A BLIND IMAGE WATERMARKING BASED ON DUAL DETECTOR

Chi-Jen Liao¹, Chaur-Heh Hsieh¹, Mao-Fu Lai²

¹Department of Information Engineering, I-Shou University, Kaohsiung, Taiwan
²Dept. of Electronic Engineering, Tong-Nan Institute of Technology, Taipei, Taiwan
Corresponding author: C-H Hsieh (hsieh@isu.edu.tw)

Abstract —This paper presents a blind image watermarking technique based on a novel detection scheme which contains two detectors aiming at the positive attack and negative attack, respectively. The experimental results indicate the proposed technique improves robustness significantly, as compared to a single detector scheme.

Key Words—blind watermarking, attacks, robustness.

I. INTRODUCTION

Digital watermarking can prove to be useful to avoid the illegal use of digital media. Generally, it can be classified into two types: robust watermarking and fragile watermarking. The requirements for robust watermarking are transparency, robustness, security, capacity, universality, and unambiguousness [1]. Among them, transparency and robustness are most important. Transparency refers to the perceptual quality of the image being protected. In other words, the watermark should be invisible over the original image. Robustness refers to the ability to detect the watermark after unintentional attack, i.e., common signal processing operations [2].

Blind watermarking means that the watermark decoder must not require the original host image for extracting the embedded watermark code. It is more useful than non-blind because the host image may not available in real-world scenarios. Generally, blind schemes are often less robust and also harder to implement than non-blind [3].

The existing blind schemes can be roughly classified into three types [3]: (a) correlation-based, (b) based on absolute modulation of individual primary or secondary elements of an image, (c) based on relative modulation of pair elements. Most of these schemes focus on embedding strategy. In addition, they do not exploit the characteristics of the attacks. In this work, we aim at the design of an efficient detection scheme that takes signal processing attacks into account.

The signal processing attack in image watermarking contains two types in which one results in valumetric distortion, and the other yields geometric distortion [2]. The valumetric distortion arises from additive noise, linear filtering, lossy compression, or amplitude scaling. In this paper, we use amplitude attack and geometric attack to represent the two distortions, respectively. In addition, only amplitude attack is considered here. Although geometric attacks are more difficult to deal with than amplitude attacks and hence have received more attentions in recent years, our study [4] indicates that most of geometric attacks yield amplitude distortions as well, thus they could be considered as generalized amplitude attack.

Detection is an inverse process of embedding. Most of detection schemes in the literature are based on the inversion only, which do not further employ the characteristics of the attacks. This paper categorizes the attacks into positive attack and negative attack, and then design two detectors accordingly.

II. PROPOSED BLIND WATERMARKING

The proposed watermarking system is shown in Fig.1. The host image is transformed by a full-domain DCT. The DC coefficient is discarded, and the remaining two-dimensional AC coefficients are converted into one-dimensional coefficient sequence $\left( k_F \right)$ via zig-zag scanning. The binary watermark is embedded into the coefficient sequence. In detection end, the received watermarked image is transformed with DCT as in the embedding end, and the resulting AC coefficients are fed into positive-attack detector and negative-attack detector simultaneously. The larger response of the two detectors is input to the thresholding device. If it is greater than the detection threshold $\rho$, we claim that the watermark exists.

The design of embedding and detection are performed on the above AC coefficient sequence, and the design details are described in the following.

2. 1 EMBEDDING

We calculate a reference (estimate) sequence from the AC coefficient sequence $F(k)$ by a slide window of $(2m+1)$; i.e.

$$\bar{F}(k) = \text{sign}(F(k)) \cdot \frac{1}{(2m+1)} \sum_{j=-m}^{m} |F(k+j)|$$

(1)

In this work, the slide window length is 5 ($m=2$). The watermark $W_i$ is embedded by

$$F_m(k) = \bar{F}(k) \left[ 1 + \text{sign}(W_i) \cdot \alpha \right]$$

(2)

where $\alpha$ is a hiding factor with the value between 0 and 1 (in this work, $\alpha = 0.5$), and sign ($x$) is a sign function.

In Eq. (2), if the watermark bit is 1, $F_m(k)$ is the sum of the reference $\bar{F}(k)$ and a positive embedded energy...
The watermark is extracted by

$$W^e(i) = \text{sign}\left( F_m^a(k) - \overline{F_m^a}(k) \right)$$

where $F_m^a(k)$ is the DCT coefficient sequence of the test image which suffers from attack, and $\overline{F_m^a}(k)$ is the corresponding reference coefficient sequence. More specifically, if $F_m^a(k) > \overline{F_m^a}(k)$, the extracted watermark bit is 1; otherwise is -1. The detection scheme has been widely used in most blinding watermarking techniques, which doesn’t consider the attack characteristics. Here, we refer it to single detector for comparison purpose.

In general, the length of a watermark, $B_L$, is much smaller than that of the AC coefficient sequence, $C_L$. To increase the security, we expand randomly the original binary watermark stream, a sequence of \{+1, -1\}, into a ternary stream, a sequence of \{+1, -1, 0\}. The coefficients corresponding to the symbol 0 are not embedded with watermark message. The length of the ternary stream, $T_L$, is obtained according to the range of the AC coefficient to be hidden. $T_L$ is larger than $B_L$ but smaller than $C_L$.

### 2.2 DETECTION

Detection is an inverse process of embedding. Most of detection schemes in the literature are based on the inversion only, which do not consider the characteristics of the attacks. This paper categorizes the attacks into positive attack (PA) and negative attack (NA), and then design two detectors accordingly. We call the detection scheme as dual detector for convenience. The scheme is designed mainly based on the analysis of the typical attacks in [4], [5], in which the signal processing attacks are classified as PA, NA and RA (random attack). In this work, the PA and NA are employed to design the dual detector.

Our observation indicates that PA would move the embedded coefficient away from the reference value. Thus the attacked coefficient is more robust to noise than the unattacked one. So PA is helpful for the extraction of watermark. On the contrary, NA will pull embedded coefficients in a direction close to the reference value. In this case, small value of noise may result in the jumping from above (or below) reference to below (above) reference, which yield the error of extraction of watermark. Furthermore, for NA, when attack energy is high, all the coefficients will be reduced to approximate zero value. In such case, the detection of watermark will fail. So, it is reasonable to say that if the value of a coefficient is very small, the coefficient is probably suffered from a large negative attack. Therefore, it is not reliable and should not be used for the detection.

As mentioned before, some of the coefficients of the attacked image are not reliable. The unreliable coefficients should not be included for watermark extraction; otherwise, it will yield errors of extraction. In this work, we present a novel detection scheme that selects the reliable coefficients for extraction. It contains two detectors that exploit the characteristics of positive attack and negative attacks, respectively.

For convenience, we define an attack magnitude deviation ratio as

$$R = \left| \left( F_m^a(k) - \overline{F_m^a}(k) \right) / \overline{F_m^a}(k) \right|$$

If no attack exists, it is easy to obtain $R = \alpha$ from Eq. (2). Therefore, if $R > \alpha$ we can say that positive attack occurs; otherwise, negative attack occurs when $R < \alpha$. The decision rule is suited for the idea case in which the reference $\overline{F_m^a}(k)$ is a fixed value. However, our investigation indicates that $\overline{F_m^a}(k)$ may fluctuate slightly with different attacks. To avoid the classification error of attack types, we narrow down the PA range by introducing the parameters $\beta$. Specifically, if $R > \beta$ ( $\beta > \alpha$), the attack is regarded as positive. Based on the concept, we design a detector for extracting the coefficients suffered from positive attack as follows.

- **Positive Attack Detector:**

  The watermark is extracted by

  $$W^e(i) = \begin{cases} 1 & \text{if } R > \beta \text{ and } \text{sign}\left( F_m^a(k) - \overline{F_m^a}(k) \right) = \text{sign}\left( F_m^a(k) - \overline{F_m^a}(k) \right) > 0 \\ -1 & \text{if } R > \beta \text{ and } \text{sign}\left( F_m^a(k) - \overline{F_m^a}(k) \right) = \text{sign}\left( F_m^a(k) - \overline{F_m^a}(k) \right) < 0 \\ 0 & \text{otherwise} \end{cases}$$

  (4)

  The extracted watermark bits for the positively-attacked (PA) coefficients are +1 or -1; otherwise are 0 for NA coefficients. Thus the detector output contains three symbols: +1, -1, and 0. The normalized correlation is defined as

  $$\rho_{m,n}\left(W,W_f^*\right) = \frac{\sum_{i} W(i) W_f^*(i)}{\sum_{i} W(i) W_f^*(i)}$$

  (5)

  The denominator is the total number symbols of +1 and -1 extracted.

  As mentioned before, negative attack may make the $R$ very small, or the reference vaule close to zero. In either case, the extraction of watermark is unreliable because it is sensitive to noise. In order to raise the reliability of watermark extraction, we delete the unreliable coefficients with $R < \varepsilon$ and $\overline{F_m^a}(k) < \sigma$, and thus obtain the negative attack detector in the following.

- **Negative Attack Detector:**

  The watermark is extracted by

  $$W^e(i) = \begin{cases} \text{sign}\left( F_m^a(k) - \overline{F_m^a}(k) \right) & \text{if } \varepsilon < R < \beta \text{ and } \overline{F_m^a}(k) > \sigma \\ 0 & \text{otherwise} \end{cases}$$

  (6)
The detector output also contains three symbols: +1, -1, and 0. The normalized correlation is
\[ \rho_{\text{max}}(W,W^*) = \frac{\sum_i W(i) W^*_i}{\sum_i W(i) W^*_i} \]  \( (7) \)

The largest of the above two detection responses is used to judge whether the watermark is present or absent by comparing it with a predefined threshold.

The values of parameters \( \varepsilon \) and \( \beta \) depend on the value of \( \alpha \). The relationship is hard to achieve theoretically. Our experience indicates that \( \beta \geq 1.25 \alpha \) and \( \varepsilon \leq 0.75 \alpha \) are good choices. In addition, the threshold \( \sigma \) is determined experimentally and \( \sigma = 20 \) is a good choice.

III. SIMULATION

The experiments are conducted on 50 natural images with size of 128 x 128. A binary watermark length with length of 1024 is obtained by generating a zero-mean pseudo-random sequence with length of 1024, and then taking the sign of each data point of the sequence. The binary watermark is expanded into a ternary stream with length 1024 x 15, and then modulated into the host images. Fifteen types of image processing attacks (listed in Table 1) including positive (e.g., sharpening), negative (e.g., blurring) and hybrid attacks, are applied to the 50 images, respectively, and thus totally 750 attacked images are obtained. The hiding factor is chosen as \( \alpha = 1.5 \), which yields good embedded picture quality (average PSNR is approximately 32 dB). The detection parameters \( \varepsilon \) and \( \beta \) are chosen as \( \varepsilon = 0.2 \) and \( \beta = 0.7 \).

To evaluate the false alarm, the 50 original images (unwatermarked) are attacked by 15 types of image processing above, thus yielding 750 images. The two set of images are used to evaluate the single detector and our proposed dual detector.

The comparison of the two detection schemes is in terms of missed detection rate and false alarm rate. The ROC curves of the two error rates shown in Fig.2 indicate that the proposed dual detector performed better than the single detector. Figs. 3 and 4 show the ROC curves under various detection threshold values. The figures provides the cue for the selection of appropriate threshold values to meet the requirements of users. It is seen from Fig.9 that if threshold is set to be 0.2 to 0.3, the dual detector gives zero missed detection rate and zero false alarm. However, Fig.3 indicates that the single detector definitely yields errors (either missed detection or false alarm or both) whatever the threshold values are chosen. Obviously, the dual detector is better than single detector from the viewpoint of threshold selection.

IV. CONCLUSIONS

A blind image watermarking technique based on the characteristics of amplitude attacks has been proposed. In the novel method, two detectors are designed which aim at extracting reliable coefficients under positive attack and negative attack, respectively. The reliable coefficients are very useful to reduce the error of watermark extraction. The results indicate that our scheme perform better significantly than the single detector that does not consider the attack characteristics. The proposed detection scheme can be applied to any existing schemes to further improve their performances. The major limitation of our scheme is that the determination of parameters values is performed experimentally. An automatic calculation mechanism for the parameters will be investigated in the future.

REFERENCES

Fig. 1 Proposed blind watermarking system.

Table 1 Image processing attacks

<table>
<thead>
<tr>
<th>Image Processing Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>01- Average (7×7)</td>
</tr>
<tr>
<td>02- Blur (75)</td>
</tr>
<tr>
<td>03- Gaussian Blur (7×7)</td>
</tr>
<tr>
<td>04- Soften (75)</td>
</tr>
<tr>
<td>05- JPEG compression (75)</td>
</tr>
<tr>
<td>06- JPEG compression (25)</td>
</tr>
<tr>
<td>07- Enhance Detail</td>
</tr>
<tr>
<td>08- Enhance Edges (75)</td>
</tr>
</tbody>
</table>
Fig. 2 The ROC performances of single detector and dual detector.

Fig. 3 The error rates vs. detection threshold values for single detector.
Fig. 4 The error rates vs. detection threshold values for dual detector.