the whole transformation process can be denoted as

$$\psi * (\mathbf{I} * A) = (\psi * \mathbf{I}) \times A^f = \mathbf{I}^\psi \times A^f, \tag{7}$$

where $\mathbf{I}^\psi$ is the wavelet transformed image in the space-frequency domain and A^f is a version of A in the frequency domain. Eq. (7) indicates that the wavelet transform of the distorted image $\mathbf{I}^A$ is equivalent to the modification (by A^f) of the wavelet transformed image $\mathbf{I}^\psi$. If A^f is a quantization operation of some compression methods, any coefficient in $\mathbf{I}^\psi$ will only be affected by itself through A^f. Because the behavior of compression like $SPIHT$ [16] is easily predicted and its corresponding tree structure is required in constructing an SDS, we will analyze its effects. $SPIHT$ is a progressive image coding scheme in which the most significant bits are transmitted first. Suppose p (a parent node) and c (a child node) form a parent-child pair in an SDS and their wavelet coefficients satisfy the relation $2^k \geq |p| \geq 2^{k-1} \geq \cdots \geq 2^{k-j} \geq |c| \geq 2^{k-(j+1)}$ with $j \geq 1$. When a $SPIHT$ compression is executed, we may encounter three different possibilities: (1) when the compression ratio is high, suppose 2^t is the threshold finally used in the dominant process [16] and $t \geq k$, the reconstructed parent-child pair, p^r and c^r, are both zeros. This means the original relationship $|p| \geq |c|$ is preserved when $p^r = c^r = 0$; (2) when the compression ratio is medium, suppose $2^{k-1} \geq 2^t \geq 2^{k-j}$, we will have $|p^r| > |c^r| = 0$. Again, the parent-child pair's relationship is preserved; (3) for a compression with a small ratio, suppose $2^{k-(j+1)} \geq 2^t$, we will have $|p^r| > |c^r| \neq 0$. Once again, the parent-child pair's relationship is preserved. From the above derivation, it is guaranteed that the proposed SDS will survive a $SPIHT$ compression at any ratio. A similar conclusion can be applied to the $JPEG$ compression.

On the other hand, if A is another incidental manipulation (excluding compressions), its behavior may not be easily analyzed because the change of a specific coefficient may be determined by its neighbors. However, it is known that an incidental manipulation tends not to destroy the semantics of an image. Based on this understanding, an SDS will not be significantly destroyed when an incidental manipulation is encountered. Therefore, one can expect that a structural digital signature is indeed a good mechanism for tolerating incidental modifications.

Another advantageous point of using SDS is its stable nature against rounding errors. The reason why this is true is due to the large chosen value of σ (by Eq. (4)). When the constituent elements of an SDS are all with a large σ, rounding errors that emerge won't influence the relationship of a parent-child pair.

3.2 False Negative due to Content Replacement

When a malicious modification like content replacement is applied to an image, its corresponding SDS will have a significant change that is very easy to detect. Therefore, we can expect the false negative probability in this case to be very low. Suppose a parent node p ($p > 0$) and a child node c is a pair in an SDS. They have the relation $|p| \geq |c|$ with $||p| - |c|| = \sigma_i$ ($\sigma_i > \sigma$). For simplicity, let p be attacked by a malicious manipulation with the modification quantity M_p. If $|p - M_p| > |c|$ holds under the condition that $|p| > |c|$, then a false negative occurs because $0 \leq M_p \leq \sigma_i$. If the effect caused by M_p forms a Gaussian distribution with variance ρ^2, then the false negative probability can be defined as $\frac{\int_{-\sigma_i}^{\sigma_i} Ce^{\frac{t^2}{\rho^2}} dt}{\int_{-\infty}^{\infty} Ce^{\frac{t^2}{\rho^2}} dt}$ (C is a constant). When a malicious distortion is applied to an image, if β ($0 \leq \beta \leq 1$) represents the proportion of the parent-child pairs that has been maliciously tampered with but still maintains their interscale relations, then the total false negative probability will be

$$P_{fn} = \Pi_{i=1}^{i=\beta\times|SDS|} \frac{\int_{-\sigma_i}^{\sigma_i} Ce^{\frac{t^2}{\rho^2}} dt}{\int_{-\infty}^{\infty} Ce^{\frac{t^2}{\rho^2}} dt}. \tag{8}$$

From Eq. (8), it is not difficult to imagine that P_{fn} will be very low. In other words, the false negative probability will be very low when a content replacement operation is applied to an image.

3.3 The Relation between σ and the Strength of Attacks

In this subsection, we will discuss an issue regarding the relationship between σ and the strength of an attack. Recall that $|SDS|$ denotes the number of parent-child pairs whose interscale relationships are recorded in a structural digital signature. Attacks can be roughly classified into two categories: incidental manipulation and malicious distortion. To simplify the analysis, we assume the strength of an attack, a, is a Gaussian distribution, $\mathcal{G}^A$, with a mean of zero. According to the Gaussian modeling of attacks [6, 12, 22], we have the following analysis. Usually, an incidental manipulation tends to have a small standard deviation ρ_I while a malicious tampering tends to have a large standard deviation ρ_M, i.e., $\rho_I < \rho_M$. Some reference values regarding ρ_I and ρ_M were provided in [7] for a specific image. Based on our scheme, a structural digital signature is constructed by selecting those parent-child pairs whose differences in magnitudes are larger than σ. The difference in magnitude, d, may have two forms: positive difference ($d \geq 0$) and negative difference ($d < 0$). The positive difference portion and the negative difference portion both form a Gaussian distribution, $\mathcal{G}^S$, without a mean of zero. Their standard deviations are denoted as ρ_S, which is usually very large (scale of hundreds) because the variance of d is large in the wavelet domain and is larger than ρ_I. The possible relationships between $\mathcal{G}^A$ and $\mathcal{G}^S$ are depicted in Fig. 1. In Fig. 1, the Gaussian distributions shown in the middle part are $\mathcal{G}^A$, whereas the right/left one is $\mathcal{G}^S$ corresponding to a positive/negative d. τ is defined as the intersection point of $\mathcal{G}^A$ and $\mathcal{G}^S$. The shaded areas, which represent the parent-child pairs with a smaller difference $|d|$ (in the tails of $\mathcal{G}^S$), are assumed to be updated based on the value in the tails of $\mathcal{G}^A$. Next, we will analyze the effect of ρ_I and ρ_M on σ, respectively.

First, let an incoming attack be an incidental one such as $JPEG/SPIHT$ compression or rescaling. The probability that the relationship of parent-child pairs may be destroyed (i.e., d's sign is changed) is denoted as p^I (the shaded areas in Fig. 1) and can be calculated by

$$\begin{aligned} p^I &= 2 \times (P\{0 < d < \tau - \sigma\} + P\{\tau < a < \infty\}) \\ &= 2 \times (P\{0 < d < \tau - \sigma\} + (1 - P\{0 < a \leq \tau\})) \\ &= 2 \times (erf(\frac{\tau - \sigma}{2\rho_S}) + [1 - erf(\frac{\tau}{2\rho_I})]), \end{aligned} \tag{9}$$

where $erf(\cdot)$ represents the error function [1] which is defined as:

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

In Eq. (9), the constant 2 represents the two symmetric $\mathcal{G}^S$'s that belong, respectively, to the positive and negative d. Because the attack under consideration is incidental, $\tau - \sigma$ is usually small. Since the standard deviation ρ_S of $\mathcal{G}_S$ is on the scale of hundreds, $\frac{\tau-\sigma}{2\rho_S}$ is, thus, very small. Under the circumstances, the first term in Eq. (9), $erf(\frac{\tau-\sigma}{2\rho_S})$, approximates zero. On the other hand, τ satisfies $\tau > \sigma$ and σ is chosen to be large (Eq. (4)), so τ is also large enough. For an incidental attack, we know the value of ρ_I is usually small. Therefore, $\frac{\tau}{2\rho_I}$ is large. As a consequence, the second term, $[1 - erf(\frac{\tau}{2\rho_I}))]$, should be very small. In summary, the above discussion explains why the probability P^I can be sufficiently small if the incoming attack is incidental with a small ρ_I. That is,

$$p^I \approx 2 \times [1 - erf(\frac{\tau}{2\rho_I})] \approx 0. \tag{10}$$

The near-optimal σ can be derived based on the condition that the incoming attack is incidental and the value of p^I is smaller than a pre-determined threshold ϵ (e.g., $\epsilon = 0.1$). Under the circumstances, the near-optimal σ can be derived by

$$p^I \approx 2 \times [1 - erf(\frac{\tau}{2\rho_I})] < \epsilon.$$

Thus, we have

$$1 - \frac{\epsilon}{2} < erf(\frac{\tau}{2\rho_I}). \tag{11}$$

Using a predetermined ϵ together with ρ_I and checking the tables of error function [1], we should be able to obtain the lower bound of τ. From this τ, the lower bound of a near-optimal σ can be approximately determined because based on the Gaussian models shown in Fig. 1 σ is close to τ.

Now, let the incoming attack such as object placement/replacement or cloning be malicious. The probability that the relationships of parent-child pairs in a structural digital signature may be destroyed is defined as

$$p^M = 2 \times (P\{0 < d < \tau - \sigma\} + P\{\tau < a < \infty\})$$

$$
\begin{aligned}
&= 2 \times (P\{0 < d < \tau - \sigma\} + (1 - P\{0 < a \leq \tau\}) \\
&= 2 \times (erf(\frac{\tau - \sigma}{2\rho_S}) + [1 - erf(\frac{\tau}{2\rho_M})]). \qquad (12)
\end{aligned}
$$

In Eq. (12), $\tau - \sigma$ is known to be small and, thus, $\frac{\tau-\sigma}{2\rho_S}$ is very small. As a consequence, the first term in Eq. (12), $erf(\frac{\tau-\sigma}{2\rho_S})$, has a value close to zero because it corresponds to an incidental modification. It is also known that ρ_M is usually large and that it may lead to a small $\frac{\tau}{2\rho_M}$. Therefore, the second term of Eq. (12), $[1 - erf(\frac{\tau}{2\rho_M}))]$, has a value which is far from zero. In general, the detection rate of regions that are maliciously tampered with is determined mainly based on the second term. If we assume P^M is large enough, and ρ_M and the tables of error function [1] are available, we will be able to determine the upper bound of τ. From the above τ, the upper bound of a near-optimal σ will be approximately obtained as in the case of incidental modifications.

To sum up, the interval where a near-optimal σ should fall can be mathematically derived from the above analysis. In Sec. 4, we will provide a numerical example to show how different values of σ affect p_I.

3.4 Security Problem

In this section, we will discuss the issues regarding (1) the positions of the elements in a structural digital signature which are known or are correctly guessed; (2) the image intensity is constantly changed.

3.4.1 Tampering at the Locations Where SDS Does not Record

If the locations of the elements in an SDS are correctly guessed, the attacker may try to tamper with those positions which are not recorded in the corresponding $SDS(\mathbf{I})$ and thus disable our method. Fortunately, the attackers cannot succeed in this case because if the parent-child pairs are not recorded in an $SDS(\mathbf{I})$, then their interscale relationships do not satisfy the condition in Eq. (4). In other words, we can verify it easily by checking the signature symbols of those parent-child pairs that are not recorded in $SDS(\mathbf{I})$ and $SDS(\tilde{\mathbf{I}})$. Let $< w_{s,o}(x,y), w_{s+1,o}(2x+i, 2y+j) >$ be a parent-child pair which is not in $SDS(\mathbf{I})$ and assume its corresponding pair $< \tilde{w}_{s,o}(x,y), \tilde{w}_{s+1,o}(2x+i, 2y+j) >$ is not in $SDS(\tilde{\mathbf{I}})$, where $0 \leq i,j \leq 1$. We can determine whether the $< w_{s,o}(x,y), w_{s+1,o}(2x+i, 2y+j) >$ pair is tampered with or not by checking $sym < \tilde{w}_{s,o}(x,y), \tilde{w}_{s+1,o}(2x+i, 2y+j) >$. If $sym < \tilde{w}_{s,o}(x,y), \tilde{w}_{s+1,o}(2x+i, 2y+j) >$ is *not* equal to V, then it has been tampered with. It is known that the condition for $sym < \tilde{w}_{s,o}(x,y), \tilde{w}_{s+1,o}(2x+i, 2y+j) >$ to belong to V is $||\tilde{w}_{s,o}(x,y)| - |\tilde{w}_{s+1,o}(2x+i, 2y+j)|| < \sigma$.

3.4.2 The Condition that Image Intensity Is Constantly Changed

Attackers may think that they can modify the image's intensity without triggering our authentication scheme. One possible method is to constantly increase or decrease the intensity of an image **I** so that the interscale relationships of all parent-child pairs are not changed. One solution to conquer this problem is to record the wavelet coefficients of the lowest frequency band because they represent the approximate information of a whole image. In addition, the high frequency bands will not be altered because a constant convolved with a wavelet will be zero due to the nature of wavelets. Once an image is tampered with by a constant update, its lowest frequency band will reflect this change. Lin and Chang [7] used a similar method to solve the above mentioned problem in the DCT domain.

4 Experimental Results

Our structural digital signature-based image authentication scheme was first tested against a Beach image with 256×256 size, as shown in Fig. 2(a). A large "umbrella" was placed in Fig. 2(a) and formed a tampered image as shown in Fig. 2(b). We used a $4-$scale wavelet transform to transform the images so that the resolution of the lowest-frequency channel had the size of 16×16. At first, the parent-child pairs whose difference d satisfying $|d| > \sigma = 256$ were chosen to construct an SDS. The detected tampering areas were shown in Figs. 2(c)$\sim$(e). Another set of detected results using $\sigma = 128$ was shown in Figs. 2(f)$\sim$(h). As we expected, the SDS with a smaller size will lose some tampered pixels. However, the integration of multiscale results was sufficient to reflect the area tampered with. Another set of experiments was conducted by placing a "small" object at the bottom-right corner of the "peppers" image. Fig. 3(a) and Fig. 3(b) show, respectively, the host image and the image tampered with. Figs. 3(c)$\sim$(e) and Figs. 3(f)$\sim$(h) show, respectively, the detected multiscale results when $\sigma = 256$ and $\sigma = 128$. The above experiments provided a good example of the compromise between robustness and fragility using two structural digital signatures with different sizes.

In the second part of our experiments, we applied several incidental distortions to Fig. 2(a) to test the robustness of our scheme. Three structural digital signatures with a different number of parent-child pairs were constructed, and their corresponding positions in the wavelet domain were shown in Fig. 4. It can be seen that the SDS with a smaller/larger $|SDS|$ (corresponding to a larger/smaller σ) would result in fewer/more elements. Table 1 shows the completeness of SDS obtained under different $SPIHT$ compression ratios using three different σ. It is obvious that when the compression ratio was smaller than 32, most of the derived $CoSDS$ were perfect. However, when the compression ratio reached 64, some fragile results emerged for $\sigma = 64$. For the $JPEG$ compression, perfect preservations of SDS (except for the results obtained from $\sigma = 64$) were obtained for quality factors ranging from 60% (7 : 1) to 10% (21.7 : 1), as shown in Table 2. Table 3 summarized the verification results obtained under other incidental distortions including rescaling, histogram equalization, blurring, median filtering, sharpening, and Gaussian noise adding. These manipulations are sometimes unavoidable in image processing and, thus, cannot be considered as malicious modifications. From Tables 1∼3, we can find that the completeness of a structural digital signature was consistently very high for incidental manipulations when $\sigma > 64$. This indicates that our method can tolerate common incidental modifications very well. However, the above conclusion is true only when the value of σ is large enough (e.g., $\sigma > 64$ in our experiments). Theoretically, a reasonable σ can be determined based on the analysis described in Sec. 3.

Next, we shall show how the value of σ influences the probability that the relationship of the parent-child pairs in an SDS is destroyed. Table 3 illustrated six incidental modifications which were used in this experiment. The minimum distance (σ) used for thresholding were 256, 128, and 64, respectively. The curves shown in Fig. 5 indicated that when σ was set to 128 or 256, the probability that the relationship of the parent-child pairs in an SDS being destroyed was zero. From Fig. 5, we found that the values obtained by theoretical analysis were not necessarily consistent with the experimental results. This phenomenon can be explained by the following potential reasons: (1) The behavior of an incidental manipulation and the elements of a structural digital signature are both assumed to be Gaussian distributed for the sake of simplicity. However, it may not be the case; (ii) We propose the shaded areas in Fig. 1 that reflect the relationship of those parent-child pairs with small $|d|$ will be destroyed, but in a practical situation this may not be true. In fact, any parent-child pair in a SDS could possibly be destroyed. We can only say that the pair with a smaller difference has a higher probability of being destroyed. Even when the ϵ of Eq. (11) is set in advance and the near-optimal σ is determined, one cannot decide whether an incoming attack is incidental or not. This is because when the regions that have been maliciously tampered with are very small, the number of destroyed parent-child pairs is small too and, thus, its value has the probability of being smaller than ϵ. Therefore, we suggest that the final decision on whether an attack is incidental or malicious still needs human intervention so that a perfect perceptual judgement can be made. Under the above circumstances, if the regions detected as having been tampered with are very small and spread over a whole image but are still recognizable and meaningful, the imposed attack should be regarded as malicious. Except for the example of a tiny content-changing modification shown in Fig. 3, our scheme is able to determine whether the imposed attack is malicious or incidental by merely comparing the value of ϵ and $1 - CoSDS(\tilde{\mathbf{I}})$.

In the following, we shall use our scheme to authenticate the images that were modified by an incidental manipulation and a malicious distortion simultaneously. Fig. 6(a) shows a beach image which was first $JPEG$ compressed with a quality factor of 10% and then an "umbrella" object was placed. The verification results obtained at $2^2 \sim 2^4$ scales using $\sigma = 128$ were shown in Figs. 6(b)∼(d), respectively. As we can see from these results, the area where the umbrella was placed could be approximately detected and the $JPEG$ compression did not affect the verification results. The experiment indicated that the structural digital signature efficiently tolerated the $JPEG$ compression while sensitively detecting object placement. Another set of experiments was shown in Fig. 6(e)∼(h). The beach image was first scaled down to 128×128 from 256×256, and then the umbrella object was placed on it. Finally, the image was rescaled to the original size 256×256, as shown in Fig. 6(e). When σ was set to be 128, Figs. 6(f)∼(h) showed the placed umbrella was detected at $2^2 \sim 2^4$ scales. It can be seen that some small fragments which were not the targets were mistakenly detected. This is because the changes of wavelet coefficients that resulted from rescaling are more liable to destroy the structural digital signature than the $JPEG$. However, we can also see that the regions belonging to the "umbrella" tend to be clustered together. By comparing the values shown in Table 2 and Table 3, it is easy to see that the $CoSDS$ values obtained by applying $JPEG$ with any quality factors are higher than those obtained by applying rescaling.

Finally, we conducted an experiment to demonstrate if malicious tampering occurred on areas which were not recorded in an SDS, then they could also be detected as we have analyzed in Sec. 3.4. In Fig. 7(a), a helicopter was placed on the sky portion of the beach image (Fig. 2(a)). As we can see from Fig. 4, the wavelet coefficients in the sky area did not belong to the structural digital signature. Using the proposed scheme, the area tampered with could be detected and shown, respectively, in Figs. 7(b)$\sim$ (d) when $\sigma = 128$. The blocky effect shown in Fig. 7(b)$\sim$ (d) was the natural result inherited from the multiresolution representation of the wavelet transform.

From the above experiments, we could make a conclusion about the selection of σ. The value of σ can be mathematically determined from the analysis described in Sec. 3. However, the assumptions used in Sec. 3 may not always hold, so we can empirically choose σ to be at least 128 which has been confirmed by several experimental results.

5 Conclusion

For image authentication, it is desired that the verification method be able to resist content-preserving modifications while being sensitive to content-changing modifications. In this paper, a new structural digital signature scheme has been proposed for image authentication. We make use of the structure of an image to construct a digital signature. The only way to destroy the structure of our digital signature is to significantly change the image's content and that would be detected as malicious. In addition, some unavoidable image processing techniques will preserve a great many of the SDS which would be detected as incidental. Performance analysis of the structural digital signature has been provided and experimental results show that our scheme is really robust to content-preserving manipulations and fragile to content-changing distortions.

Our future work will consider geometric distortions such as rotation and translation, which cannot be tolerated in this paper because the structural digital signature built in the wavelet domain is variant to rotation and translation. Another future work will focus on developing structural watermarking, which can be used for public-key detection from the viewpoint that a watermark structure can only be removed if its structure is destroyed.

Acknowledgment: The authors thank Dr. Martin Kutter for providing the beach image and the umbrella image used in the experiments.

References

[1] M. Abramowitz and I. A. Stegun, "Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables", *Dover Publications*, Inc., New York, 1965.

[2] S. Bhattacharjee and M. Kutter, "Compression Tolerant Image Authentication", *IEEE Inter. Conf. on Image Processing*, USA, pp. 435-439, 1998.

[3] 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, pp. 209-213, 1999.

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

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

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

[7] 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 (www.ctr.columbia.edu/∼cylin/auth/auth.html).

[8] C.-Y. Lin and S.-F. Chang, "Generating Robust Digital Signature for Image/Video Authentication", *Multimedia and Security Workshop at ACM Multimedia*, UK, 1998.

[9] E. T. Lin and E. J. Delp, "A Review of Fragile Image Watermarks", *Proc. of the Multimedia and Security Workshop (ACM Multimedia '99)*, Orlando, pp. 25-29, 1999.

[10] C. S. Lu, H. Y. Mark Liao, S. K. Huang, and C. J. Sze, "Cocktail Watermarking on Images", *3rd Inter. Workshop on Information Hiding*, LNCS 1768, pp. 333-347, Sept. 29-Oct. 1, 1999.

[11] 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.

[12] C. S. Lu, H. Y. Mark Liao and C. J. Sze, “Combined Watermarking for Image Authentication and Protection”, *Proc. 1st IEEE Int. Conf. on Multimedia and Expo*, USA, 2000.

[13] C. S. Lu, H. Y. Mark Liao and L. H. Chen, “Multipurpose Audio Watermarking”, *Proc. 15th Int. Conf. on Pattern Recognition*, Barcelona, Spain, Vol. III, pp. 286-289, 2000.

[14] C. S. Lu and H. Y. Mark Liao, “Structural Digital Signature for Image Authentication: An Incidental Distortion Resistant Scheme”, to appear in *Proc. Multimedia and Security Workshop at the ACM Int. Conf. on Multimedia*, Los Angeles, California, USA, 2000.

[15] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, “Handbook of Applied Cryptography”, *CRC Press*, 1997.

[16] A. Said and W. A. Pearlman, ”A New, Fast, and Efficient Image Codec based on Set Partitioning in Hierarchical Trees”, *IEEE Trans. Circuit and Systems for Video Technology*, Vol. 6, pp. 243-250, 1996.

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

[18] 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.

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

[20] L. Xie and G. R. Arce, “A Blind Wavelet Based Digital Signature for Image Authentication”, *Proc. of the EUSIPCO 98*, Island of Rhodes, Greece, Sep 1998.

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

[22] G. J. Yu, C. S. Lu, H. Y. Mark Liao and J. P. Sheu, “Mean Quantization Blind Watermarking for Image Authentication,” *Proc. IEEE Int. Conf. on Image Processing*, Vancouver, Canada, Vol. III, pp. 706-709, 2000.

[23] 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.

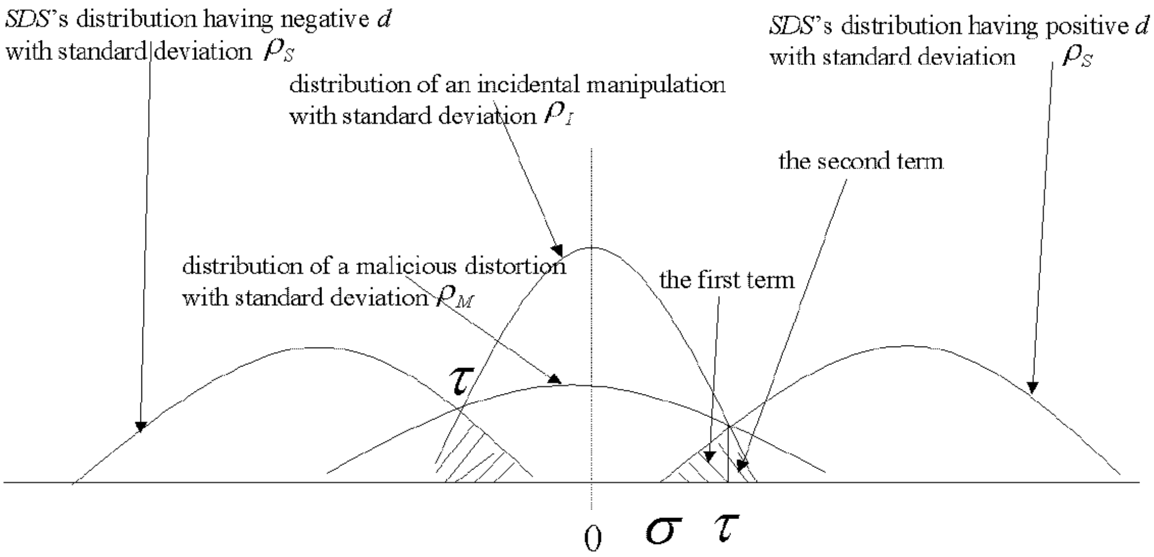

SDS's distribution having negative *d* with standard deviation ρ_S
distribution of an incidental manipulation with standard deviation ρ_I
SDS's distribution having positive *d* with standard deviation ρ_S
the second term
distribution of a malicious distortion with standard deviation ρ_M
the first term
τ
0
σ
τ

Figure 1: The relationship between the attack's distribution $\mathcal{G}^A$ (with standard deviation ρ_I or ρ_M) and the SDS's distribution $\mathcal{G}^S$ (with standard deviation ρ_S).

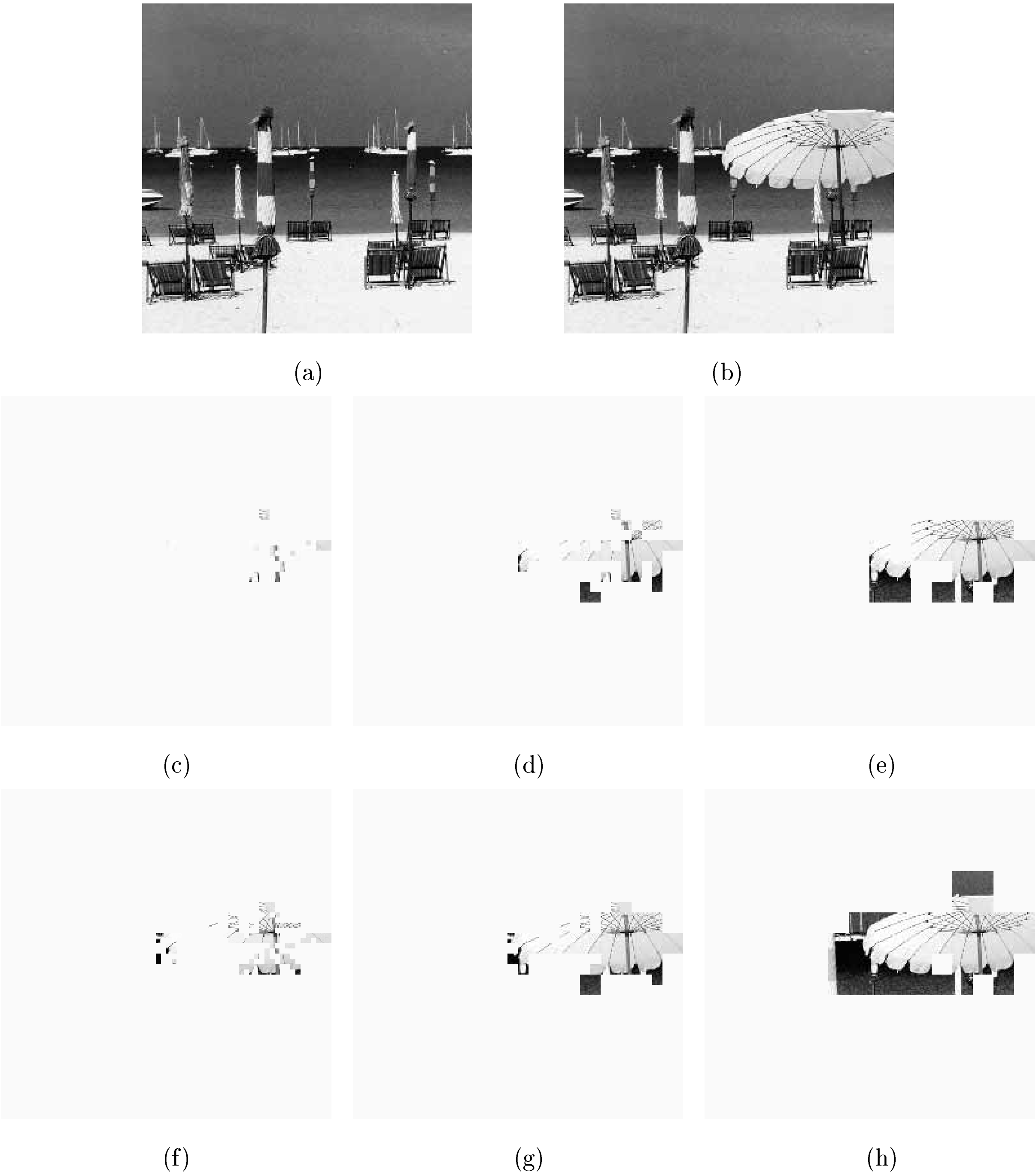
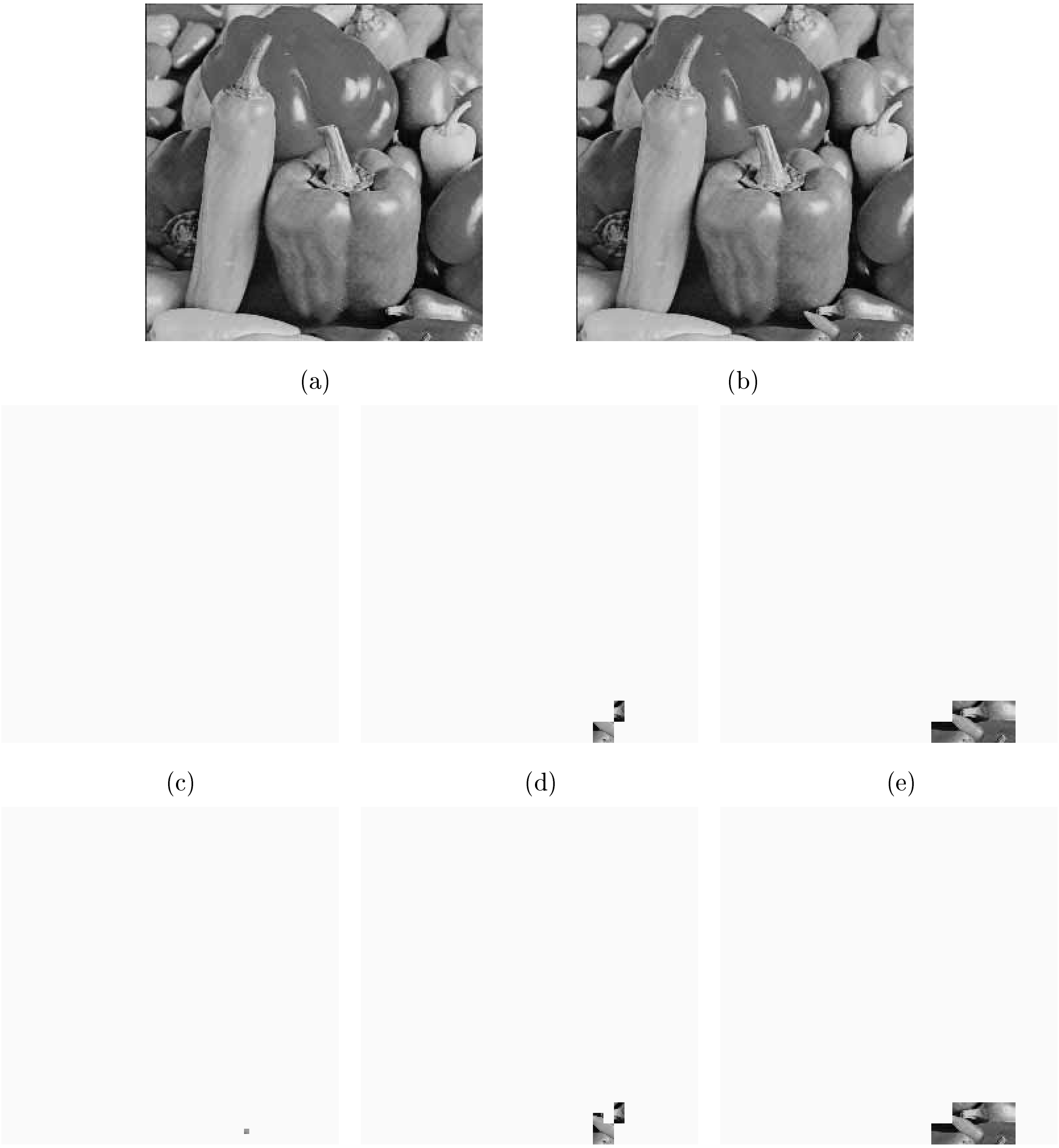


Figure 2: Content tampering: (a) host image; (b) original image with a large object placed; (c)$\sim$(e) detected results at $2^2 \sim 2^4$ scales when $\sigma = 256$; (f)$\sim$(h) detected results at $2^2 \sim 2^4$ scales when $\sigma = 128$.



(a) (b)

(c) (d) (e)

(f) (g) (h)

Figure 3: Content tampering: (a) host image; (b) original image with a small object placed at the bottom-right; (c)$\sim$(e) detected results at $2^2 \sim 2^4$ scales when $\sigma = 256$; (f)$\sim$(h) detected results at $2^2 \sim 2^4$ scales when $\sigma = 128$.

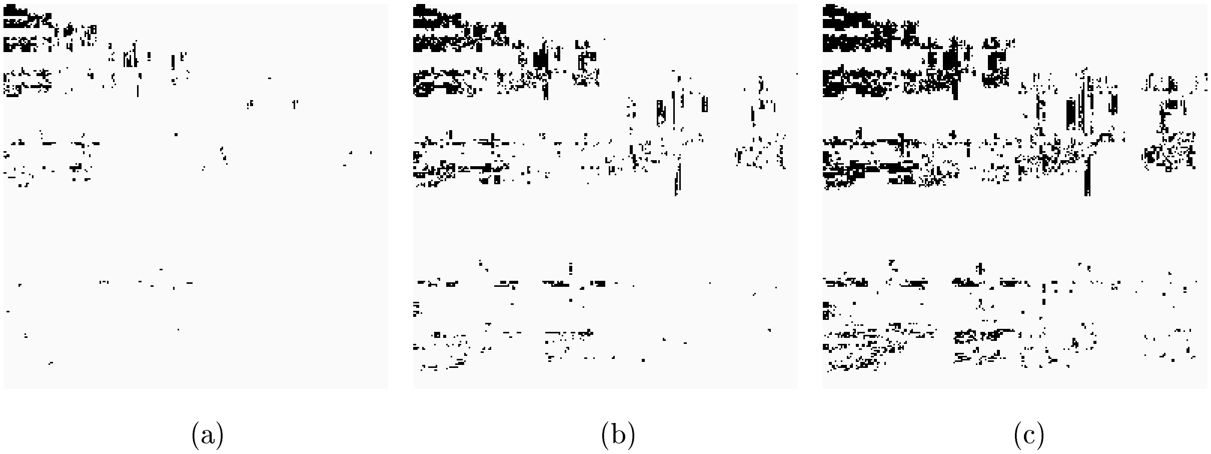


Figure 4: The positions of the elements (illustrated in black color in the wavelet domain) of an SDS constructed from Fig. 2(a) with (a) $\sigma = 256$, (b) $\sigma = 128$, and (c) $\sigma = 64$.

Table 1: *CoSDS* **of Fig. 2(a) under SPIHT with various compression ratios (CR).**

CR	*Completeness of SDS*		
	$\sigma = 256$	$\sigma = 128$	$\sigma = 64$
$8 : 1$	1.000	1.000	1.000
$16 : 1$	1.000	1.000	1.000
$32 : 1$	1.000	1.000	0.997
$64 : 1$	1.000	0.994	0.816

Table 2: *CoSDS* **of Fig. 2(a) under JPEG with various quality factors (QF).**

QF(CR)	*Completeness of SDS*		
	$\sigma = 256$	$\sigma = 128$	$\sigma = 64$
60(7.1 : 1)	1.000	1.000	1.000
50(8.2 : 1)	1.000	1.000	1.000
40(9.7 : 1)	1.000	1.000	0.999
30(11.7 : 1)	1.000	1.000	0.992
20(15.0 : 1)	1.000	1.000	0.988
10(21.7 : 1)	1.000	0.996	0.969

Table 3: *CoSDS* **of Fig. 2(a) under a set of incidental distortions (among them, sharpening and Gaussian noise adding with amount** 16 **were run using Photoshop).**

Incidental distortions	*Standard deviation* ρ_I	*Completeness of SDS*		
		$\sigma = 256$	$\sigma = 128$	$\sigma = 64$
rescaling	26.8	0.993	0.918	0.808
equalization	27.3	0.983	0.961	0.946
blurring($\mathbf{7} \times \mathbf{7}$)	22.9	0.988	0.915	0.807
medain filtering($\mathbf{5} \times \mathbf{5}$)	23.0	0.943	0.830	0.682
sharpening	23.4	1.000	0.990	0.954
Gaussian noise(**16**)	15.9	1.000	1.000	1.000

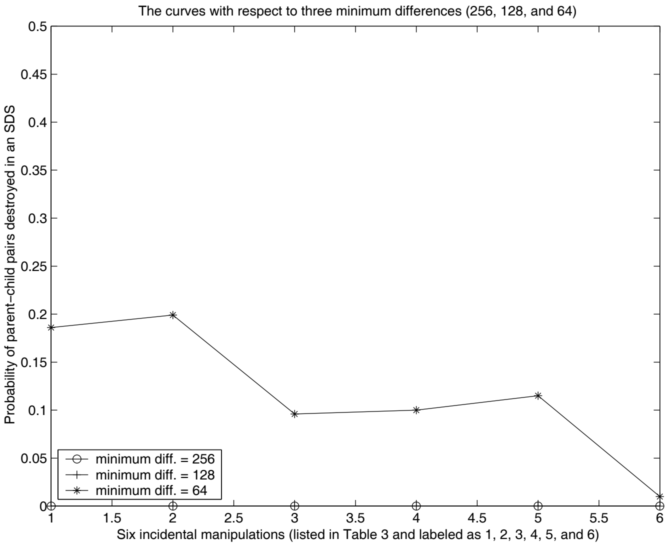


Figure 5: The probability (vertical axis) that the relationship of the parent-child pairs in an SDS might be destroyed with respect to six incidental manipulations (horizontal axis) listed in Table 3. The minimum distances (σ) used for thresholding are 256, 128, and 64, respectively.

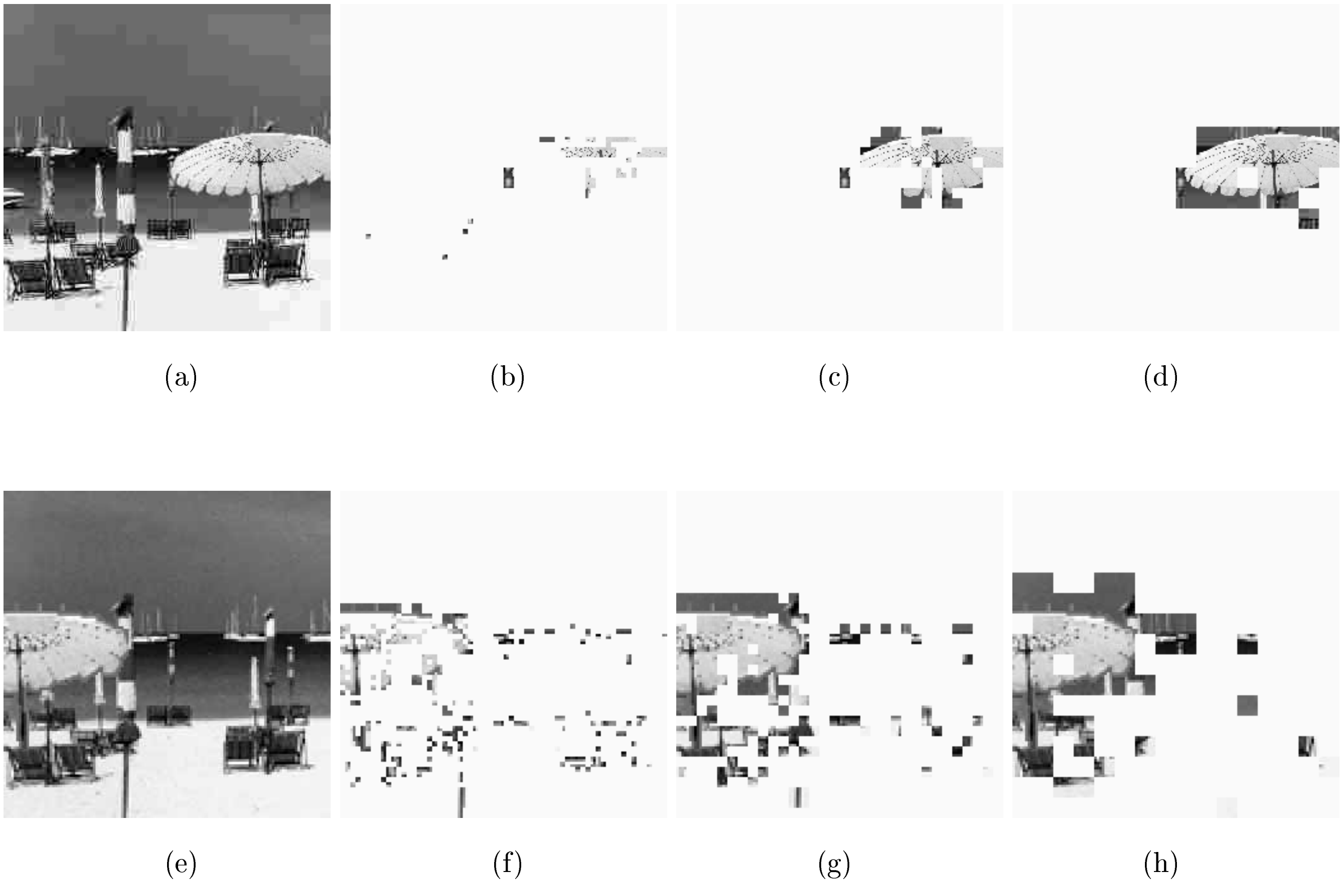


Figure 6: Combined attacks with incidental and malicious manipulations: (a) beach image after $JPEG$+"umbrella" placement; (b)$\sim$(d) detected results of (a) at $2^2 \sim 2^4$ scales when $\sigma = 128$; (e) beach image after rescaling(scaling+"umbrella" placement); (f)$\sim$(h) detected results of (e) at $2^2 \sim 2^4$ scales when $\sigma = 128$.

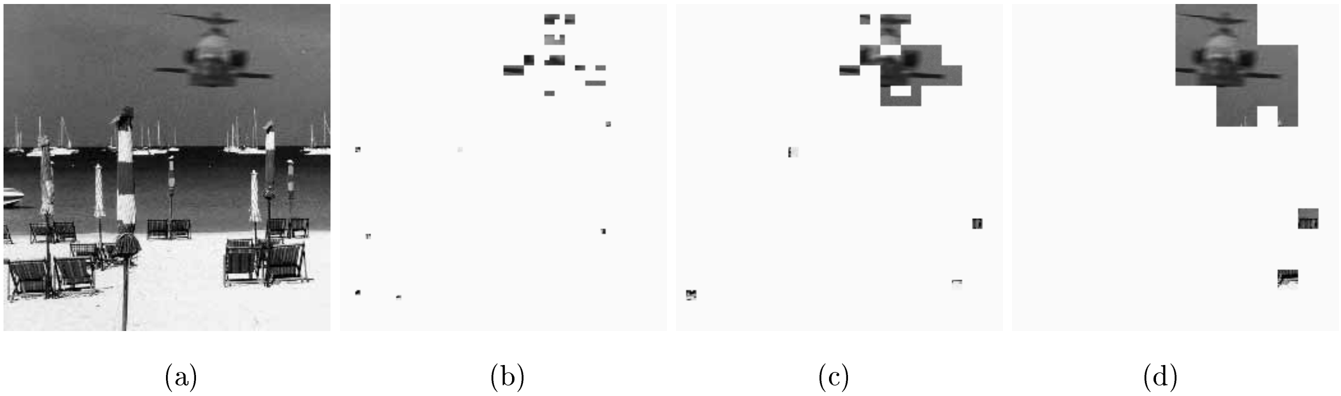


Figure 7: Malicious manipulations of non-SDS areas: (a) maliciously tampered with image with a "helicopter" in the sky; (b)$\sim$(d) detected results of (a) at $2^2 \sim 2^4$ scales when $\sigma = 128$.