Josai Mathematical Monographs
vol. 7 (2014), pp. 41-54

Digital Image Watermarking with Robust, Semifragile, and Fragile Properties

Hyunho Kang and Keiichi Iwamura

Abstract. Digital watermarking has been considered as a solution for copyright protection applications. However, in some practical applications, it is necessary to use multiple watermarks for different purposes. In this study, we embed robust, semifragile, and fragile watermarks simultaneously. To achieve this goal, we previously reported the results obtained by combining an existing scheme with our novel method. In the present study, we describe a more efficient combined method based on our recent findings, including negative correlation watermarking for a robust scheme, just-noticeable differences visual model watermarking for a semifragile method, and error-diffusion watermarking for a fragile property. The experimental evaluations showed that the proposed method is effective for multiple watermarking.

1. Introduction

We live in a digital and Internet world, where the security of multimedia data on the Internet is a challenging topic. Most of the existing watermarking schemes used to address this problem are designed for copyright protection or content authentication.

Practical applications often require the use of multiple watermarks for different purposes. In our previous study [**1**], we reported the results obtained by combining an existing scheme with our novel method. In this study, we present a more efficient multiple watermarking method based on three of our recently proposed approaches.

The first approach, called robust watermarking (RW), is used for copyright protection and the embedded watermark should be resistant to any processing that does not seriously affect the quality of the host image. To achieve this goal, the difference between the frequency coefficients and uniformly distributed real numbers is used as the embedded watermark based on our previous study [**2**].

The second approach, called semifragile watermarking (SFW), is used for soft image authentication and integrity verification. Thus, the watermark should be insensitive to mild modifications such as lossy compression, but fragile to any malicious attempt to modify the image content. To achieve this goal, we classify the nature of the attacks by counting the number of non-detected blocks.

The third approach, called fragile watermarking (FW), is used for strict image authentication and integrity verification. Thus, the watermark should not tolerate

any tampering and any changes or modifications of the image will be reflected in the hidden watermark [**3**][**4**]. In this method, we use a watermark based on an error-diffusion scheme that achieves dithering by diffusing the quantization error of a pixel to its neighboring pixels, according to the distribution coefficients [**5**].

In particular, when we consider image authentication, practical methods should be more robust to normal noise and lossy compression, which is necessary for efficient transmission over the Internet. Therefore, classical authentication techniques, such as hash-based message authentication codes [**6**] and digital signature algorithms that encrypt the hash value of the message using a public key authentication mechanism [**7**], are not appropriate for the Internet environment. Hence, soft image authentication is desired. Soft image authentication is sensitive to content modification and severe image quality tampering, whereas hard image authentication is highly sensitive and it depends on the exact values of image pixels.

The remainder of this paper is organized as follows. Section 2 presents RW with a negative correlation-based scheme. Section 3 describes SFW with permissible alterations, whereas Section 4 describes FW using a modified error diffusion scheme. Section 5 presents our experimental results. Finally, Section 6 states our conclusions.

2. Watermarking with a negative correlation-based scheme: RW

A fingerprinting application was tested using this method and our preliminary findings were reported in a previous study [**2**].

2.1. Watermark construction

In the proposed algorithm, to construct a watermark W, we compute two elements (the largest and smallest values) from the embedded area. In this study, we use the term "E_a" to refer to the embedded area after processing the image using randomization and the discrete cosine transform (DCT) (see Fig. 1).

Next, we generate a uniform distribution of random numbers from a specified interval $[max(E_a), min(E_a)]$, denoted by P_{mark}. We then obtain the embedded watermark using Eq. 1:

$$W = E_a - P_{mark}, \tag{1}$$

where P_{mark} is a sequence of uniformly distributed pseudorandom numbers.

2.2. Watermark embedding

The embedding of the watermark w_i into the host signal x_i is usually multiplicative or additive. In general, the multiplicative rule $y = x_i(1 + \alpha_i \cdot w_i)$ is used to embed the watermark. In the frequency domain, to improve watermark detectability, Barni *et al.* proposed the multiplicative rule $y_i = x_i + \alpha_i \cdot |x_i| \cdot w_i$ [**8**].

The watermark values have a negative property, thus we consider an additive

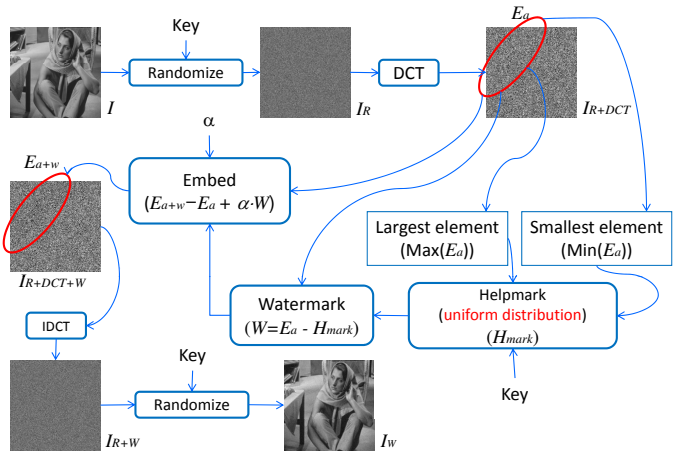

Figure 1. Embedding process including watermark construction.

embedding rule, as shown in Eq. 2. In the proposed method, a unique watermark is inserted into the DCT domain with strength α. We then define E_{a+w} as the watermarked area constructed using Eq. 2:

$$E_{a+w} = E_a + \alpha \cdot W. \tag{2}$$

Figure 1 shows the embedding process, including watermark construction.

2.3. Watermark detection

The goal of this research is to detect watermarks that are subjected to various attacks in an efficient manner. The proposed detection process is based on linear correlation and it is illustrated in Fig. 2. We obtain the pseudorandom numbers related to the watermark embedded during the embedding step using Eq. 3. Note that the numbers used to detect the watermark are not the same as those employed in its construction:

$$corr = \frac{1}{M} \sum E_{a+w} \cdot P'_{mark} \tag{3}$$

where M is the size of the embedded area, and P'_{mark} is a set of uniformly distributed pseudorandom numbers. However, P'_{mark} is a scaled and possibly shifted version of P_{mark}, although it is obtained using the same key.

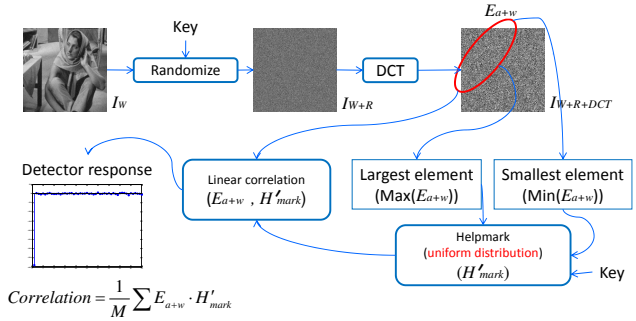

Figure 2. Watermark detection process.

3. Watermarking with permissible alterations: SFW

We propose an image block feature construction and embedding method that uses the just-noticeable differences (JND) visual model and a corresponding detection method using wavelet transforms. JND [**9**] has been used previously in image-adaptive watermarking, but our method is different because the watermark detection method uses a wavelet property and the embedding strength is image adaptive parameter based on JND and an image block feature. Research into SFW often fails to address the importance of the watermark strength. A preliminary test of this method was reported in our previous study [**10**].

3.1. Watermark and image feature

First, we construct a watermark, W, from a pseudo-random floating point sequence that comprises an array of M-by-N numbers (the same size as the original image), which have a Gaussian distribution with an average of 0 and variance of 1.

Second, we generate an image block feature, B_k, where $1 \leq k \leq t$ and t is the total number of 8×8 blocks ($t = M/8 \times N/8$, where the original image is $M \times N$), $B = b_{ij}$, $b_{ij} = 0$ or 1, $i = 1, 2, ..., 8$ and $j = 1, 2, ..., 8$. To construct an image block feature, we compare the summation of two subsets for each 8×8 block, which are denoted as "subset 1" and "subset 2" in Fig. 3. If the summation of the pixel

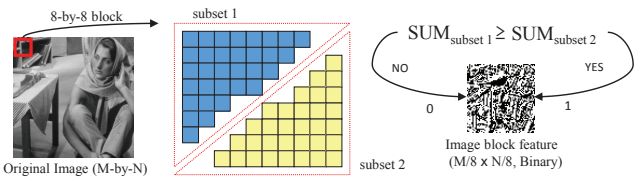

Figure 3. Generating an image block feature.

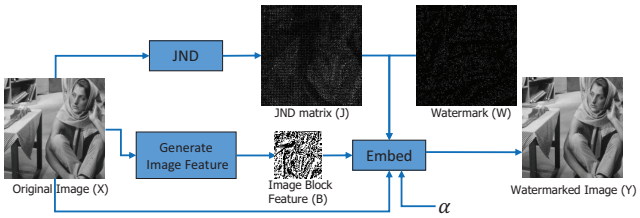

Figure 4. Embedding diagram for semi-fragile watermarking.

values in subset 1 is greater than that in subset 2, then the "image block feature" $B_k = 1$, otherwise $B_k = 0$. A block diagram of the embedding scheme is illustrated in Fig. 4.

3.2. Approach for image adaptive watermarking

In the proposed system, we embed a watermark using a threshold unit, which is often called "just-noticeable differences" or JND [**9**]. Originally, the JND scheme was applied to image compression, but it was recently introduced as an adaptive watermarking technique [**11**]. Let J_{ijk} be a threshold (JND) and the values used by the method are described as Eq. 4:

$$J_{ijk} = e_{ijk}/m_{ijk}, \tag{4}$$

where e_{ijk} is the (i,j)-th quantization error in the k-th block given by Eq. 5 and m_{ijk} is the (i,j)-th contrast masking in the k-th block given by Eq. 6.

The DCT transform is applied to each 8×8 image block and c_{ijk} is the (i,j)-th frequency component of the k-th block. Each block is then quantized by dividing it, coefficient by coefficient, using the quantization matrix q_{ij}. The quantization error e_{ijk} in the DCT domain is then:

$$e_{ijk} = c_{ijk} - (\lfloor c_{ijk}/q_{ij} + 0.5 \rfloor) q_{ij}, \tag{5}$$

$$m_{ijk} = max[t_{ijk}, |c_{ijk}|^{w_{ij}} t_{ijk}^{1-w_{ij}}], \tag{6}$$

where t_{ijk} is the (i,j)-th luminance masking in the k-th block given by Eq. 7 and w_{ij} is a number between zero and one, which we can assume has a different value for each DCT basis function. A typical empirically derived value for w_{ij} is 0.7. Let t_{ijk} be given by:

$$t_{ijk} = t_{ij}(c_{00k}/\hat{c}_{00})^{a_T} \tag{7}$$

where t_{ij} is the (i,j)-th frequency sensitivity is given by $q_{ij}/2$, c_{00k} is the DC coefficient of the DCT for block k, $\hat{c}_{00}$ is the DC coefficient that corresponds to the

mean luminance of the display, and a_T is a parameter that controls the degree of luminance sensitivity. In a previous study [**12**], it was suggested that a_T is set to 0.649.

3.3. Watermark embedding

In the proposed method, an 8-by-8 watermark is inserted in the spatial domain based on the adaptive strength α. This is particularly important in the case of SFW. We define X_k as the original image block of 8-by-8 and we define Y_k as the watermarked image block (8-by-8) given by Eq. 8. We define W_k as the watermark block and define α as the embedding strength:

$$Y_k = X_k + \alpha \cdot W_k, \quad \alpha = \begin{cases} 5, & \text{if } |DCT(X_k)| \geq |J_k| \\ & \quad \text{and } B_k = 1, \\ 1, & \text{otherwise.} \end{cases} \tag{8}$$

Figure 5 shows the original image X, watermarked image Y, and embedded information as a watermark W composed of W_k. It can be seen that the watermark is distributed in all the areas of the image.

Figure 5. Original and watermarked images.

3.4. Watermark detection

The aim of this method is to detect the presence or absence of a watermark on a block-by-block basis to evaluate the effectiveness of the proposed algorithms. We use the wavelet transform and linear correlation to detect the presence or absence of a watermark on a block-by-block basis. Each block of the attacked image is divided into low and high frequency coefficients by the discrete wavelet transform (DWT). The low frequency portion is set to zero. This signal is processed by inverse DWT (A_k).

Let I_k be an indicator variable, which is 1 if the watermark is not detected in block k and 0 if it is detected. If the correlation value ($corr_k$) is less than some threshold T, then it is a non-detected block ($I_k = 1$)(see Eq. 9). The detector counts the number of non-detected blocks in the image and this number (S) is used to estimate whether the image modification was malicious or non-malicious

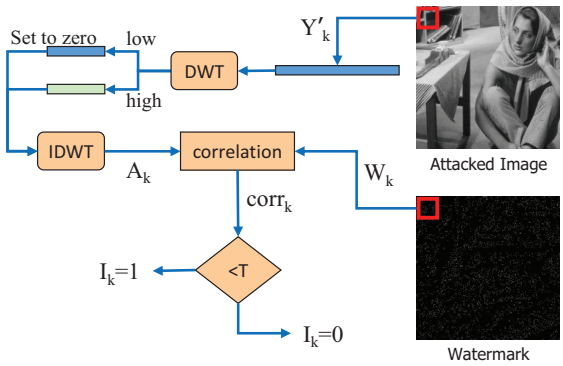

Figure 6. Detection diagram for semi-fragile watermarking.

(see Eq. 10).

$$corr_k = \frac{1}{64} \sum A_k \cdot W_k \tag{9}$$

$$S = \sum_{k=1}^{t} I_k \tag{10}$$

where t is the total number of 8×8 blocks in the watermarked image. Figure 6 shows the block diagram of the watermark detection system.

4. Watermarking using a modified error diffusion scheme: FW

A preliminary test of this method was reported in our previous study [**1**].

4.1. Watermark embedding

In the FW step, we apply a modified error diffusion scheme to bitplanes from the first bit plane (MSB) to the sixth bit plane (BP_{1-6}), or to the seventh bit plane (BP_{1-7}), to produce an image-dependent dithered image. The bitplanes (BP_{1-6} or BP_{1-7}) of input image (X_{RW+SFW}) are denoted by $b(i, j)$, where i and j denote the spatial position of the pixel.

The intensity value for every pixel in image $b(i, j)$ is converted into 0 or 255 and compared to a threshold value T(128 in experiments), which is given by Eq. 11. Let $c(i, j)$ be the converted pixel intensity value of point (i, j), then

$$c(i,j) = \begin{cases} 255, & \text{if } b(i,j) \geq T \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

The value $e(i,j)$ represents the difference between $b(i,j)$ and $c(i,j)$, which is given by Eq. 12.

$$e(i,j) = b(i,j) - c(i,j). \tag{12}$$

The value $e(i,j)$ is diffused to four different pixels after multiplication by the weighting factors (7/16, 3/16, 5/16, 1/16) and strength β (value 0.3) to obtain the error diffusion value $ed(i,j)$, which is given by Eq. 13:

$$\begin{aligned} ed(i+1,j) &= e(i,j) * 7/16 * \beta \\ ed(i-1,j+1) &= e(i,j) * 3/16 * \beta \\ ed(i,j+1) &= e(i,j) * 5/16 * \beta \\ ed(i+1,j+1) &= e(i,j) * 1/16 * \beta \end{aligned} \tag{13}$$

where $ed(i+1,j)$ is the diffusion to the next pixel, and $ed(i-1,j+1)$, $ed(i,j+1)$, and $ed(i+1,j+1)$ are the diffusions to the next lines.

The weight factors with strength β are the modified Floyd and Steinberg error diffusion coefficients [**5**], which are used to control the diffusion characteristics of the errors that appear in the detection output, and they are employed to make authentication decisions ($LP^{OR}_{1,2}$, see Fig. 9). An example of the error diffusion scheme is shown in Fig. 7. Please note that we use the value '255' to indicate a

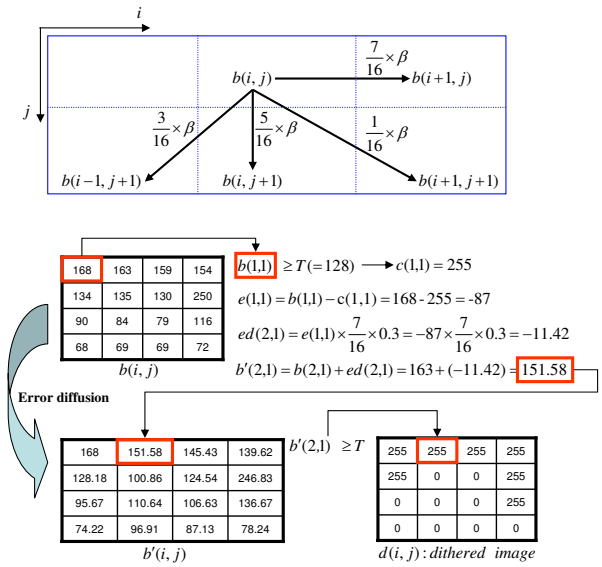

Figure 7. Example showing the production of a dithered image using an error diffusion scheme.

binary '1' in "$d(i, j)$ dithered image."

Let G be a logo with a size of $M \times N$. We tile G into a new array GI so it is the same size as the original image ($I \times J$). In addition, we apply the XOR operation to the dithered image ($d(i, j)$) and GI. Finally, the fragile watermarked image ($X_{RW+SFW+FW}$) is obtained by adding the result of the XOR operation to the least significant bit (LSB) plane of the semifragile watermarked images (X_{RW+SFW}).

4.2. Watermark detection

In the fragile watermark detection step, we can verify whether the watermarked image has been tampered with. The watermarked bitplanes (BP_{1-6} or BP_{1-7}) are processed by the error diffusion algorithm (see Fig. 7). We apply the XOR operation to the dithered image and the LSB bitplane (or the second LSB bitplane). The resulting image is processed by the XOR operation with a tiled pattern logo, which is produced from the extracted logo image or original logo image. It is easier to recognize the tampered area using an OR operation between two tiled pattern logos.

5. Experimental results

Our embedding system is summarized in Fig. 8. The watermark detection system has three steps and each can be performed independently, as shown in Fig. 9. We tested the proposed algorithm using the 512×512 Barbara grayscale image.

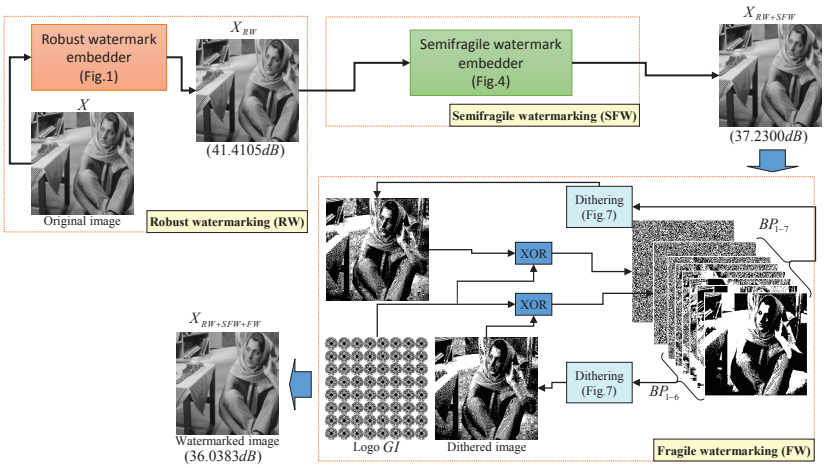

Figure 8. Multipurpose watermark embedding system.

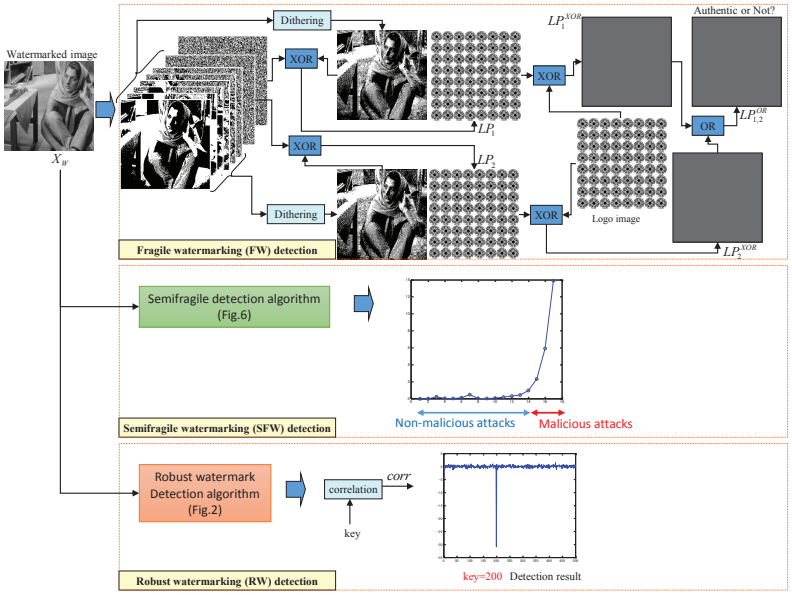

Figure 9. Multipurpose watermark detection system.

5.1. FW

The results of the tampering detection experiment obtained using FW are shown in Fig. 10 and 11. We added a spot and tampered with one region of the watermarked image. The tampered region was identified by the XOR operation and OR operation, as described in Section 4. Please refer to Fig. 9 for details of the terms LP_1, LP_2, LP_1^{XOR}, LP_2^{XOR}, and $LP_{1,2}^{OR}$.

5.2. SFW

The following manipulations were established as examples of non-malicious modifications [**13**]:

– Median filtering with a support of 3 × 3,
– Salt-and-pepper noise, up to one percent,
– Histogram equalization (uniform distribution),
– Sharpening (unsharp masking filter with coefficients [-1 -1 -1; -1 9 -1; -1 -1 -1]),
– Low-pass filtering within a support of 3 × 3 (equal weight coefficients equal to 1/9),
– Additive Gaussian noise down to a signal-to-noise ratio of 35 dB,
– Mild compression, e.g., up to 50% JPEG.

Based on experiments, we found that T=0.01 gave the best results when classifying JPEG compression less than 40% as malicious and greater than 50% as

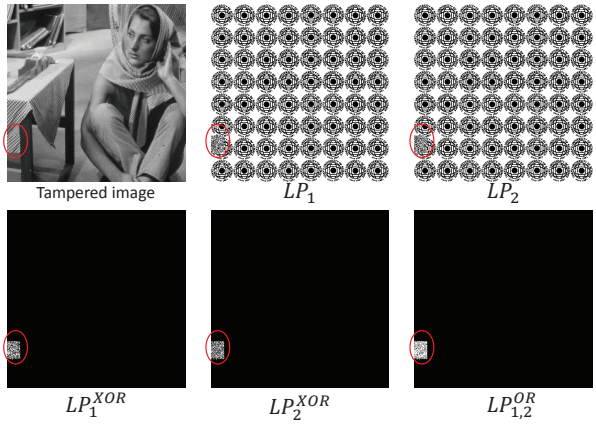

Figure 10. Results obtained with fragile watermarking after a tampering attack.

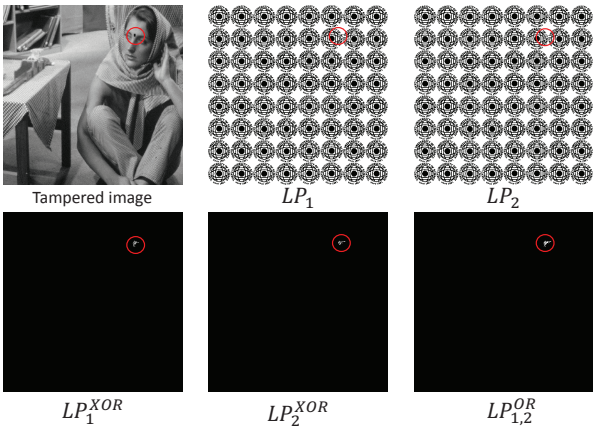

Figure 11. Results obtained with fragile watermarking after a tampering attack.

non-malicious. In Fig. 6, we have shown a block diagram that illustrates watermark detection and the counting of a non-detected block. In our experiment, if $S \geq 80$ in the attacked image, it was regarded as a malicious attack (see Table 1). White squares are used to indicate blocks where the watermark was not detected in Fig. 12.

5.3. RW

Figure 13 shows the results of negative correlation-based detection after basic image processing and JPEG compression.

Figure 12. Results obtained with semifragile watermarking. The white squares indicate non-detected blocks.

Table 1. Number of white blocks (non-detected watermark blocks) using Barbara as the test image

Attacks	Intention*	Number of non-detected blocks	Detected Intention*
Median filtering	N	9	N
Salt & pepper noise	N	1	N
Histogram equalization	N	1	N
Sharpening	N	4	N
Low-pass filtering	N	20	N
Gaussian noise	N	1	N
JPEG(100)	N	1	N
JPEG(90)	N	3	N
JPEG(80)	N	7	N
JPEG(70)	N	13	N
JPEG(60)	N	18	N
JPEG(50)	N	39	N
JPEG(40)	M	94	M
JPEG(30)	M	242	M
JPEG(20)	M	569	M

*Intention of attack: malicious attack (M) and non-malicious attack (N)

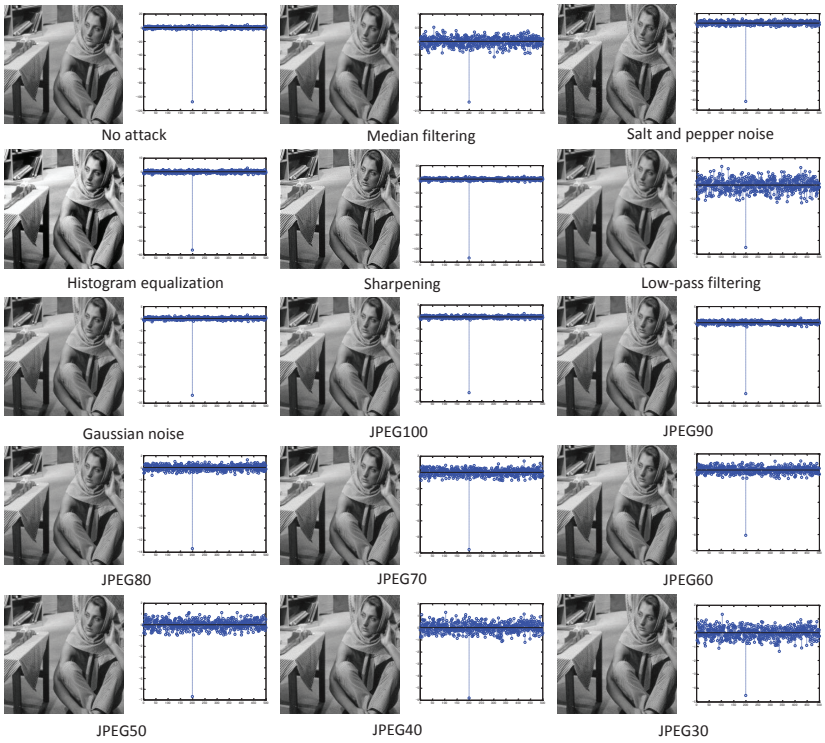

Figure 13. Results of robust watermarking after basic image processing and JPEG compression. The detector responded to 500 randomly generated watermarks and the correct watermark with a negative orientation is shown in the graph. The true key is found at number "200."

6. Conclusion

In this study, we proposed a framework that incorporates three watermarks with robust, semifragile, and fragile characteristics. A fragile watermarking algorithm based on the error diffusion scheme was also proposed. Using SFW, we can determine the nature of attacks by counting the number of non-detected blocks. Finally, we can maintain copyright ownership using a robust watermark even though the fragile and semifragile watermarks become invalid.

References

[1] H. Kang, Y. Park, B. Kurkoski, K. Yamaguchi, and K. Kobayashi, "Multiple Watermarking with Semifragile Property," Hawaii and SITA Joint Conference on Information Theory, HISC, pp. 142–147, 2007.

[2] H. Kang and K. Iwamura, "Collusion-Resistant Watermarking Using Modified Barni Method," Ubiquitous Information Technologies and Applications, Lecture Notes in Electrical Engineering 280, Springer-Verlag, 2014.

[3] M.M. Yeung and F. Mintzer, "An Invisible Watermarking Technique for Image Verification," Proc. of IEEE conf. Image Processing, Vol. 2, pp. 680–683, 1997.

[4] P.W. Wong and N. Memon, "Secret and Public Key Image Watermarking Schemes for Image Authentication and Ownership Verification," IEEE Trans. Image Processing, Vol. 10, No. 10, pp. 1593–1601, 2001.

[5] R. Floyd and L. Steinberg, "An Adaptive Algorithm for Spatial Grayscale," Proc. SID, Vol. 17, No. 2, pp. 75–77, 1976.

[6] M. Bellare, R. Canetti, and H. Krawczyk, "Keyed Hash Function for Message Authentication," Advances in Cryptology-Crypto 96 Proceedings, Lecture Notes in Computer Science Vol. 1109, Springer-Verlag, pp. 1–15, 1996.

[7] R.L. Rivest, A. Shamir, and L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," Communications of the ACM, Vol. 21, pp. 120–126, 1978.

[8] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, "Robust Watermarking of Still Images for Copyright Protection," In Proc. 13th Inter. Conf. Digital Signal Processing, vol. 2, pp. 499–502, 1997.

[9] A.B. Watson, "DCT Quantization Matrices Visually Optimized for Individual Images," Proc. of SPIE Int. Conf. Human Vision, Visual Processing, and Digital Display IV, vol. 1913, pp. 202–216, 1993.

[10] H. Kang, B. Kurkoski, K. Yamaguchi, and K. Kobayashi, "Detecting Malicious Attacks Using Semi-Fragile Watermark Based on Visual Model," Mexican Conference on Informatics Security (MCIS2006), Oaxaca, Mexico, November 2006.

[11] C.I. Podilchuk and W. Zeng, "Image-adaptive Watermarking Using Visual Models," IEEE J. Select. Areas Commun., vol. 16, no. 4, pp. 525–539, May 1998.

[12] H.A. Peterson, A.J. Ahumada, Jr., and A.B. Watson, "Improved detection model for DCT coefficient quantization," Proc. of SPIE Int. Conf. Human Vision, Visual Processing, and Digital Display IV, vol. 1913, pp. 191–201, 1993.

[13] O. Ekici, B. Sankur, B. Coskun, U. Naci, and M. Akcay, "Comparative evaluation of semifragile watermarking algorithms," Journal of Electronic Imaging, vol. 13, pp. 209–219, Jan. 2004.

Hyunho Kang

Department of Electrical Engineering, Tokyo University of Science

6-3-1 Niijuku, Katsushika-ku, Tokyo, 125-8585, Japan

kang@ee.kagu.tus.ac.jp

Keiichi Iwamura

Department of Electrical Engineering, Tokyo University of Science

6-3-1 Niijuku, Katsushika-ku, Tokyo, 125-8585, Japan

iwamura@ee.kagu.tus.ac.jp