Exploiting Self-Embedding Fragile Watermarking Method for Image Tamper Detection and Recovery

Lusia Rakhmawati¹,²* Titiek Suryani¹ Wirawan Wirawan¹ Suwadi Suwadi¹ Endroyono Endroyono¹

¹Department of Electrical Engineering, Faculty of Electrical Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
²Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Surabaya, Surabaya, Indonesia
* Corresponding author’s Email: lusiarakhmawati@unesa.ac.id

Abstract: The increasingly rapid use of digital images makes data security technology become an important issue to ensure the integrity and ownership of the received image. Watermarking technique is a technique that ensures data security including for checking the authenticity of an image. Current developments, not only can test the authenticity of an image, but also can determine the location of damage while fixing it. The method proposed in this paper is through using the block-based technique of self-embedding fragile watermarking, where the watermark is selected from feature images representing the entire image. Four authentication bits and eight recovery bits are generated in each 2x2 non-overlapping block. Authentication bits are embedded in the three least significant bits (LSBs) of the block itself while recovery bits are embedded in the three LSBs of the corresponding mapped block. Our method is efficient as it only uses simple parity check operations and comparison between average intensities as image tamper detection and use the five most significant bits (MSBs) of 2x2 pixels in average for recovery. The experimental results of the tests on cropped images show that the proposed methods allow image recovery with acceptable visual quality better than some of state-of-the-art schemes.

Keywords: Fragile watermarking, Self-embedding, Tamper detection, Tamper recovery.

1. Introduction

The use of multimedia data is increasing rapidly because of the ease in the distribution process and saving data storage space. This significant development demands the existence of a reliable security system, because it does not require the possibility of such facilities to be used by irresponsible people for the sake of bad interests. Especially if the multimedia data is an important document which has a legal value [1]. Therefore, the technique that supports image authentication is currently an interesting topic. One technique that guarantees the integrity and authenticity of images is a digital watermarking technique [1, 2].

Watermarking technique is the technique of inserting information into multimedia data before being sent. The inserted information is called a watermark. Watermarks can be inserted through pixel [3] or block based insertion patterns [4, 5]. Most block-based authentication methods, where a multimedia data is divided into blocks which later become the place of insertion, in the spatial domain [6, 7] and frequency domain [8, 9].

In addition to the application of multimedia data authentication, the current development of the watermarking technique is used to detect and localize damage and able to recover it by extracting the watermark component in the recipient’s side [10]. Therefore, the watermark component is no longer a logo, personal data, or a particular code, but...
an image feature that represents image content so that it can be used for the recovery process [11].

Most methods used for the process of digital image detection and recovery are the fragile watermarking technique [4-7], which utilizes the fragility of the watermark to be able to be easily detected when there is a modification to the image. The watermark used is divided into two components—components for detection and components for recovery. The detection component is inserted in the block itself, while the recovery component is inserted in another block of mapping results [12].

Block mapping is done to facilitate the image recovery process if damage occurs because the main problem at the time of extraction is if the block used to cover the fault is also damaged, which is called coincidence problem, so as far as possible the chaotic map method capable of spreading the blocks throughout image area is chosen. The method proposed by [2] produces a good and simple randomization process using a particular key, but it must use a key that is a prime number so that the mapping becomes a one-to-one mapping. Another method proposed by [7] uses a pseudorandom series technique that can improve data security.

Besides block mapping, block sized of each image to be inserted is important. In [4] proposed the alterable-capacity coding method which generates the alterable-length code of each block sized 8×8 pixels based upon the roughness of the block, using large block sized make the detection more sensitive even only one bit. It is necessary to develop a technique that allows to cover only error parts, without changing the value of pixels in the block as a whole [10].

Another problem that needs to be considered is the selection of the number of watermark bits used. More and more inserted watermark components can increase detection and recovery capabilities, but they will reduce the image quality of the watermark. Therefore, we must choose the balance between the two. Zhang et al. [5] propose to use reference sharing which involves the average block in each MSB used. While the method [13] uses flexible payload, but it makes the complexity of the algorithm.

Other methods proposed in [14] are as follows: selecting key information from an image by performing discrete wavelet transform (DWT), and taking the coefficients from low frequency (LL sub band) of the DWT level-1 and level-2 as watermark, and work on the block with the size of less small, 2×2 pixels, so that they can make better image resolution after the error recovery. There is a drawback in this method, that the use of the watermark sub band LL resulting from the two-level DWT is not efficient, because the information from the LL sub band level-1 is the most representative for the reconstruction error.

In this paper, we proposed the development of the electoral component watermark, and the interpolation tamper detection, so that it can be applied for tamper detection and recovery scheme. The proposed method uses two watermarks, for tamper detection uses simple operations, parity check, and for recovery bit uses average intensities of each 2×2 pixels. In addition, we used 3 LSB as the watermark insertion point since it is pointed to have better result in [10], and with 3 LSB the additional information store capacity is 12 bits, 8 bit for recovery and 4 bit for authentication. It inspects accuracy of tamper localization. The experimental results denote the accuracy of general tamper detection and localization is 100%, and the tamper recovery shows better results compared to the state of the art method. This paper is organized as follows. Section 2, describes the proposed self-embedding fragile watermarking scheme. The simulation result and discussion are described and compared with other schemes in section 3. Finally, conclusion is provided in section 4.

2. The proposed scheme

In this section we explain the proposed watermarking method, namely self-embedding fragile watermarking. Broadly speaking, this method takes the features of the image as a watermark and then pastes it before the image is sent. At the receiver’s side, the watermark is extracted along with post processing for tamper detection and recovery. Because hidden information is a part of the image content, so the possibility of a difference is very high if it is modified, and it can also improve the detection process, and using the same chaotic mapping used in [7] it is easier to retrieve lost information. The advantages of the proposed method are simple and three Least Significant Bits as the insertion space for the watermark component are used. As long as the hidden watermark data is carefully selected, it is rational to believe that this method has the potential to further improve the error recovery performance of the changed watermark image.

In general, as shown in Fig. 1, there are three stages algorithm: watermark embedding, tamper detection, and recovery. The detail of proposed technique is described in detail as follows.
2.1 Watermark embedding

We adopted embedding steps used in [12]. As shown in Fig. 1, the original image I is assumed to have size \( M \times M \), as the multiple of b, is divided into non-overlapping blocks of uniform sized \( b \times b \) forming a series of blocks. The next step is to generate a block mapping that forms a look-up table. The procedure is explained below.

Step 1: Divide the image into non-overlapping blocks of \( 2 \times 2 \) pixels and assign a sequential integer \( B, \) \( B \in \{ 1, 2, \cdots, S \} \), to each block from left to right and top to bottom, where \( S = (M/2) \times (M/2) \) is the total number of image blocks.

Step 2: For each block number \( B \), apply Eq. (1) to obtain \( B' \) to form the block mapping sequence, where \( c \) is a secret key, a prime number and \( \in [1, S] \).

\[
B' = [f(B) = (c \times B) mod S] + 1 \tag{1}
\]

For each non-overlapping block, a watermark component consisting of two detection bits (d and e), and a recovery bit (r) is generated.

As shown in Fig 1, after watermark generation, we insert the watermark of each block using the LSB technique, in this case replacing the value of 3 LSB of each block denoted as \( A_i \in [0, 255], i = 1, 2, \cdots, (M \times M) \) and \( A_i \) can be converted into a binary form, with 8 bits of binary form to indicate a recovery bit, see Eq. (2).

\[
r_{i,l} = \lceil A_i/2^{l-1} \rceil mod 2, \ l \in [1,2,\cdots,8] \tag{2}
\]

Then, authentication bits can be generated by Eq. (3), Eq. (4), and Eq. (5) which were combined to form bit vector \( [d_{i1}, d_{i2}, e_{i1}, e_{i2}] \).

\[
(d_{i1}, d_{i2}) = [(P_i)_2 \times K_i) mod 2 \tag{3}
\]

Where \((P_i)_2\) is a binary form of pixel 1, pixel 2, pixel 3, and pixel 4 in each block. For 5 MSB we have 20 bits sequence. \( K_i \) is key generated randomly of size \( 20 \times 2 \).

\[
e_{i1} = v_{i8} \oplus v_{i7} \oplus v_{i6} \oplus v_{i5} \oplus v_{i4} \tag{4}
\]

\[
e_{i2} = \begin{cases} 1, & \text{if } p_{i1} = 0 \\ 0, & \text{if } p_{i1} = 1 \end{cases} \tag{5}
\]

In the watermark embedding procedure, the 3 LSB of each block is replaced with the watermark bits, while the 5 MSB of the original image is kept unchanged. The detection bits replace the \( 2 \times 2 \) block in LSB 3, and the recovery bits replace the corresponding block in LSB 2 and LSB 1 by Eq.(1).

2.2 Tamper detection

After the watermarked image \( W \) is sent, a receiver will detect if there is any modification caused by public channel using detection bits as shown in Fig. 1. For each \( b \times b \) block in the suspicious watermarked image \( W' \), we segment the watermark extracted from its 3 LSB into two segments, i.e., recovery bits \( [r_{i,j}, j = 1-8] \) and a detection bits vector \( [d_{i1}', d_{i2}', e_{i1}', e_{i2}'] \) with the same secret key on the transmitter. Then, compared with vector \( [d_{i1}, d_{i2}, e_{i1}, e_{i2}] \). If the results of the comparison get the same value, then the block is marked as an authentic block, otherwise it is marked as an inauthentic block.

2.3 Tamper recovery

After detection process, either the authentic or inauthentic block can be identified. This proposed method only restores an inauthentic block, while authentic blocks are maintained the same. For the invalid blocks, its corresponding block is used to find the recovery information. Then, LSB 1, LSB 2 of the corresponding block are used to pad each \( 2 \times 2 \) inauthentic block. If there are a few pixels which are not recovered, the non-linear median filter is used to interpolate the residual unrecovered pixels to avoid blurring the images as described in [7].

3. Experimental results and discussions

In this section we will discuss the proposed method by conducting an assessment of invisibility, tamper detection and localization, and reconstructing the tampered area. Invisibility shows the ability of the method to protect the image quality of the watermark, which must be identical to the original under normal observation using peak signal-to-noise ratio (PSNR) of watermarked image. Tamper detection and localization performance must be able to detect any disturbance in the watermarked image using probability of false detection (PFD); the system must be sensitive to dangerous manipulations, while also verifying authentic areas. The latest performance shows the system's ability to restore manipulated areas, thus the original content is verified using PSNR of the recovered image.
Figure 1 Block diagram self-embedding scheme using the proposed watermarking: (a) encoder side and (b) decoder side.

The PSNR value for watermarked image \( W \) with respect to original image \( I \) depends on the value of \( \alpha \). Assuming the decimal value of \( \alpha \) LSB in the original image is \( \sigma_I \) and the decimal value on \( \alpha \) LSB in the watermarked image is \( \sigma_W \), where the values of \( \sigma_I \) and \( \sigma_W \) belong to \([0, 2^\alpha - 1]\). Thus, the energy distortion caused by inserting a watermark in \( \alpha \) LSB for each pixel can be seen in Eq. (6).

\[
E_W(\alpha) = \frac{1}{2^{2\alpha}} \sum_{i=0}^{(2^\alpha-1)} \sum_{j=0}^{(2^\alpha-1)} (\sigma_I - \sigma_W)^2
\]

(6)

Then, the approximate value of PSNR for the watermarked image with respect to the original image can be calculated as Eq. (7), and we can obtain the approximate PSNR value under different parameters \( \alpha \) of watermark embedding capacity in.

\[
PSNR_W(\alpha) \approx 10 \log_{10} \left( \frac{255^2}{E_W(\alpha)} \right)
\]

(7)

In order to clarify the process of watermark insertion, authentication, and recovery, we conduct an experimental evaluation and compare it with other fragile watermarking schemes. A large number of test images sized 512 × 512 are used in our experiments to demonstrate the effectiveness of the schemes. We used several standard test images as shown in Table 2. The advantage of using less LSB is in the results of watermarked image that does not show any significant change, so the value of PSNR is also high, it can be seen from Table 2.
To measure the PSNR value based on image pixels, $PSNR_W$ is computed as Eq. (8) & (9).

$$MSE = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (I(x,y) - W(x,y))^2$$

$$PSNR_W = 20 \times \log_{10}(Max_I) - 10 \times \log_{10}(MSE)$$

Here, $M$ and $N$ is the image rows and columns, $I$ and $W$ represent the original image and watermarked image respectively and $Max_I$ is the maximum possible pixel value of the image.

Furthermore, in order to assess the tamper detection and recovery performance of the proposed method, we used several state of the art methods. To evaluate the visual quality of recovered image comparison with the watermarked image, we used $PSNR_R$. To evaluate the tamper detection performance, we used the probability of false acceptance (PFA), the probability of false rejection (PFR), and the probability of false detection (PFD), which is a better detection performance if the PFD value has a low value as shown at Eq. (12).

$$PFA = 1 - \frac{N_{td}}{N_t}$$

$$PFR = 1 - \frac{N_{ab}}{(N - N_t)}$$

$$PFD = \frac{N_{ab}}{N} \times PFA + \left(1 - \frac{N_{ab}}{N}\right) \times PFR$$

Where $N$ is the block number, $N_t$ is the number of actual blocks, $N_{td}$ is the number of tampered blocks which are correctly detected, $N_{ab}$ is the number of authentic blocks which are falsely detected. We performed experiments to test the performance of our proposed scheme on both tamper detection and tampered image restoration.

To derive the analytical analysis of the detection probabilities as mention in [7], all blocks in the watermarked image, $W$ can be partitioned into four mutually exclusive sets, as depicted in Fig. 2. These four sets of regions are: (1) $R_{TB}$, the tampered blocks located on the boundary of the tampered regions, (2) $R_{TT}$ tampered blocks located in the tampered regions, (3) $R_{AT}$, the authentic blocks adjacent to the tampered block, (4) $R_{NT}$, the authentic blocks not adjacent to the tampered blocks. Let denote the four condition above, i.e., $\beta \in (TB, TT, AT, NT)$, thus the conditional PFA and PFR can be define as:

$$P(FA|\beta) = 1 - P(D|\beta), \beta \in (TB, TT)$$

$$P(FR|\beta) = P(D|\beta), \beta \in (AT, NT)$$

Table 2. Watermark embedding capacity and quality of watermarked image

<table>
<thead>
<tr>
<th>Image</th>
<th>Capacity of embedded watermark ($\alpha$)</th>
<th>Quality of watermarked image (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lena</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Peppers</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Goldhill</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Airplane</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Received: February 25, 2019

International Journal of Intelligent Engineering and Systems, Vol.12, No.4, 2019

DOI: 10.22266/ijies2019.0831.07
\[ \gamma_{TB} + \gamma_{TT} + \gamma_{AT} + \gamma_{NT} = 1 \] (15)

For general tampering, i.e. cropping attack (CA), if the watermarked image block \( W_i \) is generally modified, then we consider the cases that \( d_i, e_i, r_i, d'_i, e'_i, r'_i \) are randomly changed. The values of \( d'_i, e'_i, r'_i \) equal to each integer in the interval \([0, 2^n - 1]\) with the same probability, where \( n \) is the number of bits in the feature of watermarked image block. Thus we have the probabilities \( P\{ (d_i, e_i, r_i) \neq (d'_i, e'_i, r'_i) \} = \frac{255}{2^8} \). It follows from [7] that the conditional PFA and PFR can be expressed as

For \( \beta \in \{ TB, TT \} \)
\[
P(FA|\beta_{CA}) = \frac{1}{256} + \frac{255}{256} \times P\{ (\delta_i < \delta_j | B_i \in R_\beta) \} \] (16)

For \( \beta \in \{ AT, NT \} \)
\[
P(FR|\beta_{CA}) = \frac{255 \gamma_T}{256} \times (1 - P\{ (\delta_i < \delta_j | B_i \in R_\beta) \}) \] (17)

Where \( \delta_i \) define as the number of nonzero pixels that are adjacent to the pixel in a current block and \( \delta_j \) for corresponding block.

The Fig. 3 shows the experimental results of PFD by the proposed scheme, Huo’s scheme [4], He’s scheme [7], and Singh’s scheme [10]. It reveals that the PFD of the proposed scheme is the smallest, followed by scheme [7], [10], and [4] respectively. This may be owing to the fact that for the proposed method, the validity of image block is only determined by the four authentication-bits inserted in the same block and the block size is 2 x 2 pixels. The PFD of He’s scheme is slightly larger than that of proposed scheme because the He scheme uses a block size of 3x3 pixels. It also shows that the PFD of the proposed scheme is kept at zero when compared to He’s scheme at tamper ratios from 5 % to 50%. It means that our proposed can detect the entire tampered region. Recovery quality of 10% - 30% of cropping attacks on Lena image for proposed scheme shown in Fig. 4 (a1-a5). Fig. 4 (b1-b5) shows the results of detection for each percentage of tampered images. The corresponding recovered areas are shown in Fig. 4 (c1-c5). Similarly, the results of other images for cropping region are shown in Fig. 5. It can be seen that the curve of PSNR values with respect to tampering rate decreases smoothly, but even if the tampering rate is 30%, the recovered images have their PSNR values more than 30 dB. Therefore, we can say that the recovered image quality is quite satisfactory.

Figure 4 Recovered image Lena after tampering by cropping attacks at various tampering areas: (a1-a5) five cropped images at various tampering rates, (b1-b5) five cropped detected areas of an image, and (c1-c5) five recovered images after cropping at various tampering rates.
Fig. 5 shows the recovered results of the proposed scheme and the three schemes ([4, 7, 10]) for the four images under different tampering rates, which simulated the different missing proportions of image blocks in wireless fading channel. For comparative analysis, the watermarked images were tampered with various degrees and then recovered the image from the tampered images. The Figs. 7-10 show the graph of PSNR (dB) of the recovered image with respect to the different tampering ratio.

Figure. 5 PSNR of recovered image with respect to the tampering rate

Figure. 6 Comparison recovered image results of the schemes [4, 7, 10] and the proposed scheme respectively under 5%: (a1-a5) 10%, (b1-b5) 15%, (c1-c5) 20 %, and (d1-d5) tampered images
These plots demonstrate the effectiveness of the proposed scheme and existing schemes. PSNR values of the recovered images in the proposed scheme are effectively higher than other method. Moreover, Fig. 7 – 10 shows that the PSNRs for the Lena, Pepper, Goldhill, and Airplane images by the same watermarking scheme almost pointed at the highest value. This implies that the complexity of the image content does not have much impact on the performance of tamper detection.

Recovery in the scheme [4] gives low PSNR values, but does work with a large tampering rate because they have PSNR values more than 30 dB for a tampering rate of almost 50%, except for the Airplane image, in which it does not work well. The Huo’s scheme divides the block with a high size of pixels, 8 x 8. If one block is detected as being damaged, valid pixels can be declared invalid because the method used is block based, and then all pixels in the block are declared invalid. In the proposed scheme, recovery data is an average value of the block using a small block, 2 x 2. Thus, its restoration quality is quite good enough. Hence, this scheme effectively recovered the image and also provides high accuracy in tampered pixel localization due to the use of small size block.

4. Conclusion

A block-based self-embedding fragile watermarking scheme for image tamper detection and recovery has been described. There are two watermarks which are used for authentication and recovery bit. Authentication bit using parity check and average intensities of image for tamper detection and localization which is embedded in the same block, while recovery bit using average intensity of each block to restore the tampering region is embedded in the corresponding block based on block mapping. From the results of experiments using four standard image tests using the LSB method, it is shown that our watermarking scheme which replaces three LSB of an image, with all the tampered region of the recovered image have efficiently been detected. If compared to the same three LSB for watermarking, the results of recovered image outperform the three peer schemes. Further development can combine the proposed scheme with the frequency domain techniques to improve watermark resilience for the sake of image restoration. In addition, the future work can be further extended to obtain method to get more accurate recovered image with increased PSNR value, addressing collage attack issues, and testing more images of various types.
Acknowledgments

This research was supported by Institut Teknologi Sepuluh Nopember (ITS) through Laboratory Research Grant (Peneltitan Laboratorium) 2018.

References