Analysis and Improvement of Correlation-Based Watermarking Methods for Digital Images

Akio MIYAZAKI and Akihiro OKAMOTO
Graduate School of Information Science and Electrical Engineering, Kyushu University
6-10-1, Hakozaki, Fukuoka, 812-8581 JAPAN

ABSTRACT

This paper aims to design robust correlation-based watermarking methods for images. We first present a model of watermark embedding and extracting processes and carry out their analyses. Then we examine the robustness of the watermarking method against image processing and clarify the reason why detection errors occur in the watermark extracting process. Based on the result, we improve the watermark extracting process and design robust watermarking systems. The improvement is accomplished using the deconvolution technique. Numerical experiments using the DCT-based watermarking system show good performance as expected by us.

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 attention as a new method of protecting copyrights of digital contents. A 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 satisfy the following requirements: The embedded watermark should not spoil the quality of the original contents and should not be perceptible. It should be difficult for an attacker to remove the watermark and should be robust to signal processing and geometric distortions.

A significant number of watermarking methods have been recently reported[1],[2]. Most of these methods embed the watermark into the spectral coefficients of images by using signal transformation such as discrete cosine transformation (DCT) or discrete wavelet transformation (DWT) because embedding in the frequency domain is more tolerant to attacks and image processing than embedding in the spatial domain. However, theoretical limits to their robustness against attacks and image processing have not been given, and the watermarking methods that are robust to almost all image processing have not been reported yet. On the contrary, watermark-removal software such as StirMark[3] has succeeded in washing the watermark away for most of the known watermarking systems. Such a situation is quite discouraging but will foster new research in this field such as the analysis of the performance of watermarking systems and the development of watermarking systems with the desired robustness. Therefore, as the first stage in the development of watermarking technology, it is important to analyze the watermark embedding and extracting processes and apply the result to the design of robust watermarking systems.

In this paper, we concentrate on the watermarking of still images and deal with the correlation-baed watermarking method[2] in which a watermark code (pseudo-random sequence) is perceptually weighted and added to the spectral coefficient, and is detected by computing the correlation between the watermarked coefficient and the watermark code to be checked for the presence. We first analyze the watermark embedding and extracting processes and then investigate the robustness of the watermarking system against attacks and image processing. It is clarified how distortion of the watermark occurs by image processing and the reason why the distortion causes detection errors in the watermark extracting procedure. Based on the result, we furthermore carry out an improvement of the watermark extracting process by designing a deconvolution filter that attempts to undo the effect of watermark distortion. The design is achieved using the deconvolution technique with adaptive algorithms. Numerical experiments with the DCT-based watermarking system reveal the desired performance is achieved.

2 Analysis of Correlation-Based Watermarking Methods

In correlation-based watermarking methods, we first prepare a set of watermarks $W = \{\boldsymbol{w}_l, 1 \leq l \leq L\}$ where $\boldsymbol{w}_l = [w_l(k)]$ (B-dimensional vector) and the elements $w_l(k)$'s are random sequences of real number

drawn from the Gaussian distribution with zero-mean and variance $\sigma^2 = 1$. The watermark embedding and extracting processes are as follows.

In the watermark embedding process, an (original) image $\boldsymbol{s} = [s(n)]$ (N-dimensional vector)[1], where $s(n)$ denotes a pixel quantized to 256 levels (represented by 8 bits), is first converted into a spectral coefficient $\boldsymbol{c} = [c(n)]$ by using signal transformation such as DCT or DWT as

$$\boldsymbol{c} = \boldsymbol{T}\boldsymbol{s}, \tag{1}$$

where $\boldsymbol{T}$ is an $N \times N$ matrix created from the transfrom kernels.

Next, we select B spectral coefficients $c(i_1)$, $c(i_2)$, $\cdots, c(i_B)$ and one watermark $\boldsymbol{w} = [w(k)]$ from $\boldsymbol{c}$ and W, respectively, and embed $w(k)$ into $c(i_k)$ in the form of

$$c'(i_k) = c(i_k) + \Delta_k w(k), \tag{2}$$

where Δ_k is a positive number called embedded intensity and is usually determined by exploiting the properties of the frequency model for the human visual system (HVS) so that one can make watermarks that have higher energy perceptually invisible.

Then, the watermarked coefficient $\boldsymbol{d} = [d(n)]$ is defined as

$$d(n) = \begin{cases} c'(i_k) & , \quad n = i_k \ (1 \le k \le B) \\ c(n) & , \quad \text{otherwise} \ . \end{cases} \tag{3}$$

By defining the $N \times B$ matrix $\boldsymbol{E}_M = [e_M(m,n)]$ where $e_M(m,n) = \Delta_k$ for $(m,n) = (i_k, k)$; $= 0$, otherwise, and $1 \le k \le B$, $\boldsymbol{d}$ can be written as

$$\boldsymbol{d} = \boldsymbol{c} + \boldsymbol{E}_M \boldsymbol{w}. \tag{4}$$

By the inverse transform $\boldsymbol{T}^{-1}$ of $\boldsymbol{d}$, the watermarked image $\boldsymbol{x} = [x(n)]$ is obtained as

$$\boldsymbol{x} = \boldsymbol{T}^{-1}\boldsymbol{d} \ . \tag{5}$$

It is necessary to determine $\boldsymbol{E}_M$ so that images may not be degraded through watermark embedding[2]. From $\boldsymbol{E}_M$, we construct the $B \times N$ matrix $\boldsymbol{E}_X = [e_X(m,n)]$ where $e_X(m,n) = 1$ for $(m,n) = (k, i_k)$; $= 0$, otherwise, and $1 \le k \le B$. The set of parameters $K = \{\boldsymbol{E}_X, W\}$ is saved and used as key data in watermark detection.

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

$$\boldsymbol{d}' = \boldsymbol{T}\boldsymbol{z}, \tag{6}$$

and obtain the watermarked coefficients $\boldsymbol{u}$ as

$$\boldsymbol{u} = \boldsymbol{E}_X \boldsymbol{d}'. \tag{7}$$

Then, we calculate the correlation $r_W(l)$ $(1 \le l \le L)$ between $\boldsymbol{u}$ and all $\boldsymbol{w}_l \in W$ as

$$r_W(l) = < \boldsymbol{u}, \boldsymbol{w}_l > = \sum_{k=1}^{B} u(k) w_l(k). \tag{8}$$

Then we search for $\boldsymbol{w}^*$ that maximizes $r_W(l)$'s, that is,

$$\boldsymbol{w}^* = \arg \max_{\boldsymbol{w}_l \in W} \ r_W(l) \tag{9}$$

and determine that there exists $\boldsymbol{w}^*$ in the watermarked image.

3 Robust Watermarking Systems

We consider the robustness of the watermarking system against image processing. Let $\boldsymbol{f}$ be an image operator that represents a certain image processing, and let

$$\boldsymbol{z} = \boldsymbol{f}(\boldsymbol{x}) = [\ f_1(\boldsymbol{x}),\ f_2(\boldsymbol{x}),\ \cdots,\ f_N(\boldsymbol{x})\ ]_t, \tag{10}$$

where $\boldsymbol{x}$ is a watermarked image. Then, from Eqs. (4) and (5), we have

$$\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{T}^{-1}\boldsymbol{E}_M \boldsymbol{w}. \tag{11}$$

Since watermarked images are not degraded through watermark embedding, *i.e.*, $\|\boldsymbol{s}\| \gg \| \boldsymbol{T}^{-1}\boldsymbol{E}_M \boldsymbol{w} \|$, $\| \cdot \|$ being the norm of vectors,

$$\begin{aligned} \boldsymbol{z} &= \boldsymbol{f}(\boldsymbol{s} + \boldsymbol{T}^{-1}\boldsymbol{E}_M \boldsymbol{w}) \\ &\simeq \boldsymbol{f}(\boldsymbol{s}) + \boldsymbol{F}(\boldsymbol{s})(\boldsymbol{T}^{-1}\boldsymbol{E}_M \boldsymbol{w}) \end{aligned} \tag{12}$$

is obtained, where $f_m = f_m(\boldsymbol{x})$, $x_n = x(n)$ and

$$\boldsymbol{F}(\boldsymbol{s}) = [f_{m,n}(\boldsymbol{s})] = \left[\left. \frac{\partial f_m}{\partial x_n} \right|_{\boldsymbol{x}=\boldsymbol{s}} \right]. \tag{13}$$

Hence, from Eqs. (6) and (7), distortion of the watermarked coefficient $\boldsymbol{u}$ occurs as

$$\boldsymbol{u} = \boldsymbol{E}_X \boldsymbol{T} \boldsymbol{f}(\boldsymbol{s}) + \boldsymbol{E}_X \boldsymbol{T} \boldsymbol{F}(\boldsymbol{s}) \boldsymbol{T}^{-1} \boldsymbol{E}_M \boldsymbol{w}, \tag{14}$$

and detection errors arise in the watermark extracting procedure. It is noted that in the case of linear transformation, $\boldsymbol{z} = \boldsymbol{F}\boldsymbol{x}$, Eq. (14) can be written as

$$\boldsymbol{u} = \boldsymbol{E}_X \boldsymbol{T} \boldsymbol{F} \boldsymbol{s} + \boldsymbol{E}_X \boldsymbol{T} \boldsymbol{F} \boldsymbol{T}^{-1} \boldsymbol{E}_M \boldsymbol{w}. \tag{15}$$

We can see from Eq. (14) that the distortion of the watermarked coefficients $\boldsymbol{u}$ depends on not only the image operator $\boldsymbol{f}$ but also the original image $\boldsymbol{s}$, and the watermark $\boldsymbol{w}$ is convolved with the filter, incorporating the effect of distortion, whose impulse response is described by the elements of the $B \times B$ matrix $\boldsymbol{E}_X \boldsymbol{T} \boldsymbol{F}(\boldsymbol{s}) \boldsymbol{T}^{-1} \boldsymbol{E}_M$,

[1] An image is usually denoted by an $M \times M$ array $s = [s(m,n)]$. In this paper, for simplicity of description, putting $N = M^2$, we map $s = [s(m,n)]$ into a vector $s = [s(n)]$ of size N by row ordering.

[2] We can discuss the relationship between B embedded intensity Δ_k $(1 \le k \le B)$ and the watermarked image quality when the orthogonal transform T is used. Since $\Delta c(i_k) = c'(i_k) - c(i_k) = \Delta_k w(k)$ is the random variable with zero-mean and variance Δ_k^2, the MSE of the original and watermarked images is given by $P_n = \frac{1}{N_x N_y} \sum_{k=1}^{B} \Delta_k^2$ where we put the size of images as $N_x \times N_y$.

and furthermore the bias $\boldsymbol{E}_X\boldsymbol{T}\boldsymbol{f}(\boldsymbol{s})$ corrupts the watermark $\boldsymbol{w}$. This implies that we can design a deconvolution filter that attempts to undo the effects of the convolution filter and the bias as follows: From Eq. (14), putting $\boldsymbol{H} = (\boldsymbol{E}_X\boldsymbol{T}\boldsymbol{F}(\boldsymbol{s})\boldsymbol{T}^{-1}\boldsymbol{E}_M)^{-1}$ and $\boldsymbol{r} = -\boldsymbol{H}\boldsymbol{E}_X\boldsymbol{T}\boldsymbol{f}(\boldsymbol{s})$, we have the improved watermark extracting procedure

$$\left.\begin{array}{ll} \text{(i)} & \boldsymbol{u} = \boldsymbol{E}_X\boldsymbol{d}' \\ \text{(ii)} & \boldsymbol{v} = \boldsymbol{H}\,\boldsymbol{u} + \boldsymbol{r} \\ \text{(iii)} & r'_W(l) =< \boldsymbol{v}, \boldsymbol{w}_l > \\ & \boldsymbol{w}^* = \arg \max_{\boldsymbol{w}_l \in W} \; r'_W(l)\ . \end{array}\right\} \qquad (16)$$

4 Improvement of Watermark Extracting Process using Deconvolution Technique

In correlation-based watermarking methods, the parameter $\boldsymbol{E}_M$ is decided in consideration of the quality of watermarked images and the robustness of watermarks against basic signal processing like JPEG compression and filtering. Hence there exist some image processing techniques that destroy the watermark in image processing and attacking tools as Photoshop and StirMark. Moreover, even if the watermark is robust against a certain image processing tool at present, it may be destroyed by a new version of this tool in the future. In such a situation, a present solution is considered to be the redesign of the watermarking system to revise the watermark detector for the image processing $\boldsymbol{f}$ to which the system is not robust. Using the result is Section 3, we can carry out the revision of the watermark detector provided that we can access the original image $\boldsymbol{s}$ and get a set of input-output pairs $\{x(k), z(k)\}$ from $\boldsymbol{z} = \boldsymbol{f}(\boldsymbol{x})$. Then the revised watermark detector may be designed by application of the adaptive signal processing technique[3] because the deconvolution filter can be represented as a system described by a matrix equation and training data, *i.e.,* a set of pairs of watermark $\{w(k)\}$ and watermarked coefficient $\{u(k)\}$, can be obtained from a set of $\{x(k), z(k)\}$ of $\boldsymbol{f}$. That is, by training the system, it will automatically produce an adequate deconvolution filter in accordance with $\boldsymbol{f}$ and $\boldsymbol{s}$. The design procedure is shown briefly as follows:

We first construct the deconvolution filter ((ii) in Eq. (16)) by

$$v(k) = \sum_{l=1}^{B} h(k,l)u(l) + r(k) \quad (1 \le k \le B)\ . \qquad (17)$$

Then we train the system in order to undo the effects of the convolution filter and the bias in Eq. (14). The objective of the training process is that the weights $h(k,l)$'s and $r(k)$'s are set to the optimum values by minimizing the square of error $e(k) = w(k) - v(k)$ between various training watermarks $\{w(k)\}$ and the outputs $\{v(k)\}$ of the system. As is well known, a method for setting the weights to these values is the LMS algorithm[4], which is an iterative gradient algorithm. Training take place during many trials or runs until the weights converge to the optimum values, that is, the system (Eq. (17)) learns the characteristic of the deconvolution filter.

[3]The use of the revised watermark detector is as follows: If we cannot extract the waternark with the ordinary watermark detector (Eqs. (6)–(9)), then we utilize the revised watermark detector (Eq. (16)), with which the watermark extraction may be carried out on a trial-and-error basis for the image processings to which the system is not robust.

5 Numerical Experiments

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, and try to improve the watermark extracting process using the design technique described in Section 4. The improvement will be verified from numerical experiments using the image LENNA with the size of 128 × 128 pixels (Figure 1 (a)).

(a) (b)

Figure 1 : (a) Original image $\boldsymbol{s}$ (LENNA). (b) Watermarked image $\boldsymbol{x}$.

The watermark embedding procedure given in this experiment is based on Refs.[5] and [6]. The parameters Δ_k, B, L, and $\boldsymbol{E}_M$ are set as follows: $\Delta_k = \lambda|c(i_k)|$ ($\lambda = 0.8$), $B = 324$, $L = 1000$, and $c(i_1)$, $c(i_2)$, $\cdots$, $c(i_{99})$, $c(i_{324})$ are equal, respectively, to $c(50,50)$, $c(50,51)$, $\cdots$, $c(67,66)$, $c(67,67)$ in the 128 × 128 DCT coefficient array. (This selection of DCT coefficients is denoted by $\boldsymbol{E}_M$). Figure 1 (b) shows the watermarked image, whose MSE and PSNR are 3.18 and 43.1 [dB], respectively. Figure 2 shows the response $r_W(l)$ ($1 \le l \le L$) of the watermark detector D to the watermarked image into which we embed the 500th watermark in L randomly generated watermarks. It can be seen that one spike clearly indicates the presence of the 500th watermark embedded in this image.

Next we examine the robustness of the above watermarked image against JPEG compression with standard quality and 'StirMark' random geometric distortions from StirMark 3.1. The results are illustrated in Figures 3–5. We can see from Figures 3 and 4 that the correct (500th) watermark is detected from the JPEG compression image of standard quality, but the watermark detection from the 'StirMark' attacked image ends in failure. Thus we design the deconvolution filter in the

case of 'StirMark' random geometric distortions. Training data are 300 pairs of watermark $\{w(k)\}$ and watermarked coefficient $\{u(k)\}$, which are different from 1000 watermarks used in testing. The initial value of $h(k,l)$'s and $r(k)$'s are, respectively, $h^{(0)}(k,l) = 1$ for $k = l$; $= 0$ for $k \neq l$, and $r^{(0)}(k) = 0$. We show the result of the watermark detection with the deconvolution filter in Figure 5. As the result, the response $r'_W(l)$ has the peak whose location corresponds to that of the correct (500th) watermark. Hence the watermark extracting process is improved as expected.

6 Concluding Remarks

Using the proposed technique, we can revise the watermark detector of the DCT-based watermarking system against other image operator $\boldsymbol{f}$, e.g., filtering, sharpening, scaling, and so on. The design method can also be applied to the watermarking system using another signal transformation method such as the DWT-based watermarking method. We believe that based on the results we can develop and improve the watermarking technology. These results will be reported in forthcoming papers.

In the case that the image operator $\boldsymbol{f}$ is not known or a set of input-output pairs $\{\boldsymbol{x}, \boldsymbol{z}\}$ of $\boldsymbol{f}$ is not obtained, the design problem can be formulated as a kind of blind deconvolution problem, that is, the problem of finding $\boldsymbol{H}$, $\boldsymbol{r}$, and $\boldsymbol{v}$ in Eq. (16) from processed or attacked watermarked image(s) $\boldsymbol{z}$. It seems that this problem is difficult because the distortion (convolution) model of Eq. (14) is generally a shift-variant system[7]. The authors would like to make an attempt to solve this problem.

References

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

[2] G. C. Langelaar, I. Setyawan, and R. L. Lagendijk, "Watermarking Digital Image and Video Data," IEEE Signal Processing Magazine, vol.17, no.5, pp.20-46, Sep. 2000.

[3] StirMark: http:// www.cl.cam.ac.uk/ ˜fapp2/ watermarking/ stirmark

[4] B. Widrow and S. D. Stearns, Adaptive Signal Processing, Prentice-Hall, Englewood Cliffs, New Jersey, 1985.

[5] A. Piva, M. Barni, F. Bartoline, and V. Cappellini, "DCT-Based Watermark Recovering without Resorting to the Uncoruppted Original Image," Proc. IEEE Int. Conf. Image Processing, vol.1, pp.520-523, Oct. 1997.

[6] H. Inoue, A. Miyazaki, and T. Katsura, "An Image Watermarking Method based on the Wavelet Transform," Proc. IEEE Int. Conf. Image Processing, vol.1, pp.296-300, Oct 1999.

[7] R. Liu and L. Tong (*Guest Editor*), Special Issue on Blind System Identification and Estimation, Proc. IEEE, vol.86, no.10, Oct. 1998.

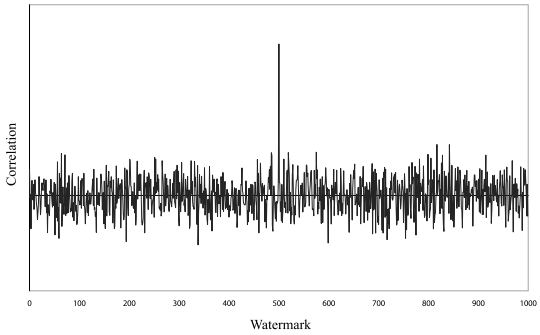

Figure 2 : Response of the watermark detector to the watermarked image.

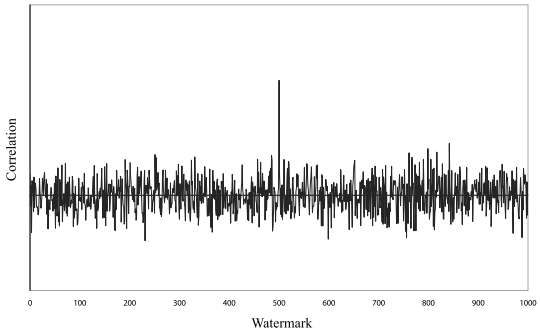

Figure 3 : Response of the watermark detector to the JPEG compression image of standard quality

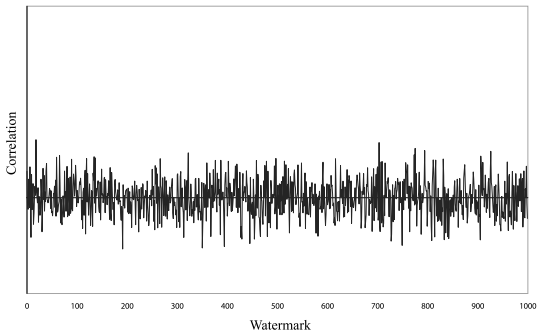

Figure 4 : Response of the watermark detector to the 'StirMark' attacked image.

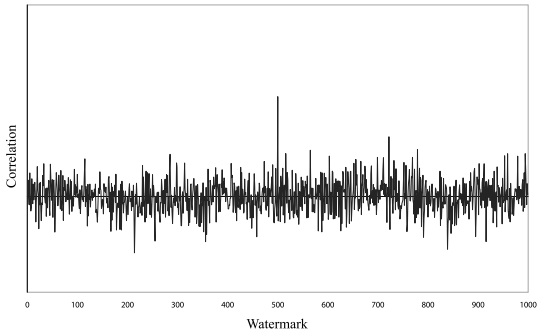

Figure 5 : Response of the watermark detector using the deconvolution filter to the 'StirMark' attacked image.