On the Robustness of Watermarking in the Frequency Domain for Still Images

Akio MIYAZAKI and Masataka EJIMA

Graduate School of Information Science and Electrical Engineering, Kyushu University
6-10-1, Hakozaki, Higashi-ku, Fukuoka, 812-8581 JAPAN
E-mail: miyazaki@is.kyushu-u.ac.jp, ejima@mobcom.is.kyushu-u.ac.jp

Abstract This paper aims to establish fundamentals of measuring and evaluating the performance of watermarking systems. We first present the general model of a watermarking method for still images. Based on this model, we propose a statistical method of measuring the performance and robustness of the watermarking system. Then, the DCT-based watermarking system is analyzed and its performance is evaluated by using the proposed method.

1. Introduction

With the rapid spread of computer networks and the further development of multimedia technologies, the copyright protection of digital contents such as audio, image and video, has been one of the most serious problems because digital copies can be made identical to the original. The digital watermark technology is now drawing the attention as a new method of protecting copyrights of digital contents. Digital watermark is realized by embedding information data, *e.g.,* owner, distributor, or recipient identifiers, transaction dates, serial number, *etc.,* directly into digital contents with an imperceptible form for human audio/visual systems, and should be satisfied the following requirements: The embedded watermark does not spoil the quality of the original contents and should not be perceptilbe. It must be difficult for an attacker to remove the watermark and should be robust to common signal processing and geometric distortions, such as digital-to-analog and analog-to-digital conversion, filtering, lossy compression, re-sampling, scaling, and cropping, *etc.*

There are mainly two methods of the digital watermark technology for still images. One is embedding in the spatial domain. The other is embedding in the frequency domain. It is generally said that embedding in the frequency domain is more tolerant of attacks and image processing than in the spatial domain. Thus, most of recently proposed methods embed the watermark into the spectral coefficients of images by using the signal transformation such as the discrete cosine transformation (DCT) and the discrete wavelet transformation (DWT).

An impressive amount of watermarking algorithms besed on the DCT or DWT have been produced in the recent contributions [1]. However, in most watermarking methods, no theoretical limits to their robustness against attacks and image processing have not been given, and the watermarking methods that are robust to almost common image processing have not appear yet. On the contrary, the watermark-removal softwares, such as StirMark [2] and unZign [3], heve succeeded in washing the watermark away for most of watermarking systems. Such a situation is quite discouraging but will foster new research in this field, that is, the analysis and evaluation of the performance and robustness of watermarking systems and the development of watermarking systems with the desired performance and robustness. Therefore, as the first stage in the development and improvement of watermarking technology, it is important to establish fundamentals of measuring and evaluating the performance of watermarking systems.

In this paper, we concentrate on the watermarking of still images and first present the general model of a watermarking method in the frequency domain. Then, using this model, we propose a statistical method of measuring the performance and robustness of the watermarking system, which can be formulated as a statistical hypothesis test. We apply the proposed method to the DCT-based watermarking method and carry out the analysis of the watermarking system and the evaluation of its performance.

2. Watermarking in the Frequency Domain

In Figure 1, we have illustrated in block diagram form the general model of the watermark embedding and extracting processes of a watermarking method in the frequency domain.

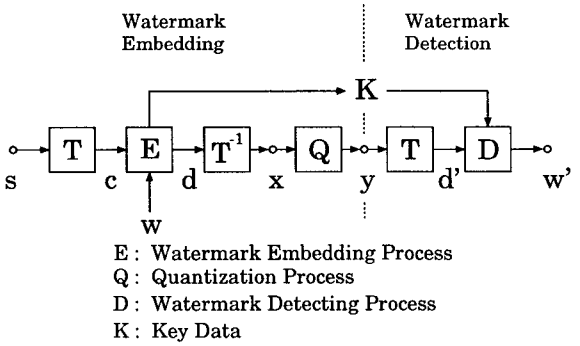


Figure 1 : General model of a watermarking method in the frequency domain.

In the watermark embedding process, an (original) image $s = [s(m,n)]$, where $s(m,n)$ denotes a pixel quantized to 256 levels (represented by 8 bits), is first converted into

a spectral coefficient $c = [c(m,n)]$ by using the singal transformation T such as the DCT or DWT as

$$c = Ts. \tag{1}$$

Then, a watermark $w = \{w_i,\ 1 \le i \le B\} \in W$, where w_i is a binary data, *i.e.*, $w_i = 0$ or 1 and W denotes the set of watermarks, is embedded into the spectral coefficient c as

$$d = E(c, w, k), \tag{2}$$

where E represents an embedding function, $d = [d(m,n)]$ is the watermarked spectral coefficient, and k denotes a set of parameters such as the embedded intensity and the information of location where data are embedded (address of embedded data), *etc.*, which is used as key data in the watermark detection. By the inverse transformation T^{-1} of the d, the watermarked image $x = [x(m,n)]$ is obtained as

$$x = T^{-1}d \tag{3}$$

and the pixels $x(m,n)$ are quantized to 256 levels (8 bits). As the result, we have the (quantized) watermarked image $y = [y(m,n)]$. It is noted that y is represented as

$$y = \text{Quantization}[x] = x + \delta, \tag{4}$$

where $\delta = [\delta(m,n)]$ denotes the quantization error whose elements are independent and probability law has a uniform probability density over the interval $[-0.5, 0.5)$.

It is necessary to decide the k so that images may not be degraded through the watermark embedding. If the condition that PSNR between y and s is greater than 40 dB is required of the watermarked image y, then we need to select the k out of elements of the set

$$K_0 = \{k \mid \text{PSNR}(y, s) \ge 40[\text{dB}]\} \tag{5}$$

In the watermark extracting, we transform the watermarked image $z = [z(m,n)]$ into the spectral coefficient $d' = [d'(m,n)]$ by the transformation T as

$$d' = Tz. \tag{6}$$

Then, using d' and k, the watermark $w' = \{w_i',\ 1 \le i \le B\}$ is extracted by

$$w' = D(d', k), \tag{7}$$

where D denotes an extracting function.

As the condition that when $z = y$,

$$w' = D(d', k) = D(Ty, k) = w \tag{8}$$

is required of the watermarking methods in the frequency domain, the performance of the watermark detector D is measured by the probability of getting a correct estimate w_i' of the i-th watermark w_i from the watermarked image y, *i.e.*,

$$P_i = \text{Prob}\{w_i' = w_i \mid s, w, k\}, \tag{9}$$

defined as the probability conditioned to given s, w, and k. Here, we consider the set of parameters (key data), for a given original image s,

$$K_{P_i} = \{k \mid P_i \simeq 1 \text{ for all } w \in W\} \tag{10}$$

and we put

$$K_P = \bigcap_{i=1}^{B} K_{P_i}. \tag{11}$$

If $k \in K_P$, then we have $w' = w$, that is, the watermark is extracted correctly by using the watermark detector D. Hence, taking Eq. (5) into consideration, we can see that in the watermarking system described above, the k should be set so as to be $k \in K_0 \cap K_P$.

Next, we consider the robustness of the watermarking system against common image processing. Let f be an image operator that represents a certain image processing, and let $z = f(y)$, where y is a watermarked image. Then, the robustness of the watermarking system to the image processing f is measured by the probability that the detector D correctly detects the i-th watermark w_i from the transformed image z *i.e.*,

$$Q_i = \text{Prob}\{w_i' = w_i \mid s, w, k, f\}, \tag{12}$$

defined as the probability conditioned to given s, w, k, and f. The measure of the robustness by means of Q_i is as follows:

(1) Let $k \in K_0 \cap K_P$ be the parameter (key data) of the watermarking system. Then, for given s and f, we can get the correct estimate w_i' of the w_i with probability $Q_i^* = E_W[Q_i]$, where $E_W[\cdot]$ means the ensemble average with respect to $w \in W$.

(2) Let $\varepsilon > 0$ be given and let's consider the set of parameters, for s and f,

$$K_{Q_i} = \{k \mid Q_i > 1 - \varepsilon \quad \text{for all } w \in W\}. \tag{13}$$

If the $k \in K_0 \cap K_P$ of the watermarking system is in K_{Q_i}, then the probability of getting $w_i' = w_i$ is greater than $(1 - \varepsilon)$. Furthermore, we can realize the watermarking system that is robust against the image processing f provided that $k \in K_0 \cap K_P \cap K_Q$, where

$$K_Q = \bigcap_{i=1}^{B} K_{Q_i}. \tag{14}$$

3. On the Robustness of the DCT-Based Watermarking Mehtod for Still Images

In this section, we focus on the DCT-based watermarking method, in which the watermark is embedded into DCT coefficients of an image, properly selected, by using a controlled quantization process, and we evaluate its performance using the measure presented in Section 2.

The watermark embedding and extracting processes of the DCT-based watermarking system are described below.

[Watermark Embedding Process]

Let $c = [c(m,n)]$ be DCT coefficients of an image $s = [s(m,n)]$, *i.e.*, $c = Ts$, and let $k = \{Q, I_B\}$ be the parameter (key data), where Q is the quantization step size, called the embedded intensity, and I_B is an operator that selects B DCT coefficients from the c. We also define an operator $J_B = I - I_B$, I being the identity operator.

Then, the watermark embedding process is as follows:
(1) $c_1 = I_B c, \quad c_2 = J_B c$.
(2) Let w_i $(1 \leq i \leq B)$ be the watermark and let $c(m_i, n_i)$ $(1 \leq i \leq B)$ be the elements of the c_1. Then, the watermark w_i is embedded into $c(m_i, n_i)$ by modifying (quantizing) $c(m_i, n_i)$, that is, let $c(m_i, n_i) \in [2lQ, (2l+1)Q)$, then

$$c'(m_i, n_i) = \begin{cases} 2lQ & \text{for} \quad w_i = 0 \\ (2l+1)Q & \text{for} \quad w_i = 1 \end{cases} \tag{15}$$

and let $c(m_i, n_i) \in [(2l-1)Q, 2lQ)$, then

$$c'(m_i, n_i) = \begin{cases} 2lQ & \text{for} \quad w_i = 0 \\ (2l-1)Q & \text{for} \quad w_i = 1 \end{cases} \tag{16}$$

Thus, we have the modified coefficients $c_1' = [c'(m_i, n_i), 1 \leq i \leq B]$.
(3) By the inverse DCT (IDCT) of the modified DCT coefficients $d = c_1' + c_2$, we obtain the watermarked image $x = [x(m, n)]$, that is, $x = T^{-1}d$.

[Watermark Extracting Process]

Let $d' = [d'(m, n)]$ be DCT coefficients of the watermarked image $z = [z(m, n)]$, *i.e.*, $d' = Tz$, and let $k = \{Q, I_B\}$ be the key data used in the watermark embedding process. Then, we have the watermark extracting process as:
(1) $d_1' = I_B d$
(2) The watermark w_i' $(1 \leq i \leq B)$ are extracted from the elements $d'(m_i, n_i)$ $(1 \leq i \leq B)$ of the d_1' as follows: If $\mathrm{int}[d'(m_i, n_i)/Q]$ is an even number, then $w_i = 0$, else $w_i = 1$, where the function $\mathrm{int}[\cdot]$ stands for the round-off operation.

Next, we calculate the probability P_i and Q_i of the DCT-based watermarking system.

[Probability P_i]

From $z = y = x + \delta$, we have $d' = Tz = d + T\delta$, and we put $e = d' - d = T\delta$ and $e_1 = I_B e = [e(m_i, n_i), 1 \leq i \leq B]$. As $w_i' = w_i$ provided that $\mid e(m_i, n_i) \mid < Q/2$, P_i is given by

$$P_i = \mathrm{Prob}\{\mid e(m_i, n_i) \mid < Q/2 \mid s, Q, I_B\} \tag{17}$$

It is noted that P_i is independent of $W = \{w\}$.

[Probability Q_i]

The d can be represented as $d = c + \Delta c$, where Δc is decomposed into $\Delta c_1 = I_B \Delta c$ and $\Delta c_2 = J_B \Delta c$, and they are expressed, respectively, as follows.
(a) Let $c_1 = I_B c = [c(m_i, n_i), 1 \leq i \leq B]$ and $\Delta c_1 = I_B \Delta c = [\Delta c(m_i, n_i), 1 \leq i \leq B]$, then, when $c(m_i, n_i) \in [2lQ, (2l+1)Q)$,

$$\Delta c(m_i, n_i) = \begin{cases} c(m_i, n_i) - 2lQ & \text{for} \quad w_i = 0 \\ (2l+1)Q - c(m_i, n_i) & \text{for} \quad w_i = 1 \end{cases} \tag{18}$$

when $c(m_i, n_i) \in [(2l-1)Q, 2lQ)$,

$$\Delta c(m_i, n_i) = \begin{cases} 2lQ - c(m_i, n_i) & \text{for} \quad w_i = 0 \\ c(m_i, n_i) - (2l-1)Q & \text{for} \quad w_i = 1 \end{cases} \tag{19}$$

(b) $\Delta c_2 = J_B c = 0$ because $J_B d = J_B c$.
Hence, we have

$$\begin{aligned} y &= T^{-1}d + \delta = T^{-1}(c + \Delta c) + \delta \\ &= s + (T^{-1}\Delta c + \delta), \end{aligned} \tag{20}$$

$$\begin{aligned} z &= f(y) = f(s + (T^{-1}\Delta c + \delta)) \\ &\simeq f(s) + F(s)(T^{-1}\Delta c + \delta) \end{aligned} \tag{21}$$

where

$$F(s) = \left[\left. \frac{\partial f(m, n)}{\partial y(k, l)} \right|_{y=s} \right] \tag{22}$$

and

$$d' \simeq Tf(s) + TF(s)(T^{-1}\Delta c + \delta). \tag{23}$$

Let $e = d' - d$ and $e_1 = I_B e = [e(m_i, n_i), 1 \leq i \leq B]$. Then $w_i' = w_i$ holds if $\mid e(m_i, n_i) \mid < Q/2$. Therefore, Q_i is estimated by

$$Q_i = \mathrm{Prob}\{\mid e(m_i, n_i) \mid < Q/2 \mid s, w, Q, I_B, f\} \tag{24}$$

where

$$\begin{aligned} e &\simeq T(f(s) - s) + T(F(s) - I)T^{-1}\Delta c \\ &+ TF(s)\delta \end{aligned} \tag{25}$$

It is noted that in the case of linear transformation, $z = Fy$, Eq. (25) can be written as

$$\begin{aligned} e &= T(F - I)s + T(F - I)T^{-1}\Delta c \\ &+ TF\delta. \end{aligned} \tag{26}$$

We have measured the performance of the DCT-based watermarking system using the image LENNA with the size of 128×128 pixels (Figure 2(a)). First, by estimating the probability P_i, it has been shown that the parameter $k = \{Q, I_B\} \in K_0 \cap K_P$ when we set $Q = 20$, $B = 100$ and $c_1 = I_B c = [c(m, n), 25 \leq m, n \leq 34]$. Figure 2(b) shows the watermarked image and, as an example, the probability density of $e(28, 28)$ is illustrated in Figure 3, from which we can see that the probability $P_{(28,28)} \simeq 1$.

(a) (b)

Figure 2 : (a) Original image s (LENNA). (b) The watermarked image y in which data of 100 bits are hidden. The root mean square (RMS) of $y - s$ is 0.98.

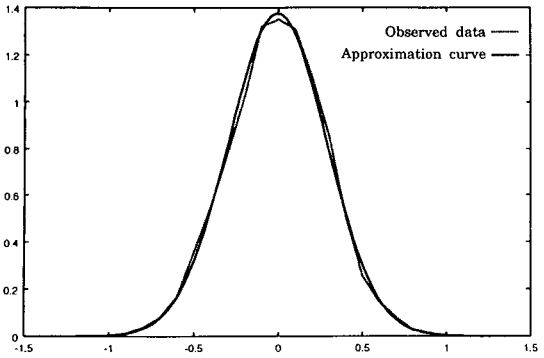


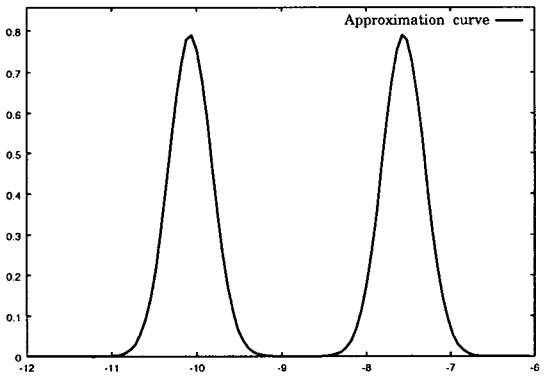
Figure 3 : Measure of probability P_i : Probability density of $e(28,28)$, which is approximated by the Gaussian probability density function with the mean $m = -5.70 \times 10^{-3}$ and the variance $\sigma^2 = 8.40 \times 10^{-2}$.

Next, we have evaluated the robustness of the watermarking system with the above key data k against the mean filter (linear filter)

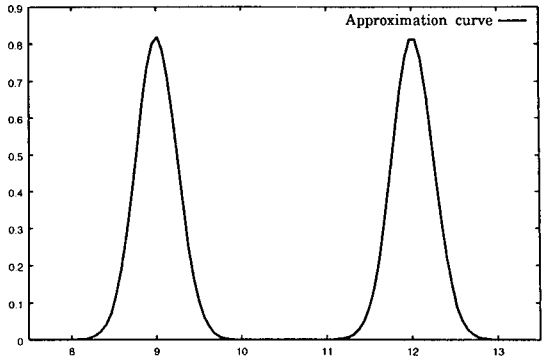
$$z(m,n) = \sum_{k,l=-1}^{1} f(k,l)\, y(m-k, n-l) \qquad (27)$$

where $f(k,l) = 0.60$ for $(k,l) = (0,0)$
$f(k,l) = 0.05$ for $(k,l) \neq (0,0)$

In this case, the root mean square (RMS) of $z - x$ is 6.48. In order to measure the robustness of the watermark embedded into the DCT coefficients $c(28,28)$ and $c(30,32)$, we have estimated the probability density of the corresponding $e(28,28)$ and $e(30,32)$ for almost all $w \in W$ and computed the probability $Q^*_{(28,28)}$ and $Q^*_{(30,32)}$. Figure 4 (a) and (b) show the probability density of $e(28,28)$ and $e(30,32)$, respectively, from which we have $Q^*_{(28,28)} = 0.69$ and $Q^*_{(30,32)} = 0.50$. Hence, we can see that the probability of getting the correct estimate of the watermark embedded into $c(28,28)$ and $c(30,32)$ is 0.69 and 0.50, respectively.



(a)



(b)

Figure 4 : Measure of probability Q_i : Probability density of (a) $e(28,28)$ and (b) $e(30,32)$. From numerical experiments, they are approximated, respectively, by the function

$$g(e) = \frac{a_1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{(e-m_1)^2}{2\sigma_1^2}\right\} + \frac{a_2}{\sqrt{2\pi}\sigma_2}\exp\left\{-\frac{(e-m_2)^2}{2\sigma_2^2}\right\}$$

where (a) $a_1 = a_2 = 0.50$, $m_1 = -10.1$, $m_2 = -7.56$, $\sigma_1^2 = \sigma_2^2 = 6.38 \times 10^{-2}$, and (b) $a_1 = a_2 = 0.50$, $m_1 = 9.01$, $m_2 = 12.0$, $\sigma_1^2 = \sigma_2^2 = 5.95 \times 10^{-2}$.

Remark : In the case that f is a linear shift-invariant filter, it has been proved that the second term in the right-hand side of Eq.(26), which we put as $\Delta c' = T(F-I)T^{-1}\Delta c$, can be expressed in the form of $\Delta c' = [\alpha(m,n)\Delta c(m,n)]$. While, for $c(m,n)$ in which the watermark is embedded, $\Delta c(m,n)$ is given by Eqs.(18) and (19). These implicitly imply that two peaks appear in the provability density of $e(28,28)$ and $e(30,32)$.

4. Concluding Remarks

Using the proposed method, we can evaluate the robustness of the DCT-based watermarking system against other image processing, *e.g.,* lossy compression such as JPEG, additive noise, reduction of grayscale level, and scaling. The proposed method can also be applied to other watermarking systems in the frequency domain such as the DWT-based watermarking system. We believe that based on the results of the evaluation of the performance of watermarking systems, we can develop and improve the watermarking technology. These results will be reported in forthcoming papers.

References

[1] B. Macq (*Guest Editor*), Special Issue on Identification and Protection of Multimedia Information, Proc. of the IEEE, Vol.87, No.7, July 1999.

[2] StirMark: http://www.cl.cam.ac.uk/~fapp2/ watermarking /stirmark

[3] unZign: http://www.altern.com/watermark