Fragile Watermarking for JPEG-2000 Images

`Wen-Nung Lie (賴文能), Tze-Liang Hsu(許澤亮), Guo-Shiang Lin (林國祥), and Wen-Jen Ho(何文捷)

Department of Electrical Engineering, National Chung Cheng University
160, San-Hsing, Ming-Hsung, Chia-Yi, 621, Taiwan, ROC.

E-mail: wnlie@ec.ccu.edu.tw

** Multimedia Technologies Laboratory, Institute for Information Industry

ABSTRACT

Fragile watermarking is designed to protect content integrity of digital data, meaning to detect any change in the content as well as localize areas that have been changed. In this paper, we propose a fragile watermarking scheme which embeds dual watermarks (one is naturally fragile and the other is actually robust) in JPEG-2000 compressed images with ROI (region of interest) selection. In this scheme, the fragile watermark is embedded in ROI coefficients at the first decomposition layer, while the robust watermark embedded in ROB (region of backgrounds) coefficients at the third decomposition layer. The combinatorial working of them is not only capable of detecting malicious changes precisely and flexibly but also estimating the degree of alteration to discriminate intended attacks from unintended ones (e.g., such normal image processing as compression, low-pass filtering, sharpening, etc). Experiments show that our functionally fragile watermarking scheme is scalable in view of detecting attack’s intension as well as accuracy in localizing the tampered areas.

1. INTRODUCTION

Digital images can be manipulated well by using image processing tools which make verification of content integrity difficult. This motivates the research of fragile watermarking [4-11,15] which embeds some characteristic data in an image in advance so that authentication of content integrity can be possible after they are modified or changed. Recent development entails a requirement that fragile watermarking should be capable of differentiating intended attacks from unintended attacks from common image processing that also modifies image content, but not severely.

JPEG-2000 is a state-of-the-art image compression standard [1,2]. Up to now, only two fragile watermarking schemes [8,15] were proposed. In this paper, we pay more attention to verify ROI (region of interest) selection. In this scheme, the fragile watermark is embedded in ROI coefficients at the first decomposition layer, while the robust watermark embedded in ROB (region of backgrounds) coefficients at the third decomposition layer. The combinatorial working of them is not only capable of detecting malicious changes precisely and flexibly but also estimating the degree of alteration to discriminate intended attacks from unintended ones.

The concept of dual watermarks had been described in [12] to deal with the deadlock problem. Kundur et al. [13] also proposed a watermarking scheme in which an auxiliary watermark is used to estimate the status of a communication channel. In [14], an additional signal is hidden and can be used to resist geometric attacks through the template matching procedure before watermark detection. Here in this paper, dual watermarks were developed with a brand new concept to achieve scalable detection of fragility.

2. ARCHITECTURE OF PROPOSED DUAL WATERMARKING

Fig. 1 illustrates the block diagram of proposed dual watermarking scheme for content integrity verification. In order to precisely detect the altered areas, we hide the first watermark \( W_1 \) which is sensitive and fragile in ROI. Basically, it cannot survive after any intended or unintended manipulations on the ROI of the targeted image. The second watermark \( W_2 \) which is composed of features of mid-frequency subbands is actually robust and devised to resist unintended manipulations. With the combinatorial interpretation of \( W_1 \) and \( W_2 \), the system is capable of differentiating unintended from intended attacks.

3. DESIGN OF DUAL WATERMARKS

3.1 The first (fragile) watermark

Taking ROI quality degradation into consideration, we propose to embed \( W_1 \) into high frequency subbands at level-1 decomposition to result in a least alteration under a given target bit-rate. First, some notations are given. Let \( I_{ro} \) be the quantized DWT coefficients of the \( l(P,Q) \) subband at the level-i decomposition. For example, \( I_{roi} \) is divided into \( N_i \) code-blocks, which are denoted as \( I_{ro,i} \), \( k = 1, 2, \ldots, N_i \), and \( I_{ro,i}(i) \) is the i-th coefficient of the \( k \)-th code-block. Fig. 2 illustrates an example of ROI masks (denoted as \( S_{ro,i} \)) for all the subbands. Where mask areas that have \( S_{ro} = 1 \) represents ROI and where \( S_{ro} = 0 \) means the ROB.

Notice that the ROI mask areas in subsequent resolution layers are not linearly proportional to the ROI area in the original image space (i.e., level-0). In practice, the ROI mask in different resolution levels can be obtained by downsizing the level-0 ROI area plus a certain extent of inflation. Let \( S_{ro,i}(i) \) be the mask value of the i-th coefficient in the k-th code-block, \( l_i \) be the last
completely encoded bit-plane of the \( k \)-th code-block, and \( I_{w1}(i) \) be the \( l \)-th bit-plane of the \( k \)-th code-block. Then the first watermark \( W1 \) is embedded into the subband HH1 by using the following rule:

\[
I_{w1}(i) = \begin{cases} 
W1(m) \in \{0,1\} & \text{if } S_{w1}(i) = 1 \\
I_{w1}(i) & \text{if } S_{w1}(i) = 0
\end{cases}
\]  

(1)

where \( W1(m) \) represents the watermark bit in \( W1 \). The above rule indicates the bit replacement strategy in the \( l \)-th bit-plane. Notice that while embedding \( W1 \), we should get the information of \( I_{w1} \) from the rate-distortion (R-D) optimization procedure in EBCOT [5]. After \( W1 \) is embedded, the number of encoded bit-planes may be changed under the same given bit-rate. Hence, the following steps are adopted to ensure that \( W1 \) can be correctly extracted at the decoder side.

Step 1: Find \( I_{w1} \) (the last completely encoded bit-plane) from the rate-distortion optimization process in EBCOT under a given bit-rate.

Step 2: Embed \( W1 \) into ROI coefficients via Eq. (1)

Step 3: Repeat Step 1 to get the updated \( I_{w1} \). Go back to step 2 to try embedding in the \((l + 1)\)-th bit-plane if \( I_{w1} \) is larger than the previous \( I_{w1} \) (or, the watermarked bit-plane will not be transmitted). Otherwise, modify the R-D value (reduce it numerously) of the \((l + 1)\)-th bit-plane so that \( I_{w1} \) remains the same after EBCOT bit allocation. In this way, the decoder can extract \( W1 \) correctly from the \( l \)-th bit-plane without the need of transmitting \( I_{w1} \).

![Fig. 2 An illustration of the ROI masks.](image)

When decoding the watermarked JPEG-2000 images, we extract \( W1 \) from the last completely encoded bit-plane \( I_{w1} \) of the coefficients \( I_{w1} \). According to the above three steps, this \( I_{w1} \) will be the same as that in the watermarking phase. We denote the extracted watermark as \( W1 \) and compare it with \( W1 \) to verify the integrity of ROI.

### 3.2 The second (robust) watermark

#### A. Generation of the second watermark

Due to fragility of the first watermark, it can support a good performance in detecting malicious attacks such as replacement of data, but it cannot differentiate reasonable image processing (such as low-pass filtering, compression, etc.) from malicious attacks. Therefore, the second watermark \( W2 \) is designed to survive after image processing and assist the above differentiation. Basically, \( W2 \) indicates inter-subband statistical relations in ROI. The change of this inter-subband statistical relations depends on the kinds and levels of external manipulations. For example, alteration of this statistical property would be observable when a considerable part of ROI image is replaced. In contrast to malicious attacks, the statistical property still remains the same or little altered after unintended manipulation. Here, the statistical relation, or called feature, is defined to the absolute differences between corresponding ROI coefficients in the LH3 and HH3 subbands. The amplitude of features is categorized into four kinds after sorting by values.

Each category is labeled by using the following rule.

\[
L_{w2}(i) = \begin{cases} 
0 & \text{if } |I_{w2}(i) - I_{w1}(i)| < T_0 \\
1 & \text{if } T_0 \leq |I_{w2}(i) - I_{w1}(i)| < T_1 \\
2 & \text{if } T_1 \leq |I_{w2}(i) - I_{w1}(i)| < T_2 \\
3 & \text{if } |I_{w2}(i) - I_{w1}(i)| \geq T_2
\end{cases}
\]

(2)

where \( T_0 \), \( T_1 \), and \( T_2 \) stand for three partition thresholds and \( L_{w2}(i) \) denotes the label for each pair of coefficients in ROI. \( L_{w2}(i) \)'s are converted into binary bits and concatenated to form the second watermark \( W2 \).

It is important to select proper thresholds \( T_i \) so as to make tradeoffs between fragility and robustness of ROI content change. For example, uniform partition of possible values of \( |I_{w2}(i) - I_{w1}(i)| \) may work. However, non-uniform partition may be capable of optimizing a criterion function (e.g., sensitivity of ROI content changes), but suffers from the need of transmitting \( T_i \)'s, as the side-information, to the receiver.

#### B. Embedding of the second watermark

From mentioned in Fig. 1, \( W2 \) is embedded into coefficients of the HL3, LH3, and HH3 (in the embedding order) subbands of ROB to improve robustness. Since the bit-planes of ROB will be scaled down, opposed to scaling-up of the bit-planes of ROI, we have to embed \( W2 \) into the higher bit-planes of ROB to avoid easy removal. Though this procedure will obviously degrade the quality of ROB significantly, it does not matter for ROB area which is normally of low quality and gains less attention from users. Nevertheless, taking image degradation into account, we will not perform bit replacement directly as in embedding \( W1 \). Instead, we select bits that belong to pass 1 (significant propagation pass), due to their higher R-D values, and do selective bit replacement. Fig. 3 indicates the bit positions selected for \( W2 \) embedding. In this way, the decoder can extract \( W2 \) after getting the information of significant bits from the higher bit-plane.

![Fig. 3 Bit positions for embedding W2.](image)
C. Extraction of the second watermark

Assume that the side information (e.g., the resolution layer where the watermarks are embedded, the subbands from which W2 are generated, T1, T2, T3, etc.) is known in advance at the decoder side. The extraction steps are described as follows.

Step 1: Calculate W2 by analyzing ROI statistical information at the embedding layer (using fixed T1, T2, T3). Denote its length as N2.

Step 2: Calculate the embedding capacity of each possible bit-plane α until C ≥ N2 is found.

Step 3: Extract W2 from the selected bit-plane α of ROB (in the order of LH3, HL3, and HH3 subbands).

We can verify the integrity of ROI by comparing W2 to W2:

\[ \Delta(i) = \left| W2(i) - W2(i) \right|, \quad (3) \]

where each W2(i) or W2(i) represents a 2-bit information (i.e., 0–3) resulting from the difference between one pair of ROI coefficients (from LH3 and HH3 subbands, respectively). If \( \Delta(i) \neq 0 \), it represents that values of the paired ROI coefficients have been changed or the second watermark embedded in ROB may have been removed. This phenomenon indicates that analyzing \( \Delta(i) \) is useful in evaluating the degree of content integrity or classifying the kind of attacks (malicious or not). We define the percentage of W2(i) samples that satisfy \( \Delta(i) \geq \gamma \), \( \gamma \in \{0, 1, 2, 3\} \), as \( p_i(\gamma) \). For example, higher values of \( p_i(2) \) and \( p_i(3) \) indicate that ROI may encounter serious attacks. The higher the value of \( p_i \), the larger probability that the image is incurred tampering.

3.3 Combining interpretations of dual watermarks

From the extraction results of W1 and W2, we can judge whether the content modification is malicious or not (from \( \Delta(i) \) ) and simultaneously localize the possible image area that is incurred modification. For the latter, there is a mapping between areas in the embedding layer (here, level-1 for W1 and level-3 for W2) and the image space. Hence, for accuracy consideration, the embedding layer should be as small as possible (but easy to be removed). For robustness consideration, the embedding layer should be as large as possible (but less localization accuracy). This is why we embed the fragile watermark W1 in level-1 while embed robust watermark W2 in level-3.

4. EXPERIMENTAL RESULTS

A. Quality Evaluation

In the experiments, gray-scale images, Lena, Baboon, and Jet, each of 256x256 pixels, were used. An ROI area of 115x115 pixels, centering around (121, 121) is defined. Fig. 4 illustrates a watermarked JPEG-2000 image with the defined ROI. The watermark W1 is embedded into ROI in the HH1 subband, while the second watermark W2 is embedded into ROB in LH3, LH3, and HH3 subbands. Table 1 shows quality degradation of these three test images. It can be observed that hiding two watermarks only results in an average quality degradation of 1.47dB.

B. Integrity Verification

Here we used three kinds of image manipulations to simulate possible attacks: replacement, JPEG compression, and smoothing. The goal is to evaluate the performance in verifying content integrity. Two criteria used to verify content integrity are described as follows.

- Detection rate: the percentage of altered pixels (partial or the full ROI) which are correctly detected. For examples, \( d_i \) represents the detection rate by W1 and \( d_i^c \) stands for the detection rate by W2 when \( \gamma = 1 \).
- False alarm rate: the percentage of pixels (with respect to the number of altered pixels) that are not altered but detected incorrectly. For examples, \( f_i \) is the false alarm rate by W1 and \( f_i^c \) denotes the false alarm rate by W2 when \( \gamma = 1 \).

The results of verifying replacement manipulations are shown in Fig. 5. Fig. 5(a) shows that the defined ROI region is replaced with another image. Fig. 5(b) shows the detection result by using W1, where black pixels represent those identified as being altered. Fig. 5(c) demonstrates the detection result by using W2 when \( \gamma = 1 \). The detection rates by our dual watermarks are summarized: \( d_i = 45.4\% , \ d_i^c = 68.8\% , \ d_i^c = 29.7\% , \ d_i^c = 6.3\% \). The false alarm rates are: \( f_i = 2.44\% , \ f_i^c = 34.4\% , \ f_i^c = 6.3\% , \) and \( f_i^c = 1.6\% \). As we can see, the detection rate of W1 is about 50%. This is because the 1-st bit-plane (less significance) of ROI in the HH1 subband is randomly altered by the attack. The false alarm rate comes from the mapping inaccuracy from a higher level subband to a lower-level one. For example, a pixel in HL3 subband corresponds to an 8x8 block in the original image space. The outer rectangle in Fig.5(c) represents the inflated area corresponding to the ROI mask at the third level. Fig. 5(d) shows another nearly seamless replacement attack (a more practical attack than Fig.5(a)). As we can see, since the last completely coded bit-plane (\( l_i \)) of ROI in the HH1 subband may be altered, the false alarm region is spread out to the whole code-block, thus increasing the \( f_i^c \) value (Fig.5(e)). Fig. 5(f) shows the detection result by using W2 with \( \gamma = 1 \).

<table>
<thead>
<tr>
<th>Bit-rate (1.2 bpp)</th>
<th>JPEG-2000</th>
<th>ROB Bit-plane shifting = 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR (dB)</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>Original</td>
<td>Marked</td>
</tr>
<tr>
<td>Lena</td>
<td>42.8</td>
<td>40.4</td>
</tr>
<tr>
<td>Baboon</td>
<td>36.4</td>
<td>34.2</td>
</tr>
<tr>
<td>Jet</td>
<td>41.3</td>
<td>39.5</td>
</tr>
</tbody>
</table>

Table 1 PSNR degradation for three JPEG-2000 images after proposed dual watermarking.
other hand, the $d'_i$ detection rate is lower for smoothing than for replacement (more malicious).

We also conduct experiments in which the watermarked JPEG-2000 images are re-encoded at a lower target bitrate. (e.g., from 1.2 bpp to 1.0 bpp). The results show that $d_i$ is approximately 0.5, as expected, and $d'_i$ is high, due to the change of $W_2$ in ROI or the removal of embedded $W_2$ in ROB.

5. CONCLUSIONS

In order to verify content integrity of JPEG-2000 images, we devise a dual watermarking scheme in which a fragile and a robust watermarks are embedded simultaneously. The first watermark is embedded in the less significant bitplanes of the ROI area in the HH1 subband to achieve more fragility, but higher accuracy in localizing altered image area. To distinguish the common image processing operations from malicious tampering, we generate the second watermark by analyzing ROI subbands to achieve robustness. Since the second watermark records the features of wavelet coefficients for verification, it’s practically a fragile watermark.

The advantage of our method is two-folds: 1) via the combination of a fragile and a robust watermarks, allowable image processing can be distinguishable from malicious attacks, 2) by varying the threshold $\gamma$, the fragility can be scalable.

ACKNOWLEDGEMENT

Thanks to NSC, ROC., with contract NSC 90-2213-E-194-021.

REFERENCES