Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2010, Article ID 341856, 17 pages
doi:10.1155/2010/341856

Research Article

Semi-Fragile Zernike Moment-Based Image Watermarking for Authentication

Hongmei Liu,[1] Xinzhi Yao,[2] and Jiwu Huang[1]

[1] *Department of Electronics and Communication, Sun Yat-sen University, Guangzhou 510006, China*
[2] *Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong*

Correspondence should be addressed to Hongmei Liu, isslhm@mail.sysu.edu.cn

Received 30 November 2009; Revised 17 May 2010; Accepted 6 July 2010

Academic Editor: Jin-Hua She

We propose a content-based semi-fragile watermarking algorithm for image authentication. In content-based watermarking scheme for authentication, one of the most challenging issues is to define a computable feature vector that can capture the major content characteristics. We identify Zernike moments of the image to generate feature vector and demonstrate its good robustness and discriminative capability for authentication. The watermark is generated by quantizing Zernike moments magnitudes (ZMMs) of the image and embedded into DWT (Discrete Wavelet Transform) subband. It is usually hard to locate the tampered area by using global feature in the content-based watermarking scheme. We propose a structural embedding method to locate the tampered areas by using the separability of Zernike moments-based feature vector. The authentication process does not need the original feature vector. By using the semi-fragilities of the feature vector and the watermark, the proposed authentication scheme is robust to content-preserved processing, while being fragile to malicious attacks. As an application of our algorithm, we apply it on Chinese digital seals and the results show that it works well. Compared with some existing algorithms, the proposed scheme achieves better performance in discriminating high-quality JPEG compression from malicious attacks.

1. Introduction

With the development of advanced image editing software, it has become easier to modify or forge digital image [1]. When the digital image contains important information, its credibility must be ensured. So a reliable image authentication system is necessary. Because the image can allow for lossy representations with graceful degradation, the image authentication system should be able to tolerate some commonly used incidental modification, such as JPEG compression and noise corruption. Therefore, the traditional bit-by-bit verification based on cryptographic hash is no longer a suitable way to authenticate the image. Image authentication that validates based on the content is desired [2].

In the literature, image authentication can be roughly classified into two categories, visual-hash-based [3–5] and watermark-based [6–22]. In visual-hash-based system, authentication information needs extra channel to transmit or store. In watermarked-based system, the authentication information is imperceptibly embedded in the image rather than appended to it, eliminating the extra storage requirements of visual-hash-based system [2]. The watermark-based system may be further divided into two categories, content-independent watermarking [6–11] and content-based watermarking [13–22]. The security of content-independent watermarking scheme is not so good. Due to the fact that the watermark in this kind of method is content independent and the detection of tampering is mainly based on the fragility of the hidden watermark, a wise malicious manipulation that does not change the watermark will cheat the scheme. For example, the algorithms in [6] and [7] cannot detect the modifications that are multiples of watermarking quantization steps, which may be exploited to pass an image with large modification as authentic [12].

In content-dependent watermarking scheme, the general framework for authentication includes the following parts.

(i) Generating feature vector from the host image.

(ii) Embedding quantized feature vector as watermark into the host image and getting the watermarked image.

(iii) Authenticating the test image by comparing the watermark extracted from the test image and the feature vector generating from the test image.

One of the most challenging issues of this framework is to define a feature vector. An ideal feature vector for authentication should have the following properties.

(i) It is computable and can capture the major content characteristics [12].

(ii) It is semi-fragile. It is robust to different incidental manipulations while fragile to malicious manipulations.

(iii) It has good discriminative capability. It is able to distinguish malicious manipulations from incidental ones.

Without these properties, the feature-based watermark will degenerate as a content-independent watermark in authentication.

A number of features have been proposed in content-based watermarking schemes for image authentication. In [13], Lin and Chang found that the magnitude relationship between two coefficients remains invariable through repetitive JPEG compression. The authentication could be verified by a 1-bit signature which represents the magnitude relationship between the two coefficients. It is an elegant algorithm. However, the drawback of the method is that once the DCT pairs are known, an attacker can easily modify DCT coefficients and keep the original relationship unchanged [14]. The algorithm in [15] extends and improves the scheme in [13] by generating the signature bit from the difference between two wavelet coefficients to which a random bias is added. The signature is inserted into the wavelet coefficients using nonuniform quantization-based method. Though the method of feature extraction increases the difficulty of the attacker to manipulate the feature, it cannot get the global information of the original image. In [16], the robust signature is cryptographically generated on the basis of invariant features called significance-linked connected component extracted from the image and then signed and embedded into the wavelet domain as a watermark using the quantization-based method. The algorithm of feature extraction produces too many bits of watermark information, which reduces the robustness. In [17], according to the approximation component and the energy relationship between the subbands of the detail components in DWT domain, global feature and local feature are both generated. Then the global watermark and local watermark are generated from global feature and local feature, respectively. This scheme has lower false positive probability than Lin and Chang's scheme in [13] and the false positive probability is 0.07% when quality factor of JPEG compression is 70. In [18], Tsai and Chien proposed an authentication scheme with recovery of tampered area. The features for watermark are generated from LL2 bands of DWT and embedded into the high-frequency bands. This method needs additional information to extract the watermark, and when recovery is achieved, the quality of the image degrades a lot. In [19], the entropy of the probability distribution of gray level values in block is used to generate binary feature mask. Positions of malicious manipulations can be localized. In [20], five features are generated and tested. Some are block-based local features, such as edge shape, standard deviation and mean value, and some are frame-based global features, such as edge shape and statistical feature. With global features, the location of attacked areas cannot be recognized. With local features, there are some problems in tolerance to the incidental operations, especially with the block-based edge shape feature. In [21], the image is partitioned into nonoverlapping 4×4 pixel blocks in the spatial domain. The mean values of these blocks form n-dimensional vectors, which are quantized to the nearest lattice point neighbors. However, it is not robust to JPEG compression. In [22], the authors proposed to extract content-based features from the DWT approximation subband to generate two complementary watermarks: edge-based watermark to detect the manipulations and content-based watermark to localize tampered regions.

In content-based watermarking scheme for image authentication, in order to locate the tampered areas, local feature is usually computed and embedded locally, just like the algorithms in [13, 15, 16, 19–22]. However, restricted by the embedding capacity and invisibility of the watermarked image, the watermark generated by local feature should be low bitrate. Thus the feature will not have the first property listed above and the algorithm is susceptible to attack, such as the feature in [13, 20]. Global feature can generate relatively lower bitrate watermark, but it is usually hard to locate the tampered areas, such as the global features in [20]. All the feature vectors in the existing schemes are assumed to have the second and third characteristics. However, they are not addressed and analyzed explicitly.

In this paper, we propose to use Zernike moments to generate feature vector. By using this global feature, we can decide whether the image is maliciously manipulated or not and locate the tampered areas. At first, we identify Zernike moments to generate feature vector and demonstrate its good semi-fragile and discriminative capability for authentication. Moments have been utilized as pattern features in many applications to achieve invariant recognition of image pattern. Of various types of moments defined in the literature, Zernike moments have been shown to be superior to the others in terms of their insensitivity to image noise, information content, and ability to provide faithful image representation [23] and thus have been used in many applications [24–28], for example, invariant watermarking [26–28] to resist RST (rotation, scale, and translation) manipulations. But there is little research on the semi-fragility and discriminative capability of Zernike moments when different kinds of manipulations are applied to the image in authentication application. In this paper, we analyze and demonstrate these properties of Zernike moments.

Then, we propose a Zernike moments-based semi-fragile watermarking algorithm in DWT domain. It is usually hard to locate the tampered areas using global feature. We propose a structural embedding method to solve this problem by using the separability of Zernike moments feature vector,

which can be separated into individual moments. The authentication process u*ses* a two-stage decision method. In the first stage, we decide if the test image is maliciously manipulated by a metric measure. In the case of malicious manipulation, we further locate the tampered areas in the second stage.

Experimental results show that the proposed authentication scheme has better performance in discriminating high-quality JPEG compression from malicious manipulations when compared with some existing methods. We also test the performance of the proposed method under the situation in which malicious manipulation is followed by other manipulations. Under this situation, the system can work well too. Our scheme can be used on different kinds of images. The experiments on Chinese digital seals support this conclusion.

The paper is organized as follows. Section 2 describes the Zernike moments and their semi-fragile property. The outline of the proposed system, content-based watermark and its structural embedding method, and how to authenticate an image are described in Section 3. Section 4 demonstrates the experimental results and the analysis. Conclusions and discussions of future works are shown in Section 5.

2. Zernike Moments Magnitudes and Semi-Fragile Property

In content-based watermarking scheme for image authentication, extraction of feature vector is one of the most challenging issues. An ideal feature vector should have three properties listed in Section 1. In this section, we propose to generate feature vector based on Zernike moments and analyze the properties of this feature vector. The invariance of Zernike moments, that is, the robustness to geometric distortions, has been investigated by the authors of [24, 26, 28]. But the semi-fragile property of Zernike moment has not been investigated in literature. In this section, we will demonstrate this property and explain how to discriminate malicious manipulations from incidental manipulations by using it. Some of the materials in the following are based on [24, 28].

2.1. Zernike Moment. In [29], Zernike introduced a set of complex polynomials that form a complete orthogonal set over the interior of the unit circle, $x^2 + y^2 = 1$. Let the set of these polynomials be denoted by $\{V_{nm}(x, y)\}$. The polynomials can be expressed as

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho) \exp(jm\theta), \quad (1)$$

where n is a non-negative integer and m is an integer such that $n - |m|$ is non-negative and even. ρ and θ represent polar coordinates over the unit circle and R_{nm} are polynomials of ρ (Zernike polynomials) given by

$$R_{nm}(\rho) = \sum_{s=0}^{n-|m|/2} \frac{(-1)^s[(n-s)!]\rho^{n-2s}}{s!((n+|m|/2)-s)!((n-|m|/2)-s)!}. \quad (2)$$

Note that $R_{n,-m}(\rho) = R_{n,m}(\rho)$. These polynomials are orthogonal and satisfy

$$\int_{x^2+y^2\le 1} [V_{nm}^*(x, y)] \times V_{pq}(x, y)\, dxdy = \frac{\pi}{n+1}\delta_{np}\delta_{mq} \quad (3)$$

with

$$\delta_{ab} = \begin{cases} 1 & a = b, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

Zernike moments are the projection of the image function onto these orthogonal basis functions. The Zernike moment of order n with repetition m for a continuous image function $f(x, y)$ that vanishes outside the unit circle is

$$A_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2\le 1} f(x, y)\, V_{nm}^*(\rho, \theta)\, dxdy. \quad (5)$$

For a digital image, we have

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x, y)\, V_{nm}^*(\rho, \theta), \quad x^2 + y^2 \le 1. \quad (6)$$

To compute the Zernike moments of a given image, the center of the image is taken as the origin and the pixel coordinates are mapped to the range of the unit circle. Those pixels falling outside the unit circle are not used in the computation. Note that $A_{nm}^* = A_{n,-m}$.

Suppose that one knows all moments A_{nm} up to order $N_{\max}$ of $f(x, y)$. Using orthogonality of the Zernike basis, we can reconstruct the image $f(x, y)$,

$$\hat{f}(x, y) = \sum_{n=0}^{N_{\max}} \sum_m A_{nm} V_{nm}(\rho, \theta) \quad (7)$$

Note that as $N_{\max}$ approaches infinity, $\hat{f}(x, y)$ will approach $f(x, y)$.

The reconstruction process is illustrated in Figure 1. For a 64×64 gray image of letter A, the reconstructed images are generated by using (7) followed by mapping the pixel value to $[0, 255]$. It shows that the lower-order moments capture gross shape information and the high-frequency details are filled in by higher-order moments.

According to the research in [24] and our experiments, Zernike moments with 12-order have a good trade-off between performance (detecting accuracy) and computation complexity, which will be illustrated in Section 2.2.

2.2. Semi-Fragile Property of Zernike Moments-Based Feature Vector. In authentication, semi-fragile means that the feature vector is robust to commonly used incidental modifications that preserve the perceptual quality while fragile to malicious manipulations. Although classification of incidental and malicious manipulations depends on a specific application, in most cases, JPEG compression and slight noise corruption are generally regarded as incidental manipulation, while cut and replace as malicious manipulations. We adopt this

(a) (b) 4-order (c) 8-order (d) 12-order (e) 15-order

Figure 1: Reconstruction of a gray image. From left to right: the original image, the reconstructed image with order 4, 8, 12 and 15, respectively.

(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 2: Some example images.

point of view and investigate the semi-fragile property of the Zernike moments-based feature vector. We also verify the robustness of Zernike moments to rotation through experiments. The moments are computed by keeping the size of manipulated image unchanged.

The semi-fragile property is described by the distance between two images. Each image is represented by a N-dimensional feature vector and the distance is computed on two feature vectors. Smaller distance means better match of the images. The distance between two feature vectors may be measured using Euclidean distance [24]. In this paper, we use absolute difference to simplify the computation. The distance SE (Simplified Euclidean distance) is defined as

$$\begin{aligned} \mathrm{SE}(f_1(x,y), f_2(x,y)) &= \mathrm{SE}(Z_1, Z_2) \\ &= \sum_{i=1}^{N} \left| ZMM_{1,i} - ZMM_{2,i} \right|, \end{aligned} \tag{8}$$

where Z_1 and Z_2 are the feature vectors of the images $f_1(x, y)$ and $f_2(x, y)$. $Z_i = (ZMM_{i,1}, ZMM_{i,2}, \ldots, ZMM_{i,N}) = (|A_{00}|, |A_{11}|, |A_{20}|, \ldots, |A_{N_{\max}N_{\max}}|)$, where $ZMM_{i,k}$ is the kth Zernike moment magnitude of the feature vector Z_i.

Assume that $f_2(x, y)$ is obtained by processing $f_1(x, y)$. We measure the distance between the feature vectors of $f_1(x, y)$ and $f_2(x, y)$. Then we address the difference of the distance when the following different kinds of manipulations are applied to $f_1(x, y)$ and get $f_2(x, y)$.

The experiments are conducted on 300 256×256 images that come from [30]. Some of them are shown in Figure 2. Each image is processed by

(i) JPEG with QF $\in$ [90, 80, 70, 60, 50, 40, 30, 20],

(ii) additive noise with varying strength $S_n \in [1, 2, 3, 4, 5, 6]$ and $[-5\ S_n, 5\ S_n]$ noises are added randomly,

(iii) rotation with increasing angle $\in [5^\circ, 15^\circ, 25^\circ, 35^\circ, 45^\circ]$,

Table 1: Comparison of 8-order, 12-order, and 15-order Zernike moments.

		8-Order	12-Order	15-Order
Distinguishing Ability	Incidental SEs identified as malicious	31	65	95
	Malicious SEs identified as incidental	366	327	316
Computation time (second) for a 256×256 image		1.6607	4.1001	6.9235

(iv) cutting out blocks at randomly chosen areas. The block sizes are 16 by 16, 24 by 24, 32 by 32, 40 by 40, and 48 by 48, respectively,

(v) Replacing the cut block by other content. The block sizes are 16 by 16, 24 by 24, 32 by 32, 40 by 40, and 48 by 48, respectively.

The first three kinds of manipulations are regarded as incidental ones, while the last two kinds of manipulations are regarded as malicious ones. Thus we get 29 processed images for each original image. Totally we have 8700 processed images. We measure the distance between Zernike moments based feature vectors of the original image and its manipulated image by (8). Zernike moments of 8-order (25 moments), 12-order (49 moments), and 15-order (72 moments) are tested in experiments. The results are shown in Figure 3. Figures 3(a), 3(c), and 3(e) demonstrate the distribution of the distances, where x-axis represents manipulations and y-axis is $\log_{10}(SE(f_1(x, y), f_2(x, y)))$. From Figures 3(a), 3(c), and 3(e), we can see that distances between the feature vectors of the original images and their incidentally manipulated images are usually much smaller than those between the feature vectors of the original images and their maliciously manipulated images, and thus can be classified into two groups. One group includes most of the distances obtained from the incidental manipulations and another includes most of those obtained from the malicious manipulations. We also give the histograms of the distances, one for the incidental manipulations and the other for the malicious manipulations, which are shown in Figures 3(b), 3(d), and 3(f), where x-axis represents the distance and y-axis is the number of occurrences of the distance. From Figures 3(b), 3(d), and 3(f), we can see that two histograms are separated clearly. Figure 3 tells that we can separate these two kinds of manipulations by using the following rule:

$$\text{decision} = \begin{cases} \text{Malicious}, & \text{SE}(f_1(x, y), f_2(x, y)) > T_1, \\ \text{Incidental}, & \text{otherwise}, \end{cases} \tag{9}$$

where T_1 is a predefined threshold, which will be given in Section 4 through experiments.

Obtained from Figure 3, we also list in Table 1 the performance of distinguishing incidental from malicious attacks for 8-order, 12-order, and 15-order Zernike moments by using the SEs. The computing time of Zernike moments for a 256×256 test image with individual order is also given. As can be seen in Table 1, when the order grows from 8 to 15, incidental SEs are more easily regarded as malicious ones while malicious SEs are less easily regarded as incidental ones; at the same time, the computing time increases gradually. Thus, 12-order Zernike moments would gain an overall better performance by considering the distinguishing ability and computing complexity, compared with 8-order and 15-order Zernike moments. In the following sections, we will adopt 12-order, 49 Zernike moments to generate the feature vector. The detailed distributions of 12-order SE*s* used in our experiments are illustrated in Figure 4.

Assume that $\mathbf{f_2}(\mathbf{x}, \mathbf{y})$ is obtained by cutting a block from $\mathbf{f_1}(\mathbf{x}, \mathbf{y})$. We also conduct the experiments to address the relationship between $\mathbf{SE}(\mathbf{f_1}(\mathbf{x}, \mathbf{y}), \mathbf{f_2}(\mathbf{x}, \mathbf{y}))$ and the size of cut block in the image. The results on the images in Figure 2 are shown in Figure 5, where x-axis is the size of the cut block and y-axis is $\mathbf{SE}(\mathbf{f_1}(\mathbf{x}, \mathbf{y}), \mathbf{f_2}(\mathbf{x}, \mathbf{y}))$. We can observe that the distance between the original image and the processed image becomes larger when the size of the cut block increases. It means that the distance of feature vector can reflect the degree of the content change of the image.

3. Proposed Authentication Algorithm

In this section, the Zernike moments-based watermarking algorithm for authentication is given. The framework, the structural embedding method of the Zernike moments-based watermark, the location of the tampered areas, and the authentication process are described.

3.1. The Framework of the Proposed Scheme. Figure 6 gives the block diagrams of the embedding and authentication processes.

The embedding steps are as follows.

(i) Compute 49 ZMMs of the host image $f_1(x, y)$. Each ZMM is quantized to 12 bits and the 9 most significant bits are selected to be part of the watermark.

(ii) Apply 3-level DWT to $f_1(x, y)$and get 10 subbands, LL_3, HL_3, LH_3, HH_3, HL_2, LH_2, HH_2, HL_1, LH_1, HH_1, where the low frequency subband LL_3 is a low pass approximation of the original image.

(iii) The watermark generated from ZMMs is structurally embedded in LL_3 subband.

(iv) IDWT is applied and the watermarked image is obtained.

The authentication steps are as follows:

(i) Compute 49 ZMMs of the test image $f_2(x, y)$.

(ii) Apply 3-level DWT to $f_2(x, y)$ and extract watermark from LL_3 subband. The watermark is restored as 49 ZMMs, which is the estimation of 49 ZMMs of the original host image $f_1(x, y)$.

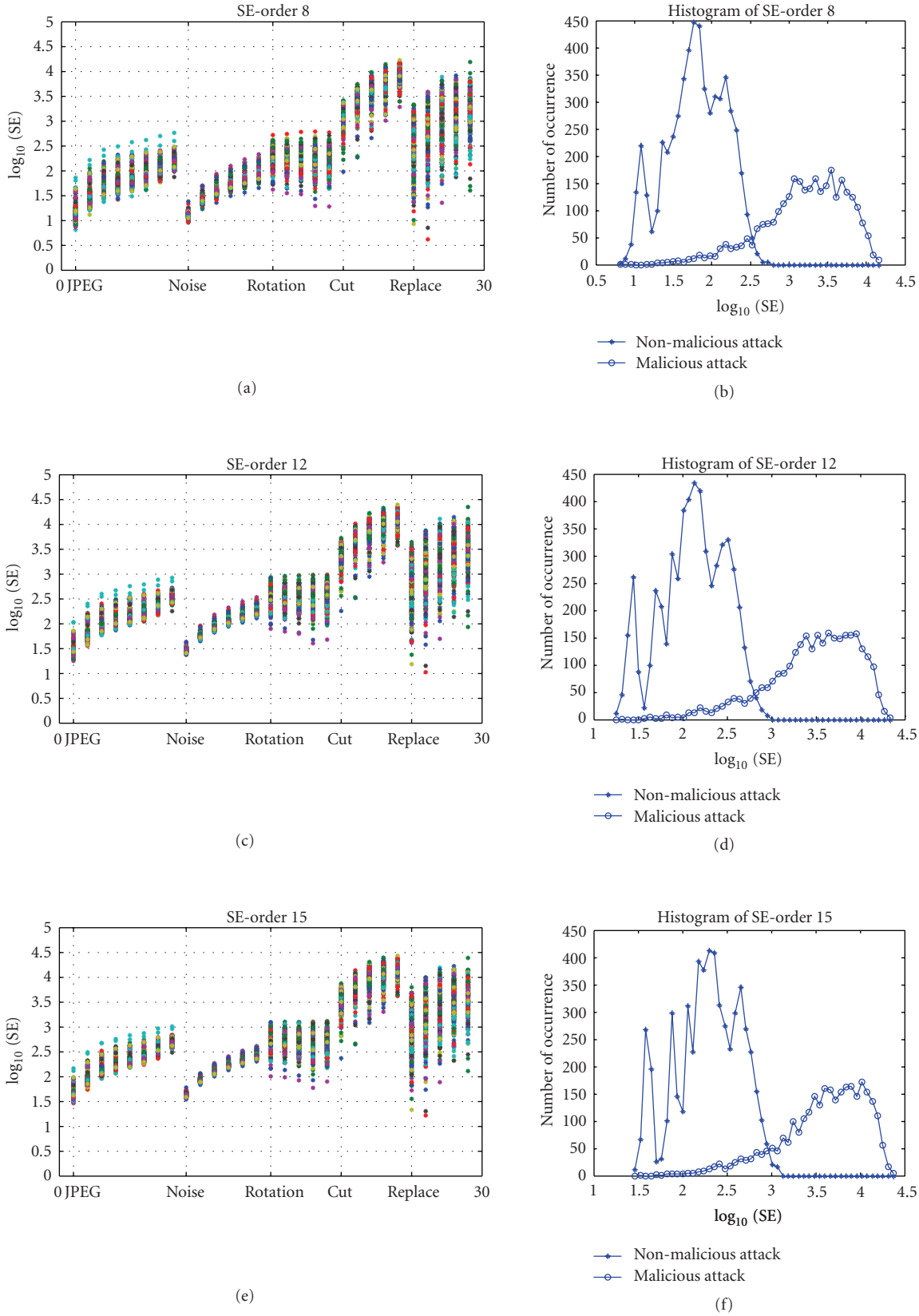

Figure 3: The distribution of the distances.

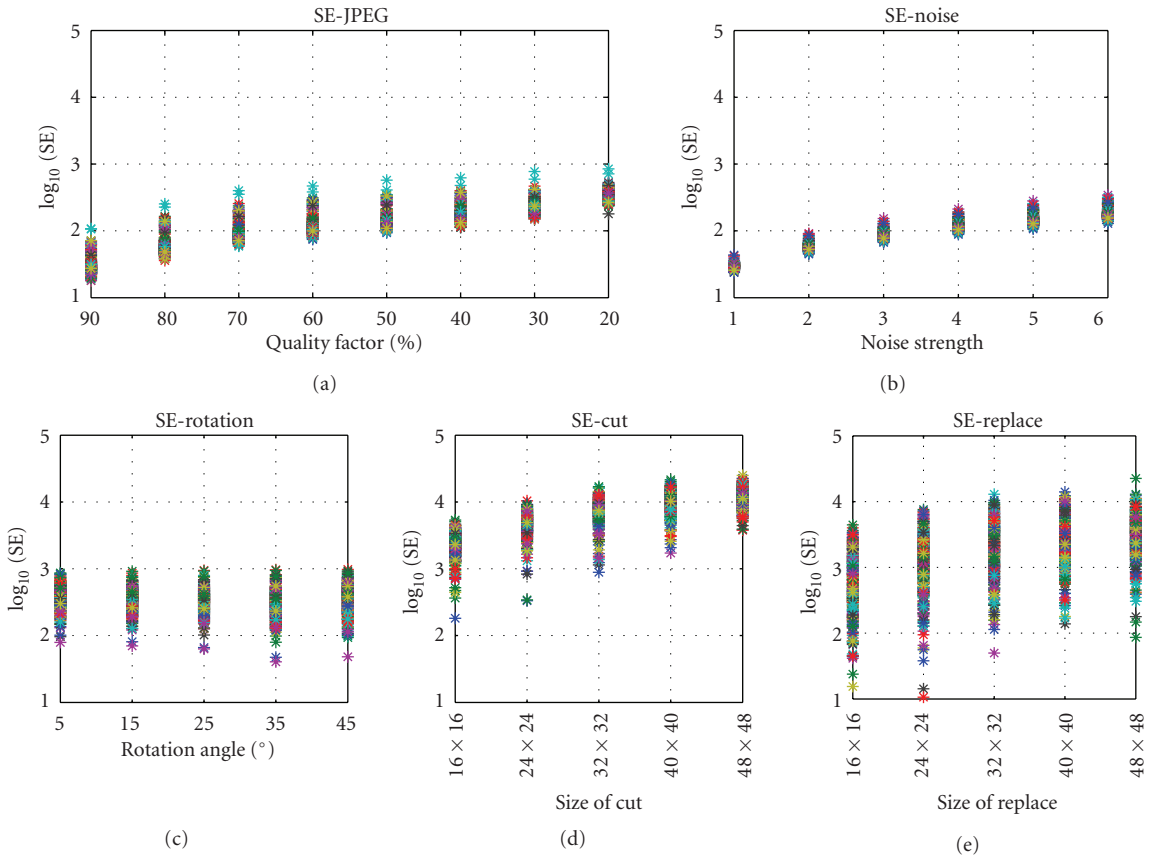

Figure 4: The distribution of SEs in order 12.

(iii) The first decision stage. Compute $SE(f_1(x, y), f_2(x, y))$and compare it with a predefined threshold to decide whether the test image is authentic or not. In the case of inauthentic, go to next step.

(iv) The second decision stage. Locate the attacked area by using the structure of the embedded watermark.

3.2. Structural Embedding Method and Location of Attacked Area. In content-based watermarking scheme, it is usually hard to locate the tampered areas by using global feature. In our system, we locate the tampered regions using the blockwise method by resorting to the separability of the Zernike moments-based feature vector and the change of watermark.

From the description in Section 2, we can know that the Zernike moments-based feature vector is composed by individual ZMMs. Each ZMM can be embedded separately into a block. When some parts of the watermarked image are changed, the ZMMs embedded in these areas will be changed and thus can be used to locate the tampered areas. The structural embedding method is as follows.

(i) LL_3 subband is segmented into nonoverlapped 3×3 blocks.

(ii) For each $ZMM_{1,i}$ in the feature vector Z_1 of $f_1(x, y)$, we randomly select a block by a secret key to embed it. If the blocks are more than ZMMs in number, then some of ZMMs can be embedded repeatedly. The secret key can be used to improve the security of the scheme.

(iii) $ZMM_{1,i}$ is embedded in the selected block with one bit in one coefficient. The embedding method we adopted can be found in [31],

$$A'(i) = A(i) - A(i) \bmod S_w + \frac{3}{4}S_w \quad \text{if } X = 1,$$
$$A'(i) = A(i) - A(i) \bmod S_w + \frac{1}{4}S_w \quad \text{if } X = 0, \tag{10}$$

where $A(i)$ and $A'(i)$ are the DWT coefficients before and after embedding, respectively. X is the watermark bit. S_w is the watermark strength which is a positive natural number. The watermark bit X'can be extracted by the following method:

$$A'(i) \bmod S_w \geq \frac{1}{2}S_w \quad \text{then } X' = 1,$$
$$A'(i) \bmod S_w < \frac{1}{2}S_w \quad \text{then } X' = 0, \tag{11}$$

Denote $ZMM_{1,i}^{(j)}$ and $\hat{Z}MM_{1,i}^{(j)}$ are the ith ZMMs in Z_1 embedded in and extracted from the selected jth block, respectively. The authentication process is as follows.

(i) Compute 49 ZMMs, $ZMM_{2,i}$ ($i = 1 - 49$), of the feature vector Z_2 of the test image $f_2(x, y)$.

(ii) Extract the watermark and get $\hat{Z}MM_{1,i}^{(j)}$ from each block of LL_3 subband of $f_2(x, y)$.

(iii) In the first stage, the authenticity of the image is decided by the following rule

$$\text{decision} = \begin{cases} \text{Malicious} & SE(f_1(x,y), f_2(x,y)) \\ & = SE\left(\tilde{Z}_1, Z_2\right) > T_1, \\ \text{Incidental} & \text{otherwise}, \end{cases} \tag{12}$$

where T_1 is a predefined threshold. $\tilde{Z}_1$ is the estimation of Z_1 and restored from the extracted watermark $\hat{Z}MM_{1,i}^{(j)}$ by averaging those with same i.

(iv) In the second stage, the tampered areas are located by the following rule:

$$\text{decision} = \begin{cases} j\text{th block is attacked} & \left|\tilde{Z}MM_{1,i}^{(j)} - \hat{Z}MM_{1,i}^{(j)}\right| \\ & > T_2, \\ j\text{th block is not attacked} & \text{otherwise}, \end{cases} \tag{13}$$

where T_2 is a predefined threshold and $\tilde{Z}MM_{1,i}^{(j)}$ are the estimation of $ZMM_{1,i}^{(j)}$. In our scheme, they are estimated from Z_2. That is, we assume that each $ZMM_{2,i}$ in Z_2 is embedded and get its corresponding block by the same secret key used in embedding side and get $\tilde{Z}MM_{1,i}^{(j)}$. We will demonstrate that it is reasonable to estimate $ZMM_{1,i}^{(j)}$ from Z_2 by an example in the following part.

There are three parameters in our schme. T_1 in (12) can be selected by the ROC (Receiver Operator Characteristic, shown in Section 4) of the scheme and the requirements of the false positive probability and the false negative probability. T_2 in (13) is set as 512 by extensive experiments and S_w is chosen to be 64.

Figure 7 demonstrates the method of locating the tampered area. Figures 7(a_1), 7(a_2), and 7(a_3) are the original image $f_1(x, y)$, the watermarked image, and the maliciously manipulated image $f_2(x, y)$. The cars on the road of Figure 7(a_2) are copied and pasted to get Figure 7(a_3). The differences between $ZMM_{1,i}$ and $ZMM_{2,i}$ of Figure 7(a_1) and Figure 7(a_3) are shown in the left image of Figure 7(a_4). X-axis represents serial number of ZMMs and y-axis represents $|ZMM_{1,i} - ZMM_{2,i}|$. The errors between the extracted watermark $\hat{Z}MM_{1,i}^{(j)}$ from jth block of Figure 7(a) and the original watermark $ZMM_{1,i}^{(j)}$ embedded in jth block are shown in the right image of Figure 7(a_3). X-axis represents the serial number of the block in LL_3 subband and y-axis represents $|ZMM_{1,i}^{(j)} - \hat{Z}MM_{1,i}^{(j)}|$. From Figure 7($a_4$), we can observe that malicious manipulation introduces much greater changes to the embedded watermarks in the tampered blocks than to the individual components of the feature vector. So using the estimated watermark $\tilde{Z}MM_{1,i}^{(j)}$ in (13) will not affect the locating of tampered areas too much. The error between the extracted watermark $\hat{Z}MM_{1,i}^{(j)}$ and the estimated watermark $\tilde{Z}MM_{1,i}^{(j)}$ is shown in Figure 7(a_5). X-axis represents the serial number of the block in LL_3 subband and y-axis represents $|\tilde{Z}MM_{1,i}^{(j)} - \hat{Z}MM_{1,i}^{(j)}|$. We can observe that the bursts in the right image of Figure 7(a_4) are still kept in Figure 7(a_5). Figure 7(a_6) shows the location result by comparing the errors in Figure 7(a_5) with T_2. From Figure 7, we can see that the structural embedding method is effective in locating the tampered areas by resorting to the location of the changed watermark.

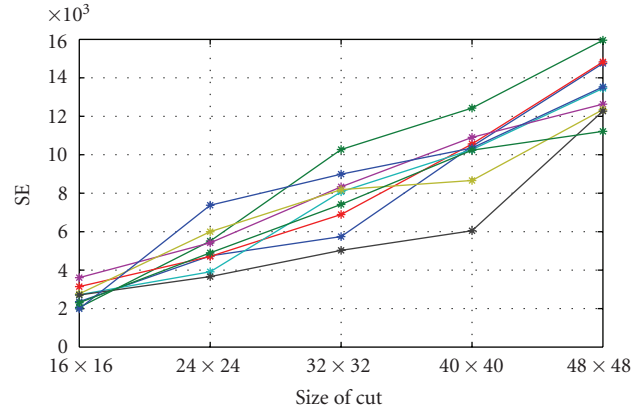

Figure 5: The relationship between distance and the size of cut block.

3.3. The Robustness of Watermark to Incidental Manipulations. The robustness of watermark to incidental manipulations is very important in authentication, because the extracted watermark is used to estimate original feature vector of the image and decide if the test image is authentic. We measure the robustness of the watermark by computing the distance between the original feature vector of the image and the estimated feature vector from the extracted watermark by (8). The experiments are conducted on the 300 images used in Section 2.2. Each watermarked image is processed by

(i) JPEG with QF $\in [90, 80, 70, 60, 50]$,

(ii) additive noise with varying strength $S_n \in [1, 2, 3, 4, 5]$.

The histogram of the distance is shown in Figure 8, where x-axis represents the distance and y-axis is the occurrence number of the distance. From Figure 8, we can see that most of the distance is zero. It means that the extracted watermark is equal to the embedded watermark in most cases and thus the watermark is robust to high-quality JPEG compression and noise.

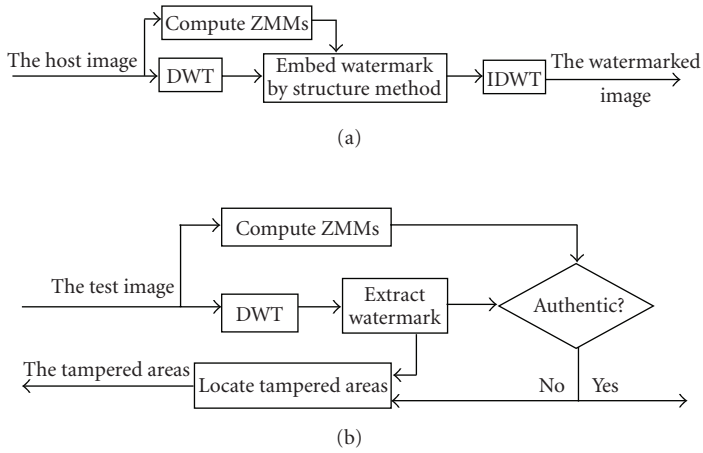

Figure 6: The framework of the proposed scheme: (a) embedding process (b) authentication process.

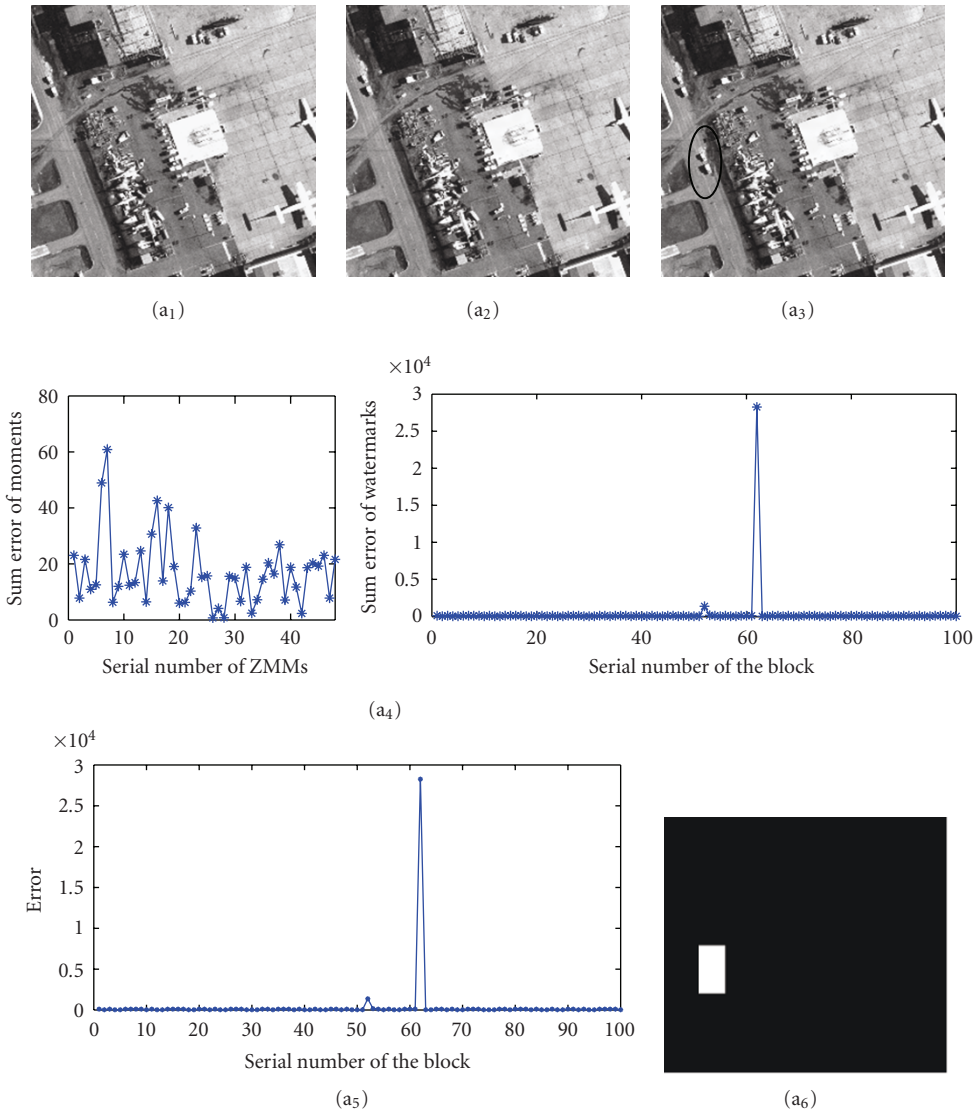

Figure 7: Demonstration of the location method of the attacked area.

Table 2: Some P_{fp} and P_{fn}.

T_1	Number of the false negative image	P_{fn}	Number of the false positive image	P_{fp}
2680	7	0.0023	77	0.0257
2820	10	0.0033	69	0.0230
3000	14	0.0047	65	0.0217
3320	17	0.0057	63	0.0210
3940	20	0.0067	61	0.0203
4300	25	0.0083	60	0.0200
4900	30	0.0100	59	0.0197
6700	44	0.0147	49	0.0163
8000	56	0.0187	47	0.0157
9000	70	0.0233	44	0.0147

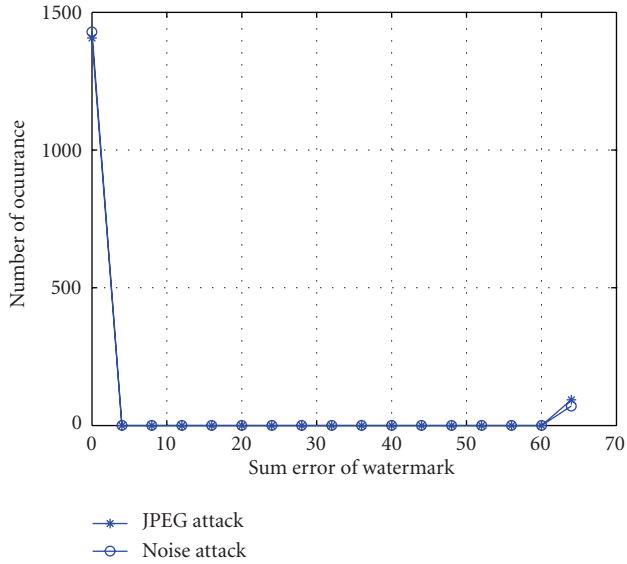

Figure 8: The robustness of watermark to incidental manipulations.

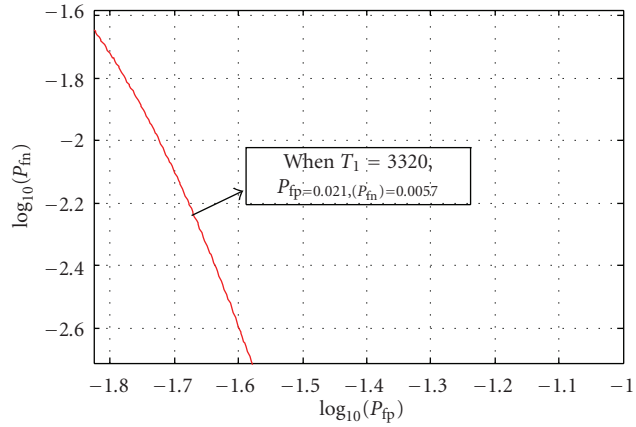

Figure 9: ROC curve.

4. Experimental Results

To demonstrate the power of our authentication system, we study the ROC of the scheme and set the threshold $\mathbf{T_1}$. Then we present some results obtained by applying only malicious or incidental manipulation on standard test images and Chinese seal images. We also demonstrate the results of locating the tampered areas when the image is processed by combining malicious manipulation with JPEG compression, sharpening, or blurring. Comparisons with some existing schemes will also be presented.

4.1. ROC and Threshold. Experiments are performed on 300 images that come from [30], which do not include the images used in Section 2. All of these images are watermarked and then processed by two kinds of manipulations as follows.

(i) Malicious attacks. Adding, erasing, and replacing something with different sizes.

(ii) Non-malicious manipulations. Compressing by JPEG with QF $\in$ [90, 80, 70, 60, 50] and adding Gaussian noise with strength $S_n \in [1, 2, 3, 4, 5]$.

We generate 6000 processed images. Among them 3000 images are produced by incidental manipulations and 3000 images are generated by malicious attacks. P_{fp} and P_{fn} are used to represent the false positive probability and the false negative probability, respectively. Some P_{fp} and P_{fn} under different thresholds are shown in Table 2. Our observation shows that the false positive image usually is the JPEG compressed image with QF 50 and the false negative image is usually the maliciously manipulated image with small size content change. The ROC of the scheme is shown in Figure 9, where x-axis is $\log_{10}(P_{\text{fp}})$ and y-axis is $\log_{10}(P_{\text{fn}})$. The thresholds are between 2680 and 9000. T_1 is set as 3320 in our experiments because we can get relatively low P_{fp} and P_{fn} at the same time by using this threshold.

4.2. Authentication Results When Single Attack Is Applied. The experiments are firstly conducted on the standard test images in Figure 10. The PSNRs of their watermarked images are shown in Table 3. Table 4 lists the authentication results when JPEG compressions are applied to their watermarked images. Figure 11 shows the tamper localization results when malicious attacks are applied to some of them. Then we

Figure 10: The test images.

Table 3: PSNRs obtained by watermarking the images in Figure 10.

Image in Figure 10	I_{01}	I_{02}	I_{03}	I_{04}	I_{05}	I_{06}	I_{07}	I_{08}	I_{09}	I_{10}
PSNR (dB)	42.6	42.5	42.3	42.2	42.4	42.6	42.7	42.0	42.3	42.8
Image in Figure 10	I_{11}	I_{12}	I_{13}	I_{14}	I_{15}	I_{16}	I_{17}	I_{18}	I_{19}	I_{20}
PSNR (dB)	42.1	42.6	42.4	42.3	42.8	42.7	42.3	42.9	42.5	42.9

conduct experiments on Chinese seal images in Figure 12 and show the authentication results when malicious attacked are applied to the watermarked images. Table 5 lists the authentication results when JPEG compressions are applied to their watermarked images. From Tables 4 and 5, we can see that our system can successfully pass almost all the JPEG-compressed images with QF as low as 40. As for the additive Gaussian noise, our scheme can tolerate noisy images with PSNR as low as 33.6 dB. From the experiment results, we can see that the proposed scheme is robust to JPEG compression while sensitive to malicious manipulations with good capability in locating the attacked areas.

4.3. Authentication Results When Combined Attacks Are Applied. The objective of this section is to check whether our scheme can successfully detect and locate a malicious manipulation when some other manipulations are applying to the image simultaneously. We apply two-stage decision method. The authenticity of the test image is firstly decided. We observe that combined manipulations introduce more changes to the watermark and the feature vector than single manipulation. In first stage, $\mathrm{SE}(f_1(x,y), f_2(x,y)) > T_1$ in (12) is true and the image is regarded as maliciously manipulated. Figure 13 shows the tampering and location results in the second stage. The manipulations following

Figure 11: Authentication results when some standard test images are maliciously manipulated where I: original standard image, WI: watermarked image, TI: tampered watermarked image and the oval highlights the tampered part, LI: location of the attacked areas.

malicious tampering include JPEG compressions, blurring, and sharpening. In order to compare with the algorithm in [8], we adopt the same symbols. We can see that our scheme can work well in most cases. In the case of a combined manipulation involving JPEG, Figure 13 indicates that when the quality factor is as low as 40, the detection result is still good. In the case of a combined manipulation involving blurring, the detection result is good when window size is 3×3 and becomes worse when window size increases. In the case of a combined manipulation involving sharpening, the results are good when the sharpening factor is smaller than 50. When the sharpening factor exceeds 50, the result becomes worse when the factor increases.

4.4. Performance Comparison. In authentication, one of the most important issues is discriminating the incidental

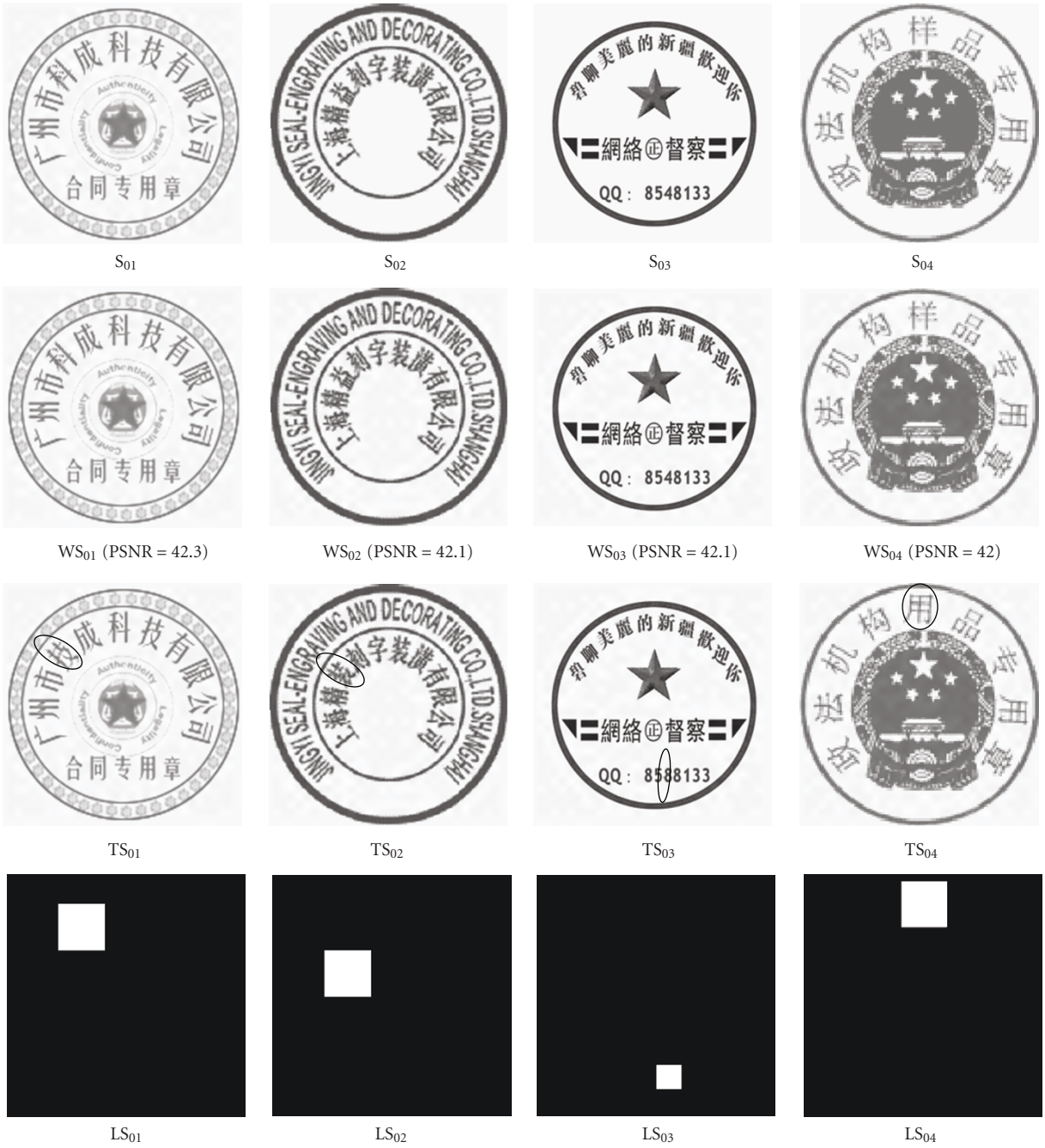

Figure 12: Authentication results when some Chinese digital seal images are maliciously manipulated where S: original seal image, WS: watermarked seal image, TS: tampered watermarked seal image and the oval highlights the tampered part, LS: location of the attacked areas.

and malicious attacks. Conventional content independent watermarking approaches, such as the schemes in [7, 8, 11], do not provide a rational metric measure for the discriminating. They use the detected attacked areas to decide whether the image is maliciously attacked. Because incidental manipulations can introduce error of watermark which may be mistaken as the result of maliciously attack, sometime the scheme does not work well. For example, in [8], the scheme works very well on 11 of 12 test images in Figure 10 and passes JPEG compressed images with QF as low as 30 as authentic. But for image I_{20} in Figure 10, the JPEG compressed image with QF as high as 70 is still mistaken as maliciously attacked image. The scheme in this paper gives a two-stage scheme and a metric measure for the discriminating. For 20 images in Figure 10, this measure can pass most JPEG compressed images with QF as low as 40. The comparison between our algorithm and that in [8] can be found in Tables 6 and 7, where Table 6 demonstrates the performance of discriminating when only JPEG compression is applied to the images and Table 7

Figure 13: The detection results when combined attacks are applied to watermarked image. W_image and T_image denote the watermarked image and the tampered watermarked image, respectively. The oval in T_image highlights the tampered part. The symbols T, J, B and S denote malicious tampering, JPEG-compression, blurring and sharpening, respectively. + means followed by. The number following each symbol is the parameter adopted by the manipulation in Photoshop.

Table 4: Authentication results when JPEG compressions are applied to the corresponding watermarked images. $\surd$ means that our scheme regards the manipulation is incidental and $\times$ means that our scheme regards the manipulation is malicious.

	Image in Figure 10									
Manipulation	I_{01}	I_{02}	I_{03}	I_{04}	I_{05}	I_{06}	I_{07}	I_{08}	I_{09}	I_{10}
JPEG ($QF > 40$)	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
JPEG ($QF = 40$)	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\times$	$\surd$	$\surd$
JPEG ($QF = 30$)	$\times$	$\times$	$\times$	$\surd$	$\times$	$\surd$	$\times$	$\times$	$\surd$	$\times$
	Image in Figure 10									
Manipulation	I_{11}	I_{12}	I_{13}	I_{14}	I_{15}	I_{16}	I_{17}	I_{18}	I_{19}	I_{20}
JPEG ($QF > 40$)	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
JPEG ($QF = 40$)	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
JPEG ($QF = 30$)	$\times$	$\surd$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$

shows the detection results when combined manipulations are applied to the images. From Table 6 and the experimental results in Sections 4.2 and 4.3, we can see that our scheme is more stable in discriminating high-quality JPEG compression from malicious attacks than the approach in [8] and can be used on different kinds of images. Table 7 shows that our scheme can give similar detection results for the maliciously attacked areas as the approach in [8], but the scheme in [8] uses the original watermark in the authentication process.

Comparisons with some other content-independent [7, 9–11] and-dependent [22] watermarking approaches for authentication are listed in Table 8. From Table 8, we can see that the performance of discriminating JPEG from the

Table 5: Authentication results when JPEG compressions are applied to the watermarked seal images. √ and × have the same meanings as those used in Table 4.

	Images in Figure 12			
Manipulation	WS_{01}	WS_{02}	WS_{03}	WS_{04}
JPEG $(QF > 50)$	√	√	√	√
JPEG $(QF = 50)$	√	√	√	√
JPEG $(QF = 40)$	√	√	×	√
noise $(S_n \leq 5)$	√	√	√	√
noise $(S_n = 6)$	√	×	√	×

Table 6: Comparison with the scheme in [8] on images I_{18} and I_{20} in Figure 10.

	The scheme in [8]	The proposed scheme	The scheme in [8]	The proposed scheme
	Images in Figure 10			
Manipulation	I_{18}	I_{18}	I_{20}	I_{20}
JPEG $(QF >= 80)$	√	√	√	√
JPEG $(QF = 70)$	√	√	×	√
JPEG $(QF = 60)$	√	√	×	√
JPEG $(QF = 50)$	√	√	×	√
JPEG $(QF = 40)$	√	√	×	√
JPEG $(QF = 30)$	√	×	×	×
JPEG $(QF = 20)$	×	×	×	×

Table 7: The comparison of the performance of detecting attacked areas when combined manipulations are applied to the image I_{18} in Figure 10. The symbols and the numbers have the same meanings as those in Figure 12.

	$T + B$	$T + J$	$T + S$
Proposed algorithm	3×3	40	40
The scheme in [8]	3×3	40	40

malicious manipulation of our scheme is superior to those of the algorithms in [7, 9–11, 22].

Comparisons with some content-based watermarking approaches in [13, 17] are shown in Table 9, where the data in the last two columns come from [17]. From Table 9, we can see that our scheme has better robustness to high-quality JPEG compression.

5. Conclusion and Future Works

In this paper, we propose a content-based watermarking scheme for image authentication. The contributions of this paper are as follows:

(1) to have found the semi-fragile property of the Zernike Moments-based feature vector.

(2) to have proposed to use Zernike feature vector as the feature in image authentication. Extensive experiments show that Zernike moments have good robustness and discriminating capability for authentication,

(3) to have proposed a two-stage decision method in authentication process and a metric measure for discriminating the incidental manipulations from malicious attacks,

(4) by using the separability of Zernike moments-based feature vector, a structural embedding method for the ZMMs-based watermark is given. Extensive experiments show that this method can locate the attacked area effectively. It can locate the altered blocks even if the altered image has been lossy compressed, blurred, or sharpened with medium strength,

(5) the proposed authentication scheme has better performance of discriminating high-quality JPEG compression from malicious attacks than some existing schemes. The scheme does not need the original feature vector for authentication process,

(6) the proposed scheme can be used on different kinds of images. The experiments on Chinese seal images with a very homogeneous background support this conclusion.

The feature vector of Zernike moments can also work well to authenticate binary images [32] like documents and CAD images. It can also be used in video authentication. Our extensive experiments show that this feature vector has good semi-fragile characteristics for video processing. Some preliminary results on video authentication by using Zernike moments-based feature vector has been published in [33].

Our future works include researching on the embedding algorithm robust to geometric distortions and improving the precision in locating the altered areas. Recovery [18] of the tampered area will be also studied in our future work.

Table 8: Comparisons with other methods in [7, 9–11, 22] on image "Lena."

	Proposed algorithm	Kundur's scheme in [7]	Lu'sscheme in [9]	Bao's scheme in [10]	Yang's scheme in [11]	Qi's scheme in [22]
PSNR(dB)	42.8	43.0	30.5	40.5	36.34	39.46
Robustness to JPEG (QF)	40	50	80	80	60	50

Table 9: Comparisons with some content-based watermarking methods (P_{fp}%).

	Proposed algorithm	Wang's scheme in [17]	Lin's scheme in [13]
Robustness to JPEG with QF 70	0	0.07	3.1
No-attack	0	0	0.2

Acknowledgment

This work is supported by 973 Program (2011CB302204), GDIID Program (GDIID2008IS046), and Guangdong Science and Technology program (2009B090300345).

References

[1] J. Fridrich, "Security of fragile authentication watermarks with localization," in *Security and Watermarking of Multimedia Contents IV*, vol. 4675 of *Proceedings of SPIE*, pp. 691–700, San Jose, Calif, USA, 2002.

[2] C. Fei, D. Kundur, and R. H. Kwong, "Analysis and design of secure watermark-based authentication systems," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 1, pp. 43–55, 2006.

[3] A. Swaminathan, Y. Mao, and M. Wu, "Robust and secure image hashing," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 215–230, 2006.

[4] C. D. Roover, C. D. Vleeschouwer, F. Lefebvre, and B. Macq, "Robust image hashing based on radial variance of pixels," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '05)*, vol. 3, pp. 77–80, Genova, Italy, 2005.

[5] C.-H. Lin and W.-S. Hsieh, "Semi-fragile image authentication method for robust to JPEG, JPEG2000 compressed and scaled images," in *Information Hiding and Application*, vol. 227 of *Studies in Computational Intelligence*, pp. 141–162, Springer, Berlin, Germany, 2009.

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

[7] D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1167–1180, 1999.

[8] G.-J. Yu, C.-S. Lu, and H.-Y. M. Liao, "Mean-quantization-based fragile watermarking for image authentication," *Optical Engineering*, vol. 40, no. 7, pp. 1396–1408, 2001.

[9] Z.-M. Lu, D.-G. Xu, and S.-H. Sun, "Multipurpose image watermarking algorithm based on multistage vector quantization," *IEEE Transactions on Image Processing*, vol. 14, no. 6, pp. 822–831, 2005.

[10] P. Bao and X. Ma, "Image adaptive watermarking using wavelet domain singular value decompostion," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 1, pp. 96–102, 2005.

[11] H. Yang and X. Sun, "Semi-fragile watermarking for image authentication and tamper detection using HVS model," in *Proceedings of the International Conference on Multimedia and Ubiquitous Engineering (MUE '07)*, pp. 1112–1117, April 2007.

[12] B. Zhu, M. D. Swanson, and A. H. Tewfik, "When seeing isn't believing," *IEEE Signal Processing Magazine*, vol. 21, no. 2, pp. 40–49, 2004.

[13] C.-Y. Lin and S.-F. Chang, "Semi-fragile watermarking for authenticating JPEG visual content," in *Security and Watermarking of Multimedia Contents II*, vol. 3971 of *Proceedings of SPIE*, pp. 140–151, San Jose, Calif, USA, 2000.

[14] R. Radhakrishnan and N. Memon, "On the security of the SARI image authentication system," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '01)*, pp. 971–974, October 2001.

[15] K. Maeno, Q. Sun, S.-F. Chang, and M. Suto, "New semi-fragile image authentication watermarking techniques using random bias and nonuniform quantization," *IEEE Transactions on Multimedia*, vol. 8, no. 1, pp. 32–45, 2006.

[16] Q. Sun and S.-F. Chang, "Semi-fragile image authentication using generic wavelet domain features and ECC," in *Proceedings of the International Conference on Image Processing (ICIP' 02)*, pp. 901–904, September 2002.

[17] J. Wang, S. Lian, Z. Liu, R. Zhen, and Y. Dai, "Multimedia data authentication in wavelet domain," in *Independent Component Analyses, Wavelets, Unsupervised Smart Sensors, and Neural Networks IV*, vol. 6247 of *Proceedings of SPIE*, pp. 3–12, 2006.

[18] M. J. Tsai and C. C. Chien, "Authentication and recovery for wavelet-based semifragile watermarking," *Optical Engineering*, vol. 47, no. 6, p. 067005, 2008.

[19] S. Thiemert, H. Sahbi, and M. Steinebach, "Using entropy for image and video authentication watermarks," in *Security, Steganography and Watermarking of Multimedia Contents VIII*, vol. 6072 of *Proceedings of SPIE*, pp. 1–10, 2006.

[20] J. Dittmann, "Content-fragile watermarking for image authentication," in *Security and Watermarking of Multimedia Contents III*, vol. 4314 of *Proceedings of SPIE*, pp. 175–184, 2001.

[21] M. Schlauweg, D. Profrock, T. Palfner, and E. Muller, "Quantization-based semi-fragile public-key watermarking for secure image authentication," in *Mathematics of Data/Image Coding, Compression, and Encryption VIII, with Application*, vol. 5915 of *Proceedings of SPIE*, pp. 1–11, 2005.

[22] X. Qi, X. Xin, and R. Chang, "Image authentication and tamper detection using two complementary watermarks," in

Proceedings of the International Conference on Image Processing (ICIP '09), pp. 4257–4260, 2009.

[23] C. Teh and R. T. Chin, "On image analysis by the methods of moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 496–513, 1988.

[24] A. Khotanzad and Y. H. Hong, "Invariant image recognition by Zernike moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 489–497, 1990.

[25] C. Maaoui, H. Laurent, and C. Rosenberger, "2D color shape recognition using Zernike moments," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '05)*, vol. 3, pp. 976–979, September 2005.

[26] Y. Xin, S. Liao, and M. Pawlak, "A multibit geometrically robust image watermark based on Zernike moments," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, vol. 4, pp. 861–864, August 2004.

[27] P. Amin and K. P. Subbalakshmi, "Rotation and cropping resilient data hiding with zernike moments," in *Proceedings of the International Conference on Image Processing (ICIP '04)*, vol. 4, pp. 2175–2178, October 2004.

[28] H. S. Kim and H.-K. Lee, "Invariant image watermark using Zernike moments," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 8, pp. 766–775, 2003.

[29] V. F. Zernike, "Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode," *Physica*, vol. 1, no. 7–12, pp. 689–704, 1934.

[30] http://www.cs.washington.edu..

[31] M.-J. Tsai, K.-Y. Yu, and Y.-Z. Chen, "Joint wavelet and spatial transformation for digital watermarking," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 1, pp. 241–245, 2000.

[32] X. Yao, H. Liu, W. Rui, and J. Huang, "Content-based authentication algorithm for binary images," in *Proceedings of the International Conference on Image Processing (ICIP '09)*, pp. 2893–2896, 2009.

[33] H. Liu, L. Zhu, and J. Huang, "A hybrid watermarking scheme for video authentication," in *Proceedings of the International Conference on Image Processing (ICIP '06)*, pp. 2569–2572, 2006.