Received: May 25, 2020. Revised: June 16, 2020.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Blind Robust and Self-Embedding Fragile Image Watermarking for Image Authentication and Copyright Protection with Recovery Capability

Lusia Rakhmawati[1, 2*] **Suwadi Suwadi[1]** **Wirawan Wirawan[1]**

[1]*Department of Electrical Engineering, Faculty of Intelligent Electrical and Informatics Technology,*
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2] *Department of Electrical Engineering, Faculty of Engineering,*
Universitas Negeri Surabaya, Surabaya, Indonesia
* Corresponding author's Email: lusiarakhmawati.17071@mhs.its.ac.id

Abstract: This paper proposes a dual watermarking scheme that can integrate the functions of authentication, copyright protection, and image recovery in the same cover image. The robust watermarking utilizes a single watermark using the discrete cosine transform (DCT) domain for copyright protection, while the fragile scheme utilizes two self-embedding watermarks in a spatial domain for authenticating and restoring digital image content. The mechanism of the two schemes is carried out sequentially and uses a block-based approach. A new blind robust watermarking is proposed using a quantitative property of intermediate frequency coefficient for embedding and adaptive embedding strength selection to balance transparency and robustness. Fragile watermarks are inserted into a robust watermarked image based on the improved replacement approach of the least significant bits. Experimental results show that the proposed method can withstand various processing attacks; enhance the accuracy of deceptive localization with good visual quality of image recovery.

Keywords: Watermarking, Authentication, Copyright protection, Tamper detection and recovery.

1. Introduction

The protection of multimedia content is becoming important for its dissemination on the Internet. In addition to protecting the public interest from the actions of others who misuse electronic information and/or electronic documents, this protection may provide evidence of ownership of multimedia content. One effective way of protecting multimedia content is through watermarking techniques [1, 2]. This technique evolved not only to insert ownership data but also to insert additional data taken from the image features themselves, which are representations of original multimedia content better known as a self-embedding watermarking scheme [3 -4]. The embedded data can then be extracted with the decoder to demonstrate the validity and can also be used to identify the degradation area and to enhance the multimedia content.

There are a number of self-embedding watermarking techniques that have been proposed [5]. Either uses fragile watermarking or robust watermarking techniques. The fragile watermarking technique proposed by [6] consider the watermarked image's visual qualities on the basis of the data interleaving mechanism, where this scheme uses flexible numbers from the most significant bit (MSB) layer to produce reference bits that are inserted for recovery of content, and also uses the number of LSB layer variables to accommodate watermark bits. Another fragile method is proposed by [7] which uses the average value of each overlapping block through interleaving information as a reference bit, and in the same way, uses the least significant bit (LSB) technique for the insertion process. Unlike the two methods, the one developed by [8] uses two watermarks, it includes the average block information value itself, while the second contains information about block authentication. This work's main contribution is the use of Arnold's transformation to hide information that can improve the efficiency against geometric attacks, ensuring that the neighbor bits of the watermark is embedded in the remote blocks. Overall, these three approaches are effective, both in the process of damage identification and in the repair of information content. Nonetheless, the watermark inserted will be impaired when talking about digital picture processing widely used in applications such as Joint Photographic Experts Group (JPEG) compression, filters, or contrast values.

Robust watermarking practices: discrete wavelet transform (DWT) [9, 10], DCT [1, 11, 12], DCT-DWT [13, 14], are preferred for common applications such as JPEG compression. The use of a discrete cosine domain followed by an embedding watermark is chosen in several methods [1, 11, 15]. In [1], a non-linear, chaotic map is used to divide the DCT integer coefficient into various not superposing blocks and a circulating matrix is generated in order to embed the watermark with a singular value decomposition. In addition, the method [11] uses a combined pattern by extracting certain pattern feature information efficiently and using varying levels based on blocks and being able to adaptively produce reference watermarks that are short enough to be embedded in one LSB. DCT transform is used by [16] which divides the image into small blocks and DCT is done on each block. The watermark bit is embedded either on the direct current (DC) or alternating current (AC) part of the DCT at a low frequency via Chinese Remainder Theorem (CRT) usage; the results achieved indicate strong resistance to JPEG compression.

Problems arising from some of the proposed methods can be outlined in two types, namely problems of coincidence, where the inserted reference data also experiences damage and excess watermark components, in the sense that the inserted watermark is not used. The MSB implemented in [17] is the most important bit. This method is, however, only successful if less than 6.6 percent of the area destroyed. In another method [18], a watermark derived from two other blocks is performed by each block of the host image; in other words, there are two copies of the watermark for each non-overlapping block hidden in two different blocks, which leads to a second chance of the blocks being restored in one copy being destroyed. By using this approach, the probability of simultaneously damaging coincidences does not increase the cost of the waste watermark.

Although many schemes have already been proposed, much progress is still required. An authentication check for integrity was introduced in [19]. It located the tampers, but the precision of the location was impaired. In [20] present an effective singular values (SV)-based semi-fragile watermarking scheme for image content authentication with tamper localization but it could not perform recovery of the altered regions. Schemes with dual functionalities are therefore now more preferred. Performance evaluation for tamper localization and recovery was proposed in [2]. This method proposes a solid, semi-blind watermarking method based on linear interpolation, which works for precise manipulation and high-quality image recovery within the space domain. Furthermore, existing research combines two widely used watermarking methods with the aim of only one or two of them. Therefore, there are certain contributions within the proposed scheme: (1) The detection and localization of tampered regions is based on a block-wise mechanism, which is using small block 2×2 pixels with additional two secure key bits in modified LSB method, thus the small pixels used to detect malfunctioning blocks will minimize false detection and the keys boost security; (2) By using the average intensity of the original image as watermark, the recovered image can be enhanced; (3) Adaptive embedding strength determination with new coefficient selection approach for watermark robust under compression attacks and the extraction of a robust watermark is blind; (4) a dual watermarking scheme provides simultaneously all the three functionalities of copyright protection, tamper detection with recovery capability. The remainder of the paper is structured accordingly. Section 2 corresponds with the relevant works. Section 3 contains information about the approach proposed. Section 4 offers experimental results. Section 5 ultimately sets out the conclusions.

2. Related works

In this section, we briefly describe the previous literature relating to fragile watermarking, robust watermarking, and dual watermarking. All literature references have been used extensively in several applications or cited by other studies. Fragile watermarking is mainly used to authenticate image integrity; such authentication must be very sensitive to changes in the host image signal, and as such is largely based on the spatial domain. The fragile watermarking state allows you to tamper it. One of the initial fragile image authentication watermarking schemes was proposed by [21]; in their scheme, a secret key is used to produce a binary value function. The result of watermarked image shows satisfactory. However, they focus primarily on detecting whether they have tampered or not using a 3 × 3 block neighborhood, which means that if the tampered area is smaller than 3 × 3 block, it recovered the whole pixels. Thus, with the use of small pixels, it can minimize false detection, and using the mean intensity of the original image, the recovered images can be improved.

In the spatial domain scheme [4], watermarks are inserted in the host image through a pixel value change. This has the advantage of low complexity and easy implementation, but not too strong to withstand affine transformations and some image processing attacks. However, the frequency domain scheme [5] usually makes it more difficult to see and provides resistance to too many common attacks. In the frequency domain watermarking scheme, the robustness and imperceptibility of the host picture are in the balance. Recently, robust watermarking performance has been further enhanced by several computational intelligence-based techniques, such as genetic algorithms (GAs) and differential evolution. Therefore, the adaptive frequency domain watermarking scheme is presented in this paper to balance the resilience and imperceptibility of the watermark image adaptively, while maintaining the advantage of satisfying computational speed with embedding strength determination.

The scheme presented by Ariatmanto and Ernawan [22] described the embedding intensity of each selected imaging block as having a different embedding force. Based upon the effects of the select middle frequency DCT coefficients on the average DCT coefficients of his block, a new embedding technique was developed. For each block selected, the proposed scheme generates specific built-in powers. Lai's method [12] propose an improved SVD watermarking technique, which takes account of HVS characteristics, where a block-based watermarking approach is used. A more complex block has been used to insert a watermark to achieve high imperceptibility for a block-based watermarking scheme. Therefore, the characteristics of an image should be considered to identify blocks to be implemented to achieve the required levels of performance for the proposed watermarking scheme.

The dual watermarking technique is proposed by embedding different watermarks for multipurpose image protection. Lin et al. [22] concentrated on presenting a dual watermarking scheme for intensive copyright protection. In their scheme, the visible watermark image is directly added to the spatial domain of the host image, and the watermark image that is not visible is embedded in the frequency domain by utilizing technical distortion that is only visible. Liu et.al scheme [23] presents a blind dual watermarking mechanism for color images. Existing methods emphasize one or two functions in one image, while the principal contribution of this paper is that our scheme can protect copyright and authenticate the image with also recovery capability simultaneously and that the extraction of watermarks from the protected image can take place blindly without the original host and watermarks.

3. The proposed method

The proposed scheme consists of three main phases: robust watermark embedding, fragile watermarks embedding, and watermarks extraction. Fig. 1 shows the watermark embedding process which includes dual watermarking scheme: robust watermarking and fragile watermarking scheme. In the following subsections, we explain each component in detail.

3.1 Robust watermark embedding

Robust watermark used as copyright protection must be resistant to various types of attacks, including filters, noise, geometric distortion, and JPEG compression. The scheme is agreed on in the DCT frequency domain, which is the same domain used by the JPEG compression algorithm.

Assume that an original gray scale cover image I_O has M_1 rows and M_2 columns where M_1 and M_2 are multiples of 8, and N represents the total number of pixels ($N = M_1 \times M_2$). Divide the original image into non-overlapping 8×8 pixel blocks, similar to that used by the JPEG algorithm [11], and denote the gray value of each pixel in I_O as $p_k(x, y) \in [0, 255]$, and p_k can be represented by 8 binary bits in Eq. (1) as $f_{k,t}(x, y)$ where $1 \leq k \leq N/64, 1 \leq x \leq 8,$ and $1 \leq y \leq 8,$

$$f_{k,t}(x,y) = \lfloor p_k(x,y)/2^t \rfloor \, mod \; 2, \quad t = 0, 1, 2, \ldots, \tag{1}$$

Thus for 8×8 pixel blocks, the forward DCT (F_k) is given by Eq. (2):

$$F_k(u,v) = \frac{1}{4} C(u)C(v) \sum_{x=1}^{8} \sum_{y=1}^{8} f_k(x,y) \times cos\left(\frac{(2x+1)\pi u}{16}\right) cos\left(\frac{(2y+1)\pi v}{16}\right) \tag{2}$$

where $C(u), C(v) = 1/\sqrt{2}$ for $u, v = 1$ and $C(u), C(v) = 1$ other $f_k(x, y)$ is the kth block image pixel value, and $F_k(u, v)$ is the transform coefficients.

An important parameter of watermarking is to determine the embedding position for the watermark in the host image. In [16] some of the changes introduced in the field of the DCT coefficient subsisting on JPEG compression under certain conditions. The energy of the image is concentrated in the low-frequency region and only a small fraction of the total energy is represented by high-frequency components. The watermark can cause undetectable distortions of the picture within the DCT coefficients. JPEG quantifier is used not to add the limit but instead to break the high-frequency domain corresponding coefficients. The JPEG quantifier cut off the coefficients corresponding to the high-frequency domain. Changes in low frequencies may, on the other hand, cause significant changes in the image. In the mid-frequency spectrum (F_M), as shown in Fig. 2a, the most appropriate frequencies can be used with a new coefficient selection strategy in Fig. 2.

As used in the work of Huang et al. [24] which choose coefficient +1 and -1 in each block as an embedding field. An illustration using the Lena image of 512 x 512, we can see the histogram for all the quantized dc and ac coefficients of the Lena image with quality factor 80 at figures 2a and 2b. Fig. 2(a) shows that the host image could be distorted considerably by invalid shifting. Meanwhile in Fig. 2(b) appears that the size of the image file may increase significantly if some zero coefficients are changed in the embedding process. Thus, we can calculate the number of coefficients "1" and "-1" as a capacity of embedding, coefficients "0" are remain unchanged, while others as a total distortion of image. Fig. 2(d) described of the ac position (1-63) in each block. As shown in Fig. 2(d), we can choose the coefficients included in the middle frequency that have a low distortion capacity. One middle-frequency coefficient is chosen to embed one bit of the watermark. Each watermark bit w is inserted according to the additive white Gaussian noise (AWGN) embedding as mentioned in the Eq. (3) for $w(x,y) = 0$ and Eq.(4) for $w(x,y) = 1$, the picture embedded message association blind recognition.

$$F_k(u,v) = F_k(u,v) + \alpha \cdot r_k(u,v) \tag{3}$$

$$F_k(u,v) = F_k(u,v) - \alpha \cdot r_k(u,v) \tag{4}$$

where $u, v \in H_k$, and $F_k(u,v) = F_k(u,v)$ for others. H_k represents the best coefficient location as a result of a coefficient selection strategy. The embedding strength coefficient is more than zero. The reference pattern r_k is an array of pixel intensities the same size as the cover image which contains a white Gaussian noise. We used the pseudo-random vector chosen in compliance with this experiment, $N(0,1)$.

Upon modifications to the DCT coefficient values, it is reconstructed based on the inverse transformation of DCT, using Eq. (5) the following:

$$f_k(x,y) = \frac{1}{4}\sum_{x=1}^{8}\sum_{y=1}^{8} C(u)C(v)\,F_k(u,v) \times \cos\left(\frac{(2x+1)\pi u}{16}\right)\cos\left(\frac{(2y+1)\pi v}{16}\right) \tag{5}$$

where $f_k(x,y)$ are the kth block pixel values of the robust watermark image I_R.

The inputs to the detector are the image and the watermark key where the output is the watermark's information. We use a linear correlation between the image and the reference pattern as shown in Eq. (6) for calculating the detection.

$$L(I_O, r_k) = \frac{I_O \circ r_k}{\|r_k\| \cdot \|r_k\|} \tag{6}$$

where ∘ denote the dot product, and || || is vector norm. Therefore, we can determine if a watermark seems to have a threshold, τ, so the binary one is embedded if $L(I_O, r_k) \geq \tau$, binary zero is embedded if $L(I_O, r_k) \geq -\tau$, and nothing are embedded if $|L(I_O, r_k)| < \tau$.

The dimension and energy of the image (or part of the image), in which the AWGN watermark is formed, determine the effectiveness of its embedding strength. Such coefficient properties determine the compression strength of the predicted compression. Nonetheless, the problem of robustness when embedding in part of image coefficients is a much more complex one than embedding in the whole image. The domain in which we insert watermarks is often not the same as the domain for lossy compression. Besides, it is generally not known when embedding which compression our image will be subjected to. General advice on embedding intensity cannot, therefore, be given in this case.

When discussing DCT embedding [25], we will use the term image sub-channel. Sub-channel is the vector that has coordinates in DCT-components block, with the same index in rows. Sub-channels are ordered by zigzag order. Sub-channel 1 thus consists of all DC block elements; sub-channel 22 consists of all block elements at position 22 (zigzag-order).

For our original image I_O and reference pattern, the linear correlation is Eq. (7).

$$L(I_O, r_k) = \frac{I_O \circ r_k}{\|r_k\| \cdot \|r_k\|} = \frac{\sum_{i=1}^{N} I(i) r_k(i)}{N} \tag{7}$$

As a result, the value $L(I_O, r_k)$ for fixed F_k and reference pattern r_k (addresses from N (0, 1)), has normal distribution values as well, $N(0, \sigma^2)$, we can measure standard deviations in Eq. (8) for the entire image, σ, is.

$$\sigma = \left(\frac{I_O(1)^2+I_O(2)^2+I_O(3)^2+\cdots+I_O(N)^2}{N}\right)^{1/2} \qquad (8)$$

We can see convergence in the first 32 sub-channels as incorporation into the sub-image, $s = \{s(1), s(2), s(3), \ldots, s(32)\}$ thus

$$\sigma_s \approx \left(\frac{I_O(1)^2+I_O(2)^2+I_O(3)^2+\cdots+I_O(N)^2}{N}\right)^{1/2} /2 \qquad (9)$$

In the case of successful embedding in the first 32 sub-channels, Eq. (9) can be approximate based on the reference in [25], a double force is required compared to embedding over the entire image, i.e. $\alpha_s = 2 \times \alpha$. To evaluate the real impact of watermark embedding on image fidelity, we will notice that $\|r_s\| = \|r_k\|/\sqrt{2}$. If we "standardize" the pattern of r_s, that is, lead it to the norm of reference pattern r, thus $\|r_s\| = \sqrt{2} \cdot \|r_s\|$. Therefore, efficient integration force in the first 32 sub-channels must be performed $\sqrt{2}$ times more strongly in order to be effective in comparison with the integration of the entire image as alternatively $\alpha_s = \sqrt{2} \times \alpha$ as the adaptive embedding strength of robust watermarking.

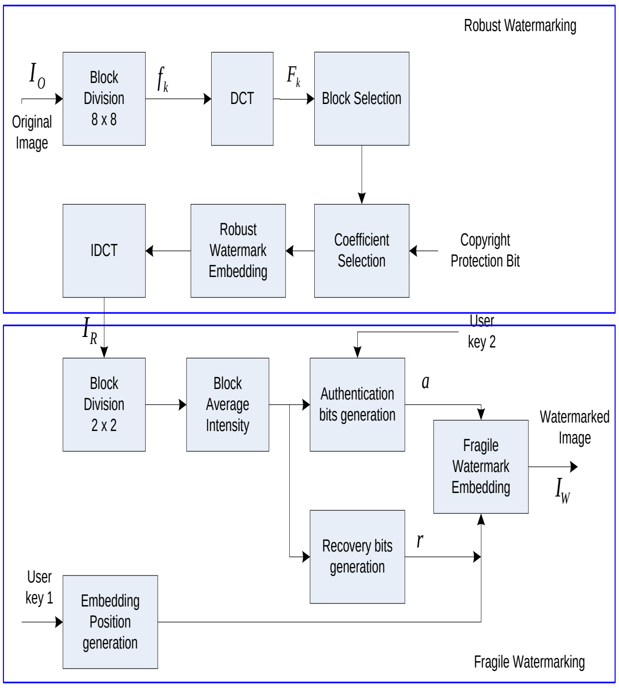
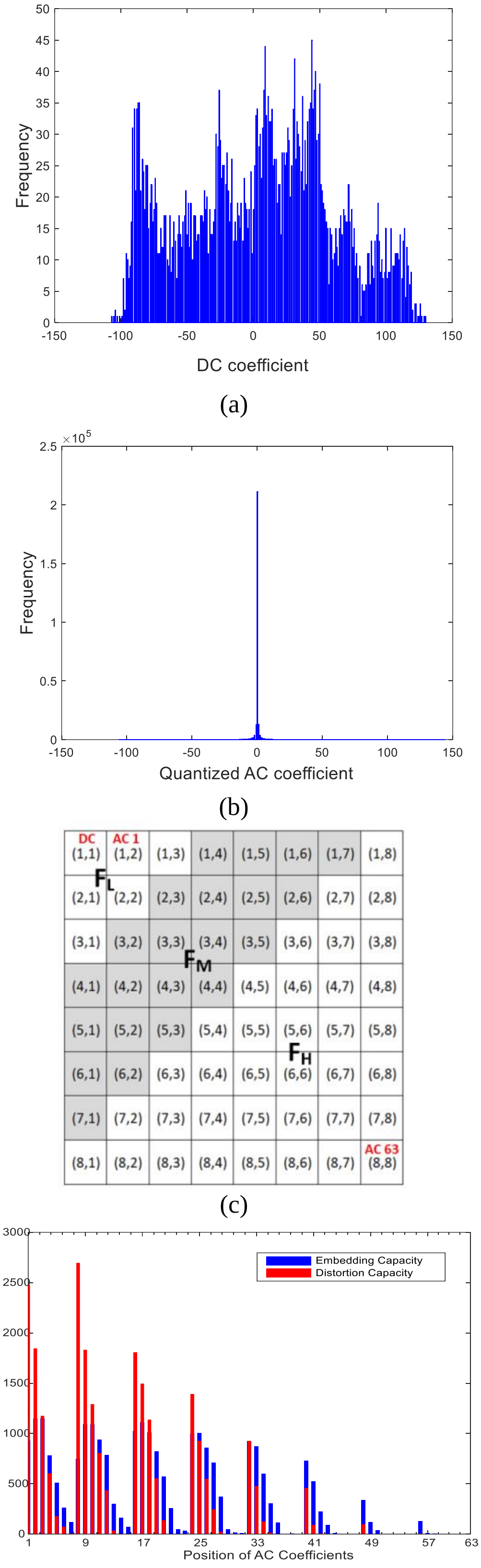


Figure. 1 The watermark embedding scheme



(a)

(b)

DC (1,1)	AC 1 (1,2)	(1,3)	(1,4)	(1,5)	(1,6)	(1,7)	(1,8)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)	(2,7)	(2,8)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)	(3,7)	(3,8)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)	(4,7)	(4,8)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)	(5,7)	(5,8)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)	(6,7)	(6,8)
(7,1)	(7,2)	(7,3)	(7,4)	(7,5)	(7,6)	(7,7)	(7,8)
(8,1)	(8,2)	(8,3)	(8,4)	(8,5)	(8,6)	(8,7)	AC 63 (8,8)

(c)

(d)

Figure. 2 The coefficient selection strategy of the Lena image with QF 80: (a) Histogram of dc coefficients, (b)Histogram of ac coefficients, (c) middle-frequency domain, and (d) The total capacity of embedding and distortion

3.2 Fragile watermark embedding

Denote the size of robust watermarked image I_R is the same as the cover image which has the total number of pixels, N. In the design of the proposed scheme, the detection of the tampered region is based on each non-overlapping image block-sized $b \times b$. Thus, for simplicity, M_1 and M_2 are both assumed as the multiples of b.

For each block, we adopt method [26] to generate two parts that will be respectively used for tampered area localization and content recovery. Assume that the number of MSB layers used for the generation of recovery bits is denoted as m. The l LSB layers of robust watermarked images are used to accommodate the watermark data. For each non-overlapping block, we allocate a authentication bits for tamper detection, thus the number of recovery bits for content recovery is $(l \times b^2 - a)$ bits. In our scheme, the authentication bits are built into the block rather than the mapped one, reducing the error of false detection an also used the small size of block image, 2×2, and using l LSB to embed the watermark. The details of watermark generation and embedding phases of our proposed mechanism are described in the following.

Step 1: We segment the robust watermarked image into $b \times b$ non-overlapping blocks. For each non-overlapping block, we generate four authentication bits $(a_{i,1}, a_{i,2}, a_{i,3}, a_{i,4})$ and eight recovery bits $(r_{i,1}, r_{i,2}, r_{i,3}, r_{i,4}, r_{i,5}, r_{i,6}, r_{i,7}, r_{i,8})$ using Eq. (10) – (13).

Step 2: We compute the mean value of the m MSB of each block, B_i, and converted to binary form to get eight recovery bits.

$$r_{i,j} = \lfloor B_i / 2^{j-1} \rfloor \, mod \, 2, \; j = 1,2, \dots 8 \tag{10}$$

Step 3: We generate $a_{i,1}, a_{i,2}$ using Eq. 11. For m MSB we have $m \times b^2$ sequence, where C_i is randomly key generated with the sized of $2 \times m \times b^2$ and $(A_i)_2$ is a binary form of every m MSB in each block.

$$(a_{i1}, a_{i2}) = ((A_i)_2 \times C_i) \, mod \, 2 \tag{11}$$

Step 4: We compute $a_{i,3}, a_{i,4}$ using Eq. 12 and 13 with ex-or operation ($\oplus$) of every m MSB in each block.

$$a_{i,3} = A_{i,8} \oplus A_{i,7} \oplus A_{i,6} \oplus A_{i,5} \oplus A_{i,4} \tag{12}$$

$$a_{i,4} = \begin{cases} 1, if \;\; a_{i,3} = 0 \\ 0, if \;\; a_{i,3} = 1 \end{cases} \tag{13}$$

Step 5: We insert watermark bits using LSB technique. Authentication bits embedded the l LSB of current block, whereas recovery bits embedded in corresponding block which was obtained randomly using 1-D linear transformation [22].

3.3 Watermark extraction

In the proposed extraction procedure, the robust watermark and the fragile watermark can be extracted separately for copyright protection and image authentication with recovery capability, respectively.

The extraction process is a step by step procedure to extract the binary watermark as copyright protection from the received image as follows. Step 1: The received image is segmented into 8 × 8 non-overlapping blocks. Step 2: Extracted each block and measured using a two-dimensional DCT. Step 3: Selected DCT coefficients were crossed into a vector through a zig-zag scan. In the middle frequency, we pick a DCT coefficient the same as in the encoder. Step 4: The watermark bits were extracted by following set rules: $w'_k(u, v) \leq 0$ if we embed binary zero and $w'_k(u, v) > 0$ if we embed binary one. Step 5: Generate the recovery of watermarks to protect copyright.

Concerning image authentication, after the watermarked image I_W is sent, the receiver can detect any changes caused by the public channel using detection bit. For each $b \times b$ block in the suspicious watermarked image $I_W{}'$, with the same secret key on the encoder, extracted watermarks from its l LSB will be segmented into two parts, i.e., recovery bits $[r_{i,j}, \; j = 1 - 8]$ and an authentication bits vector $[\; a_{i,1}{}', a_{i,2}{}', \; a_{i,3}{}', \; a_{i,4}{}']$. Then, compared with vector $\left[a_{i,1}, a_{i,2}, a_{i,3}, a_{i,4}\right]$. The block will appear as a true block if the comparison outcomes are of the same magnitude, otherwise, it will be classified as an untrue block.

The authentic or inauthentic block can be recognized after the detection process. Only an inauthentic block is restored, while genuine blocks are retained. Its block is used to find the information about retrieval for the invalid blocks. To pad any $b \times b$ inauthentic block, LSB 1 is then used, LSB 2 of the corresponding row. If a few pixels are not recovered, a nonlinear median filter is used to interpolate the remaining pixels to enhance the results of the recovered watermark bits image.

4. Experimental results

The proposed dual watermarking mechanism is explained in this section. The experiment was carried out by using a gray image with five commonly used images, namely, "Lena," "Airplane," "Baboon," "Peppers," and "Lake." All images have the same size, that is, 512 × 512. The embedded fragile watermark is a random binary bits stream that results from the authentication bit calculation described earlier. For robust watermark, we used 32 x 32 binary logo images.

4.1 Imperceptibility results

This experiment used two metrics of image quality assessment (IQA) [23], peak signal-to-noise ratio (PSNR) as described in Eq. (15), where mean square error (MSE) calculated from Eq. (14) and structural similarity (SSIM) in Eq. (16). PSNR is a conventional IQA metric that operates directly at the image-based stage. The famous SSIM brought IQA to the structural stage, based on the hypothesis that the perceptual system was highly adapted for bringing structural information from a visual perspective, from a conventional pixel-based stage. Note, however, that SSIM values are always in the range 0-1. Two IQA metrics are demonstrated to be appropriate and widely used for watermarked images for a given test image.

$$MSE(I_o, I_w) = \frac{1}{M_1 \times M_2} \sum_{x=1}^{M_1} \sum_{y=1}^{M_2} \left(I_O(x,y) - I_W(x,y)\right)^2 \quad (14)$$

$$PSNR(I_O, I_W) = 10 \times log_{10}(MAX_I/MSE) \quad (15)$$

Here, I_O represent the watermarked image, and I_W is recovered imag, and MAX_I is the image’s maximum possible pixel value.

$$SSIM(I_O, I_W) = \frac{\left(2\mu_{I_O}\mu_{I_W} + c_1\right)(2v + c_2)}{\left(\mu_{I_O}^2 + \mu_{I_W}^2 + c_1\right)\left(\sigma_{I_O}^2 + \sigma_{I_W}^2 + c_2\right)} \quad \begin{cases} c_1 = (k_1 L)^2 \; k_1 = 0.01 \\ c_2 = (k_2 L)^2 \; k_2 = 0.03 \end{cases} \quad (16)$$

Table 1 summarizes the two IQA average values for the five test images after the watermark is embedded, where the "Robust watermarked image" column presents the image results after robust watermarking only, the "Fragile watermarked image" column is the IQA value of the image after the fragile watermarking process, and the "Robust-fragile watermarked images" represent the results of robust and fragile watermarking processes. The average PSNR values of robust watermarked, fragile watermarked, and dual watermarked images are 41.83, 37.27, and 35.69 dB, respectively. This is under the calculations performed by He [21], where we can assume that a uniform distribution is the original data in the LSB field. The integrated watermark bit emits the average energy distortion, α, as

$$E_d = \frac{1}{2^{2\alpha}} \sum_{x=1}^{2^\alpha} \sum_{y=1}^{2^\alpha} (x - y)^2 \quad (17)$$

therefore, from Eq. (15) and Eq. (17), the approximate PSNR value of the fragile watermarked image with relation to the original one is

$$PSNR \approx 10 \; log_{10}\left(\frac{255^2}{E_D}\right) \quad (18)$$

Table 2 indicates the fragile watermarking technique produces a higher PSNR value than the robust watermarking technique because the insertion procedure through LSB has an insignificant visual impact. This occurs because the replacement operation, in this case, uses 3 LSB which only slightly changes the pixel value, which is a maximum of seven pixels. Compared to the method [23], the final value of the combination of the robust-fragile watermark method in which on average is almost the same as the fragile watermarking method. In the case of SSIM, the average value is around 0.94, which is quite satisfying. Besides, there are very few SSIM variants for different watermark images.

The system has been tested under various forms of attack. Standard Normalized Cross-correlation (NC) to measure the watermark solidity extraction. The NC is defined by:

$$NC = \frac{\sum_{x=1}^{M_1} \sum_{y=1}^{M_2} w(x,y).w'(x,y)}{\sqrt{\sum_{x=1}^{M_1} \sum_{y=1}^{M_2} w(x,y)^2 . \sum_{x=1}^{M_1} \sum_{y=1}^{M_2} w^1(x,y)^2}} \quad (19)$$

where $w'(x,y)$ is the extracted watermark and $w(x,y)$ is the original watermark. M_1 and M_2 denote the row and column sizes of the watermarked image.

4.2 Fragile watermarks extraction

In particular, for authenticating images, location of distractions, and tamper recovery we have developed our proposed fragile watermark. The self-recovery watermarking systems allow the detection or substitution of a watermarked image. The main distinction lies in the precision of the localization of

Table 1. PSNR and SSIM for five test images

Image	Robust Watermarked		Fragile Watermarked		Robust-Fragile Watermarked	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM
Lena	41.659	0.962	37.615	0.930	36.134	0.895
Airplane	44.490	0.976	37.033	0.924	36.160	0.901
Baboon	41.659	0.996	37.186	0.974	35.545	0.970
Peppers	41.252	0.967	37.110	0.931	35.347	0.900
Sailboat	40.086	0.968	37.425	0.948	35.249	0.917
Average	41.829	0.9737	37.274	0.942	35.687	0.916

Table 2. The Estimated PSNR (dB) of The Watermarked Image in Comparison to the initial image under various α

Capacity	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$	$\alpha = 4$	$\alpha = 5$
$E_W(\alpha)$	0.5	2.5	10.5	42.5	170.5
$PSNR_W(\alpha)$	51.1	44.15	37.92	31.85	25.81

distorted images and their consistency. The quality of a recovered image depends heavily on the amount of tampered regions, as indicated in [19]. The image content complexity and precise positioning also affect the quality of the image that is being recovered. Fig. 3 shows the attacked images, including the collage attack, addition object attack, substitution attack, and the cropping attack, as well as their corresponding authentication performance. Figure 3 (column a) shows an example of various attacks. Fig. 3(a1) illustrates a collage attack, which changes the lips of the "Lena" image by superimposing the lips of a "blonde woman" image.

Fig.3 (a2) shows an example of general tampering which adds an object as a visible watermark to the watermarked image. Fig. 3 (a3) shows an example of a substitution attack which replacing some pixels of the image with other pixels in the same watermarked image. Fig. 3 (a4) shows an example of a cropping attack which crop the center of the watermarked image with a white pixel. Fig.3 (column b) describes the extracted watermark for copyright protection which will be discussed further in section 4.3; the NC values are 1, 1, 0.9902, and 1 respectively. Fig. 3 (column c) shows the corresponding tamper localizations which is shown in yellow. Fig. 3 (column d) shows the watermarks extraction for image recovery and Fig. 3 (column e) shows the corresponding recovered images with PSNR= 47.25 dB, 28.61 dB, 41.11 dB, 30.47 dB and SSIM=0.9978, 0.9537, 0.9908, 0.9152 respectively.

Taking into account the collage effect, the collaged image was created by copying and pasting some areas of "Barbara" onto the image of "Lena" and its relative spatial places were retained. As the quantitative measurements [2] and the tamper detection efficiency of collage manipulation are shown, the probabilities of false rejection (PFR), the probabilities of false acceptance (PFA), and the probabilities of false detection (PFD) are used and the results are shown in Fig. 4. PFR and PFD all appear to be lower than 0.05 when the tamper ratio is less than 70%, however, the PFA results show the lower tampered region, 45%. However, our method can detect tampered block with more than 99% probability under collage attacks, a PFA shows less than 0.9 %.

Fig. 5 shows that the proposed PSNRs at various interference levels are significantly greater than those of other schemes, from 5 to 75%. The PSNR values of the proposed scheme are larger than 30 dB as long as the collaged region is not greater than 40% of the host image. In these two schemes, the quality of the images recovered is also very poor, as shown by the low PSNRs. This shows that He's scheme and the scheme of Lin are unable to withstand the collage assault. As describes in Fig. 5, our PSNRs are higher than 25 dB from 5 to 65% of the host image at different tampered areas.

4.3 Robust watermark extraction

Robustness is a significant concern for copyright protection mechanisms. In this section, we divided the experiment into two major parts, namely geometrical attacks and image processing attacks that were performed on the test image "Lena" to demonstrate the robustness of our watermarking scheme as shown in Fig. 6. The results of several geometrical attacks are shown at Fig. 7, which include cropping, rotate, and translation attacks. Meanwhile, signal processing attacks are shown in Fig. 8 and Fig. 9. The signal processing attacks that were performed in our experiments were Gaussian noise, salt and pepper, Poisson noise, speckle noise, Gaussian blurring, Wiener filtering, histogram equalization, sharpen, adjustment, and JPEG compression.

Fig. 7 shows the performances of our proposed scheme, Lai's scheme, and Ariatmanto and Ermawan's scheme [7] in terms of copyright protection of six geometrical attacks categories such as center cropped, row cropped, column cropped, rotate, resize, and translation. We can see that our scheme had the highest NC values, which almost 1 except in rotate and translation attacks. The Ariatmanto and Ermawan's scheme followed, where there were nearly equal proportions of center cropped and resizing attacks. Lai's method had the lowest NC values for all types' geometrical attacks, except in translation attack. In general, the statistics show that the robustness under image processing attacks proposed scheme performed better in all attacks than the existing schemes.

Besides, we also tested the proposed scheme under the image processing attacks which illustrates in Fig. 8 e.g. Gaussian noise, Salt and Pepper noise, Poisson noise, speckle noise, Gaussian blurring, media filtering, Wiener filtering, histogram equalization, sharpen, and adjustment. A prominent feature is that a significantly low NC value of robustness holds by Lai's method Most of the types

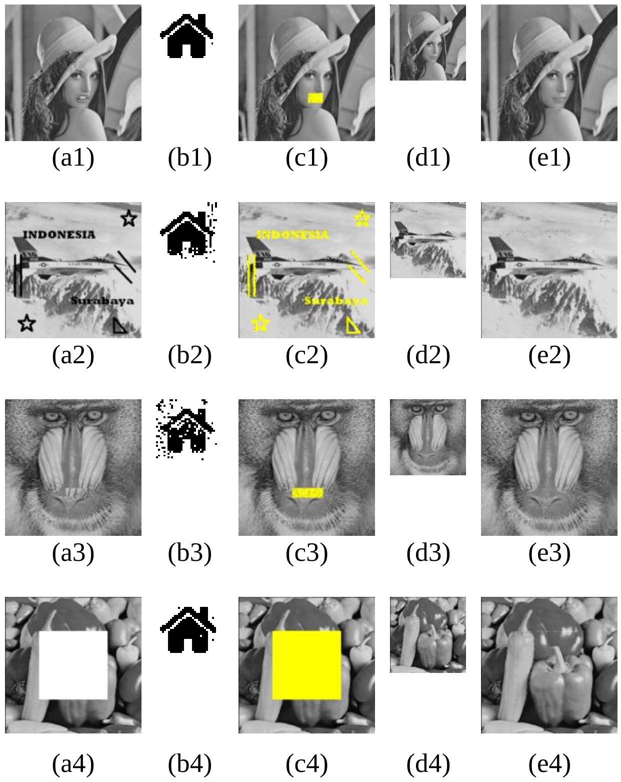


Figure. 3 Tampering test: (a1-a4) collage attack, additional object attack, substitution attack, cropping attack respectively; (b1-b4) their corresponding watermark extraction for copyright protection (NC=1, 0.9836, 0.9395, 0.9984 respectively) , (c1-c4) their corresponding tampering detection result, (d1-d4) their corresponding watermark extraction for authentication and recovery, and (e1-e4) their corresponding recovered image (PSNR= 47.25 dB, 28.61 dB, 41.11 dB, 30.47 dB; SSIM=0.9978, 0.9537, 0.9908, 0.9152 respectively

of image processing attacks hold better robustness in both Ariatmanto and Ermawan's scheme and proposed scheme. There is a significant similarity in the NC values of geometrical attacks at all categories between the two methods except in Gaussian noise; while this number is 0.9902 in Ariatmanto and Ermawan's method, in the proposed scheme is a mere 0.9305. Other results for robustness were roughly equivalent in adjustment attacks for three methods. In other image processing attacks, however, there were some significant differences especially for Lai's method

To further demonstrate the efficacy of the proposed scheme for image processing attack, the cover image has tampered JPEG compression with varying tampering quality as shown in Fig. 9 for five test images. The graph indicates the changes in the NC values of five images in JPEG compression over quality factor (QF) from 20 to 90. "Airplane" had the highest NC values of JPEG compression attacks with a stable around 1 and this figure had grown slightly to about 0.02 by quality factor 25. "Lena" was in second place with around 1 of NC values, although by QF 20 "Lena" had the largest NC value at around 0.97. The lowest is owned by "Baboon" which had the raw texture than others. It is also evident from Table 3 that the NC value of the proposed method compared to existing methods placed the highest value. It can be seen that the amount of NC value varied considerably across the two groups: JPEG and JPEG2000. The higher the JPEG quality factor the higher the NC value. Conversely, the higher the compression ratio (CR) of JPEG2000, the lower the NC value.

4.4 Dual watermarking schemes compared

The following section emphasizes higher performance in comparison to the associated dual-watermarking mechanisms [6, 23], and [27] of the dual watermark method proposed.

The biggest distinction among the four schemes is the invisible hybrid watermark for three functionalities: copyright protection and image authentication with recovery capability of our suggested scheme, while the scheme[6] and the scheme[23] rely on copyright protection and image authentication without respect for image recovery, the other scheme [27] is intensely watermarking for images authentification and image recovery without copyright protection.

Previously, the scheme [6] and the scheme [27] do not identify the violation until the credibility of the secure image has been compromised. The violation can be accurately detected by using the image authentication method via our proposed scheme and the scheme [23]. All four of the dual watermarking systems will endure some may attack to copyright protection. Nevertheless, the original host image must be separated in [23], which is a non-blind watermark scheme, from the second stable watermark in the color space of the RGB system. The method proposed achieved the optimum result (almost 41 dB) of the five images in terms of the PSNR value of the watermarked image. Table 4 displays the different characteristics of the dual watermarking methods.

Furthermore, our proposed scheme can reliably detect the changed areas, regardless of the source. However, given the precision of the position of usual tampered areas, the scheme [6] cannot detect all the changed regions and the detection rate is lower than 50%, which is not satisfactory, while scheme [27] has the highest detection rate, because they used two fragile watermarks. Nonetheless, based on the discussion in section 4.2, it is clear that the proposed scheme will successfully identify all the modified

Received: May 25, 2020. Revised: June 16, 2020.

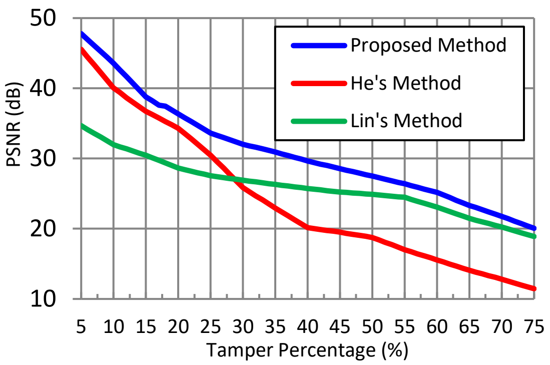


Figure. 4 Tamper detection performance under collage tampering of the proposed method

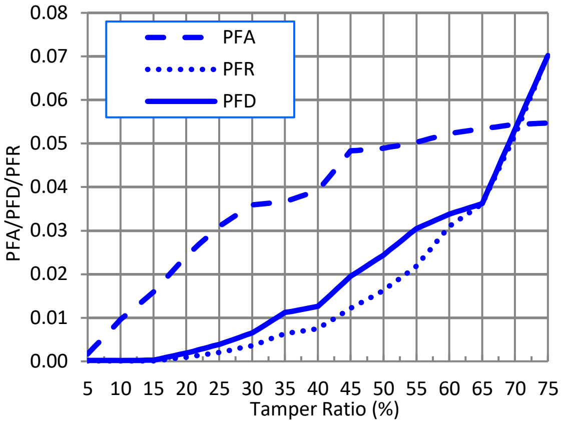


Figure. 5 Performance comparison of PSNR recovered image under the collage attack

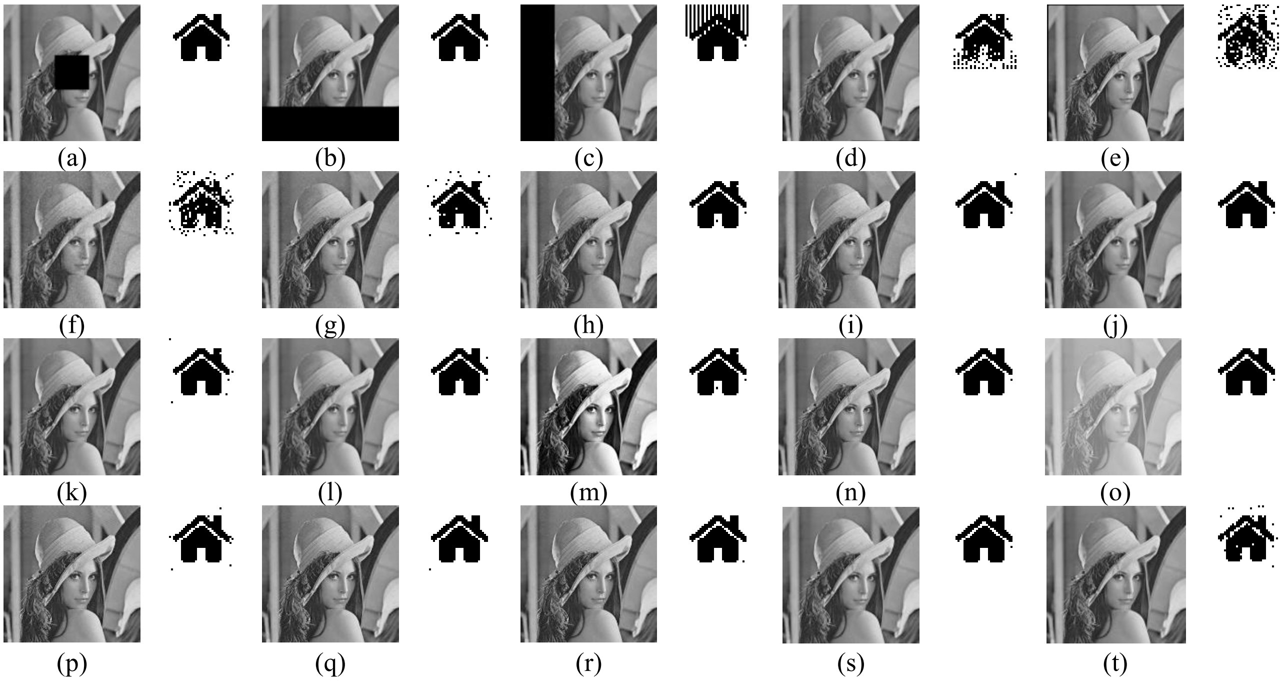


Figure. 6 Results of robust watermarked Lena image under various attacks and extracted watermark image under: (a) 25% center cropped, (b) 25% row cropped, (c) 25% column cropped, (d) Rotate 5^0, (e) translation [5,5], (f) Gaussian Noise 0.001, (g) Salt & Pepper Noise 0.01, (h) Poisson Noise, (i) Speckle noise 0.003, (j) Gaussian Blurring 3x3, (k) Median Filtering 3x3, (l) Wiener Filtering 3x3, (m) Histogram equalization, (n) Sharpen, (o) Adjustment, (p) JPEG Q20, (q) JPEG Q30, (r) JPEG Q50, (s) JPEG2000 CR 0, and (t) JPEG2000 CR18

regions. These findings indicate that the proposed scheme is outstanding among the related watermarking schemes during image authentication.

However, the watermark extraction of the schemes [6] is non-blind with respect to the performance of copyright security, with a strongly embedded watermark created by the host image itself instead of a predefined image on the logo. This means that, during copyright authentication, the ownership of the image can only be checked by fuzzy detection instead of obtaining a recognizable watermarked file. These recognizable watermarks are very popular in many practical applications. Therefore, after considering the practical global existence of the different mechanisms, the proposed dual watermarking mechanism is superior.

5. Conclusion

We have proposed dual watermarking methods which used to combine three functions of watermarking: authentication, copyright protection, and image recovery with superior robust watermark and recovery quality. A robust watermarking scheme uses DCT based on optimum embedding strengths and reference patterns. Selecting the DCT

Received: May 25, 2020. Revised: June 16, 2020.

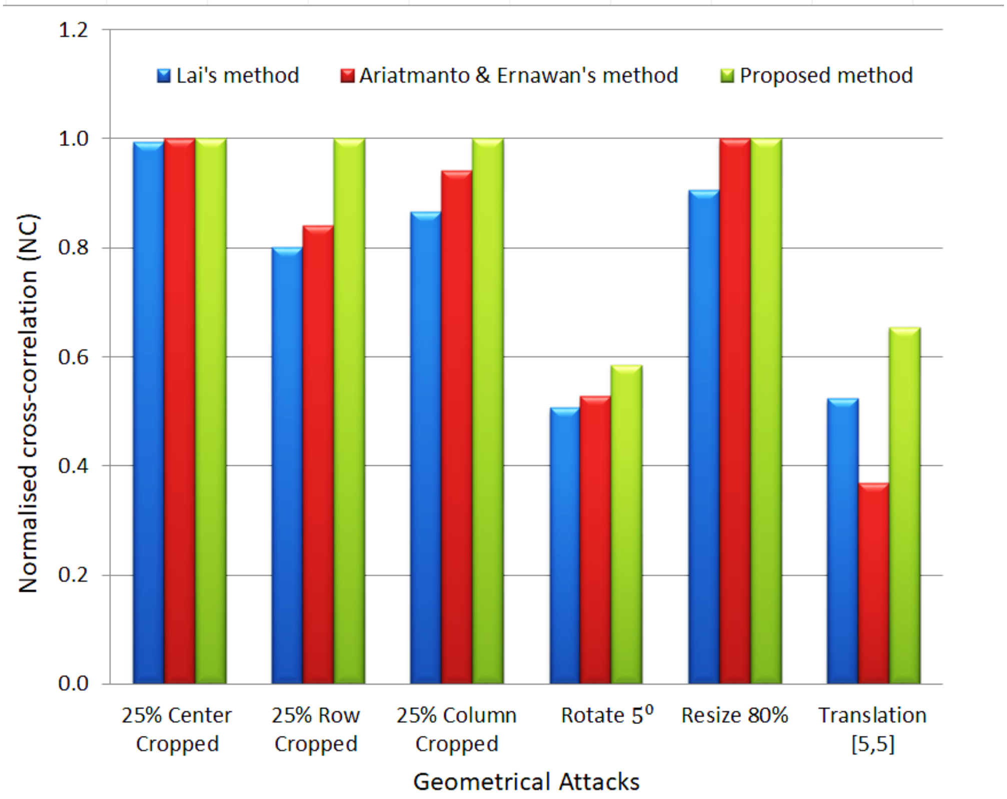


Figure. 7 Performance comparison under image geometrical attacks

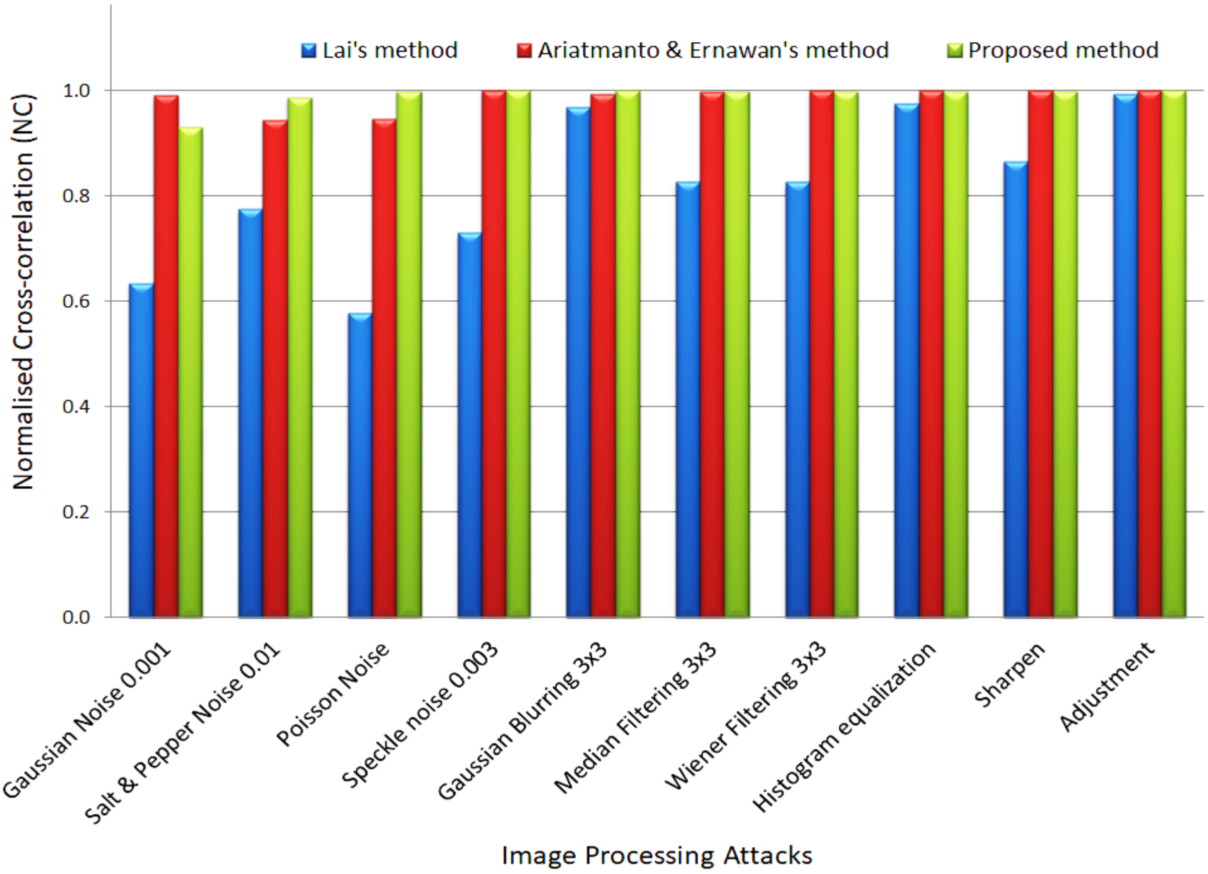


Figure. 8 Performance comparison under image processing attack

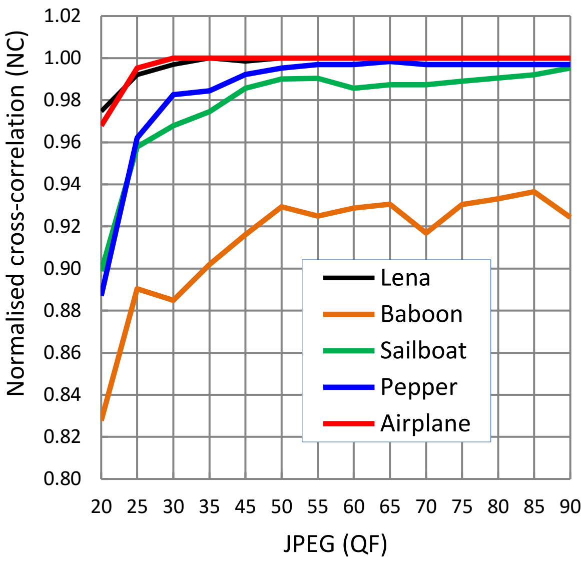


Figure. 9 NC values of the derived watermark of five reference images of varying JPEG qualities

Table 3. Performance NC values comparison of the extracted watermark with various JPEG and JPEG2000 qualities

Image Compression	Lai's Metho d	Ariatmanto & Ernawan's Method	Proposed Method
JPEG Q20	0.1438	0.6095	0.9612
JPEG Q30	0.7717	0.8917	0.9984
JPEG Q40	0.9642	0.9862	1.0000
JPEG Q50	0.9902	1.0000	1.0000
JPEG2000 CR 2	1.0000	1.0000	1.0000
JPEG2000 CR 6	1.0000	0.9990	1.0000
JPEG2000 CR 10	0.9715	0.9941	1.0000
JPEG2000 CR 14	0.8120	0.9254	0.9921
JPEG2000 CR 18	0.6417	0.8405	0.9858

Table 4. Comparisons of the characteristics of our dual watermarking schemes and the associated dual watermarking methods

Schemes	Hurrah et al [6]	Liu et al [23]	Bolourian et al[27]	Our schemes
Watermarks Type	Binary	Gray	Gray	Gray + Binary
Type Dual Watermarking	Fragile + Robust	Fragile + Robust	Fragile + Fragile	Fragile + Robust
Embedding Domain	Spatial + DCT	Spatial + DWT	Spatial + Spatial	Spatial + DCT
Imperceptibility (SSIM)	~1	~1	~1	~1
Watermarked image (PSNR)	~40 dB	~40 dB	~40 dB	~41 dB
Detection process	Not Blind + Blind	Not Blind + Blind	Blind + Blind	Blind + Blind
Tamper Detection Accuracy	High	Low	Very high	High
Robustness	High	Low	Very Low	High
Watermark Security (Map)	Low	Low	High	High
Copyright Protection	Yes	Yes	No	Yes
Image authentication	Yes	Yes	Yes	Yes
Image recovery	No	No	Yes	Yes

coefficients by certain rules generated the amount of embedding watermark. A Fragile watermarking scheme generates the block-mapping sequence randomly by the secret key and adopts the neighbourhood characterization to design an automatic.

Analytical expressions of false acceptance probabilities and false rejection of the proposed tamper detection system have been extracted and analysed in various malicious manipulations with good results. The self-embedding watermarking method can improve the quality of the recovered image, especially for collage attacks; experiments show the proposed method can increase the PSNR value by an average above 30 dB for the percentage of damage of more than 40%. Robust watermark can be obtained with good results for various compression attacks. Compared with existing image authentication algorithms, the proposed scheme can simultaneously implement tamper verification, identification of content, tamper localization, and recovery.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The first author mainly developed and evaluated the theory analysis, experimental evaluation and wrote this manuscript. The second author managed this research project, format analysis, validation this manuscript. The third author supported by creating a framework for thinking and reviewed this manuscript. All authors read and approved the final manuscript.

Acknowledgments

This work was supported by the *Kemenristekdikti* Indonesia under Grant No.6/E/KPT/2019, Grant No. 5/E1/KP.PTNH/20I9, and 737/PKS/ITS/2019.

References

[1] S. P. Singh and G. Bhatnagar, "A new robust watermarking system in integer DCT domain", *Journal of Visual Communication and Image Representation,* Vol. 53, No. February 2017, pp. 86–101, 2018.

[2] L. Laouamer and O. Tayan, "Performance Evaluation of a Document Image Watermarking Approach With Enhanced Tamper Localization and Recovery", *IEEE Access*, Vol. 6, pp. 26144–26166, 2018.

[3] D. Singh and S. K. Singh, "Effective self-embedding watermarking scheme for image tampered detection and localization with recovery capability", *Journal of Visual Communication and Image Representation*, Vol. 38, pp. 775–789, 2016.

[4] K.-C. Liu, "Self-embedding watermarking scheme for colour images by bi-level moment-preserving technique", *IET Image Processing*, Vol. 8, No. 6, pp. 363–372, 2014.

[5] L. Rakhmawati, W. Wirawan, and S. Suwadi, "A recent survey of self-embedding fragile watermarking scheme for image authentication with recovery capability," *EURASIP Journal on Image and Video Processing*, Vol. 2019, No. 1, 2019.

[6] C. Qin, H. Wang, X. Zhang, and X. Sun, "Self-embedding fragile watermarking based on reference-data interleaving and adaptive selection of emb e dding mode", *Information Sciences*, Vol. 373, No. 516, pp. 233–250, 2016.

[7] F. Cao, B. An, J. Wang, D. Ye, and H. Wang, "Hierarchical Recovery for Tampered Images Based on Watermark Self-Embedding Correspondence", *Displays*, Vol. 46, No. January, pp. 52–60, 2017.

[8] N. N. Hurrah, S. A. Parah, N. A. Loan, J. A. Sheikh, M. Elhoseny, and K. Muhammad, "Dual watermarking framework for privacy protection and content authentication of multimedia", *Future Generation Computer Systems*, Vol. 94, pp. 654–673, 2019.

[9] W. Alakk, H. Al-ahmad, and A. Kunhu, "A New Watermarking Algorithm for Scanned Grey PDF Files Using DWT and Hash Function", In: *Proc. of the 9th International Symposium on Communication Systems, Networks & Digital Sign (CSNDSP),* Manchester, UK, pp. 690–693, 2014.

[10] R. B. and M. R. G. Prabakaran, "A Robust QR-Code Video Watermarking Scheme Based On SVD and DWT Composite Domain", In: *Proc. of International Conference on Pattern Recognition, Informatics and Mobile Engineering*, Salem, India pp. 251–257, 2013.

[11] M. Hamid and C. Wang, "Adaptive Image Self-Recovery Based on Feature Extraction in the DCT Domain", *IEEE Access*, Vol. 6, pp. 67156–67165, 2018.

[12] C. C. Lai, "An improved SVD-based watermarking scheme using human visual characteristics", *Optical Commuications*, Vol. 284, No. 4, pp. 938–944, 2011.

[13] W. Chow, Y. Susilo, W. Tonien, and J. Zong, "A QR Code Watermarking Approach based on the DWT-DCT Technique A QR Code Watermarking Approach based on the DWT-DCT", *Lecture Notes in Computer Sccience 10343 314-331*, Auckland, New Zealand, 2017.

[14] L. Rzouga Haddada, B. Dorizzi, and N. Essoukri Ben Amara, "A combined watermarking approach for securing biometric data", *Signal Processing: Image Communications*, Vol. 55, No. March, pp. 23–31, 2017.

[15] S. Ong, S. Li, K. Wong, and K. Tan, "Fast recovery of unknown coefficients in DCT-transformed images Fast Recovery of Unknown Coefficients in DCT-Transformed Images", *Signal Processing: Image Communications*, Vol. 58, pp. 1-3, 2017.

[16] J. C. Patra, J. E. Phua, and C. Bornand, "A novel DCT domain CRT-based watermarking scheme for image authentication surviving JPEG compression", *Digital Signal Processing*, Vol. 20, No. 6, pp. 1597–1611, 2010.

[17] X. Zhang, Z. Qian, Y. Ren, and G. Feng, "Watermarking With Flexible Self-Recovery Quality Based on Compressive Sensing and Compositive Reconstruction", *IEEE Transactions on Information Forensics and Security*, Vol. 6, No. 4, pp. 1223–1232, 2011.

[18] C.-M. Wu and Y.-S. Shih, "A Simple Image Tamper Detection and Recovery Based on Fragile Watermark with One Parity Section and Two Restoration Sections", *Optics and Photonics Journal*, Vol. 3, No. 2, pp. 103–107, 2013.

[19] N. Bhargava, M. M. Sharma, A. S. Garhwal, and M. Mathuria, “Digital Image Authentication System Based on Digital Watermarking”, In: *Proc. of International Conference on Radar, Communication and Computing (ICRCC)*, pp. 185–189, 2012.

[20] X. Qi and X. Xin, “A singular-value-based semi-fragile watermarking scheme for image content authentication with tamper localization,” *Journal of Visual Communication and Image Representation*, Vol. 30, pp. 312–327, 2015.

[21] H. He, F. Chen, H. Tai, S. Member, T. Kalker, and J. Zhang, “Performance Analysis of a Block-Neighborhood- Based Self-Recovery Fragile Watermarking Scheme”, *IEEE Transactions on Information Forensics and Security*, Vol. 7, No. 1, pp. 185–196, 2012.

[22] P. L. Lin, C. K. Hsieh, and P. W. Huang, “A hierarchical digital watermarking method for image tamper detection and recovery”, *Pattern Recognitions*, Vol. 38, No. 12, pp. 2519–2529, 2005.

[23] X. L. Liu, C. C. Lin, and S. M. Yuan, “Blind Dual Watermarking for Color Images’ Authentication and Copyright Protection”, *IEEE Transactions on Circuits Systems and Video Technology*, Vol. 28, No. 5, pp. 1047–1055, 2018.

[24] S. Kim, F. Huang, and H. J. Kim, “Reversible data hiding in JPEG images using quantized DC”, *IEEE Transactions on Circuits Systems and Video Technology*, Vol. 26, No. 9, pp. 1610–1621, 2016.

[25] V. Vučković, “Embedding strength criteria for AWGN watermark, robust against expected distortion”, *Computer Informatics*, Vol. 29, No. 3, pp. 357–387, 2010.

[26] L. Rakhmawati, “Exploiting Self-Embedding Fragile Watermarking Method for Image Tamper Detection and Recovery”, *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 4, pp. 62-70, 2019.

[27] B. Bolourian Haghighi, A. H. Taherinia, and A. Harati, “TRLH: Fragile and blind dual watermarking for image tamper detection and self-recovery based on lifting wavelet transform and halftoning technique”, *Journal of Visual Communication and Image Representation*, Vol. 50, No. December 2016, pp. 49–64, 2018.