A SOLUTION TO THE WATERMARK DETECTION PROBLEM BASED ON BAYESIAN ESTIMATION AND EM ALGORITHM

Akio Miyazaki
Faculty of Information Science, Kyushu Sangyo University, Fukuoka 813-8503, Japan
phone: +81-92-673-5421, fax: +81-92-673-5454, email: miyazaki@is.kyusan-u.ac.jp

ABSTRACT

We have recently discussed digital watermarking techniques based on modifying the spectral coefficient of an image, and have presented a model of the watermark embedding and extracting processes and a model of watermark distortion caused by image processing and attack of watermarked images. In this paper, based on these models, we formulate the watermark detection problem as a kind of blind deconvolution problem. Then, as a solution to this problem, we propose a watermark detection method using Bayesian estimation and EM algorithm.

1. INTRODUCTION

The digital watermark technology is now drawing the attention as one of the useful methods of protecting copyright and security of digital contents. Digital watermark is realized by embedding information data directly into digital contents with an imperceptible form for human audio/visual systems, and should satisfy the following requirements: The embedded watermark does not spoil the quality of the original contents and should not be perceptible. It must be difficult for an attacker to remove the watermark and should be robust to common signal processing and geometric distortions.

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 [1] [2] embed the watermark into the spectral coefficients of images by using the signal transformation such as the discrete cosine transformation (DCT) or the discrete wavelet transformation (DWT). Some of them have been incorporated into signal and image processing software.

In the development of watermark technology, it is important to analyze the watermark embedding and extracting processes and apply the result to the design of robust watermarking systems. As an approach to the analysis and design, communication-based models of watermarking, in which watermarking is in essence a form of communication, that is, a watermark (message) is communicated from the embedding system to the extracting one, are well known [3]–[7]. Using these models, we can clarify how distortion of the watermark occurs by image processing and the reason why the distortion causes detection errors in the watermark extracting procedure, and discuss the performance of the watermarking systems from the view point of quality of the watermarked image, quantity of the embedded watermark data, and robustness of the watermarking system. In the several papers like [4]–[6], the discussion mainly focuses on the correlation-based watermarking method in which a watermark (pseudorandom sequence) is detected by computing detection statistics such as the correlation of the embedded and extracted sequences, the performance of the watermarking method is evaluated and the robust watermarking method is developed with the probability of detection errors calculated from the detection statistics.

In this paper, as another method using these models, we investigate the design of robust watermarking systems from a viewpoint of detecting a watermark with accuracy by revising the watermark extracting process adaptively in accordance with attacks and image processing to watermarked images. From this investigation, we can expect that a watermark is detected accurately even in the case that watermarked images suffer damage from attacks and image processing against which the watermarking system is not robust at the present time or which threaten the robustness of the watermarking system in the future. We first formulate the watermark detection problem as a kind of blind deconvolution problem using a watermark distortion model presented in [7]. Then, as a solution to this problem, we propose a watermark detection method using Bayesian estimation. In the case that the distortion model is not known but its statistical characteristic like its mean and variance is evaluated or estimated, we present an algorithm for watermark detection using EM algorithm. It is demonstrated from numerical experiments with the DWT-based watermarking system that the proposed watermark detection method shows good improvement of the rate of watermark detection as wished by us.

2. MODEL OF WATERMARKING SYSTEMS IN THE FREQUENCY DOMAIN

Figure 1 illustrates the model of a watermarking system in the frequency domain in block diagram form. We explain briefly the watermark embedding and extracting processes in Fig. 1. Here the symbols $\boldsymbol{T}$ and $\boldsymbol{T}^{-1}$ denotes the image transformation such as DCT or DWT and the inverse transformation such as IDCT or IDWT, respectively.

(1) Watermark embedding process

$\boldsymbol{x} = [x(m)]$ is a host (original) image, for simplicity of description, denoted by an M-dimensional vector. An M-dimensional vector $\boldsymbol{c} = [c(m)]$ is a spectral coefficient obtained by the transformation of $\boldsymbol{x}$, that is, $\boldsymbol{c} = \boldsymbol{T}\boldsymbol{x}$. $\boldsymbol{w} = [w(k)]$ is a K-dimensional watermark vector (embedded data vector), where $K < M$ and $w(k)$ is a binary digit. K spectral coefficients $c(i_k)$ $(1 \leq k \leq K)$ are selected adequately from $\boldsymbol{c}$. Then $w(k)$ is embedded into $c(i_k)$ by a controlled thresholding or quantization process and the watermarked spectral coefficient $\boldsymbol{d} = [d(m)]$ (M-dimensional vector) is created. By the inverse transformation of $\boldsymbol{d}$, the watermarked image $\boldsymbol{y} = [y(m)]$ is obtained, that is, $\boldsymbol{y} = \boldsymbol{T}^{-1}\boldsymbol{d}$. Here some of parameters used in watermark embedding and the set $P = \{i_k, 1 \leq k \leq K\}$, whose element stands for the position of the watermarked coefficients, are saved and used as key data in watermark detection.

(2) Watermark extracting process

Let an M-dimensional vector $\boldsymbol{z} = [z(m)]$ be the watermarked image. If $\boldsymbol{z}$ does not suffer from attacks and image processing, then it is represented as $\boldsymbol{z} = \boldsymbol{y}$, else as $\boldsymbol{z} = \boldsymbol{f}(\boldsymbol{y}) + \boldsymbol{n}$, in which $\boldsymbol{f}$ is an image operator that denotes a certain attack and image processing and $\boldsymbol{n}$ is an additive noise vector. $\boldsymbol{z}$ is transformed into the spectral coefficient $\boldsymbol{d}' = [d'(m)]$ by $\boldsymbol{d}' = \boldsymbol{T}\boldsymbol{z}$. Then the watermarked coefficient $c'(i_k)$ $(1 \leq k \leq K)$ is extracted from $\boldsymbol{d}'$ and the watermark $\boldsymbol{w}'$ is detected from $\{c'(i_k)\}$ by using key data.

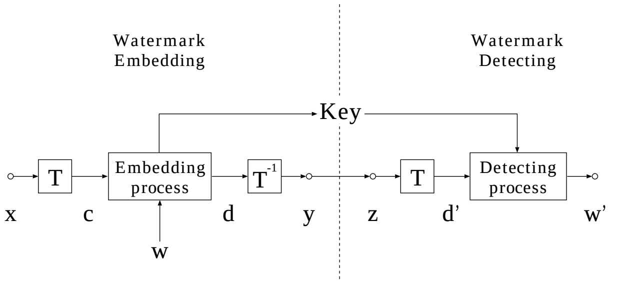


Fig. 1. Model of a watermarking system in the frequency domain.

We have analyzed the watermark embedding and extracting processes and investigated how the watermark is distorted by attacks and image processing. As a result, we have indicated that the watermark $\boldsymbol{w}$ is distorted in the form of

$$\boldsymbol{w}' = \boldsymbol{A}\boldsymbol{w} + \boldsymbol{b} \tag{1}$$

under the condition that the watermarked image is not degraded through the watermark embedding, and the attack and image processing to the watermarked image, i.e.,

$$\| \boldsymbol{y} - \boldsymbol{x} \| \ll \| \boldsymbol{x} \| \quad \text{and} \quad \| \boldsymbol{y} - \boldsymbol{z} \| \ll \| \boldsymbol{y} \|,$$

$\| \cdot \|$ being the norm of vectors.

Putting $N = K + 1$, defining $\boldsymbol{w}$ afresh as $\boldsymbol{w} = [\ w(1)\ w(2)\ \cdots\ w(K)\ 1\ ]^t$ (N-dimensional vector), and putting $\boldsymbol{u} = [\ w'(1)\ w'(2)\ \cdots\ w'(K)\ 0\ ]^t$ (N-dimensional vector) and

$$\boldsymbol{H} = \left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{b} \\ \hline \boldsymbol{0} & 0 \end{array}\right] \quad (N \times N \text{ matrix}) , \tag{2}$$

$\boldsymbol{0}$ being K-dimensional zero vector and the suffix t denoting the transposition of matrices, the distortion model, Eq.(1) can be written as

$$\boldsymbol{u} = \boldsymbol{H}\boldsymbol{w} \tag{3}$$

Here it is noted that $\boldsymbol{H} = [h_{ij}]$, that is, $\boldsymbol{A}$ ($K \times K$ matrix) and $\boldsymbol{b}$ (K-dimensional vector) depends on the host image $\boldsymbol{x}$, the attack $\boldsymbol{f}$, and the noise $\boldsymbol{n}$. (These are explained in [7] in detail.)

3. WATERMARK DETECTION BASED ON BAYESIAN ESTIMATION

The problem of watermark detection is formulated as that of estimating the watermark $\boldsymbol{w}$ embedded into a host image $\boldsymbol{x}$ from an observation $\boldsymbol{r}$ of the distorted watermark $\boldsymbol{u}$. In order to minimize the error of watermark detection in consideration of the distortion model described by Eq.(3), we solve the problem with the following strategy. We first compute the probability density function (pdf) $p(\boldsymbol{s}|\boldsymbol{r})$ for a candidate $\boldsymbol{s}$ of $\boldsymbol{w}$ conditioned on $\boldsymbol{r}$, and then find $\boldsymbol{s}$ that maximizes $p(\boldsymbol{s}|\boldsymbol{r})$. By Bayes formula:

$$p(\boldsymbol{s}|\boldsymbol{r}) = \frac{p(\boldsymbol{r}|\boldsymbol{s})p(\boldsymbol{s})}{p(\boldsymbol{r})} \tag{4}$$

the maximization of the conditional pdf $p(\boldsymbol{s}|\boldsymbol{r})$ with respect to $\boldsymbol{s}$ for given $\boldsymbol{r}$ is equivalent to that of $p(\boldsymbol{r}|\boldsymbol{s})p(\boldsymbol{s})$. If $p(\boldsymbol{s})$ is uniform, then the maximum likelihood estimation that maximizes the likelihood function $p(\boldsymbol{r}|\boldsymbol{s})$ or the log-likelihood function $\log p(\boldsymbol{r}|\boldsymbol{s})$ is used.

We first consider the case that the distortion model $\boldsymbol{H} = [h_{ij}]$ is known or estimated. If the pdf of $\boldsymbol{r}$ and $\boldsymbol{s}$ is the Gaussian pdf, then

$$\begin{aligned} p(\boldsymbol{r}|\boldsymbol{s}) &= p(\boldsymbol{r}|\boldsymbol{H}, \boldsymbol{s}) \\ &= \frac{1}{\sqrt{(\pi\sigma^2)^N}}\ e^{-\frac{1}{\sigma^2}(\boldsymbol{r}-\boldsymbol{H}\boldsymbol{s})^t(\boldsymbol{r}-\boldsymbol{H}\boldsymbol{s})} \end{aligned} \tag{5}$$

$$\log p(\boldsymbol{r}|\boldsymbol{s}) \doteq -\frac{1}{\sigma^2}(\boldsymbol{r} - \boldsymbol{H}\boldsymbol{s})^t(\boldsymbol{r} - \boldsymbol{H}\boldsymbol{s}) \tag{6}$$

is obtained and the estimation of watermark is carried out by maximizing the above equations with respect to $\boldsymbol{s}$ for given $\boldsymbol{r}$. Here the symbol $\doteq$ denotes equivalent for optimization purposes. Therefore the estimation problem results in the deconvolution problem of finding $\boldsymbol{s}$ such that $\boldsymbol{H}\boldsymbol{s} = \boldsymbol{r}$ for given $\boldsymbol{r}$.

The above procedure is applied to the case that the image operator $\boldsymbol{f}$ is known or estimated and we can access the host image $\boldsymbol{x}$ [7]. The reason is the following. We get a set of input-output pairs $\{\boldsymbol{y}, \boldsymbol{z}\}$ from $\boldsymbol{x}$ and $\boldsymbol{z} = \boldsymbol{f}(\boldsymbol{y})$. Form the set of $\{\boldsymbol{y}, \boldsymbol{z}\}$, we create a set of pairs of watermark $\boldsymbol{w}$ and distorted watermark $\boldsymbol{u}$, which utilizing as training data, we can estimate $\boldsymbol{H}$ corresponding to $\boldsymbol{f}$ and its inverse (pseudoinverse) and then obtain the watermark estimation $\boldsymbol{s}$.

In the case that $\boldsymbol{x}$ or $\boldsymbol{f}$ is unknown, the deconvolution problem becomes a kind of blind one. In the following, we treat this blind deconvolution problem. We consider the case that the distortion model $\boldsymbol{H} = [h_{ij}]$ is unknown but statistics of $\boldsymbol{H}$, like mean and variance of $\boldsymbol{H}$, are evaluated, and present a solution to this problem using the EM (Expectation-Maximization) algorithm[8], which consists of two steps called E-step (Expectation Step) and M-step (Maximization Step) and repeats these steps until convergence.

When the distortion model $\boldsymbol{H}$ is unknown, the dependence of $p(\boldsymbol{s}|\boldsymbol{r})$ on the random distortion model is

$$p(\boldsymbol{s}|\boldsymbol{r}) = E[p(\boldsymbol{s},\boldsymbol{H}|\boldsymbol{r})] = \int_{\boldsymbol{H}} p(\boldsymbol{s},\boldsymbol{H}|\boldsymbol{r})d\boldsymbol{H} \tag{7}$$

where

$$p(\boldsymbol{s},\boldsymbol{H}|\boldsymbol{r}) = \frac{p(\boldsymbol{r},\boldsymbol{H}|\boldsymbol{s})p(\boldsymbol{s})}{p(\boldsymbol{r})} \tag{8}$$

As the observation data is only $\boldsymbol{r}$ and incomplete data, that is, $\boldsymbol{H}$ is the hidden data, we cannot directly find out $\boldsymbol{s}$ that maximizes $p(\boldsymbol{s},\boldsymbol{H}|\boldsymbol{r})$ or $p(\boldsymbol{r},\boldsymbol{H}|\boldsymbol{s})$. Accordingly, we first represent $p(\boldsymbol{r},\boldsymbol{H}|\boldsymbol{s})$ as

$$p(\boldsymbol{r},\boldsymbol{H}|\boldsymbol{s}) = p(\boldsymbol{H}|\boldsymbol{r},\boldsymbol{s})p(\boldsymbol{r}|\boldsymbol{s}) \tag{9}$$

$$\log p(\boldsymbol{r}|\boldsymbol{s}) = \log p(\boldsymbol{r},\boldsymbol{H}|\boldsymbol{s}) - \log p(\boldsymbol{H}|\boldsymbol{r},\boldsymbol{s}) \tag{10}$$

and then calculate the conditional expectation of Eq.(10) with respect to $\boldsymbol{H}$ conditioned on an observation $\boldsymbol{u} = \boldsymbol{r}$ and an watermark estimation $\boldsymbol{w} = \hat{\boldsymbol{s}}$.

$$\begin{aligned}\log p(\boldsymbol{r}|\boldsymbol{s}) &= E[\log p(\boldsymbol{r},\boldsymbol{H}|\boldsymbol{s})|\boldsymbol{r},\hat{\boldsymbol{s}}] \\ &- E[\log p(\boldsymbol{H}|\boldsymbol{r},\boldsymbol{s})|\boldsymbol{r},\hat{\boldsymbol{s}}] \end{aligned} \tag{11}$$

When we define

$$L(\boldsymbol{s}) = \log p(\boldsymbol{r}|\boldsymbol{s}) \tag{12}$$

$$Q(\boldsymbol{s}|\hat{\boldsymbol{s}}) = E[\log p(\boldsymbol{r},\boldsymbol{H}|\boldsymbol{s})|\boldsymbol{r},\hat{\boldsymbol{s}}] \tag{13}$$

$$R(\boldsymbol{s}|\hat{\boldsymbol{s}}) = E[\log p(\boldsymbol{H}|\boldsymbol{r},\boldsymbol{s})|\boldsymbol{r},\hat{\boldsymbol{s}}] \tag{14}$$

then Eq.(11) can be written as

$$L(\boldsymbol{s}) = Q(\boldsymbol{s}|\hat{\boldsymbol{s}}) - R(\boldsymbol{s}|\hat{\boldsymbol{s}}) \tag{15}$$

It is noted that we would like to find $\boldsymbol{s}$ to maximize the log-likelihood $L(\boldsymbol{s})$ for $\boldsymbol{r}$, but we do not have the complete data to compute $L(\boldsymbol{s})$. So instead, we maximize $Q(\boldsymbol{s}|\hat{\boldsymbol{s}}) - R(\boldsymbol{s}|\hat{\boldsymbol{s}})$ given the observation $\boldsymbol{r}$ and our current estimate $\hat{\boldsymbol{s}}$. Here if $Q(\boldsymbol{s}|\hat{\boldsymbol{s}}) > Q(\hat{\boldsymbol{s}}|\hat{\boldsymbol{s}})$, then $L(\boldsymbol{s}) > L(\hat{\boldsymbol{s}})$ because $R(\boldsymbol{s}|\hat{\boldsymbol{s}}) \leq R(\hat{\boldsymbol{s}}|\hat{\boldsymbol{s}})$ holds according to Jensen's inequality. This is the basic idea behind the EM algorithm, from which we have the following algorithm expressed in two steps.
(a) E-step : Compute

$$\begin{aligned} Q(\boldsymbol{s}|\boldsymbol{s}^{(k)}) &= E[\log p(\boldsymbol{r},\boldsymbol{H}|\boldsymbol{s})|\boldsymbol{r},\boldsymbol{s}^{(k)}] \\ &= \int_{\boldsymbol{H}} [\log p(\boldsymbol{r},\boldsymbol{H}|\boldsymbol{s})] p(\boldsymbol{H}|\boldsymbol{r},\boldsymbol{s}^{(k)})d\boldsymbol{H} \end{aligned} \tag{16}$$

for the observation $\boldsymbol{u} = \boldsymbol{r}$ and the watermark estimate $\boldsymbol{w} = \boldsymbol{s}^{(k)}$ at the k-th iteration.
(b) M-step : Let $\boldsymbol{s}^{(k+1)}$ be the value of $\boldsymbol{s}$ which maximize $Q(\boldsymbol{s}|\boldsymbol{s}^{(k)})$, that is, find

$$\boldsymbol{s}^{(k+1)} = \arg\max_{\boldsymbol{s}} Q(\boldsymbol{s}|\boldsymbol{s}^{(k)}) \tag{17}$$

The EM algorithm consists of choosing $\boldsymbol{s}^{(0)}$, then performing the E-step and the M-step successively until convergence. Convergence may be determined by examining when the parameters quit changing, *i.e.*, stop.

In order to present the watermark estimation algorithm, we compute $Q(\boldsymbol{s}|\boldsymbol{s}^{(k)})$ of Eq.(16). We suppose that the pdf of $\boldsymbol{r}$, $\boldsymbol{s}$, $\boldsymbol{s}^{(k)}$, and $\boldsymbol{H}$ is the Gaussian pdf and compute $Q(\boldsymbol{s}|\boldsymbol{s}^{(k)})$. Let the mean and the variance of the elements h_{ij}'s of $\boldsymbol{H}$ be $E[h_{ij}] = e_{ij}$ and $E[(h_{ij} - e_{ij})(h_{kl} - e_{kl})] = \sigma_{ij}^2$ $(i = k$ and $j = l)$; $= 0$ (otherwise), respectively, and let the pdf of $\boldsymbol{H}$ be the Gaussian pdf

$$p(\boldsymbol{H}) = \frac{1}{\sqrt{\pi^{N^2} \prod_{i=1}^{N} |\boldsymbol{\Lambda_i}|}} \prod_{i=1}^{N} e^{-(\boldsymbol{h_i}-\boldsymbol{e_i})^t \boldsymbol{\Lambda_i}^{-1} (\boldsymbol{h_i}-\boldsymbol{e_i})} \tag{18}$$

in which $\boldsymbol{h_i} = [h_{ij}]$ (column vector), $\boldsymbol{e_i} = [e_{ij}]$ (column vector), and $\boldsymbol{\Lambda_i} = E[(\boldsymbol{h_i} - \boldsymbol{e_i})(\boldsymbol{h_i} - \boldsymbol{e_i})^t]$ (diagonal matrix).

Under the assumption that $\boldsymbol{H}$ and $\boldsymbol{s}$ are statistically independent mutually, the conditional pdf $p(\boldsymbol{r},\boldsymbol{H}|\boldsymbol{s})$ can be written as

$$\begin{aligned} p(\boldsymbol{r},\boldsymbol{H}|\boldsymbol{s}) &= \frac{p(\boldsymbol{r},\boldsymbol{H},\boldsymbol{s})}{p(\boldsymbol{s})} = \frac{p(\boldsymbol{r}|\boldsymbol{H},\boldsymbol{s})p(\boldsymbol{H},\boldsymbol{s})}{p(\boldsymbol{s})} \\ &= \frac{p(\boldsymbol{r}|\boldsymbol{H},\boldsymbol{s})p(\boldsymbol{H})p(\boldsymbol{s})}{p(\boldsymbol{s})} = p(\boldsymbol{r}|\boldsymbol{H},\boldsymbol{s})p(\boldsymbol{H}) \end{aligned} \tag{19}$$

while the conditional pdf $p(\boldsymbol{H}|\boldsymbol{r},\boldsymbol{s})$ can be expressed as

$$\begin{aligned} p(\boldsymbol{H}|\boldsymbol{r},\boldsymbol{s}) &= \frac{p(\boldsymbol{r}|\boldsymbol{H},\boldsymbol{s})p(\boldsymbol{H},\boldsymbol{s})}{p(\boldsymbol{r},\boldsymbol{s})} \\ &= \frac{p(\boldsymbol{r}|\boldsymbol{H},\boldsymbol{s})p(\boldsymbol{H})p(\boldsymbol{s})}{p(\boldsymbol{r},\boldsymbol{s})} = \frac{p(\boldsymbol{r}|\boldsymbol{H},\boldsymbol{s})p(\boldsymbol{H})}{p(\boldsymbol{r}|\boldsymbol{s})} \end{aligned} \tag{20}$$

From Eq.(19) and Eq.(20), $Q(\boldsymbol{s}|\boldsymbol{s}^{(k)})$ can be written as

$$\begin{aligned} Q(\boldsymbol{s}|\boldsymbol{s}^{(k)}) &\doteq \int_{\boldsymbol{H}} [\log p(\boldsymbol{r}|\boldsymbol{H},\boldsymbol{s})] p(\boldsymbol{r}|\boldsymbol{H},\boldsymbol{s}^{(k)})p(\boldsymbol{H})d\boldsymbol{H} \\ &+ \int_{\boldsymbol{H}} [\log p(\boldsymbol{H})] p(\boldsymbol{r}|\boldsymbol{H},\boldsymbol{s}^{(k)})p(\boldsymbol{H})d\boldsymbol{H} \end{aligned} \tag{21}$$

As the second term of the right-hand side of Eq.(21) is independent of $\boldsymbol{s}$ and is eliminated in M-step, we have

$$Q(\boldsymbol{s}|\boldsymbol{s}^{(k)}) \doteq \int_{\boldsymbol{H}} [\log p(\boldsymbol{r}|\boldsymbol{H},\boldsymbol{s})] p(\boldsymbol{r}|\boldsymbol{H},\boldsymbol{s}^{(k)})p(\boldsymbol{H})d\boldsymbol{H} \tag{22}$$

Using Eq.(5) and Eq.(18), we obtain

$$\begin{aligned} p(\boldsymbol{r}|\boldsymbol{H},\boldsymbol{s})p(\boldsymbol{H}) &= \frac{1}{\sqrt{\pi^{N^2} \prod_{l=1}^{N} |\boldsymbol{\Delta_l}|}} \\ &\times \prod_{l=1}^{N} e^{-(\boldsymbol{h_l}-\boldsymbol{d_l})^t \boldsymbol{\Delta_l}^{-1} (\boldsymbol{h_l}-\boldsymbol{d_l})} \end{aligned} \tag{23}$$

where

$$\boldsymbol{d_l} = E[\boldsymbol{h_l}|\boldsymbol{r},\boldsymbol{s}] = \boldsymbol{\Delta_l} \left(\frac{r_l}{\sigma^2}\boldsymbol{s} + \boldsymbol{\Lambda_l}^{-1}\boldsymbol{e_l} \right) \tag{24}$$

$$\boldsymbol{\Delta_l} = \left(\frac{1}{\sigma^2}\boldsymbol{s}\boldsymbol{s}^t + \boldsymbol{\Lambda_l}^{-1} \right)^{-1} \tag{25}$$

Substituting Eq.(5) and Eq.(23) into Eq.(22) and calculating $Q(\boldsymbol{s}|\boldsymbol{s}^{(k)})$, we have

$$\begin{aligned} Q(\boldsymbol{s}|\boldsymbol{s}^{(k)}) \;&\doteq\; -\sum_{l=1}^{N} r_l^2 + 2\boldsymbol{s}^t\left(\sum_{l=1}^{N} r_l \boldsymbol{d}_{\boldsymbol{l}}^{(\boldsymbol{k})}\right) \\ &-\; \boldsymbol{s}^t\left\{\sum_{l=1}^{N}\left(\boldsymbol{d}_{\boldsymbol{l}}^{(\boldsymbol{k})}\boldsymbol{d}_{\boldsymbol{l}}^{(\boldsymbol{k})^t} + \boldsymbol{\Delta_l}^{(k)}\right)\right\}\boldsymbol{s} \end{aligned} \tag{26}$$

where

$$\boldsymbol{d_l}^{(k)} = E[\boldsymbol{h_l}|\boldsymbol{r},\boldsymbol{s}^{(k)}] = \boldsymbol{\Delta_l}^{(k)}\left(\frac{r_l}{\sigma^2}\boldsymbol{s}^{(k)} + \boldsymbol{\Lambda_l}^{-1}\boldsymbol{e_l}\right) \tag{27}$$

$$\boldsymbol{\Delta_l}^{(k)} = \left(\frac{1}{\sigma^2}\boldsymbol{s}^{(k)}\boldsymbol{s}^{(k)t} + \boldsymbol{\Lambda_l}^{-1}\right)^{-1} \tag{28}$$

In Eq.(27), $\boldsymbol{d_l}^{(k)} = E[\boldsymbol{h_k}|\boldsymbol{r},\boldsymbol{s}^{(k)}]$ $(k = 1,2,\cdots,N)$ stands for the distortion model estimated from $\{\boldsymbol{r},\boldsymbol{s}^{(k)}\}$. Form these equations, we can obtain the updated value $\boldsymbol{s}^{(k+1)}$ of $\boldsymbol{s}$ that maximizes $Q(\boldsymbol{s}|\boldsymbol{s}^{(k)})$ of Eq.(26) by $\partial Q(\boldsymbol{s}|\boldsymbol{s}^{(k)})/\partial\boldsymbol{s} = 0$ where

$$\begin{aligned} \frac{\partial Q(\boldsymbol{s}|\boldsymbol{s}^{(k)})}{\partial\boldsymbol{s}} \;&=\; -2\left\{\sum_{l=1}^{N}\left(\boldsymbol{d}_{\boldsymbol{l}}^{(\boldsymbol{k})}\boldsymbol{d}_{\boldsymbol{l}}^{(\boldsymbol{k})^t} + \boldsymbol{\Delta_l}^{(k)}\right)\right\}\boldsymbol{s} \\ &+\; 2\left(\sum_{l=1}^{N} r_l \boldsymbol{d}_{\boldsymbol{l}}^{(\boldsymbol{k})}\right) \end{aligned} \tag{29}$$

From this, we have the procedure to estimate $\boldsymbol{s}$ from given $\boldsymbol{r}$ as follows.

(1) Set the parameters σ^2, $\boldsymbol{e_l}$, and $\boldsymbol{\Lambda_l}$. Put $k = 0$ and choose an initial $\boldsymbol{s} = \boldsymbol{s}^{(0)}$.

(2) Compute $\boldsymbol{\Delta_l}^{(k)}$ and $\boldsymbol{d_l}^{(k)}$ using

$$\boldsymbol{\Delta_l}^{(k)} = \left(\frac{1}{\sigma^2}\boldsymbol{s}^{(k)}\boldsymbol{s}^{(k)t} + \boldsymbol{\Lambda_l}^{-1}\right)^{-1} \tag{30}$$

$$\boldsymbol{d_l}^{(k)} = \boldsymbol{\Delta_l}^{(k)}\left(\frac{r_l}{\sigma^2}\boldsymbol{s}^{(k)} + \boldsymbol{\Lambda_l}^{-1}\boldsymbol{e_l}\right) \tag{31}$$

(3) Update the value of $\boldsymbol{s}$ using

$$\begin{aligned} \boldsymbol{s}^{(k+1)} \;&=\; \left\{\sum_{l=1}^{N}\left(\boldsymbol{d}_{\boldsymbol{l}}^{(\boldsymbol{k})}\boldsymbol{d}_{\boldsymbol{l}}^{(\boldsymbol{k})^t} + \boldsymbol{\Delta_l}^{(k)}\right)\right\}^{-1} \\ &\times\; \left(\sum_{l=1}^{N} r_l \boldsymbol{d}_{\boldsymbol{l}}^{(\boldsymbol{k})}\right) \end{aligned} \tag{32}$$

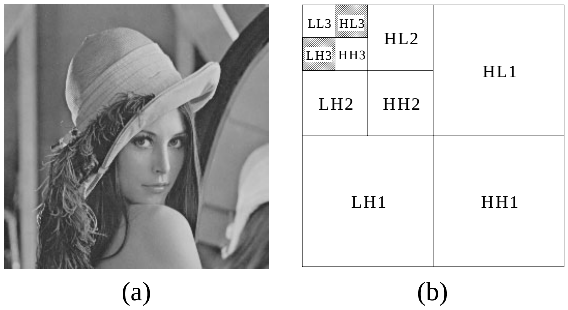
(4) If $\|\boldsymbol{s}^{(k+1)} - \boldsymbol{s}^{(k)}\| \geq \epsilon$ for some $\epsilon(> 0)$, then set $k := k+1$ and go to step (2), else put $\boldsymbol{s}^* = g(\boldsymbol{s}^{(k+1)})$ and decide that $\boldsymbol{s}^*$ is the embedded watermark. Here $g(\cdot)$ is an appropriate step function, by which we judge the watermark as a binary digit.

4. NUMERICAL EXPERIMENTS

In this section, we carry out numerical experiments using the DWT-based watermarking method [9], in which we will investigate the robustness of watermark against image processing and verify the improvement of watermark detection using the proposed method. In this experiment, we use the test image LENNA with 256 × 256 pixels and 8 bit/pixel shown in Fig.2 (a).

The watermark embedding process is the following. The image LENNA is decomposed into ten subbands for three scales by using the Daubechies wavelet with 8-tap as shown in Fig.2 (b). L DWT coefficients $c(i_k)$ $(1 \leq k \leq L)$ with smaller magnitude are selected from the elements of the multiresolution representation components HL3 and LH3, and then $c(i_k)$ is modified according to a data bit $b(k)$ as $c(i_k) = Q$ for $b(k) = 1$ or $c(i_k) = -Q$ for $b(k) = 0$.

In this experiment, putting $K = 4$, we prepare 4-dimensional watermark vector $\boldsymbol{w} = [w(k)]$ $(1 \leq k \leq 4)$ and make a 5-dimensional code vector $\boldsymbol{v} = [v(l)]$ $(1 \leq l \leq 5)$ consisting of four data bits $v(l) = w(l)$ $(1 \leq l \leq 4)$ and one parity-check bit $v(5)$. (Hence we have 16 code vectors $\boldsymbol{v_i}$ $(1 \leq i \leq 16)$.) Then by putting $L = 25$ and $Q = 10$, one code vector $\boldsymbol{v}$ is embedded into LENNA five times, that is, $b(5k + l) = v(l)$ $(0 \leq k \leq 4, 1 \leq l \leq 5)$. The watermarked image quality PSNR is 53.7 [dB].



(a) (b)

Fig. 2. (a) Test image LENNA. (b) Wavelet decomposition of image.

In the watermark extracting process, L DWT coefficients $c'(i_k)$ $(1 \leq k \leq L)$ are extracted from the watermarked and attacked LENNA and four data bits $v(l)$, *i.e.*, watermark $w(l)$ $(1 \leq l \leq 4)$ is detected from $c'(i_k)$'s as follows.

(I) The conventional watermark detection method

Checking the sign of $c'(i_k)$'s, we set $b'(k) = 1$ for $c'(i_k) \geq 0$ or $b'(k) = 0$ for $c'(i_k) < 0$. Next, for all l, we determine four data bits and one parity-check bit $v'(l)$ from $\{b'(5k + l), 0 \leq k \leq 4\}$ with decision by majority. If there is no error in $v'(l)$'s by parity-check, we decide that $v'(l)$ $(1 \leq l \leq 4)$ are embedded into LENNA, else we proceed to the next step of estimating data bits embedded into LENNA.

(II) The proposed watermark detection method

We compute the average $r(l)$ of $\{c'(i_{5k+l}) + Q\}/2Q$, $(0 \leq k \leq 4)$ for all l, i.e.,

$$r(l) = \frac{1}{10Q}\sum_{k=0}^{4}\{c'(i_{5k+l}) + Q\} \quad (0 \leq k \leq 4) \tag{33}$$

and set the observation data as $\boldsymbol{r} = [\ r(1)\ r(2)\ r(3)\ r(4)\ 0\ ]^t$. Next we estimate the embedded data bits from $\boldsymbol{r}$ using the proposed method with setting $N = 5$, in which the initial setting of parareters is as follows. Putting $\boldsymbol{\Lambda_i} = \sigma_\Lambda^2\boldsymbol{I}$ $(i = 1,2,\cdots,N)$ and $\sigma^2 = \sigma_\Lambda^2$, we set the value of σ as $\sigma =$

$\sigma^{(k)} = \|\boldsymbol{r} - \boldsymbol{s}^{(k)}\|/N$ using the observation $\boldsymbol{r}$ and the k-th estimate $\boldsymbol{s}^{(k)}$, that is, σ is updated in the iteration. The setting method is determined by the advance experiment. Since degradation of the watermarked image by attacks and image processing is small, $\boldsymbol{e_i} = [e_{ij}]$ is set as $e_{ii} = 1$ and $e_{ij} = 0$ for $i \neq j$, that is,

$$\boldsymbol{E} = E[\boldsymbol{H}] = \left[\begin{array}{c|c} \boldsymbol{I} & \boldsymbol{0} \\ \hline \boldsymbol{0} & 0 \end{array}\right] \quad (N \times N \text{ matrix}), \qquad (34)$$

$\boldsymbol{I}$ and $\boldsymbol{0}$ being the identity matrix and the zero vector, respectively. The initial estimate of $\boldsymbol{s}$ is $\boldsymbol{s}^{(0)} = [\ v'(1)\ v'(2)\ v'(3)\ v'(4)\ 1\ ]^t$, whose elements are $v'(l)$'s obtained in step (I).

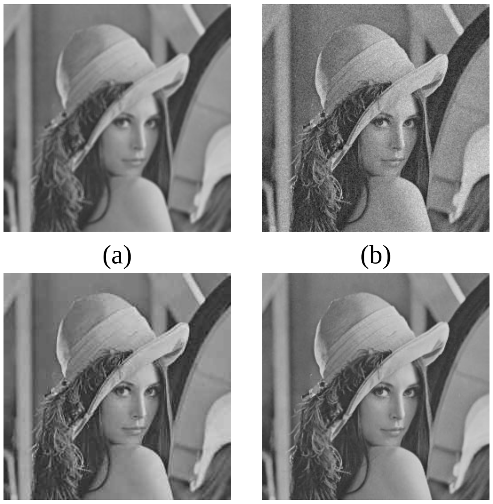
We first examine the robustness of the above watermarking system against smoothing, adding noise, and JPEG compression, in which watermark embedding and extracting are carried out for each of 16 code words $\boldsymbol{v_i}$'s.

We investigate the robustness by smoothing the watermarked LENNA with the mean filter

$$z(i,j) = \sum_{l,m=-1}^{1} \lambda_{l,m} y(i-l, j-m) \qquad (35)$$

where $\lambda_{l,m} = \eta$ for $(l,m) = (0,0)$; $= (1-\eta)/8$ for $(l,m) \neq (0,0)$ and $\eta = 0.1$. The smoothed image quality PSNR is 29.8[dB]. Figure 3 (a) shows the smoothed image, of which PSNR is 29.8[dB].

In the case of smoothing, we could detect all of data bits perfectly with the conventional method and did not need the proposed method.



(a) (b)

(c) (d)

Fig. 3. Watermarked images degraded and distorted through (a) smoothing, (b) adding noise, (c) JPEG compression, and (d) geometric distortion (StirMark)

We evaluate the robustness by adding Gaussian noise with the zero mean and the variance $\sigma^2 = 100$ to the watermarked LENNA. PSNR of the noisy image is 28.1[dB]. Figure 3 (b) illustrates the noisy image, whose PSNR is 28.1[dB].

In the case of adding noise, we could not detect 4 code words out of 16 ones with the conventional method. But applying the proposed method to these 4 code words, we could estimate the data bits correctly.

We test the robustness by compressing the watermarked LENNA using JPEG with parameter of 30 % quality. The compression image quality PSNR is 31.2[dB]. Figure 3 (c) shows the compression image, of which PSNR is 31.2[dB].

In the case of JPEG compression, we needed the estimation using the proposed method for half of 16 code words. As a result, the data bits could be estimated correctly, too.

We next examine the robustness of the above watermarking system against random geometric distortion in StirMark Attack [10]. The average PSNR of attacked images is 25.7 [dB]. The attacked image, whose PSNR is 25.7 [dB], is shown in Figure 3 (d). As is well known, this attack is strong and destroys almost all of data bits. Hence the conventional method ends in failure. So we prepare 20 observation data for one code word, that is, 320 ones in total, which cause detection error. Then we estimate the embedded data bits using the proposed method for each of 320 observation data. As a result, we could estimate the data bits for 310 observation data correctly, i.e., with accuracy of 97 %.

5. CONCLUSION

We have indicated that the watermark detection problem is formulated as a kind of blind deconvolution problem and proposed the watermark detection method using Bayesian estimation. Then we have considered the case that the distortion model of watermark is unknown but its statistics like mean and variance is evaluated or estimated, and presented the watermark estimation and decision algorithm using the EM algorithm. The effectiveness has been shown from the experiments using the DWT-based watermarking system.

The future work is to investigate the distortion model of watermark and develop how to estimate its statistics, and so on. These related problems will be discussed in forthcoming papers.

REFERENCES

[1] I. J. Cox, M.L.Miller, and J.A.Bloom, Digital Watermarking, p.542, Academic Press, 2002.

[2] M. Arnold, M.Schmucker, and S.D.Wolthusen, Techniques and Applications of Digital Watermarking and Content Protection, p.274, Artech House, Inc., 2003.

[3] I. J. Cox, et. al., "Watermarking as Communications with Side Information," Proc. IEEE, Vol.87, No.7, pp.1127-1141, 1999.

[4] M. Ramkumar and A. N. Akansu, "A Robust Oblivious Watermarking Scheme," Proc. IEEE ICIP, Vol.2, pp.61-64, Sep. 2000.

[5] M. Ramkumar and A. N. Akansu, "Capacity Estimates for Data Hiding in Compressed Images," IEEE Trans. Image Processing, Vol.10, No.8, pp.1252-1263, Aug. 2001.

[6] Q. Cheng and T. S. Huang, "Robust Optimum Detection of Transform Domain Multiplicative Watermarks," IEEE Trans. Signal Processing, Vol.51, No.3, pp.960-924, April 2003.

[7] A. Miyazaki, "Digital Watermarking for Images –Its Analysis and Improvement Using Digital Signal Processing Technique–," IEICE TRANS. FUNDAMENTALS, Vol.E85–A, No.3, pp.582–590, March 2002.

[8] G. J. McLachlan and T. Krishnan, The EM Algorithm and Extensions, p.274, John Wiley & Sons, Inc., 1997.

[9] H. Inoue, et. al., "A Digital Watermark Technique Based on the Wavelet Transform and its Robustness on Image Compression and Transformation," IEICE Trans. on Fundamentals, Vol.E82-A, No.1, pp.2-9, Jan. 1999.

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