# Wavelet-Based Semi-Fragile Watermarking with Tamper Detection

Jun-Hyuk Lee, Hun Jung, Yeung-Su Seo, Chun-Gun Yu, Hae-Woo Park
*Graduate School of Computer Engineering, YeungNam University*

## Abstract

*In this letter, a novel wavelet-based semi-fragile watermarking scheme is presented which exploiting the time-frequency feature of chaotic map. We also analyze the robustness to mild modification and fragility to malicious attack of our scheme. Its application includes tamper detection, image verification and copyright protection of multimedia content. Simulation results show the scheme can detect and localize malicious attacks with high peak signal-to-noise ratio (PSNR), while tolerating certain degree of JPEG compression and channel additive white Gaussian noise (AWGN)*

*Key words: watermarking, semi-fragile, wavelet, tamper detection*

## 1. Introduction

Digital watermarking plays an important role in image verification and copyright protection. In many watermarking schemes, they are designed to be very robust to image processing. Even if the watermarked image is severely attacked, the copyright information is still reserved. This is called robust watermarking. Another type of watermarking is fragile watermarking. They are usually very fragile to image processing. A mild modification can prevent from watermark extraction. They offers tamper proofing, thus greatly intensifies the credibility and integrity of multimedia content. Semi-fragile watermarking is robust to mild modification such as JPEG compression and channel additive white Gaussian noise (AWGN) but fragile to malicious attacks [1], [2]. The pioneer fragile watermarking techniques [4], [5], [7], [8] were usually based on the sensitivity of hash function to tampering. They can localize tamper, but they treat JPEG compression and channel AWGN as malicious attacks. Reference [1] presents a novel semifragile watermarking method based on index constrained vector quantization (VQ). However, it can only tolerate a limited degree of JPEG compression and the peak signal-tonoise ratio (PSNR) is low. Moreover, the codebook of vector quantization must be known in both watermark embedding and extraction side. Reference [2] presents a fragile watermarking based on fusion of multi-resolution. Tamper detection is obtained through the time-frequency feature of DWT.761-1784.

The Watson's quantization matrix and the human visual system (HVS) are elaborately used in quantization process. In Ref. [3], the authors alter the wavelet domain coefficients according to the binary watermark stream to meet the statistical parameter equity relationship based on the Gaussian mixture statistical model. They tell mild modification from malicious attack through parameter trends. But they treat AWGN as malicious attack. The time-frequency feature of DWT is also adopted in our letter, but we combine this with the chaotic map to achieve better performance.

In this letter, we propose a wavelet-based chaotic semifragile watermarking scheme aiming at better performance. In Sect. 2, we first introduce the chaotic theory. Then we investigate in detail both watermark embedding and watermark detection process. In Sect. 3, we give the performance analysis of the scheme. In Sect. 4, we show the simulation results of the proposed scheme. Finally we draw the conclusion in Sect. 5.

## 2. The Proposed Scheme

We first perform 2-level 2-D wavelet decomposition of the host image. Thus we obtain $\{LL_2,HL_2, LH_2,HH_2, HL_1,LH_1,HH_1\}$. The gray-scale value in $LL_2$ sub-band is mapped as the initial value of chaotic map. After iterating several times, we get the binary watermark then we perform odd-even quantization to gray-scale value in $\{HL_2,LH_2,HH_2\}$ sub-bands, which is in accordance with the sensitivity of the human visual system (HVS). The scheme can detect and localize malicious attack thanks to the high sensitivity on initial value of chaotic map. By properly choosing two quantization parameters $\Delta_1$ and $\Delta_2$, we get the robustness against mild modification such as AWGN and JPEG compression. We describe the proposed scheme in detail as follows.

### 2.1 Chaotic Map

In our scheme, chaos is used to generate a pseudo-random sequence. Chaos is a special solution for non-linear equation. Its random output is decided by a determinate equation. We can control a set of parameters of chaotic system as the key of the proposed scheme. Many research on this issue shows that it is more secure than traditional ways to generate a pseudo-random sequence. Consider a 1D discrete chaotic map

$$f : U \rightarrow U, U \subset R:$$

$$z(n+1) = f(\lambda, z(n)),\ \lambda \in R,\quad z(n) \in U \tag{1}$$

### 2.2 Watermark Embedding

The diagram of watermark embedding scheme is shown in Fig. 1. Unlike usual watermarking scheme, watermark is created according to the host image in our algorithm. The binary watermark is embedded into the host image by using a simple odd-even quantization. Define odd-even quantization function:

$$y = g(x, b, d)\quad x \in R\quad b \in \{0, 1\}\quad d \in Z^{+} \tag{2}$$

which performs quantization on x into odd-even region according to a binary number b. d is the quantization parameter. We calculate

$$I = \begin{cases} 0 & \lfloor x/d \rfloor \ \ is \ \ even \\ 1 & \lfloor x/d \rfloor \ \ is \ \ odd \end{cases} \tag{3}$$

$y$ is obtained as follows:

$$y = \begin{cases} \lfloor x/d \rfloor \cdot d + d/2 + x_r & if \quad I = b \\ y' + x_r & if \quad I \neq b \end{cases} \tag{4}$$

where

$$y' = \begin{cases} (\ \lfloor x/d \rfloor - 1) \cdot d + d/2 \quad if \\ \qquad x \in \left[\ \lfloor x/d \rfloor \cdot d,\ \ \lfloor x/d \rfloor \cdot d + d/2\ \right) \\ (\lfloor x/d \rfloor + 1) \cdot d + d/2 \quad if \\ \qquad x \in \left[\ \lfloor x/d \rfloor \cdot d + d/2,\ \ \lfloor x/d \rfloor \cdot d + d\ \right) \end{cases} \tag{5}$$

$$x_r = sgn(x)(|x| \bmod 1) \tag{6}$$

[ · ] denotes the floor function. We quantize the integral part while remaining the decimal part untouched, which highly promotes the invisibility of our scheme.

Watermark embedding algorithm:

S1. Select parameters: $K_1$, $K_2$, $\Delta_1$, $\Delta_2$. $K_1$ and $K_2$ are the private keys of the scheme. $\Delta_1$ and $\Delta_2$ are the quantization parameters.

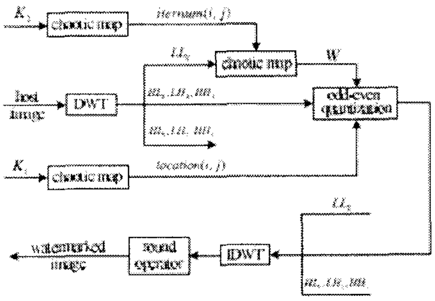


**Figure. 1 Diagram of watermark embedding scheme.**

S2. 2-level 2-D wavelet decomposition of host image:

$\{LL_2, HL_2, LH_2, HH_2, HL_1, LH_1, HH_1\}$,

$r \times c$ is the size of $LL_2$.

S3. Create two-dimensional pseudo-random arrays location(i, j) $\in \{1, 2, 3\}$, $1 \leq i \leq r\ 1 \leq j \leq c$ and iternum(i, j), $1 \leq i \leq r\ 1 \leq j \leq c$ separately using private keys $K_1$ and $K_2$.

S4. Map Qnum =[ $LL_2(i, j)/\Delta_1$ ] as the initial value of the chaotic map. After iterate iternum(i, j) times, we obtain W(i, j) $\in \{0, 1\}$. Applying such operation to every coefficient in $LL_2$, we get the binary watermark W of size $r \times c$.

S5. Perform quantization on wavelet coefficients as follows:

```
FOR i =1 to r
FOR j =1 to c
SWITCH location(i, j)
CASE 1:$HL_2(i, j) = g(HL_2(i, j), W(i, j), \Delta_2)$
CASE 2:$LH_2(i, j) = g(LH_2(i, j), W(i, j), \Delta_2)$
CASE 3:$HH_2(i, j) = g(HH_2(i, j), W(i, j), \Delta_2)$
```

S6. The reconstructed image is obtained through 2-level 2-D wavelet reconstruction. Because the DWT and inverse DWT (IDWT) are both floating-point operation and the gray-scale value of digital image is integer, thus the ultimate watermarked image is obtained by performing round operator on the reconstructed image.

### 2.3 Watermark Detection

The diagram of watermark detection scheme is shown in Fig. 2, which is similar to the first part of watermark embedding. For a possible watermarked image, applying step S2-S4, we easily get the extracted watermark *W'*. On the other hand, we may obtain another version of extracted watermark. Find the sub-band and the quantized coefficient, say u(i, j), according to location(i, j). The second one may be obtained by the following formula:

$$W''(i, j) = (\lfloor u(i, j)/\Delta_2 \rfloor) \bmod 2 \qquad (7)$$

Having obtained the two versions of extracted watermark, we define the tamper detection matrix:

$$\mathbf{T} = |W' - W''| \qquad (8)$$

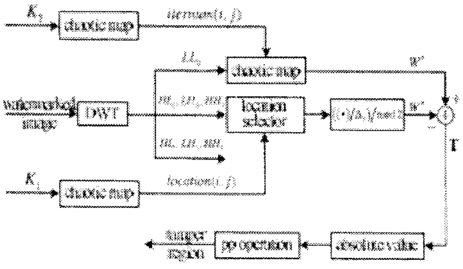


**Figure.2 Diagram of watermark detection scheme.**

If *W'* = *W''*, then T = 0. This means that the watermarked image was not tampered. Otherwise, the '1' elements in the tamper detection matrix denote the pixels that were tampered. Note that the size of tamper detection matrix is r ×c, which is about 1/16 of the host image. Thus one element in the matrix denotes a corresponding 4 × 4 block in the host image. Because of the floating-point operation of DWT and IDWT, round operator must be performed before the watermarked image is obtained. Thus the tamper detection matrix will not be equal to 0 even if there is no attack. Our proposed scheme is blind because the original image is not required during the watermark detection process.

Since our scheme is designed to be robust to mild modification in all cases, it is inevitable that we do not detect all malicious attack in pixel-wise. For a practical case such as image-content-add attack, we assume that malicious attack always be applied in a certain region in host image. Thus tamper pixels are always continuous. For a certain tamper detection matrix element T(i, j), if the number of tampered element in the neighboring elements of T(i, j) is greater than a given threshold, we thus consider T(i, j) to be tampered,too. The following equation is the summary of such postprocessing (PP) operation of tamper detection matrix:

$$T'(i, j) = \begin{cases} 1, & \sum_{u=-\alpha}^{\alpha} \sum_{v=-\alpha}^{\alpha} T(i+u, j+v) > \beta \\ 0, & \sum_{u=-\alpha}^{\alpha} \sum_{v=-\alpha}^{\alpha} T(i+u, j+v) \le \beta \end{cases} \qquad (9)$$

## 3. Performance Analysis of the Scheme

Semi-fragile watermarking can localize malicious tamper while tolerating certain degree of JPEG compression and channel AWGN. It is extremely suitable for wireless image communications where the channel is corrupted by AWGN.

According to quantization parameters $\Delta_1$ and $\Delta_2$, when the difference between quantized coefficient value varies in $[0, \Delta_2)$ and that in $LL_2$ sub-band varies in $[0, \Delta_1)$, correct detection is guaranteed and robustness is obtained. Otherwise, it will be viewed as malicious tamper. That is the case when cropping, image content addition or other malicious attack happens. Thanks to the pseudo-random characteristic of the chaotic map, when tampers exceed the above region, the probability of correct detection is 0.5. Thus 50% of such tampering coefficients will be correctly detected from the statistical aspect. Moreover, they will be distributed in the tampered region randomly and the profile of tampered region is clear. By applying the PP operation, the tamper detection matrix will be more effective to malicious attack. Thus fragility and tamper localization are guaranteed.

$\Delta_1$ partly controls the sensitivity and security of proposed scheme. The larger $\Delta_1$, the less number of initial values of the chaotic map will be. This leads to the reduction of the sensitivity and security of the whole scheme. On the other hand, if we choose smaller $\Delta_1$, mild modification will change the initial value of the chaotic map, so we choose a moderate value to guarantee a certain degree of robustness.

$\Delta_2$ controls the robustness together with $\Delta_1$. Larger leads to more robustness. While smaller $\Delta_2$ leads to more fragility. Thus a tradeoff exists. Note that only $\Delta_2$ influences the PSNR of the watermarked image. For embedding process, the computational complexity is mainly decided by the total number of embedding position and the chaotic iteration number of each embedding position, which is O(E(iternum) × r × c), where E(.) denotes the expectation operator. This is the same for watermark detection process. The computational complexity is much higher in Ref. [2] since the vector quantization in their scheme needs code book design, training and a VQ encoder.

## 4. Simulation Results

To check the validity of our scheme, many attacks have been conducted. 512 × 512 × 8 'Peppers' gray-scale image is adopted as illustration purpose. We first perform 2-level 2-D wavelet decomposition of the 'Peppers' image. Let $\Delta_1 = 30$, $\Delta_2 = 16$. Figure 3(a) shows the watermarked image with PSNR=41.6 dB. It is noted that the PSNR is only 31.85 dB with the scheme in [1] and 38.43 dB in [2]. We add two peppers in the watermarked image which is shown in Fig. 3(b). The tampered region is shown in Fig. 3(c). If we consider the situation when the image is mildly modified such as severe AWGN while on the other hand it is maliciously attacked such as image content addition attack, Fig. 4 shows the results of PP operation when using different parameters. It is very like the erosion operation in image processing. When $\alpha = 1$, $\beta = 3$, we can tell the malicious attack from the mild modification very clearly. To be simplicity, we choose $\alpha = 1$, $\beta = 3$ in the following experiments. This means we define one-order neighborhood, and set the threshold in the middle of total element number. The results of above experiments show that our algorithm is capable of tamper detection with relative precise localization

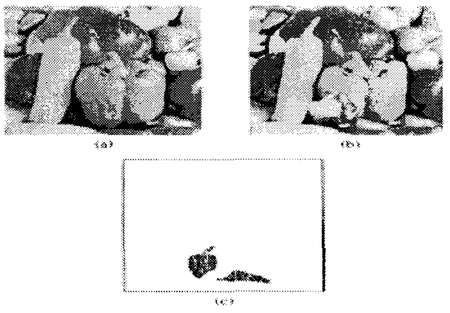


**Figure.3 Result of tamper detection. (a) The watermarked image. (b) Im-age after attack. (c) The tampered region**

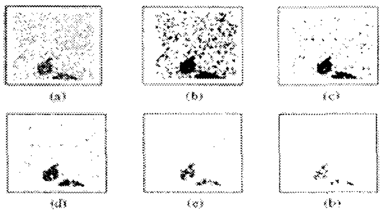
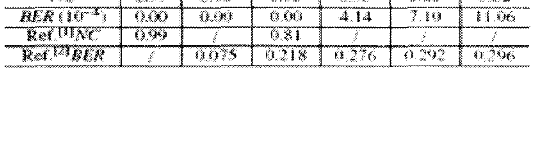


**Figure. 4 PP operation on tamper detection matrix. (a)No PP operation (b) $\alpha = 1, \beta = 1$ (c) $\alpha = 1, \beta = 2$ (d) $\alpha = 1, \beta = 3$(e) $\alpha = 1, \beta = 4$(f) $\alpha = 1, \beta = 5$**

**Table 1. Robustness against AWGN and JPEG compression.**



| AWGN:$\sigma^2$ | 6 | 12 | 18 | 24 | 30 | 36 |
|---|---|---|---|---|---|---|
| NC | 0.97 | 0.94 | 0.91 | 0.87 | 0.83 | 0.79 |
| BER ($10^{-4}$) | 0.00 | 2.37 | 4.14 | 7.10 | 10.32 | 13.73 |
| JPEG:QF | 100 | 90 | 80 | 70 | 60 | 50 |
| NC | 0.99 | 0.98 | 0.95 | 0.93 | 0.88 | 0.82 |
| *BER* ($10^{-4}$) | 0.00 | 0.00 | 0.00 | 4.14 | 7.19 | 11.06 |
| Ref.[1]NC | 0.99 | / | 0.81 | / | / | / |
| Ref.[2]*BER* | / | 0.075 | 0.218 | 0.276 | 0.292 | 0.296 |

Table 1 shows the NC value and BER after AWGN with different variance σ2 and JPEG compression with different quality factor (QF). We also list the NC values in Ref. [1]. When the QF is 80%, the NC value of our scheme is 0.95 while that of Ref. [1] 0.81.

From the results of this experiment we can see that our scheme can tolerate more JPEG compression than Ref. [1]. The BER values in Ref. [2] are listed at the bottom of Table 1. It can be found that our scheme is more robust than Ref. [2]

## 5. Conclusion

A novel wavelet-based chaotic semi-fragile watermarking is presented. The main characteristic of our scheme is as follows. (i) Robust to mild modification and fragile to malicious attack. (ii) Relative precise localization. (iii) High invisibility. (iv) Low computational complexity. Further research can lay emphasis on theoretical analysis and optimization of watermark embedding position.

## 6. References

[1] 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., vol.39, no.1, pp.35–36, 2003.

[2] 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., vol.38, no.24, pp.1512–1523, 2002.

[3] H. Yuan and X.P.Zhang, "A multiscale fragile watermarking based on the Gaussian mixture model in the wavelet domain," Proc. 2004 Int. Conf. on Acoustics, Speech and Signal Processing, vol.3, pp.413–416, Montreal, QC, Canada, May 2004.

[4] M.U. Celik, G.Sharma, E.Saber, and A.M.Tekalp, "Hierarchical watermarking for secure image authentication with localization," IEEE Trans. Image Process., vol.11, no.6, pp.585–595, 2002.

[5]P.W.Wong, "Public key watermark for image verification and authentication," Proc. 1998 Int. Conf. on Image Processing, vol.1, pp.455–459, Chicago, IL, Oct. 1998.

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

[7] H.T. Lu, R.M. Shen, and F.L.Chung, "Fragile watermarking scheme for image authentication," Electron. Lett., vol.39, no.12, pp.898–900, 2003.

[8] H.Tamori, N.Aoki, and T.Yamamoto, "A fragile digital watermarking technique by number theoretic transform," IEICE Trans. Fundamentals, vol.E85-A, no.8, pp.1902–1904, Aug. 2002.