Tamper Proofing and Attack Identification of Corrupted Image By using Semi-fragile Multiple-watermarking Algorithm

Soo-Chang Pei
Department of Electrical Engineering
National Taiwan University
886-2-23635251-321
pei@cc.ee.ntu.edu.tw

Yi-Chong Zeng
Graduate Institute of Communication Engineering
National Taiwan University
d89942010@ntu.edu.tw

No.1, Sec. 4, Roosevelt Rd., Taipei, Taiwan, 10617, R. O. C

ABSTRACT

We propose a novel semi-fragile multiple-watermarking algorithm based on quantization index modulation. This algorithm utilizes two quantization steps to yield the non-uniform intervals in the real-number axis. Each interval corresponds to one binary symbol, includes stable-zero (S_0), unstable-zero (U_0), stable-one (S_1), and unstable-one (U_1). In addition, visual cryptography is integrated with the watermarking algorithm to increase the watermark capacity. Therefore, the host image is embedded the multiple watermarks, and then we extract the watermarks from the corrupted image. According to the extracted watermarks, the algorithm achieves the tamper proofing and attack identification. From the experimental result, it shows single and multiple tampered areas are detected and demonstrates that the amount of test images will not influence the accuracy of attack identification.

Categories: D.2.11 [Software Engineering]: Software Architectures–; K.6.5 [Management of Computing and Information Systems]: Security and Protection–;

Keywords: semi-fragile watermarking, multiple-watermark, visual cryptography, tamper proofing, attack identification.

1. INTRODUCTION

The digital multimedia widely spread in the commercial, entertainment, art, etc. However, the pirates illegally copy, tamper and edit the media, threaten to the media industry. For this reason, the research workers study the various schemes to protect the products copyright and its authorization, and watermarking technique was developed. Most of watermarking techniques work on the spatial, frequency, wavelet and other domains [1-17].

Fragile, semi-fragile and robust watermarking schemes have different capability of signal security. Fragile watermarking is weak against any attack/distortion, but the robust watermarking ought tolerate all kind of distortions. The capacity of semi-fragile watermarking is defined between fragile and robust watermarking schemes. For the capability of semi-fragile watermarking algorithm, it is robust against a selection of distortions (such as, JPEG compression), but is weak against another distortions (such as, media filtering, lowpass filtering, sharpening, etc). Because the most of images are delivered thought Internet by JPEG compression; if the watermarked image is corrupted by JPEG compression, the extracted watermark of corrupted image must be survived as allowed distortion.

Previous studies concentrate on the single watermark approach, the difficulty of the multiple-watermark is the tradeoff between the capability of attack tolerance and the image quality. Hsu's method implements the multiple-watermark embedding in the middle frequency of DCT coefficients [1]. Shieh et al. propose a method to hide several watermarks in vector quantization and discrete cosine transform domains [14].

Fridrich [15] develops a watermarking technique to detect the tampers, she announces it need small memory and computational requirements to implement in digital camera. Furthermore, Fridrich suggests a hybrid watermarking scheme for tamper detection [16], this method is implemented by using robust and fragile watermarks. In order to improve the robustness of the watermarking scheme, Kundur et al. [17] adopt the reference and robust watermarks to embed in the host signal, and then characterize the attacks to improve the robust watermarking method. Besides tamper proofing, Macq et al. [18] discuss various benchmarking approaches of watermarking algorithms and the risk evaluation of delivery scenarios for digital eights management.

The visual cryptography has been addressed in many papers [19-25]. Naor suggests decoding the concealed images without any complex computations [19]. They not only generate the random shares, but also generate the meaningful shares to hide the secret information [20]. Moreover, Naor and Pinkas develop the visual authentication and identification [21]. Ateniese et al. proposed the general access structure of visual cryptography [22]. The conventional visual cryptography uses two or more secret shares to construct a significant image. In a *t*-out-*n* method of visual cryptography, a secret image is encoded into *n* random shares [23]. A halftone visual cryptography is discussed in [24], Zhou et al. use blue-noise dithering principles to construct halftone shares. Hou et al. [25] develop an asymmetric watermarking method based on visual cryptography, which integrates watermarking technology and visual cryptography. It encodes the watermark to two random shares, one share is

ASIACCS'06, March 21–24, 2006, Taipei, Taiwan.

embedded in the image and the other is a secret key for extracting watermark. Our algorithm will integrate the watermarking approach with visual cryptography to increase the watermark capacity.

In this paper, the proposed algorithm can embed bi-watermark and tri-watermark. For the definition of semi-fragile watermarking in our algorithm, it is robust against JPEG/JPEG-2000 compression, Gaussian noise, image rotation, frequency mode Laplacian removal, salt and pepper noise, and region modification; however, it is fragile against median filtering, Gaussian blurring, lowpass filtering, and image scaling. For JPEG compression, the average compression rate ranges from 0.77 bits/pixel (quality factor, QF=40%) to 6.01 bits/pixel (QF=100%). For JPEG-2000 compression, the compression rate is 2 bits/pixel applied to the tested image.

This paper is organized as follows. The previous works are described in Section 2. Section 3 will introduce the proposed algorithm. The experimental results are shown in Section 4. The conclusions are made in Section 5.

2. PREVIOUS WORKS

2.1 Quantization-base Watermarking Technique

The quantization-index-modulation (QIM) watermarking technique, requires low complexity than the other techniques, has been presented in [8-13]. The conventional approach divides the real-number axis into the uniform intervals by one quantization step, and then sets watermark symbols to these intervals. Given a quantization step Q, the sum value t is located at the p-th interval is represented as $p=\lfloor t/Q \rfloor$. In the watermark embedding, the host data Y is modified to ensure its sum value t_Y located at the specified interval. During the watermark extracting, we measure $t_{Y'}$ of watermarked data Y' and then extract the watermark symbol.

The quantization step size influences the watermarked image quality and attack tolerance. The small quantization step preserves the higher image quality than the large one. However, the watermarked image is robust against various attacks with large quantization step, but is weak with the small one. To overcome the drawback of attack tolerance in the watermarking scheme, the mean-quantization methods are suggested to increase the watermark robustness. Yu et al. [8, 9] adopt the mean quantization base watermarking approach to achieve the image authentication and detect the malicious tampering, the method is performed in the wavelet domain. The similar idea is addressed in [10], Chen et al. present a mean quantization approach to achieve the copyright protection of digital image in the wavelet domain as well. Eggers et al. investigate a watermarking scheme [11], which uses the dithered quantization and combines fingerprinting, for distinguishing the copies of multimedia document. Chen et al. develop the quantization index modulation to achieve the information embedding [12]. Moreover, Mihcak et al. have proposed the multiple non-uniform quantization steps to embed more symbols [13].

2.2 Visual Cryptography

Visual cryptography is a secret sharing method that uses human viewing to get the secret information. A well-known 2-out-of-2 visual threshold method encodes the pixel by two arrays of subpixels. Subsequently, the k-out-of-n visual threshold methods were discussed in [19, 20]. The secret image encrypts to derive two shares, and both shares are the random binary images in the previous studies. However, Naor et al. present an extension to construct a method, which adopts the special 2×2 arrays of subpixel to yield the meaningful binary shares.

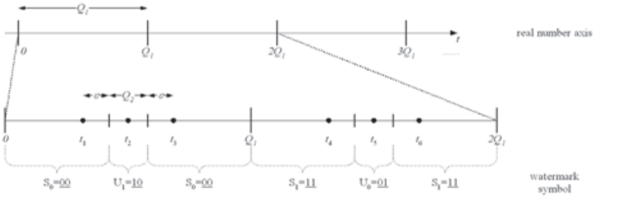


Figure 1: The non-uniform quantization subintervals

Hou et al. suggest an asymmetric watermarking method based on visual cryptography [25]. They encrypted a secret image to yield two shares, one share is embedded in the host image by watermarking technique and the other is treated as the secret key for extracting watermark. To integrate the watermarking technique and visual cryptography for increasing the watermark capacity is our search purpose, and we will extend the method to different applications.

3. PROPOSED ALGORITHM

3.1 Quantization-base Watermarking via Non-uniform Intervals

Most of previous studies for the quantization-base watermarking technique are usually to use a single quantization step to derive uniform interval, and then assign binary symbol to each interval periodically. Furthermore, the method embeds only one watermark in the host image at a time.

In order to improve the watermark approach, we apply two quantization steps to divide the real-number axis into the non-uniform subintervals as shown in Figure 1. The binary symbols collocate two states to obtain four kinds of symbols: stable-zero (S_0), unstable-zero (U_0), stable-one (S_1) and unstable-one (U_1). The first quantization step Q_1 is similar to the single quantization step of conventional approach, and further divides the quantized uniform interval into some non-uniform subintervals by small quantization step. The second quantization step Q_2, which is smaller than Q_1, determines the subinterval width of unstable symbols (U_0 and U_1). The stable symbols (S_0 and S_1) adjoin the sides of unstable symbol. Six specified quantized values are defined by,

$$\begin{array}{l} t_1 = t_2 - \tfrac{1}{2}Q_2 - \varepsilon,\ \ t_2 = \tfrac{1}{2}Q_1 + 2\alpha Q_1,\ \ t_3 = t_2 + \tfrac{1}{2}Q_2 + \varepsilon, \\ t_4 = t_5 - \tfrac{1}{2}Q_2 - \varepsilon,\ \ t_5 = \tfrac{3}{2}Q_1 + 2\alpha Q_1,\ \ t_6 = t_5 + \tfrac{1}{2}Q_2 + \varepsilon, \end{array}, \quad (1)$$

where $\alpha=\lfloor t/2Q_1 \rfloor$, ε is a scale value and $0<\varepsilon\leq(Q_1-Q_2)/4$. t is defined the sum of pixel values in one divided-block. t_1 and t_4 are left side of the specified quantized values of S_0 and S_1 near the intervals of unstable symbols. t_3 and t_6 are right side of specified quantized values of S_0 and S_1 near the intervals of unstable symbols as well. t_2 and t_5 are specified quantized values of U_1 and U_0 respectively.

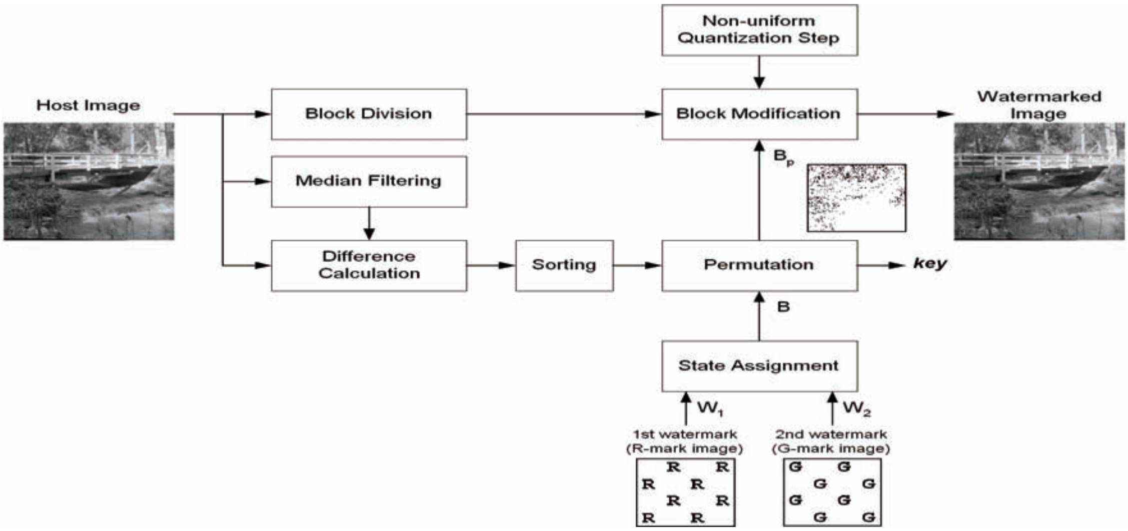


Figure 2: The block diagram of bi-watermark embedding algorithm

3.2 Watermark Embedding Algorithm

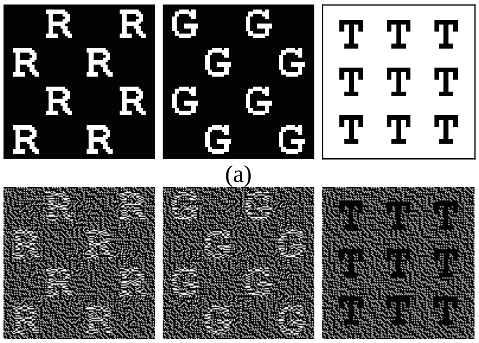
The basis of proposed multiple-watermarking scheme is the bi-watermarking algorithm. Assume that the first watermark ($\mathbf{W}_1$) and second watermark ($\mathbf{W}_2$) are meaningful binary image of size $M \times N$. An $M \times N$ state watermark $\mathbf{W}_S$, is composed of $\mathbf{W}_1$ and $\mathbf{W}_2$, will be embedded into the host image. A size $W \times H$ host image is divided into several $M \times N$ blocks of the size $a \times b$, where $a=W/M$ and $b=H/N$. $\mathbf{W}_1(i,j)$, $\mathbf{W}_2(i,j)$ and $\mathbf{W}_S(i,j)$ denote the watermark bit of 1st, 2nd and state watermarks at (i,j)-th position, respectively. The definition is listed as following,

$$\mathbf{W}_S(i,j) = \begin{cases} \mathrm{S}_0 = 00 \;, \text{if } \mathbf{W}_1(i,j)=0 \text{ and } \mathbf{W}_2(i,j)=0 \\ \mathrm{S}_1 = 11 \;, \text{if } \mathbf{W}_1(i,j)=1 \text{ and } \mathbf{W}_2(i,j)=1 \\ \mathrm{U}_0 = 01 \;, \text{if } \mathbf{W}_1(i,j)=0 \text{ and } \mathbf{W}_2(i,j)=1 \\ \mathrm{U}_1 = 10 \;, \text{if } \mathbf{W}_1(i,j)=1 \text{ and } \mathbf{W}_2(i,j)=0 \end{cases}, \tag{2}$$

where $1 \leq i \leq M$ and $1 \leq j \leq N$.

During bi-watermark embedding procedure of Figure 2, we utilize a 3×3 median filter to blur the host image, and calculate the absolute difference image between host and blurred images. The difference image is divided into several $M \times N$ blocks, and calculates the variance of each block. We rearrange that the unstable $\mathbf{W}_S(i,j)$'s are embedded to the blocks with larger variance, and the stable $\mathbf{W}_S(i,j)$'s are embedded to the rest ones. The watermark rearrangement provides the perceptual invisibility and achieves the watermark permutation. The similar approach has been proposed in [1]. The permutation is termed as the secret key, whose space is $(M \times N)!$, and it records the original position of the permuted state watermark bits. Subsequently, modifying the pixels value of block ensures the quantized value t locates at the appropriate interval, and t alters to the suitable quantized value in Eq.(1). For example, let t_1=53 (S_0), t_2=55 (U_1), t_3=57 (S_0), t_4=123 (S_1), t_5=125 (U_0), t_6=127 (S_1), and the quantized value t of one block is 100. If the block will embed U_0, hence, we modify the pixels value of block to make $t=t_5$.

In order to increase the watermark capacity, the bi-watermarking algorithm integrates with visual cryptography, is termed tri-watermarking algorithm. Before the embedding procedure, three meaningful binary images ($\mathbf{I}_1$, $\mathbf{I}_2$ and $\mathbf{I}_3$) are encrypted and transformed into crypto-watermarks ($\hat{\mathbf{W}}_1$, $\hat{\mathbf{W}}_2$ and $\hat{\mathbf{W}}_3$) shown in Figure 3. The relation of crypto-watermarks is defined by,



(a)

(b)

Figure 3: (a) Three binary images of size 64×64, and (b) three crypto-watermarks of size 128×128 are encrypted from Figure 3a.

$$\hat{\mathbf{W}}_1 + \hat{\mathbf{W}}_2 = \hat{\mathbf{W}}_3, \tag{3}$$

where the symbol '+' denotes the or-logical operation. $\hat{\mathbf{W}}_1$ and $\hat{\mathbf{W}}_2$ are the shared watermarks, and $\hat{\mathbf{W}}_3$ is the desired watermark. The encryption algorithm is referred to Naor' method [19], which encodes a pixel by two 2×2 arrays of subpixel. The watermarks W_1 and W_2 replace by $\hat{\mathbf{W}}_1$ and $\hat{\mathbf{W}}_2$ in the bi-watermark embedding, and the host image embeds the third desired watermark $\hat{\mathbf{W}}_3$ simultaneously.

3.3 Watermark Extracting Algorithm

While receiving a watermarked image, we divided it into several $M \times N$ blocks with size $a \times b$. The quantized values of blocks are measured to extract the watermark $\mathbf{B}_p$. After depermuting $\mathbf{B}_p$ to $\mathbf{B}$ by secret key, the recovered watermark $\mathbf{B}$ is also named the first watermark, $\mathbf{W}^*_1$. We interchange the binary values (0→1, 1→0) of unstable bits in $\mathbf{W}^*_1$, and the result is named the second watermark, $\mathbf{W}^*_2$. The block diagram of bi-watermark extracting

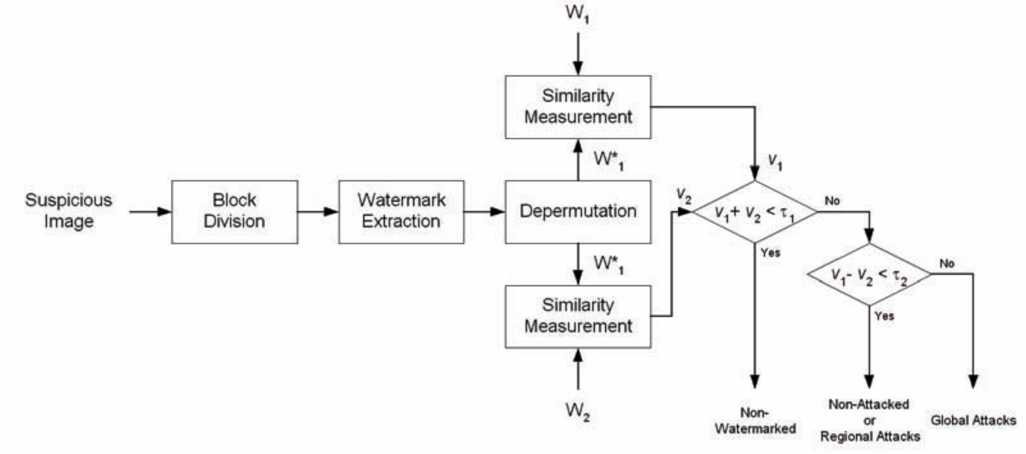


Figure 4: The block diagram of bi-watermarking and attack discrimination procedures

algorithm is shown in Figure 4. The similarity (SIM) between two watermarks is defined by,

$$\mathrm{SIM}_{k,l} = \frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N} b_{k,l}(i,j)\,, \tag{4}$$

$$b_{k,l}(i,j) = \begin{cases} 1 & ,\text{if } \mathbf{W}_k^*(i,j) = \mathbf{W}_l(i,j) \\ 0 & elsewise \end{cases}. \tag{5}$$

$\mathbf{W}^*_k$ and $\mathbf{W}_l$ represent the *k*-th extracted watermark and *l*-th original watermark, respectively. $b_{k,l}$ is a binary image, it locates the same watermark bits between extracted watermark $\mathbf{W}^*_k$ and original one $\mathbf{W}_l$.

If the watermarked image hides tri-watermark, the bi-watermark extracting algorithm firstly extracts two watermarks, $\mathbf{W}^*_1$ and $\mathbf{W}^*_2$, and then derives the third watermark $\mathbf{W}^*_3$ by the or-logical operation in Eq.(3). To measure the similarity between original and extracted watermarks, we define five similarities, v_1=SIM$_{1,1}$, v_2=SIM$_{1,2}$, v_3=SIM$_{2,2}$, v_4=SIM$_{3,2}$, v_5=SIM$_{3,3}$, which are formulated in Eq.(4). The values v_1, v_3, and v_5 typically represent the similarity between original and extracted watermarks. The purpose of v_2 and v_4 is to measure the characteristic of attack, which is either global tampering or regional tampering.

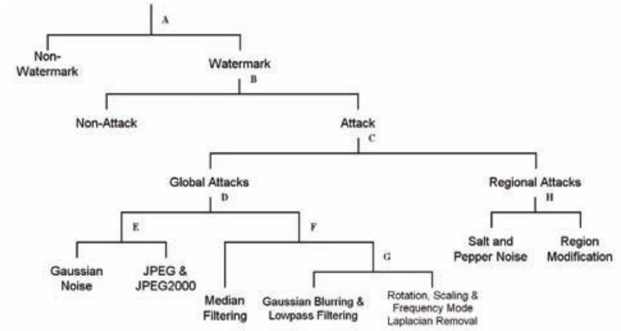


Figure 5: The decision tree

3.4 Attack Identification for Tri-watermarking Algorithm

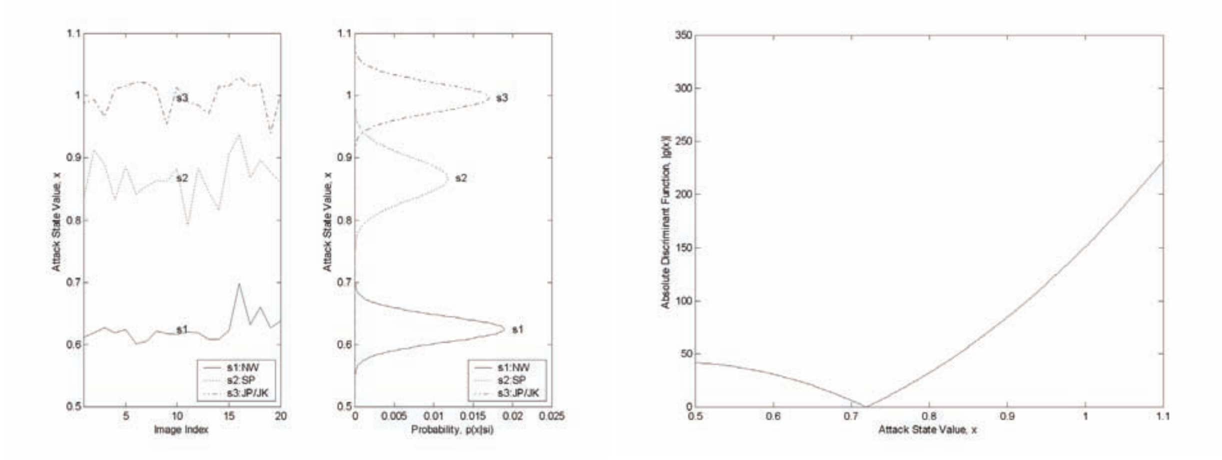
Five similarities of tri-watermarking algorithm are applied to the attack classification and identification. A decision tree of attack identification is shown in Figure 5. The reason for adopting these attacks is that they are easily and frequently applied to image processing throughout the existing commercial software, e.g. Adobe Photoshop, PhotoImpact, Paint Shop Pro, etc. In our study, the image is corrupted by 25 different attacks for 11 classes, including: (1) JPEG (JP) with QF=100%, 90%, 80%, 70%, 60%, 50%, 40%; (2) JPEG 2000 (JK) with 2 bit/pixel compression ratio; (3) Gaussian noise (GN) with zero mean and variance σ^2=80; (4) 3×3 median filtering (MD); (5) 3×3 lowpass filtering (LP); (6) 3×3 Gaussian blurring (GB); (7) image rotation (RT) with 1°, 5°, 10° and 20°; (8) image scaling (SC) with 50%, 75%, 150% and 200% of image size; (9) frequency mode Laplacian removal [26] (FM) with γ=0.03 and α=0.05, 0.5 and 2; (10) salt and pepper noise addition (SP) with density of 5%; and (11) region modification (RM) with 33% of image size. The image cutting, image cropping, pattern inserting, and some distortions perform on the regional areas, are included in the region modification. There are 11 classes classified into 7 categories. JK is regarded as JP attack, GB is regarded as LP attack, and RT, SC and FM are involved to the same category (Stirmark [27]). Meanwhile, we add the non-watermarked (NW) and non-attacked (NA) categories, hence, there are totally 9 categories in our algorithm. The identified equation at each node of decision tree is formulated as,

$$\begin{aligned} & v_1(i,j)\cdot\omega_1 + v_2(i,j)\cdot\omega_2 + \\ & \quad v_3(i,j)\cdot\omega_3 + v_4(i,j)\cdot\omega_4 + v_5(i,j)\cdot\omega_5 = a(i,j) \end{aligned}, \tag{6}$$

where $v_k(i,j)$ denotes the similarity v_k under *j*-th attack at *i*-th training image, $1 \le i \le M$, $1 \le j \le 27$ and $k=\{1,2,\ldots,5\}$. ω_k and $a(i,j)$ represent the weighting value and the attack state respectively. We employ *M* training images to estimate the weight values. All of the identified equations in Eq.(6) are rewritten to the matrix form,

$$\begin{bmatrix} v_1(1,1) & v_2(1,1) & v_3(1,1) & v_4(1,1) & v_5(1,1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ v_1(M,1) & v_2(M,1) & v_3(M,1) & v_4(M,1) & v_5(M,1) \\ v_1(M,2) & v_2(M,2) & v_3(M,2) & v_4(M,2) & v_5(M,2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ v_1(M,27) & v_2(M,27) & v_3(M,27) & v_4(M,27) & v_5(M,27) \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \\ \omega_5 \end{bmatrix} = \begin{bmatrix} a(1,1) \\ \vdots \\ a(M,1) \\ a(M,2) \\ \vdots \\ a(M,27) \end{bmatrix}, \tag{7}$$
$$\Leftrightarrow \mathbf{V\Omega} = \mathbf{A}$$

where matrices $\mathbf{V}$, $\mathbf{\Omega}$ and $\mathbf{A}$ are the matrices of size 5×(27×M), 1×5 and 1×(27×M), respectively. An example of global/regional attack classification (at node C), we set the initial $a(i,j)=\alpha$ for



(a) (b)

Figure 6: (a) The state values of 20 training images for 3 attacks and probability density distributions are illustrated in left and right parts, respectively; (b) the absolute discriminate function |*g*(*x*)| and the appropriate threshold is 0.719.

global attack and initial $a(i, j)=\beta$ for regional attack. The variable α is 0 or −1, and β is 1 in our algorithm. The problem of Eq.(7) is solved by using the pseudoinverse operation, which is based on minimum squared-error (MSE) method [28]. The solution $\mathbf{\Omega}'$ is formulated as,

$$\mathbf{\Omega}' = \mathbf{V}^{\dagger}\mathbf{A}, \qquad (8)$$

where $\mathbf{V}^{\dagger}$ is called the pseudoinverse of matrix $\mathbf{V}$ defined as,

$$\mathbf{V}^{\dagger} = \lim_{\gamma \to 0}\left(\mathbf{V}^{t}\mathbf{V} + \gamma\,\mathbf{I}\right)^{-1}\mathbf{V}^{t}, \qquad (9)$$

where $\mathbf{V}^{t}$ denotes the transpose of matrix $\mathbf{V}$. Recalculate the $a(i,j)$‘s by substituting $\mathbf{\Omega}'$ in Eq.(7), we estimate a threshold τ to classify the attacks into two groups.

To find an appropriate threshold is considered as two-category classification program. For example, three attacks (NW, SP and JP/JK) will be classified into two categories: one is NW, and the other is the group of SP and JP/JK attacks. In the left diagram of Figure 6a, it illustrates the attack state values of three attacks. Assume that each category is Gaussian distribution, we calculate the means μ and variances σ^2 of attacks, and the distributions are illustrated in right diagram of Figure 6a. In order to simply the classifying procedure, we find two nearest distributions by distance function. The distance function between s_i and s_j is defined as,

$$d = \left|(\mu_i + 2\sigma_i) - (\mu_j - 2\sigma_j)\right|, \qquad (10)$$

where $\mu_i < \mu_j$. Hence, we find that two nearest distributions are s_1 (NW) and s_2 (SP) in Figure 6a. Eq.(10) is a simple L_1-norm function. The reason of choosing the position of 2σ from μ is the cumulative distribution function (CDF) of the Gaussian distribution at 2σ is 0.99. It implies the position could represent the corresponding distribution. Therefore, Eq.(10) is not only to calculate the distance between two positions, but it is also to measure the distance between two distributions. The discriminant function $g(x)$ [28] is used for finding the proper separating point for two distributions, and is formulated as

$$g(x) = \ln\frac{p(x \mid s_1)}{p(x \mid s_2)} + \ln\frac{p(s_1)}{p(s_2)}, \qquad (11)$$

where x is the state value, $p(s_i)$ and $p(x|s_i)$ represent the probability and conditional probability of distribution s_i. The diagram of absolute discriminant function $|g(x)|$ is shown in Figure 6b. Consequently, the threshold is the state value corresponds to the minimum $|g(x)|$, and τ is 0.719. At each node of decision tree, we implement the above-mentioned classification method to estimate the weighting values and thresholds.

The decision tree is valid. There are four reasons: firstly, when we receive a suspected image, the primary question is whether the image embeds watermark. Therefore, the discrimination between non-watermarked and watermarked images proceeds in the first stage. Secondary, the similar question for watermarked image is whether the image encounters the attacks. Due to the similarities for non-attacked image are constant (e.g., v_1=1, v_2=0.865, v_3=1, v_4=0.935 and v_5=1), the discrimination between non-attacked and attacked images proceeds in second stage. Thirdly, in order to increase the accuracy of attack identification, it classifies the attacks into the global and the regional attacks. Fourthly, the discrimination in the global/regional attack is based on the characteristic between the attacks. For instance, lowpass filtering, median filtering, Gaussian blurring and Laplacian removal are performed with a 3×3 mask, and the functionality of four attacks is the noise-cleaning process. The characteristic of these attacks are different to JPEG/JPEG 2000 compression and Gaussian noise, therefore, they are immediately classified at node D after the global/regional attack classification.

4. EXPERIMENTAL RESULTS

In this section, we will show the experimental results, including: multiple-watermark extraction, tamper proofing and attack classification/identification. We first use 540 corrupted images (20 training images encountered under non-watermarked, non-attacked and 25 attacks to result in 540 corrupted images) for training the parameters. For 1st experiment in the test procedure, we use 810 corrupted images (30 tested images encountered under non-watermarked, non-attacked and 25 attacks to result in 810 corrupted images). Additionally, there are totally 1620 images (previous 810 corrupted image in 1st experiment and additional extra 810 corrupted images) are employed for 2nd experiment in the test procedure. The parameters of all attacks have been described in Section 3.4.

4.1 Multiple-watermark Extraction

Dividing the watermarked image into 128×128 blocks with size 4×4, which embed one bi-watermark bit in one block. Meanwhile, two 128×128 meaningful binary images are considered as watermarks, one is full of R-marks named 1st watermark and the other one is full of G-marks named 2nd watermark. The parameters Q_1, Q_2, and ε are set to 70, 1, 1.5, respectively. The bi-watermarked image is PSNR=40.39dB. The extracted bi-watermarks of 12 corrupted mages are shown in Figure 7. In clearly indicates that our method achieves the better attack tolerances in JP, JK, GN, RT, FM, SP and RM, and preserves the 2nd watermark after attacking. It shows the proposed semi-fragile watermarking is weakly in MD, GB, LP and SC. The 1st watermarks carry the special characteristic. While the watermarked image is corrupted by the global attack, the unstable bit of 1st watermark is changed to the neighboring stable bit. Hence, The 1st watermark will similar to 2nd watermark for multiple-watermarking algorithm. In addition, the PSNR of tri-watermarked image is 40.18dB, and the similarities of extracted tri-watermark are listed in Table 1.

4.2 Tamper Proofing

The experiment result of tamper proofing is shown in Figure 8, and Figure 8a illustrates the watermarked image. The image $\mathbf{I}_{kl}$ marks the unequal watermark bits between $\mathbf{W}^*_k$ and $\mathbf{W}_l$, and then $\mathbf{I}_{kl}$ permutes to a tampered-mark image $\mathbf{I}'_{kl}$ by *secret key*. The extracted bi-watermark ($\mathbf{W}^*_1$ and $\mathbf{W}^*_2$) and tampered-mark images ($\mathbf{I}'_{11}$ and $\mathbf{I}'_{22}$) for single RM attack, which modifies the center region of watermarked image, are displayed in the Figure 8b. Moreover, we implement multiple attacks, including: JPEG compression (QF=80%), two region modifications (at the center and the upper-left corner of the image). The bi-watermark and two tampered-mask images are shown in Figire 8c. $\mathbf{I}'_{11}$ illustrates the mixture-tampered areas, however, $\mathbf{I}'_{22}$ illustrates distinctly the tampered areas of regional attacks at the upper-left corner and the center of the image.

In addition, we apply thirty 512×512 gray-scale images to implement the multiple attacks. First, tested images encounter 10 region modification attacks with different sizes (range 0.5% to 2% of the image size), these attacks appear in the random positions of the watermarked images. The accuracy of tampered area detection for multiple attacks (AC_{MA}) is defined by,

$$AC_{\text{MA}} = \frac{\text{The detect - tampered area size}}{\text{The actual - tampered area size}}\,. \qquad (12)$$

Hence, the average accuracy AC_{MA} of 30 attacked images is 0.693. Twelve possible attack combinations for the regional tampered areas detection experimental results are listed in Table 2. The tamper proofing is successful implemented in RM with JK, GN, RT, FM, SP, RM and JP (QF≥40%) attacks. However, it fails in RM with GB, LP, MD, SC or JP (QF<40%) attacks.

4.3 Attack Classification/Identification

For tri-watermarking scheme, we employ 540 corrupted images of size 512×512 to train the parameters, including: τ, ω_1, ω_2, ω_3, ω_4, ω_5 and ω_6 at each node of decision tree, and the appropriate parameters are listed in Table 3. Subsequently, 810 images are tested for the 1st experiment of the attack identification, and the accuracies of 9 categories are listed in Table 4. In addition, there are 1620 images are tested for the 2st experiment of the attack identification, and the accuracies are listed in Table 4 as well. Besides JP/JK, MD, GB/LP and RT/SC/FM, the accuracies of the others are 1.

In Table 5, it lists the function's comparisons among the existing 9 watermarking schemes and our method. To compare the results, there are several schemes can embed multiple watermarks. Moreover, most schemes can achieve the tampered proofing and need secrete key to improve the watermarking security. For the attack analysis, Kundur's [17] and our method analyze the attack characterization to improve the robust watermarking method and perform attack classification respectively. *Our method is superior to the others in the function of attack identification*, and it can classify/identify the attacks to 9 attack categories, which are never provided in the other methods. For subjective tests, we employ a group of members, include specialists, artist, non-specialists, to evaluate the difference between original and watermarked images. Therefore, the results are satisfactory with subjective tests for transparent evaluation.

5. CONCLUSION

In this paper, we present a novel semi-fragile multiple-watermarking algorithm. It's based on the quantization-base watermarking, and can embed two and three watermarks by bi-watermarking and tri-watermarking algorithms respectively. The experimental results show that the proposed method successfully locates the single and multiple tampered areas for tamper proofing. Moreover, the attack classification uses mean square-error method to classify the attacks into 9 categories. The experimental results also show that the bi-watermarking technique robust against JPEG and JPEG 2000 compression, Gaussian noise, image rotation, frequency mode Laplacian removal, salt and pepper noise, and region modification, but is weak against median filtering, Gaussian blurring, lowpass filtering, and image scaling.

6. REFERENCES

[1] Hsu, C. T., and Wu, J. L. Hidden digital watermarks in images. *IEEE Trans. on Image Proc., 8, 1*, (Jan. 1999), 58-68.

[2] Lin, C. Y., and Chang, S. F. Semi-fragile watermarking for authenticating JPEG visual content. *SPIE International Conf. on Security and Watermark of Multimedia Contents II*, 3971, 13, (San Jose, USA, Jan 2000).

[3] Ko, H. H., and Park, S. J. Semi-fragile watermarking for telltale tamper proofing and authenticating. *ITC-CSCC 2002*, (July 2002), 623-626.

[4] Lu, Z. M., Lin, C. H., Xu, D. G., and Sun, S. H., Semi-fragile image watermarking method based on index constrained vector quantization. *Electronics Letters*, 39, 1, (Jan. 2003), 35-36.

[5] Kundur, D., and Hatzinakos, D. Digital watermarking using multiresolution wavelet decomposition. *ICASP*, 5, (Seattle, Washington, USA, May 1998), 2969-2972.

[6] Kundur, D., and Hatzinakos, D. Digital watermarking for telltale tamper proofing and authentication. *Proceedings of the IEEE*, 87, 7, (July 1999), 1167-1180.

[7] Paquet, A. H., and Ward, R. K. Wavelet-based digital watermarking for image authentication. *Proc. 2002 IEEE Canadian Conf. Electrical and Computer Engineering*, 2, (Winnipeg, Canada, 2002), 879-884.

[8] Yu., G.-J., Lu, C.-S., Liao, Mark, H.-Y., and Sheu, J.-P. Mean quantization blind watermarking for image authentication. *ICIP 2000*, 3, (Vancouver, BC, Canada, 10-13, Sept. 2000), 706-709.

[9] Yu, G.-J., Lu, C.-S., and Liao, Mark, H.-Y. Mean quantization based fragile watermarking for image authentication. *Optical Engineering*, 40, 7, (July 2001), 1396-1408.

[10] Chen, L.-H., and Lin, J.-J. Mean quantization based image watermarking. *Image and Vision Computing*, 21, 8, (Aug. 2003), 717-727.

[11] Eggers, J. J., and Girod, B. Quantization watermarking. *Proceedings of SPIE, Security and Watermarking of Multimedia Contents II*, 3971, (San Jose, Jan. 2000), 60-71.

[12] Chen, B., and Wornell, G. W. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. *IEEE Trans. on Information Theory*, 47, 4, (May 2001), 1423-1443.

[13] Mihcak, M. K., Venkatesan, R., and Kesal, M. Watermarking via optimization algorithms for quantizing randomized statistics of image regions. in *Proc. 40th Allerton Conference on Communications Control and Computing*, (Monticello, Illinois, Oct. 2002).

[14] Shieh, C.-S., Huang, H.-C., and Wang, F.-H. An embedding algorithm for multiple watermarks. *Journal of Information Science and Engineering*, 19, 2, (March 2003), 381-395.

[15] Fridrich, J. Image watermarking for tamper detection. *ICIP'98*, 2, (4-7 Oct. 1998), 404-408.

[16] Fridrich, J. A hybrid watermark for tamper detection in digital images. *ISSPA'99*, 1, (22-25 Aug. 1999), 301-304.

[17] Kundur, D., and Hatzinakos, D. Improved robust watermarking through attack characterization. *The International Electronic Journal of Optics*, 3, 12, (7 Dec. 1998), 485-490.

[18] Macq, B., Dittmann, J., and Delp, E. J. Benchmarking of image watermarking algorithms for digital rights management. *Proceedings of The IEEE*, 92, 6, (June 2004), 971-984.

[19] Naor, M., and Shamir, A. Visual cryptography. *Eurocrypt'94, Lecture Notes in Computer Science*, 950, (Springer-Verlag, 1995), 1-12.

[20] Naor, M., and Shamir, A. Visual cryptography II: improving the contrast via the cover base. *in Proceedings of the International Workshop on Security Protocols*, (Springer-Verlag, 1997), 69-74.

[21] Naor, M., and Pinkas, B. Visual authentication and identification. *Crypto'97, Lecture Notes in Computer Science*, 1294, (1997), 322-336.

[22] Ateniese, G., Blundo, C., Santis, A. De, and Stinson, D. R. Visual cryptography for general access structures. *Information and Computation*, 129, 2, (15 Sept. 1996), 86-106.

[23] Stinson, D. Visual cryptography and threshold methods. *IEEE Potentials*, 18, 1, (Feb.-March 1999), 13-16.

[24] Zhou, Z., Arce, G. R., and Crescenzo, G. Di. Halftone visual cryptography. *ICIP' 2003*, 1, (14-17 Sept. 2003), 521-524.

[25] Hou, Y.-C., and Chen, P.-M. An asymmetric watermarking method based on visual cryptography. *WCCC-ICSP 2000*, 2, (21-25 Aug. 2000), 992-995.

[26] Barnett, R., and Pearson, D. E. Frequency mode LR attack operator for digitally watermarked images. *Electronics Letters*, 34, 2, (Sept. 1998), 1837-1839.

[27] Stirmark 4.0, http://www.petitcolas.net/fabien/watermarking/stirmark/.

[28] Duda, R. O., Hart, P. E., and Stork, D. G. *Pattern classification*, John Wiley & Sons Inc., NY, 2001.

Table 1: There are 5 similarities (SIM) of the watermarked image corrupted by 13 attacks.

	NW	NA	GN	JP	JK	MD	GB	LP	RT	SC	FM	SP	RM
v_1	0.493	1	0.759	0.831	0.835	0.795	0.834	0.782	0.845	0.767	0.847	0.862	0.837
v_2	0.492	0.865	0.850	0.951	0.949	0.815	0.917	0.833	0.953	0.785	0.936	0.763	0.759
v_3	0.491	1	0.871	0.979	0.997	0.832	0.918	0.839	0.953	0.784	0.933	0.858	0.831
v_4	0.498	0.935	0.869	0.969	0.984	0.827	0.919	0.833	0.953	0.782	0.932	0.812	0.792
v_5	0.494	1	0.823	0.912	0.923	0.813	0.873	0.812	0.901	0.772	0.912	0.866	0.833

Table 2: Twelve attack combinations of the regional tampered areas detection. The tamper proofing is successful implemented in RM with JK, GN, RT, FM, SP, RM and JP (QF≥40%) attacks. However, it fails in RM with GB, LP, MD, SC or JP (QF<40%) attacks

Attacks / Regional Modification	JP		JK	GN	RT	FM	SP	RM	GB	LP	MD	SC
	(QF≥40%)	(QF<40%)										
RM	√	×	√	√	√	√	√	√	×	×	×	×

Table 3: The parameters are set at each node of the decision tree

Node	α	β	τ	ω_1	ω_2	ω_3	ω_4	ω_5
A	0	1	0.620	-0.750	5.199	6.343	-10.804	1.109
B	-1	1	-1	-1/3	0	-1/3	0	-1/3
C	0	1	0.487	-4.483	13.648	20.315	-42.583	13.320
D	-1	1	0.7135	3.869	2.050	-4.704	0	0
E	0	1	0.5	5.950	-17.346	-37.200	42.7316	7.027
F	0	1	0.3335	-18.627	-37.355	-0.837	13.079	43.872
G	0	1	0.514	-38.266	-26.785	6.168	-15.900	70.800
H	0	1	0.365	-43.371	-136.813	-94.006	63.786	195.454

Table 4: There are 9 average accuracies of attack identification for 810 corrupted tri-watermarked images, and 1620 corrupted tri-watermarked images with 128×128 watermark.

Attacks / Watermark Size	NW	NA	GN	JP&JK	MD	GB&LP	RT&SC & FM	SP	RM
128×128 (810 corrupted images)	1	1	1	0.925	0.967	0.9	0.997	1	1
128×128 (1620 corrupted images)	1	1	1	0.91	0.983	0.892	0.998	1	1

Table 5: The function's comparisons among the existing 9 watermarking schemes and our method are listed. DCT is discrete cosine transform, and VQ is vector quantization.

	Hsu [1]	Ko [3]	Lu [4]	Kundur [6]	Paquet [7]	Yu [9]	Shieh [14]	Fridrich [16]	Kundur [17]	Our Method
Multiple Watermark	Yes	No	No	No	No	No	Yes	Yes	Yes	Yes
Watermark Category	Robust	Semi-Fragile	Semi-Fragile	Fragile	Fragile	Fragile	Robust	Fragile+ Robust	Robust	Semi-Fragile
Domain	DCT	Wavelet	VQ	Wavelet	Wavelet	Wavelet	VQ+DCT	Spatial+DCT	Wavelet	Spatial
Tamper Proofing	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Attack Analysis	No	No	No	No	No	No	No	No	Yes	Yes
Attack Identification	No	No	No	No	No	No	No	No	No	Yes
Secrete Key	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes

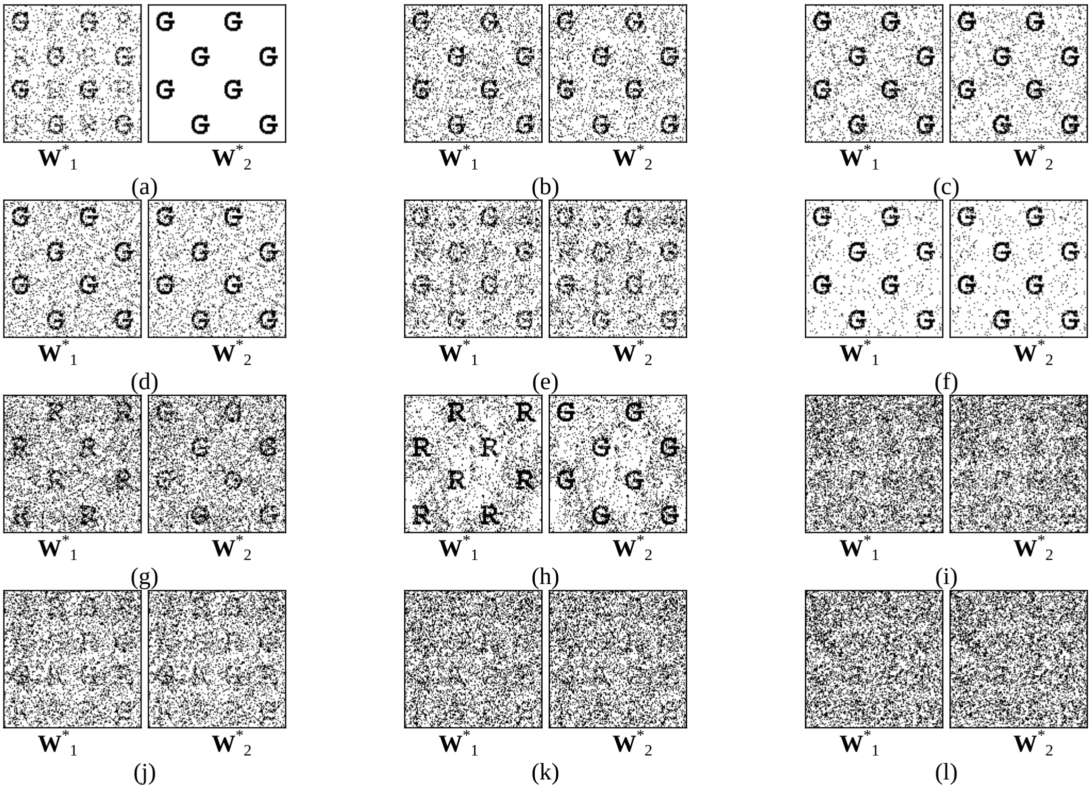


Figure 7: The extracted bi-watermark of corrupted images by (a) JPEG with QF=100% (v_1=0.822, v_3=1); (b) JPEG with QF=40% (v_1=0.753, v_3=0.862); (c) JK (v_1=0.754, v_3=0.903); (d) GN (v_1=0.743, v_3=0.872); (e) RT (v_1=0.765, v_3=0.804); (f) FM (v_1=0.823, v_3=0.949); (g) SP (v_1=0.739, v_3=0.739); (h) RM (v_1=0.836, v_3=0.837); (i) MD (v_1=0.656, v_3=0.663); (j) GB (v_1=0.718, v_3=0.723); (k) LP (v_1=0.663, v_3=0.661); and (l) SC (v_1=0.643, v_3=0.645).

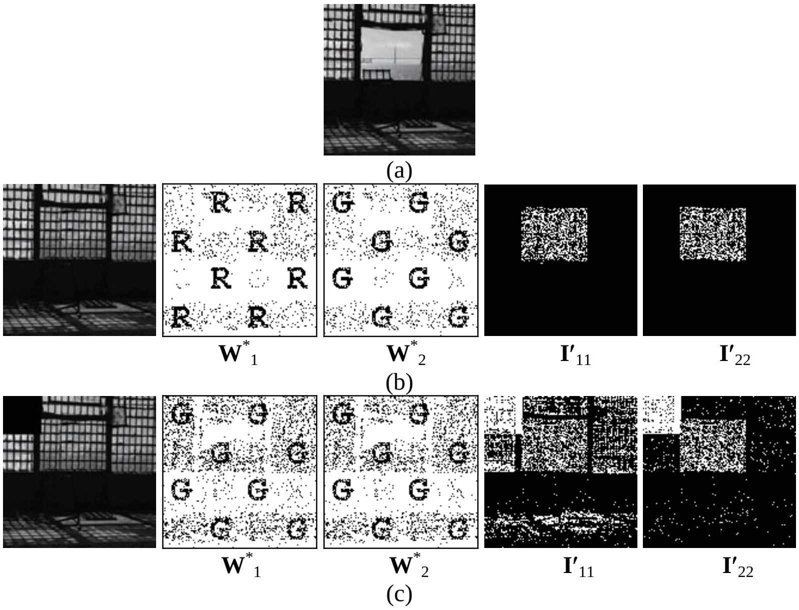


Figure 8: (a) The watermarked image; (b) the image is only corrupted by single RM, and (c) the image is corrupted by multiple RM attacks and JPEG compression. In Figure 8a and Figure 8c, the corrupted image, two extracted watermarks and two tampered-mask images rank from left to right.