Authentication and recovery for wavelet-based semifragile watermarking

Min-Jen Tsai
Chih-Cheng Chien
National Chiao Tung University
Institute of Information Management
1001 Ta-Hsueh Road
Hsin-Chu 300, Taiwan
E-mail: mjtsai@cc.nctu.edu.tw

Abstract. We propose a novel image authentication and recovery scheme based on discrete wavelet transform (DWT). By using the property of a DWT multiresolution structure, we generate a semifragile watermark from low-frequency bands and embed the recovery information into the high-frequency bands based on the human visual system (HVS) approach. The image authentication system is able to locate precisely any malicious alteration made to the image and restore the altered or destroyed regions based on the recovery mechanism. Therefore, the requirements of ownership protection and tampering detection of digital right management (DRM) are met, and the legal usage of digital content is available. In addition, robustness to mild modifications like JPEG compression and channel additive white Gaussian noise (AWGN) is also achieved with high recovery image quality. © *2008 Society of Photo-Optical Instrumentation Engineers.* [DOI: 10.1117/1.2947580]

Subject terms: authentication; digital right management; human vision system; semifragile watermarking; tamper detection; tamper recovery.

Paper 080002R received Jan. 7, 2008; revised manuscript received Apr. 7, 2008; accepted for publication Apr. 8, 2008; published online Jun. 25, 2008.

1 Introduction

Digital watermarking has been extensively studied and regarded as a potentially effective means for protecting copyright of digital right management (DRM) systems. Digital watermarking describes methods and technologies that allow hiding information, for example, a sequence of numbers or recognizable patterns, in digital media, such as images, video, and audio. A lot of digital watermarking techniques have been proposed by many researchers and can be divided into various categories and in various ways.[1]

For content authentication and integrity verification, fragile (or semifragile)watermarks are used because they become fragile with certain alterations and modifications of the authenticated multimedia. Some fragile watermarking techniques[2–5] were usually based on the concept of checksum produced by secure hash functions (e.g., MD5, SHA160) to verify the completeness of an image. They can detect and localize tamper correctly, but they treat admissible manipulations such as JPEG compression and channel additive white Gaussian noise (AWGN) as malicious attacks. Therefore, semifragile watermarking techniques[6–8] are more practically applied than fragile watermarking, since they are robust to some mild modifications such as JPEG compression and channel AWGN caused during the process of exchange and storage, but fragile to malicious attacks like image cropping, which crops objects from a source and pastes them onto a target. According to the conveyance of authentication data, fragile (or semifragile) watermarks can be classified into two main categories: labeling-based authentication schemes and watermarking-based authentication schemes. The watermarking-based authentication schemes embed the data into the original multimedia content, and labeling-based authentication stores the authentication data in a separate file. Consequently, the authentication data becomes the integral part of the original multimedia and can be transmitted more efficiently and securely. In this work, we focus on the semifragile watermarking-based authentication scheme. Some necessary requirements like the sensitivity of manipulations, tolerance of information loss, localization of altered region, blind extraction (no need for the original source), visibillty, robustness, and security must be preserved.[9] In addition, this research not only achieves the tamper authentication but also the content recovery for reconstruction of the altered regions.

The goal of this work is to propose a novel scheme for image tampering authentication and recovery for wavelet-based semifragile watermarking. The rest of this work is organized as follows. Related works about image authentication and image tamper recovery are introduced briefly in Sec. 2. The details of the proposed algorithm are explained in Sec. 3. Section 4 demonstrates the experimental results and discussion, and the conclusion is in Sec. 5, respectively.

2 Related Works

2.1 *Image Authentication and Tamper Detection*

Semifragile watermarks can be embedded in the spatial domain or the transformed domain. The schemes operating in the spatial domain are simpler than the ones using transform modulation by utilizing the least significant bit (LSB) of data. However, the schemes that embed watermarks in the transformed domain offer a higher degree of robustness.[10] Recently, many semifragile methods are based on the wavelet transform domain, since it applies image processing operations to obtain the highest degree of

0091-3286/2008/$25.00 © 2008 SPIE

robustness, and allows the method to have spatial and frequency localization of digital data by the nature of multiresolution discrete wavelet decomposition.

Kundur and Hatzinakos[11] proposed one of the first approaches to semifragile watermarking called telltale tamper proofing. They embed a watermark in the discrete wavelet domain of the image by quantizing the corresponding coefficients. They claim their tamper detection, determined both in localized spatial and frequency regions, is unlike previously proposed techniques embedding a watermark in the spatial domain, which only provides information on the spatial location of the changes and fails to give a more general characterization of the type of distortion applied to the signal. They also use a statistics-based tamper assessment function as measurement for tamper proofing and authentication. Alexandre and Rabab[12] proposed a novel technique for content authentication of digital images by quantizing wavelet packet coefficients and adopting characteristics of the human visual system to maximize the embedding weights for improved good imperceptibility of watermarked images. According to experimental results, their method is able to detect and localize malicious image modifications while offering a certain degree of robustness to image compression. A similar concept was also proposed in Ref. 13, where they proposed a discrete wavelet transform-based image semifragile watermarking scheme base on fusion of multiresolution. The Watson's quantization matrix[14] and the features of the human visual system (HVS) are clearly adopted in the quantization process to achieve good quality of the watermarked image. Liu et al.[15] presented a semifragile image watermarking technique based on index constrained vector quantization (VQ). However, the peak signal-to-noise ratio (PSNR) of their watermarked image is low and their scheme would waste storage and not be flexible for the codebook of vector quantization that should be known in both watermark embedding and extraction processes. Yuan and Zhang[16] proposed a novel semifragile watermarking method based on image modeling using the Gaussian mixture model (GMM) in the wavelet domain. They modify selected wavelet coefficients according to the GMM parameters obtained through an EM algorithm. In experimental results, their scheme achieves minimum watermarking distortion and identifies mild modification from malicious attacks, but it treats AWGN as a malicious attack.

Ding et al.[17] propose a wavelet-based chaotic semifragile watermarking scheme based on chaotic map and odd-even quantization. Their scheme can detect and localize malicious attacks with high peak signal-to-noise ratio (PSNR), while allowing mild JPEG compression and channel AWGN tolerance. However, they did not disclose what chaotic map applied in their simulation data. In Chu et al.,[18] the authors presented a semifragile watermarking scheme for authenticating the region of interest (ROI) of an image. First, the reference mask is obtained by Poisson matting. Then, they embed watermark according to the reference mask, representing the region of interest of the image.

2.2 *Image Tamper Recovery*

Currently, there is a great need for tamper recovery techniques, since there are not many references in this area. The recovery information can be embedded in either the transform domain or spatial domain. For example, Lin and Chang[19] proposed a semifragile algorithm that is conceived to tolerate, in particular, JPEG-style compression of the watermarked image. It is based on two properties of the discrete cosine transform (DCT) coefficient quantization, namely, the order invariance, where the order relation of the DCT coefficient pairs remains unaltered after JPEG compression, if not set equal; and the coefficient invariance, where if a coefficient is quantized to an integer multiple of the step size, its value is not changed after JPEG compression with a smaller step size. The first one is used to generate the authentication bits, and the other is used to embed the signature. The authors have proposed some improvement such as recovery bits. The advantage of these overhead bits is two-fold: they allow an approximation of the original block to be reconstructed, and they help to locate precisely the zones of the images that were really faded. The recovery bits are generated from a downsampled and compressed version of the original image. They are then embedded into four blocks. The embedding process of recovery bits is similar to that of authentication bits.[8]

Lin, Hsieh, and Huang[20] proposed a hierarchical digital watermarking method for image tamper detection and recovery in the spatial domain. It uses simple operations such as parity checks and comparison between average intensities. For example, with a 4×4 block named A, the intensity feature will be embedded into another 4×4 block named B, while the one-to-one block mapping relationship for the whole image can be uniquely decided. For each block A of 4×4 pixels, they further divide it into four subblocks of 2×2 pixels. The watermark in each subblock is a tuple (u,p,r), where both u and p are 1-bit authentication watermark, and r is a six-bit recovery watermark for the corresponding subblock within block A mapped to sub-block within block B. The 8-bit watermarks (u,p,r) are embedded onto the two LSBs of each pixel within the subblock of B. This scheme provides us with the capability of tamper recovery by trading off the quality of the watermarked images by about 5 dB.

Lin and Chang's[19] algorithm performs very well in the presence of JPEG compression, but otherwise it is very fragile against signal-processing attacks. The image tamper recovery scheme in Ref. 20 is only suitable for fragile-watermark schemes. If we perform any compression or attack, the recovery information is lost.

Since little research is on topics about an image tampering recovery mechanism based on semifragile watermarks, it is also our motivation to propose a novel image authentication and tampering recovery algorithm that is robust against JPEG compression and channel additive white Gaussian noise for wavelet-based semifragile watermarks. The detailed description is in the next section.

3 Proposed Algorithm

Since the goal of this work is to develop a complete architecture to effectively verify the integrity of an authorized grayscale image, detect the tampered region, and restore the content, the proposed algorithm can be categorized into four subalgorithms and is explained next.

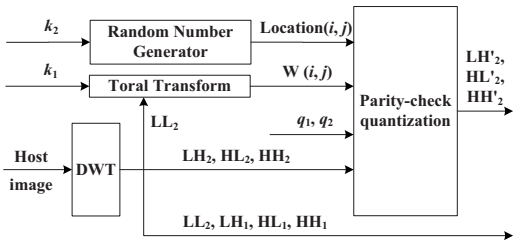

Fig. 1 Flow diagram of the watermark generation and embedding scheme.

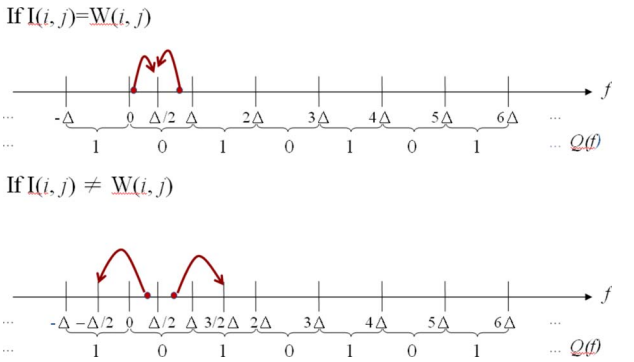

Fig. 2 Diagram of how parity-check quantization works.

- Semifragile watermark generation and embedding algorithm
- image recovery information embedding algorithm
- semifragile watermark authentication and tamper detection algorithm
- image tamper recovery algorithm.

3.1 *Semifragile Watermark Generation and Embedding Algorithm*

The host image first performs 2-D wavelet decomposition, and the flow chart is shown in Fig. 1. The semifragile watermark generation and embedding procedure is as follows.

1. Calculate the two-level wavelet coefficients of the host image; $r \times c$ is the size of subband LL_2.
2. Select parameters: keys k_1 and k_2 are the private keys of the embedding scheme. q_1 and q_2 are the quantization parameters.
3. To get higher security, we refer to total automorphisms[21] (total transform) as the chaotic system for random scrambling, and the modified formula is shown in Eq. (1). Using $S(i,j)=\lfloor \mathrm{LL}_2(i,j)/q_1 \rfloor$ and k_1 as controlling parameters for Eqs. (1) and (2), we obtain the binary watermark $W(i,j) \in \{0,1\}$, $1 \leqslant i \leqslant r$, $1 \leqslant j \leqslant c$. Note: $\lfloor \cdot \rfloor$ denotes the floor function.

$$A_r(i,j,k_1): \begin{cases} i' = (i+j) & \mod r \\ j' = \{[S(i,j)+k_1]\cdot i + [S(i,j)+k_1+1]\cdot j\} & \mod c \end{cases}, \tag{1}$$

$$W(i,j) = (i'+j') \bmod 2, \tag{2}$$

where (i,j) and (i',j') are the pixel location before and after the total transform.

4. For the watermark embedding location corresponding to subbands $\{\mathrm{HL}_2, \mathrm{HH}_2, \mathrm{LH}_2\}$, key k_2 is applied as a random seed to create a pseudorandom array $location(i,j) \in \{1,2,3\}, 1 \leqslant i \leqslant r, 1 \leqslant j \leqslant c$. The pseudocodes of the operation are $location(i,j) = [\mathrm{rand}(k_2)\%3]+1$, where $location(i,j)=1$ means the subband HL_2, $location(i,j)=2$ means the subband HH_2, and $location(i,j)=3$ means the subband LH_2. Rand(seed k) is a function that uses the seed k to return a pseudorandom integral number.
5. The binary watermark is embedded into the image by parity-check quantization, which is modified from the odd-even quantization,[19] and the operation is demonstrated in Fig. 2. Parity-check quantization is essentially a scalar quantization used in Ref. 21, where the uniform quantizer is used in this study. We define parity-check quantization function f in Eq. (3) and the detailed procedures in Eqs. (4)–(6). The function performs quantization on decomposed wavelet coefficients $x(i,j)$ into odd-even regions according to the binary watermark $W(i,j)$, (i,j) indicates the spatial location, and $x \in \{\mathrm{LH}_2, \mathrm{HL}_2, \mathrm{HH}_2\}$. The inputs of function f are either $\mathrm{LH}_2(i,j)$, $\mathrm{HL}_2(i,j)$, or $\mathrm{HH}_2(i,j)$, with $W(i,j)$ and q_2. The output of function f is the result after the process of Eqs. (4)–(6).

$$y(i,j) = f[x(i,j), W(i,j), q_2], \quad x(i,j) \in R, \quad W \in \{0,1\}, \quad q_2 \in Z^+, \tag{3}$$

$$I(i,j) = \begin{cases} 0, & \lfloor x(i,j)/q_2 \rfloor \text{ is even} \\ 1, & \lfloor x(i,j)/q_2 \rfloor \text{ is odd} \end{cases}, \tag{4}$$

$$y(i,j) = \begin{cases} \lfloor x(i,j)/q_2 \rfloor \times q_2 + q_2/2, & \text{if } I = W(i,j) \\ y'(i,j), & \text{if } I \neq W(i,j) \end{cases}. \tag{5}$$

$y'(i,j)$ is obtained as follows:

$$y'(i,j) = \begin{cases} \lfloor x(i,j)/q_2 - 1 \rfloor \times q_2 + q_2/2 & \text{if } \lfloor x(i,j)/q_2 \rfloor \times q_2 \leqslant x < \lfloor x(i,j)/q_2 \rfloor \times q_2 + q_2/2 \\ \lfloor x(i,j)/q_2 + 1 \rfloor \times q_2 + q_2/2 & \text{if } \lfloor x(i,j)/q_2 \rfloor \times q_2 + q_2/2 \leqslant x < \lfloor x(i,j)/q_2 \rfloor \times q_2 + q_2 \end{cases}. \tag{6}$$

6. Perform the parity-check quantization function f on the selected wavelet coefficients according to the location array for subband HL_2, LH_2, and HH_2.

For $i=1$ to r
for $j=1$ to c
process each location(i,j)
case 1: $HL_2'(i,j)=f[HL_2(i,j),W(i,j),q_2]$
case 2: $HH_2'(i,j)=f[HH_2(i,j),W(i,j),q_2]$
case 3: $LH_2'(i,j)=f[LH_2(i,j),W(i,j),q_2]$.

3.2 *Image Recovery Information Embedding Algorithm*

To recover the content of the tampered region if the modification is performed, low-frequency components representing important visual information are extracted and embedded into high-frequency coefficients as the recovery data. The recovery value of $LL_2(i,j)$ is obtained after the floor function $LL_2(i,j)$ divided by the quantization value of function pow$(2,\lambda)$ first. If the value after the floor function can be represented as a binary format as $(b_4,b_3,b_2,b_1,b_0)_2$, the decimal value will be equivalent to $(b_4\cdot 2^4+b_3\cdot 2^3+b_2\cdot 2^2+b_1\cdot 2^1+b_0)\times 2^\lambda$ for the recovery, which will provide a wide range of reconstruction capability. Therefore, each recovery value of $LL_2(i,j)$ will need at least 5 bits to record the (b_4,b_3,b_2,b_1,b_0) value and each b_4, b_3, b_2, b_1, b_0 is either 1 or 0, which is similar to the watermark of semifragile watermarking designed in Sec. 3.1. After intensive study using the commonly available images from the USC image database,[22] the wavelet LL_2 band coefficients using the biorthogonal filters[23] usually fall within the range 0 to 31 after scaling ($\lambda=5$). Under such circumstances the parity-check quantization mentioned in Sec. 3.1 can be used again for embedding each (b_4,b_3,b_2,b_1,b_0) value for DWT recovery coefficients. Even the proposed design deals only with the positive wavelet LL_2 band coefficients. However, it could be easily modified to include the sign bit from (b_4,b_3,b_2,b_1,b_0) to a (*sign bit*, b_3, b_2, b_1, b_0) format if negative wavelet coefficients are important.

The location of embedding recovery information for $LL_2(i,j)$ is decided by a mapping function [Eqs. (7) and (8)] demonstrated in Fig. 3, where location α ($\alpha=i\cdot c+j$) is mapped to location β (the detailed mapping procedures are explained in the algorithm section). Due to the security concern, the recovery value for each location β is embedded in either the β_1 group (HL_2 and HH_1), β_2 group (HH_2 and HH_1), or β_3 group (LH_2 and LH_1) components according to the computed result from control parameter key k_1.

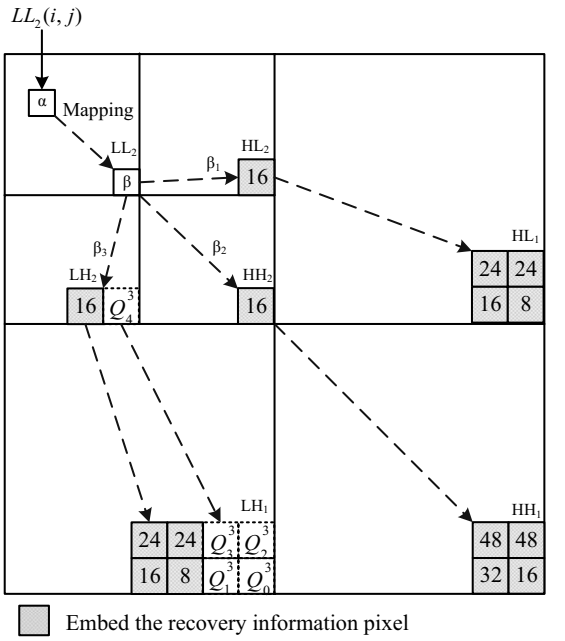

Fig. 3 Demonstration of embedding the recovery information. The recovery location α of $LL_2(i,j)$ is mapped to either β_1, β_2, or β_3. The parent-children relationship is subsequently illustrated for bands HL_2 and HL_1, HH_2 and HH_1, LH_2 and LH_1. The odd-even quantization parameter q_2 values for each bit b_4, b_3, b_2, b_1, b_0 are listed in the blocks for the associated bands.

Since there exists the parent-children relationship between the wavelet subbands, as shown in Fig. 3, the β_1, β_2, or β_3 group each has five units, with one unit in level 2 and four units in level 1. Using this characteristic, the recovery value can be represented in the binary format with 5 bits, and bit b_4 will be embedded in either the HL_2, HH_2, or LH_2 band, and b_3, b_2, b_1, b_0 will be embedded in either the HL_1, HH_1, or LH_1 band according to whether the β_1, β_2, or β_3 group is selected. If the coefficients of subband HL_2, HH_2, or LH_2 have already embed the authentication bit during Sec. 3.1, there will be a location switch mechanism to bypass the selected band for other bands if the outcome from the random generator picks the same band for the recovery information embedding. After the embedded location for bit b_4 is decided, the embedded location for bit b_3, b_2, b_1, b_0 will also be determined by the parent-children relationship.

To efficiently embed the recovery information, the parity-check quantization in Sec. 3.1 is applied again here. The associated quantization parameter value q_2 for each embedded bit location at different subbands is different according to the study of subband noise visibility.[14] For ex-

Table 1 The quantization table (Q table) of recovery bits.

q_2 values for $k=1,2,3$	Q_4^k	Q_3^k	Q_2^k	Q_1^k	Q_0^k
q_2 of β_1 group, $k=1$	16	24	24	16	8
q_2 of β_2 group, $k=2$	16	48	48	32	16
q_2 of β_3 group, $k=3$	16	24	24	16	8

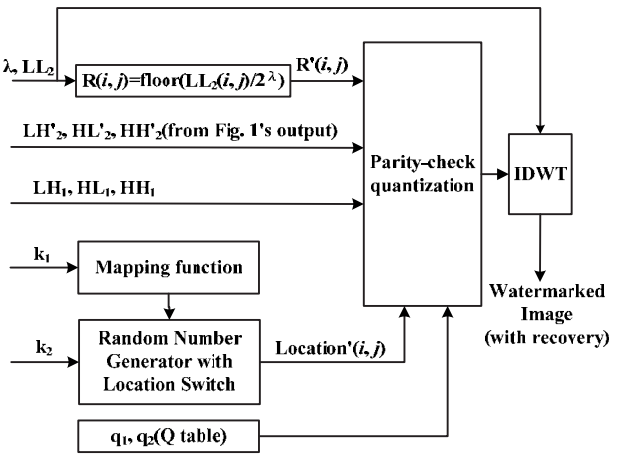

Fig. 4 Diagram of the image recovery information embedding scheme.

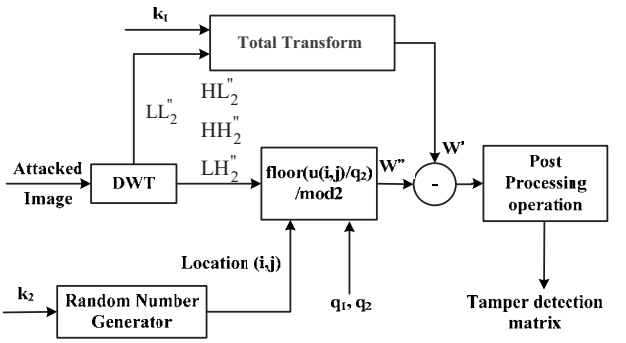

Fig. 5 Diagram of the image authentication and tamper detection scheme.

ample, the q_2 value for the bit b_4 of the β_3 group in band LH_2 will be 16, and the q_2 values for bit b_3, b_2, b_1, b_0 of the β_3 group in band LH_1 will be 24, 24, 16, and 8, respectively, shown in Fig. 3. An empirical study for the quantization step size called the Q table is shown in Table 1 for the band of HL_2, HH_2, LH_2, HL_1, HH_1, and LH_1 which is used for the β_1, β_2, or β_3 groups. Q_i^k represents each q_2 value of the parity-check quantization for the β_1, β_2, or β_3 groups, and $i=1,2,3,4,5$ and $k=1,2,3$. An example of q_2 values, including Q_4^3, Q_3^3, Q_2^3, Q_1^3, Q_0^3 for β_3 groups, are shown in Fig. 3.

The flow chart of recovery information embedding is shown in Fig. 4, and the detailed algorithm is as follows.

1. Input or use the prestored private keys k_1 and k_2. Input the quantization parameters q_1 and Q table for q_2.
2. k_2 is applied as a random seed to create the 2-D psuedorandom array $location(i,j) \in \{1,2,3\}, 1 \leqslant i \leqslant r, 1 \leqslant j \leqslant c$ that will designate the subband for authentication pixel embedding.
3. For security concerns, the location of the recovery value of $LL_2(i,j)$ will be mapped into different locations. A mapping function[20] of Eq. (8) will map the location from α to γ.

$$g(\alpha) = \gamma = (\kappa \times \alpha) \bmod N, \tag{7}$$

$$\beta = \text{switch}[\gamma, \text{location}(i,j)], \tag{8}$$

where $1 \leqslant i \leqslant r, 1 \leqslant j \leqslant c$, κ is the key k_1, and N is the total number of coefficients of subband LL_2 of the image. If the band to embed the authentication bit and the recovery information is the same, the location switch mechanism will be applied to make them embedded in different bands. The pseudocode of the switch function is as follows:

$s = [\text{rand}(k_2) \% 3] + 1$

if $[s \neq \text{location}(i,j)]$ $\beta = \beta_s = \gamma$

else

$s = (s+1) \% 3 + 1, \quad \beta = \beta_s = \gamma.$

If location α is mapped to location β after the location switch, each location β will be embedded in either the β_1 group (HL_2 and HL_1), the β_2 group (HH_2 and HH_1), or the β_3 group (LH_2 and LH_1) components. Therefore, a location list $location'(i,j) \in \{1,2,3\}, 1 \leqslant i \leqslant r, 1 \leqslant j \leqslant c$ will be obtained for subbands $\{HH_2, LH_2, HL_2\}$.

4. Adjust the control variable λ so the value of floor $[LL_2(i,j)/2^{\lambda}]$ will be mapped within the region of (0, 31). Therefore, 5-bit binary representation will be used for parity-check quantization to embed the recovery information.
The function performs quantization on $x(i,j)$ into odd-even regions according to the bit value of $R'(i,j)$; (i,j) indicates the spatial location and $x \in \{LH_2, HL_2, HH_2, LH_1, HL_1, HH_1\}$. The values of quantization parameter q_2 in different subbands are listed in the Q table of Table 1.

$$y(i,j) = f[x(i,j), R'(i,j), q_2], \tag{9}$$

$$R(i,j) = \lfloor LL_2(i,j)/2^{\lambda} \rfloor, \tag{10}$$

$$R'(i,j) = \begin{cases} 0 & \text{if } R(i,j) < 0 \\ 2^{\lambda} - 1 & \text{if } R(i,j) \geqslant 2^{\lambda} \\ R(i,j) & \text{else} \end{cases}. \tag{11}$$

5. Perform parity-check quantization on wavelet coefficients as follows.

For $i=1$ to r
for $j=1$ to c
Process each $location'(i,j)$
case 1: $HL''_2(i,j) = f[HL'_2(i,j), b_4, Q_4^1]$
$HL''_1(i,j) = f[HL'_1(2i,2j), b_3, Q_3^1]$
$HL''_1(i,j) = f[HL'_1(2i,2j+1), b_2, Q_2^1]$
$HL''_1(i,j) = f[HL'_1(2i+1,2j), b_1, Q_1^1]$
$HL''_1(i,j) = f[HL'_1(2i+1,2j+1), b_0, Q_0^1]$
case 2: $HH''_2(i,j) = f[HH'_2(i,j), b_4, Q_4^2]$

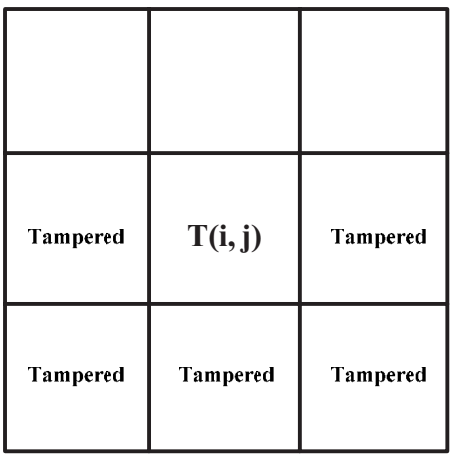

Fig. 6 The authentication pixel and its eight neighboring pixels. $T(i,j)=1$ if $\tau=1$, $\delta=4$.

$$HH''_1(i,j)=f[HH'_1(2i,2j),b_3,Q_3^2]$$
$$HH''_1(i,j)=f[HH'_1(2i,2j+1),b_2,Q_2^2]$$
$$HH''_1(i,j)=f[HH'_1(2i+1,2j),b_1,Q_1^2]$$
$$HH''_1(i,j)=f[HH'_1(2i+1,2j+1),b_0,Q_0^2]$$

case 3: $LH''_2(i,j)=f[LH'_2(i,j),b_4,Q_4^3]$

$$LH''_1(i,j)=f[LH'_1(2i,2j),b_3,Q_3^3]$$
$$LH''_1(i,j)=f[LH'_1(2i,2j+1),b_2,Q_2^3]$$
$$LH''_1(i,j)=f[LH'_1(2i+1,2j),b_1,Q_1^3]$$
$$LH''_1(i,j)=f[LH'_1(2i+1,2j+1),b_0,Q_0^3]$$

6. The watermarked image is obtained after the inverse wavelet transform.

3.3 *Semifragile Watermark Authentication and Tamper Detection Algorithm*

Figure 5 shows the flow chart of watermark authentication and tamper detection scheme, which is similar to part of semifragile watermark embedding. The procedures are as follows:

1. Input keys k_1, k_2, q_1, and q_2 as the private keys of the scheme. The values of k_1, k_2, q_1, and q_2 should be the same in embedding and extraction processes.
2. Compute the two-level 2-D wavelet coefficients of the watermarked image; $r\times c$ is the size of LL''_2.
3. Use k_1 and k_2 to create the 2-D pseudorandom arrays; $W'(i,j)\in\{0,1\}$, $1\leqslant i\leqslant r$, $1\leqslant j\leqslant c$, and $location(i,j)\in\{1,2,3\}, 1\leqslant i\leqslant r\ 1\leqslant j\leqslant c$.
4. According to $location(i,j)$, the extracted watermark will be calculated by Eq. (12), where the subband coefficient value is defined as $u(i,j)$.

$$W''(i,j)=\{[[u(i,j)/q_2]]\}\bmod 2. \tag{12}$$

5. After obtaining two watermarks W' and W'', we define the tamper detection matrix as Eq. (13). If $W'=W''$, then $T=0$. It means the image was not tampered. Otherwise, the "1" element in the tamper detection matrix indicates pixels that were tampered.

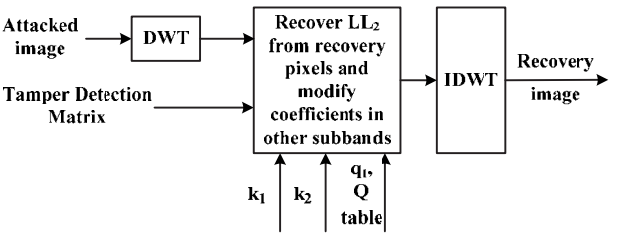

Fig. 7 Diagram of the image recovery scheme.

$$T=|W'-W''|. \tag{13}$$

6. Since the algorithm is designed to be a semifragile watermarking scheme, which would want to be robust to mild modifications in all cases, it is inevitable that we cannot detect all malicious attacks pixel-wise. However, for practical cases such as watermark removal using neighbor pixels and image cropping that crops objects from a source and pastes them onto a target, the malicious attacks are always applied in a certain region in the watermarked image. That is to say, we assume tamper pixels are always continuous. Therefore, for a certain tamper detection matrix element $T(i,j)$, if the number of tampered neighboring elements for $T(i,j)$ is greater than a given threshold, we can regard $T(i,j)$ as a tampered one. The summary of such postprocessing operations of the tamper detection matrix is shown as Eq. (14) and the demonstration is in Fig. 6:

$$T'=\begin{cases}1, & \sum_{k=-\tau}^{\tau}\sum_{l=-\tau}^{\tau}T(i+k,j+l)>\delta\\ 0, & \sum_{k=-\tau}^{\tau}\sum_{l=-\tau}^{\tau}T(i+k,j+l)\leqslant\delta\end{cases}. \tag{14}$$

Note that τ is the width of window, and δ is threshold.

7. Rescale the tamper detection matrix to have the same size of the watermarked image and obtain the tamper detection image.

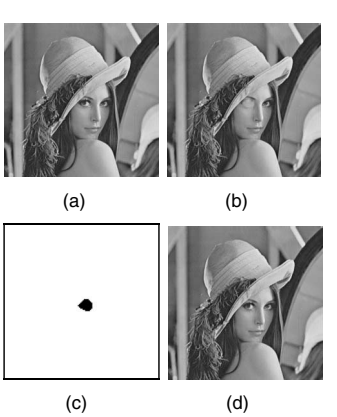

Fig. 8 Robust authentication and recovery of the tampered Lena image. (a) watermarked Lena image. (b) Tampered image. (c) Tamper detection. (d) Recovered image.

3.4 *Image Tamper Recovery Algorithm*

After the image authentication stage, all the wavelet coefficients of subband LL″ are marked either valid or erroneous by tamper detection matrix T. We only need to recover the erroneous wavelet coefficients and leave the other coefficients unchanged. If the location of erroneous coefficient is α, the mapped location embedding the recovery coefficient will be at either the β_1, β_2, or β_3 group, which is the same mapping relationship as shown in Fig. 3 by Eqs. (7) and (8). The associated Q table values of parity-check quantization parameter q_2 for each bit b_4, b_3, b_2, b_1, b_0 at different bands-are the same as shown in Table 1 and are adopted here for the recovery. The image tampering recovery procedures (Fig. 7) for each erroneous coefficient is described as follows:

1. After the image authentication scheme, we can get a tamper detection matrix T that tells where there has been tampering.
2. For each wavelet coefficient in subband LL''_2, $1 \leqslant i \leqslant r$, $1 \leqslant j \leqslant c$, if $T(i,j)=1$, which means it is tampered, the coefficients related with location (i,j) based on the parent-children relationship across subbands are also tampered. Therefore, they need to be set to zero first, since the tampered values are not useful for recovery. Then perform the recovery scheme by steps 3, 4, and 5. Else if $T(i,j)=0$, go to step 6.
3. Use the same mapping function from Eqs. (7) and (8) during the embedding procedure to get the mapping relationship between location β and location α, where the location switch function is also considered. If $T(\beta)=1$, skip the recovery operation for this coefficient since the recovery info is also tampered.
4. Use the watermark extraction scheme by parity-check quantization to get the 5-bit recovery information $\{b'_4(i,j), b'_3(i,j), b'_2(i,j), b'_1(i,j), b'_0(i,j)\}$.

For each $location'(i,j)$

case 1: $b'_4(i,j)=(\lfloor(\mathrm{HL}''_2(i,j)/Q_4^1)\rfloor)\mathrm{mod}2$
$b'_3(i,j)=(\lfloor(\mathrm{HL}''_1(2i,2j)/Q_3^1)\rfloor)\mathrm{mod}2$
$b'_2(i,j)=(\lfloor(\mathrm{HL}''_1(2i,2j+1)/Q_2^1)\rfloor)\mathrm{mod}2$
$b'_1(i,j)=(\lfloor(\mathrm{HL}''_1(2i+1,2j)/Q_1^1)\rfloor)\mathrm{mod}2$
$b'_0(i,j)=(\lfloor(\mathrm{HL}''_1(2i+1,2j+1)/Q_0^1)\rfloor)\mathrm{mod}2$

case 2: $b'_4(i,j)=(\lfloor(\mathrm{HH}''_2(i,j)/Q_4^2)\rfloor)\mathrm{mod}2$
$b'_3(i,j)=(\lfloor(\mathrm{HH}''_1(2i,2j)/Q_3^2)\rfloor)\mathrm{mod}2$
$b'_2(i,j)=(\lfloor(\mathrm{HH}''_1(2i,2j+1)/Q_2^2)\rfloor)\mathrm{mod}2$
$b'_1(i,j)=(\lfloor(\mathrm{HH}''_1(2i+1,2j)/Q_1^2)\rfloor)\mathrm{mod}2$
$b'_0(i,j)=(\lfloor(\mathrm{HH}''_1(2i+1,2j+1)/Q_0^2)\rfloor)\mathrm{mod}2$

case 3: $b'_4(i,j)=(\lfloor(\mathrm{LH}''_2(i,j)/Q_4^3)\rfloor)\mathrm{mod}2$
$b'_3(i,j)=(\lfloor(\mathrm{LH}''_1(2i,2j)/Q_3^3)\rfloor)\mathrm{mod}2$
$b'_2(i,j)=(\lfloor(\mathrm{LH}''_1(2i,2j+1)/Q_2^3)\rfloor)\mathrm{mod}2$
$b'_1(i,j)=(\lfloor(\mathrm{LH}''_1(2i+1,2j)/Q_1^3)\rfloor)\mathrm{mod}2$
$b'_0(i,j)=(\lfloor(\mathrm{LH}''_1(2i+1,2j+1)/Q_0^3)\rfloor)\mathrm{mod}2.$

5. According to the extracted recovery bits, the recovered value $R''(\alpha)$ of the LL_2 wavelet coefficient at location α will be calculated by Eq. (15):

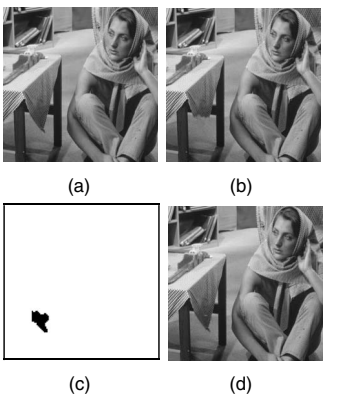

Fig. 9 Robust authentication and recovery of the tampered Barbara image. (a) Watermarked Barbara image. (b) Tampered image. (c) Tamper detection. (d) Recovered image.

$$R''(\alpha) = [b'_4(i,j)\cdot 2^4 + b'_3(i,j)\cdot 2^3 + b'_2(i,j)\cdot 2^2 + b'_1(i,j)\cdot 2^1 + b'_0(i,j)\cdot 2^0] \times 2^{\lambda}. \quad (15)$$

Use the recovered info $R''(\alpha)$ as the value of the $\mathrm{LL}''_2(i,j)$ wavelet subband coefficient.

6. After all LL''_2 coefficients are all recovered, perform the inverse 2-D wavelet transform to obtain the reconstructed image.

4 Experiments and Discussion

The proposed robust authentication and recovery semifragile watermarking algorithm has been implemented and intensively tested by using the commonly available image database. For illustration purposes, four widely available images of 512×512 Lena, Barbara, Peppers, and F16 are tampered in Figs. 8–11, and are recovered based on the approach of the proposed algorithm. During the simulation, the parameters of robust authentication are $q_1=30$, $q_2=16$, and $\lambda=5$, which result in the watermarked images being above 39 dB compared to the original images. Table 2 lists the PSNR values of Figs. 8–11 and we can tell that the images including recovery information will be degraded about 9 dB as the payload. In addition, PSNR values of the

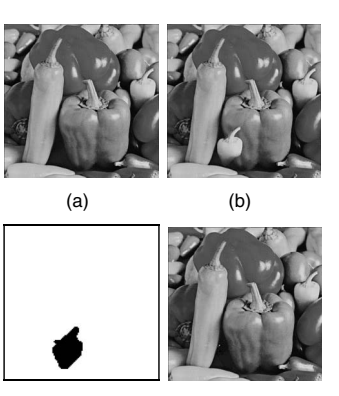

Fig. 10 Robust authentication and recovery of the tampered Peppers image. (a) Watermarked Peppers image. (b) Tampered image. (c) Tamper detection. (d) Recovered image.

Table 2 PSNR values (dB) of the watermarked images versus the original images for Figs. 8–11.

Image	Lena	Barbara	Peppers	F16
With recovery	30.7	30.5	30.6	30.8
No recovery	39.5	39.4	39.4	39.4

tampered and recovery images with watermarked and original images for Figs. 8–11 are tabulated in Table 3. According to our design, the robust authentication can correctly indicate the location of the tampered areas in Figs. 8(c), 9(c), 10(c), and 11(c), and the recovery mechanism can also successfully reconstruct the images in Figs. 8(d), 9(d), 10(d), and 11(d), respectively. From Fig. 8, the eye portion of Lena is reconstructed with no visible difference. The reconstructed Barbara image in Fig. 9 is also very similar to the original one. While the tampered area increases, the PSNR values are decreasing in Table 3, and the rough recovery image will disclose the discrepancy, which can be seen in Fig. 11(d).

Since the extracted watermark in this study is the binary sequence, the correlation value should be evaluated with the original watermark to judge the robustness. Therefore, the normalized cross-correlation (NC) function is adopted here as the criteria for comparison. The equation is as follows:

$$\mathrm{NC} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c} W(i,j)W'(i,j)}{\sum_{i=1}^{r}\sum_{j=1}^{c} |W(i,j)|^2}. \tag{16}$$

To compare with the other semifragile approach, Table 4 tabulated the NC values of the tampered watermarked Peppers image, set at the same PSNR as tested in Ref. 17. In our approach, 10−18 biorthogonal wavelet filters[23] are used for our experiments, even though there are no limitations for the selection of wavelet filters for the proposed design. Although there is no information about the wavelet filters and chaotic map used in Ref. 17, we believe the authors tried for the best results. From Table 4, we can see

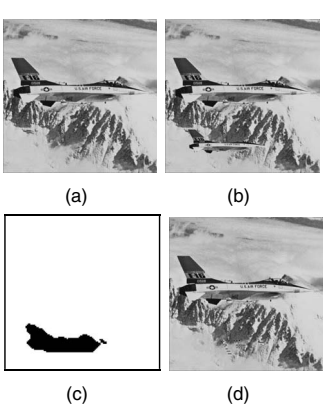

Fig. 11 Robust authentication and recovery of the tampered F16 image. (a)Watermarked F16 image. (b) Tampered image. (c) Tamper detection. (d) Recovered image.

Table 3 PSNR values (dB) for Figs. 8–11. T means tempered, W means watermarked, R means recovery, and O means original image.

Image	Lena	Barbara	Peppers	F16
PSNR(T,W)	35.9	35.0	22.7	23.8
PSNR(R,W)	39.3	44.0	39.8	33.8
PSNR(T,O)	29.6	29.3	22.1	23.0
PSNR(R,O)	30.2	30.4	30.4	29.3

the NC values for the proposed algorithm are higher than the data from Ref. 17. In addition, the recovery scheme of the proposed algorithm can recover the tampered area, which is a novel approach among existing techniques. Figure 12 demonstrates the images when Peppers is under both the JPEG compression at QF=80 and the tampered attack. Figure 13 shows the images when Peppers is under both AWGN noise $\sigma^2=12$ and the tampered attack. Based on Figs. 12(d) and 13(d), our scheme has been able to recover most of the distorted information with high image quality. Therefore, we can conclude that our robust authentication and recovery semifragile watermarking can resist mild attacks like JPEG and AWGN.

To compare with other recovery schemes in the transform domain, Lin and Chang's[19] method is essentially a good fit for analysis. In Fig. 14, the watermarked Lena images by Lin and Chang[19] and the proposed algorithm are illustrated for comparison. They are both set at the same PSNR value=32 dB. The pin of Lena's hat is removed for both watermarked images. Since Lin and Chang's[19] method is a DCT-based approach, the outcome of the tamper detection is also block oriented for recovery. However, the proposed algorithm is a wavelet-based technique, which has the property of multiresolution characteristics. Therefore, the recovery information is embedded in multiresolution subbands that offer better recovery image quality than the block-based approach. From Figs. 14(g) and 14(h), we can

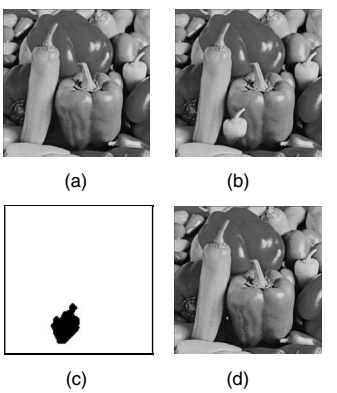

Fig. 12 Tamper detection and recovery of the Peppers image under both JPEG compression QF=80 (compression ratio=4.3:1) and tampering attack. (a) Watermarked Peppers image. (b) JPEG compression and tampering. (c) Tamper detection. (d) Recovered image.

Table 4 Roubustness against JPEG compression and AWGN for Peppers image.

JPEG quality factor	100	90	80	70	60	50
NC	0.99	0.99	0.98	0.96	0.93	0.88
NC of Ref. 17	0.99	0.98	0.95	0.93	0.88	0.82
AWGN: σ^2	6	12	18	24	30	36
NC	0.98	0.95	0.92	0.89	0.86	0.82
NC of Ref. 17	0.97	0.94	0.91	0.87	0.83	0.79

easily distinguish the image content and verify that the image quality of the pin in Fig. 14(h) is superior to that of Fig. 14(g). It is another advantage of the wavelet-based scheme.

Further research could try to perform more levels of wavelet transforms. It would reduce the number of embedded watermarks, so as to reduce the quantization extent of images to enhance the watermarked image quality. But the size of the detection unit would be enlarged according to more levels of wavelet transform. Because some of the subtle distortion of the tampered image could not be detected, it is still an issue for detection efficiency and recovery capability.

After intensive performance comparison, the results of different attacks of tamper detection, JPEG and AWGN noise, and visual quality analyses demonstrate that the proposed wavelet-based semifragile watermarking is more robust with better image quality. In summary, we are convinced that the proposed complete architecture and algorithm is a superior scheme among the existing techniques.

5 Conclusion

A novel semifragile watermarking-based technique for copyright protection of DRM and robust authentication with a recovery algorithm is presented. For authentication and verification of the integrity for the watermarked images, we apply a semifragile watermark algorithm that can detect and localize malicious attacks effectively yet tolerate mild modifications such as JPEG compression and channel additive white Gaussian noise (AWGN). As compared with other methods, our approach is not only superior in tamper detection and localization, it also provides the capability of tamper recovery. According to the simulation results, the watermarked image can be recovered successfully under mild attacks with higher image quality than DCT-based techniques.

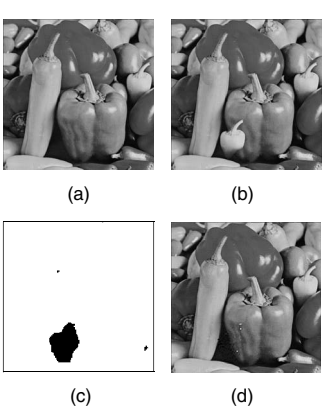

Fig. 13 Tamper detection and recovery of the Peppers image under AWGN noise $\sigma^2=12$ and tampering attack. (a) Watermarked Peppers image. (b) AWGN attack and tampered image. (c) Tamper detection. (d) Recovered image.

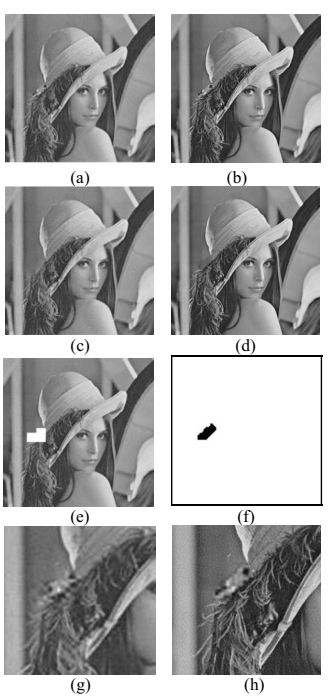

Fig. 14 Recovery performance comparison with the algorithm in Ref. 19 at the same PSNR value=32 dB. (a) is the watermarked image by the method of Ref. 19. (b) is the watermarked image by the proposed approach. (c) is the tampered image of (a), where the pin of Lena's hat is removed. (d) is the tampered image of (b), where the pin of Lena's hat is removed. (e) is the tamper detection using Ref. 19, where the white blocks indicate the tampered area. (f) is the tamper detection by the proposed approach. (g) is a close-up of the recovery image by Ref. 19. (h) is a close-up of the recovery image by the proposed approach.

Acknowledgments

This work was supported by the National Science Council in Taiwan under NSC95-2416-H009-027 and NSC96-2416-H009-015.

References

1. S. J. Lee and S. H. Jung, "A survey of watermarking techniques applied to multimedia," *Proc. IEEE Intl. Symp. Industrial Electron.* **1**, 272–277 (2001).
2. C. C. Chang, Y. S. Hu, and T. C. Lu, "A watermarking-based image ownership and tampering authentication scheme," *Pattern Recogn. Lett.* **27**, 439–446 (2006).
3. S. Walton, "Information authentication for a slippery new age," *Dr. Dobbs J.* **20**(4), 18–26 (1995).
4. H. Yuan and X. P. Zhang, "A secret key based multiscale fragile watermark in the wavelet domain," *Proc. IEEE Intl. Conf. Multimedia Expo*, pp. 1333–1336 (2006).
5. R. B. Wolfgang and E. J. Delp, "A watermark for digital images," *Proc. IEEE Intl. Conf. Image Process.* **3**, 219–222 (1996).
6. D. Zou, Y. Q. Shi, Z. Ni, and W. Su, "A semi-fragile lossless digital watermarking scheme based on integer wavelet transform," *IEEE Trans. Circuits Syst. Video Technol.* **16**(10), 1294–1300 (2006).
7. C. Fei, D. Kundur, and R. Kwong, "Analysis and design of secure watermark-based authentication systems," *IEEE Trans. Inf. Forensics Security* **1**, 43–55 (2006).
8. Ö. Ekici, B. Sankur, B. Coskun, U. Nazi, and M. Akcay, "Comparative evaluation of semifragile watermarking algorithms," *J. Electron. Imaging* **13**(1), 209–216 (2004).
9. C. Rey and J. L. Dugelay, "A survey of watermarking algorithms for image authentication," *EURASIP J. Appl. Signal Process.* **6**, 613–621 (2002).
10. I. J. Cox, J. Kilian, J. F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.* **6**(12), 1673–1687 (1997).
11. D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," *Proc. IEEE* **87**(7), 1167–1180 (1999).
12. H. P. Alexandre and K. W. Rabab, "Wavelet-based digital watermarking for image," *Proc. IEEE Canadian Conf. Electric. Computer Eng.* **1**, 879–884 (2002).
13. J. Q. Hu, J. W. Huang, D. R. Huang, and Y. Q. Shi, "Image fragile watermarking based on fusion of multi-resolution tamper detection," *Electron. Lett.* **38**(24), 1512–1523 (2002).
14. A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Trans. Image Process.* **6**(8), 1164–1175 (1997).
15. Z. M. Lu, C. H. Liu, D. G. Xu, and S. H. Sun, "Semi-fragile image watermarking method based on index constrained vector quantization," *Electron. Lett.* **39**(1), 35–36 (2003).
16. H. Yuan and X. P. Zhang, "A multiscale fragile watermarking based on the Gaussian mixture model in the wavelet domain," *Proc. IEEE Intl. Conf. Acoust. Speech Signal Process.* **3**, 413–416 (2004).
17. K. Ding, C. He, L. G. Jiang, and H. X. Wang, "Wavelet-based semi-fragile watermarking with tamper detection," *IEICE Trans. Fundamentals* **E88-A**(3), 787–790 (2005).
18. Y. Chu, Y. Zhang, S. Zhang, and X. Ye, "Region of interest fragile watermarking for image authentication," *Proc. First Intl. Multi-Symp. Computer Comput. Sci. imsccs* **1**, 726–731 (2006).
19. C. Y. Lin and S. F. Chang, "Semi-fragile watermarking for authenticating JPEG visual content," *Proc. SPIE* **3971**, 140–151 (2000).
20. P. L. Lin, C. K. Hsieh, and P. W. Huang, "A hierarchical digital watermarking method for image tamper detection and recovery," *Pattern Recogn.* **38**(12), 2519–2529 (2005).
21. M. J. Tsai, K. Y. Yu, and Y. Z. Chen, "Joint wavelet and spatial transformation for digital watermarking," *IEEE Trans. Consum. Electron.* **46**(1), 241–245 (2000).
22. USC SIPI—The USC-SIPI Image Database, see http://sipi.usc.edu/services/database/Database.html.
23. M. J. Tsai, J. D. Villasenor, and F. Chen, "Stack-run image coding," *IEEE Trans. Circuits Syst. Video Technol.* **6**, 519–521 (1996).

Min-Jen Tsai received his BS degree in electrical engineering from National Taiwan University in 1987, his MS degree in industrial engineering and operations research from University of California at Berkeley in 1991, and his engineer and PhD degrees in electrical engineering from University of California at Los Angeles in 1993 and 1996, respectively. He served as a second lieutenant in the Taiwan army from 1987 to 1989. From 1996 to 1997, he was a senior researcher at America Online, Incorporated. In 1997, he joined the Institute of Information Management at National Chiao Tung University in Taiwan, and is currently an associate professor. His research interests include multimedia systems and applications, digital rights management, digital watermarking and authentication, digital forensics, and enterprise computing for electronic commerce applications. He is a member of IEEE, ACM, IEICE, and Eta Kappa Nu.

Chih-Cheng Chien received his BS degree from the Department of Information Management at National Central University in 2005, and his MS degree from the Institute of Information Management at National Chiao Tung University in 2007. He is now serving his military service duty in Taiwan. His research interests are in the fields of multimedia and image security.