Vol. 47 No. 8 IPSJ Journal Aug. 2006

Regular Paper

Use of Statistically Adaptive Accumulation to Improve Video Watermark Detection

Isao Echizen,† Yasuhiro Fujii,† Takaaki Yamada,†
Satoru Tezuka† and Hiroshi Yoshiura††

Redundant coding is a basic method of improving the reliability of detection and survivability after image processing. It embeds watermarks repeatedly in every frame or region and can thus prevent errors due to the accumulation of frames or regions during watermark detection. Redundant coding, however, is not always effective after image processing because the watermark signal may be attenuated by the accumulation procedure when image processing removes watermarks from specific frames or regions. We therefore propose a method of detection to prevent the attenuation of the watermark signal by accumulating a subset of the regions so that the accumulated region has a minimal bit-error rate, which is estimated from the region. Experimental evaluations using actual motion pictures have revealed that the new method can improve watermark survivability after MPEG encoding by an average of 15.7% and can widely be used in correlation-based watermarking.

1. Introduction

Digital video is being made available through various media such as the Internet, digital broadcasting, and DVD because of its advantages over analog content. It requires less space, is easier to process, and is not degraded by age or repeated use. A serious problem, however, is that the copyrights for digital videos are easily violated because they can easily be copied and sent illegally over the Internet. Video watermarking, which helps protect the copyrights for digital video by embedding copyright information, is therefore becoming important.

Video watermarking can be used to embed copyright and copy-control information in video frames and can therefore be used in DVD players and recorders as well as in digital broadcasting equipment such as set-top boxes [1),2)]. Video providers subject the watermarked pictures to various kinds of image processes–such as compression (using MPEG or other compression technology), resizing, and filtering–and these image processing procedures can also be exploited by illegal users who want to remove the embedded information. The watermarks (WMs) should nonetheless be robust enough to be reliably detected after any of these kinds of processes. WM survivability is thus essential and methods for improving it have been studied in both the pixel and frequency domains. Methods using redundant coding [3)∼5)] and spread spectrum coding [6),7)] of embedded information have been established, as have those for embedding information in perceptually significant parts of the pictures [8),9)] and in elements that will not be greatly affected by expected image processes [10),11)]. Using human visual models to embed more WMs without degrading picture quality [12),13)] have also been proposed.

Redundant coding is a basic method of improving reliability of detection and survivability after image processing. It embeds WMs repeatedly in every frame or region and can thus prevent errors by accumulating frames or regions coded repeatedly during WM detection. Redundant coding, however, is not always effective after image processing procedures that remove WMs from specific frames or regions because the WM signal may be attenuated by increasing image noise during the accumulation procedure. Detection methods should therefore select frames or regions where WMs remain and should accumulate these.

We thus propose a method of detection to prevent the attenuation of WM signals by accumulating a subset of regions so that the accumulated region has a minimal bit-error rate (BER). The BER is estimated from each watermarked region using inferential statistics. Section 2 of this paper describes one previous method and the problems with it. Section 3 describes our method, Section 4 reports experimental evaluations confirming that it can

† Systems Development Laboratory, Hitachi, Ltd.
†† Faculty of Electro-Communication, The University of Electro-Communications

withstand image processing, and Section 5 concludes the paper.

2. Previous Method and Problems

2.1 Previous Method

Redundant coding methods can be classified into two types:

(a) Redundant coding within a frame: WM embedding is done by dividing the frame into several regions and applying the same WM to each region. In detection, WMs are extracted by accumulating all the divided regions in a frame [3),4)].

(b) Redundant coding over frames: WM embedding is done by applying the same WM to each consecutive frame of a video. In detection, WMs are extracted by accumulating all the frames or a specific number of sequential frames [4),5)].

One basic WM scheme employing both (a) and (b) is treated in Kalker et al. [4)]. When M-bit information is embedded in a video, a WM-pattern image representing this information is added to each region of the video frames. When the information is detected, M-bit values are determined from M statistical values calculated by correlating the WM-pattern with the accumulated region. To simplify an explanation, we will first introduce a 1-bit-WM schema and then introduce a multiple-bit-WM schema.

2.1.1 One-bit-WM Schema

(1) WM embedding

The luminance set of the fth frame consisting of N pixels is $\mathbf{y}^{(f)} = \{y_i^{(f)} \mid 1 \le i \le N\}$. The process flow for 1-bit-WM embedding is described below (See **Fig. 1**):

Step E1: Do the following steps over $f = 1, 2, \ldots$.

Step E2: Input the original frame, $\mathbf{y}^{(f)}$, and divide $\mathbf{y}^{(f)}$ into R regions $\mathbf{y}^{(f,r)}$s ($r = 1, \ldots, R$) consisting of the corresponding pixels: $\mathbf{y}^{(f,r)} = \{y_i^{(f,r)} \mid 1 \le i \le \lfloor N/R \rfloor\}$, which satisfy $\mathbf{y}^{(f)} = \bigcup_r \mathbf{y}^{(f,r)}$, $\mathbf{y}^{(f,r)} \bigcap_{r \ne r'} \mathbf{y}^{(f,r')} = \emptyset$.

Step E3: Generate each watermarked region $\mathbf{y}'^{(f,r)}$ by adding the WM pattern, $\mathbf{m} = \{m_i \in \{-1, +1\} \mid 1 \le i \le \lfloor N/R \rfloor\}$, comprising a pseudo random array, ± 1 s, to the original region, $\mathbf{y}'^{(f,r)}$, according to the embedding bit, b:

$$\mathbf{y}'^{(f,r)} = \begin{cases} \mathbf{y}^{(f,r)} + \mu^{(f,r)}\mathbf{m} & \text{if } b{=}1 \\ \mathbf{y}^{(f,r)} - \mu^{(f,r)}\mathbf{m} & \text{if } b{=}0, \end{cases} \tag{1}$$

where $\mu^{(f,r)}$ is the WM strength of the region $\mathbf{y}^{(f,r)}$.

Step E4: Output the watermarked frame, $\mathbf{y}'^{(f)}$.

(2) WM detection

Calculate the statistical value, v, of the accumulated region by correlating the WM pattern, $\mathbf{m}$, with the accumulated region (See **Fig. 2**):

Step D1: Do the following steps over $f_0 = 1, F+1, 2F+1, \ldots$.

Step D2: Input F watermarked frames $\mathbf{y}'^{(f)}$s ($f = f_0, \ldots, f_0 + F - 1$) and divide FR regions $\mathbf{y}'^{(f,r)}$s ($f = f_0, \ldots, f_0 + F - 1, r = 1, \ldots, R$).

Step D3: Accumulate the FR regions, $\mathbf{y}'^{(f,r)}$s, in region $\tilde{\mathbf{y}} = \{\tilde{y}_i \mid 1 \le i \le \lfloor N/R \rfloor\}$. The $\tilde{y}_i$ for the accumulated region $\tilde{\mathbf{y}}$ is given by

$$\tilde{y}_i = \frac{1}{FR} \sum_{f=f_0}^{f_0+F-1} \sum_{r=1}^{R} y'^{(f,r)}_i. \tag{2}$$

Step D4: Calculate statistical value v, which is obtained by correlating WM pattern $\mathbf{m}$ with accumulated region $\tilde{\mathbf{y}}$. That is,

$$\begin{aligned} v &= \frac{1}{\lfloor N/R \rfloor} \sum_{i=1}^{\lfloor N/R \rfloor} m_i \tilde{y}_i \\ &= \frac{1}{FR\lfloor N/R \rfloor} \sum_{f,r,i} m_i y_i^{(f,r)} \pm \mu, \end{aligned} \tag{3}$$

where μ is the WM signal given by $\mu = 1/FR \sum_{f,r} \mu^{(f,r)}$. Since $m_i y_i^{(f,r)}$ is considered to be an independent stochastic variable with a mean of 0 [14)], each v fol-

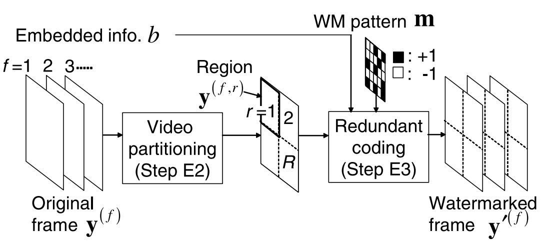


Fig. 1 Overview of WM embedding.

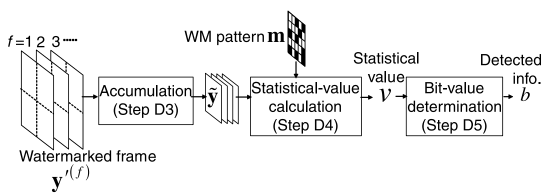


Fig. 2 Overview of WM detection.

lows a normal distribution with mean $\pm\mu$ and variance σ^2 if the number of $m_i y_i^{(f,r)}$s, $FR\lfloor N/R \rfloor$ is sufficiently large. That is,

$$v \sim \begin{cases} N(\mu, \sigma^2) & \text{if } b = 1 \\ N(-\mu, \sigma^2) & \text{if } b = 0. \end{cases} \qquad (4)$$

Step D5: Determine the embedded bit, b, by comparing v with a threshold value, T (> 0):

$$b = \begin{cases} 1 & \text{if } v \geq T \\ 0 & \text{if } v \leq -T \\ \text{“not detected”} & \text{if } -T < v < T. \end{cases} \qquad (5)$$

2.1.2 Multiple-bit-WM Schema

For M-bit-WM embedding ($M > 1$), each region, $\mathbf{y}^{(f,r)}$, is divided into M subregions, $\mathbf{y}_k^{(f,r)}$s ($k = 1, \ldots, M$), and the 1-bit embedding schema is applied to each subregion.

(1) **WM embedding**

The luminance set of the fth frame consisting of N pixels is $\mathbf{y}^{(f)} = \{y_i^{(f)} \mid 1 \leq i \leq N\}$. The process flow for M-bit-WM embedding is described below:

Step E1: Do the following steps over $f = 1, 2, \ldots$.

Step E2: Input the original frame, $\mathbf{y}^{(f)}$, and divide $\mathbf{y}^{(f)}$ into RM subregions, $\mathbf{y}_k^{(f,r)}$s ($r = 1, \ldots, R$, $k = 1, \ldots, M$), consisting of the corresponding pixels: $\mathbf{y}_k^{(f,r)} = \{y_{k,i}^{(f,r)} \mid 1 \leq i \leq \lfloor N/(RM) \rfloor\}$, which satisfies $\mathbf{y}^{(f)} = \bigcup_{r,k} \mathbf{y}_k^{(f,r)}$, $\mathbf{y}_k^{(f,r)} \bigcap_{r \neq r'} \mathbf{y}_k^{(f,r')} = \emptyset$, and $\mathbf{y}_k^{(f,r)} \bigcap_{k \neq k'} \mathbf{y}_{k'}^{(f,r)} = \emptyset$ (See example in **Fig. 3**).

Step E3: Generate each watermarked subregion, $\mathbf{y'}_k^{(f,r)}$, by adding the WM pattern, $\mathbf{m}_k = \{m_{k,i} \in \{-1, +1\} \mid 1 \leq i \leq \lfloor N/(RM) \rfloor\}$, comprising a pseudo random array, ±1 s, into the original region, $\mathbf{y'}_k^{(f,r)}$, according to the embedding bit, b_k:

$$\mathbf{y'}_k^{(f,r)} = \begin{cases} \mathbf{y}_k^{(f,r)} + \mu^{(f,r)} \mathbf{m}_k & \text{if } b_k = 1 \\ \mathbf{y}_k^{(f,r)} - \mu^{(f,r)} \mathbf{m}_k & \text{if } b_k = 0, \end{cases} \qquad (6)$$

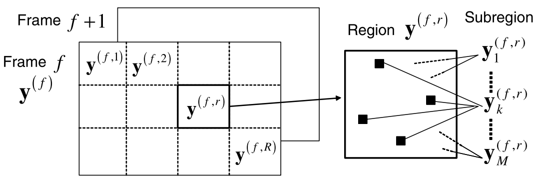


Fig. 3 Example of video partitioning.

Step E4: Output the watermarked frame, $\mathbf{y'}^{(f)}$.

(2) **WM detection**

Calculate the M statistical values, v_ks, of the accumulated region by correlating the WM pattern, $\mathbf{m}_k$, with the accumulated subregion:

Step D1: Do the following steps over $f_0 = 1, F+1, 2F+1, \ldots$.

Step D2: Input F watermarked frames, $\mathbf{y'}^{(f)}$s ($f = f_0, \ldots, f_0 + F - 1$), and divide FRM subregions, $\mathbf{y'}_k^{(f,r)}$s ($f = f_0, \ldots, f_0 + F - 1, r = 1, \ldots, R,\ k = 1, \ldots, M$).

Step D3:

Step D3: For each k ($k = 1, \ldots, M$), accumulate the FR subregions, $\mathbf{y'}_k^{(f,r)}$s, in the subregion, $\tilde{\mathbf{y}}_k = \{\tilde{y}_{k,i} \mid 1 \leq i \leq \lfloor N/(RM) \rfloor\}$. The $\tilde{y}_{k,i}$ of the accumulated subregion, $\tilde{\mathbf{y}}_k$, is given by

$$\tilde{y}_{k,i} = \frac{1}{FR} \sum_{f=f_0}^{f_0+F-1} \sum_{r=1}^{R} {y'}_{k,i}^{(f,r)}. \qquad (7)$$

Step D4: Calculate the set consisting of the M statistical values, $\mathbf{v} = \{v_k \mid 1 \leq k \leq M\}$. Statistical value v_k is obtained by correlating WM pattern $\mathbf{m}_k$ with accumulated subregion $\tilde{\mathbf{y}}_k$. That is,

$$\begin{aligned} v_k &= \frac{1}{\lfloor N/(RM) \rfloor} \sum_{i=1}^{\lfloor N/(RM) \rfloor} m_{k,i} \tilde{y}_{k,i} \\ &= \frac{1}{FR\lfloor N/(RM) \rfloor} \sum_{f,r,i} m_{k,i} y_{k,i}^{(f,r)} \pm \mu, \end{aligned} \qquad (8)$$

As in the 1-bit-WM schema, each v_k follows a normal distribution if the number of $m_{k,i} y_{k,i}^{(f,r)}$s, $FR\lfloor N/(RM) \rfloor$ is sufficiently large. That is,

$$v_k \sim \begin{cases} N(\mu, \sigma^2) & \text{if } b_k = 1 \\ N(-\mu, \sigma^2) & \text{if } b_k = 0. \end{cases} \qquad (9)$$

Step D5: Determine M embedded bits, b_ks, by comparing v_k with threshold value T:

$$b_k = \begin{cases} 1 & \text{if } v_k \geq T \\ 0 & \text{if } v_k \leq -T \\ \text{“not detected”} & \text{if } -T < v_k < T. \end{cases} \qquad (10)$$

2.2 Problems with Previous Method

WMs on images must be able to survive various kinds of image processing procedures. The accumulation in Step D3 in Section 2.1.2, however, may not always be effective because the WM signal, μ, in formula (8) could be attenuated by accumulating regions from which WMs had been removed during image processing. For example, in accumulating (averaging) two regions with the same noise and same WM signal, the noise of region σ is reduced to $\frac{1}{\sqrt{2}}\sigma$. The WM signal, μ, however, is reduced to $\frac{1}{2}\mu$ if WMs from one region out of two are removed. Consequently, the S/N ratio worsens due to accumulation. Accumulation with the previous method could thus cause the signals of WMs, μ, to decrease so much that the embedded bits could not be reliably detected.

To solve this, methods of detection should select regions where WMs remain and should accumulate these. We therefore propose a method that accumulates a subset of regions so that the accumulated region has a minimal degree of WM removal, which is estimated from the region.

3. Using Statistically Adaptive Accumulation to Improve the Detection of Video WMs

3.1 Principle behind Proposed Method

The method of detecting WMs we propose controls the accumulation of the regions so that the accumulated region has a minimal degree of WM removal and detects WMs from the accumulated region. We use a bit-error rate (BER) as a measure of the degree of WM removal and the BER is estimated from each watermarked region. Our method can prevent a decrease in the S/N ratio in the accumulated regions by using the estimated BERs of the watermarked regions to control the accumulation in a way that minimizes the BERs of the accumulated regions.

Figure 4 outlines the process for the method we propose, which involves calculating the statistical values, estimating the BER, sorting, controlling accumulation, and determining the bit values:

Statistical-value calculation: Calculates the statistical values from the region.

BER estimation: Estimates the BER from the statistical values of the region.

Sorting: Sorts regions (actually corresponding sets of statistical values) with corresponding BERs.

Accumulation control: Accumulates, one by one, regions (set of the statistical values) in ascending order of the corresponding BER and re-estimates the BER of the accumulated region in each accumulation step. Selects the accumulated region having the smallest BER.

Bit-value determination: Determines bit-values by comparing the threshold value with the set of statistical values of the accumulated region selected by accumulation control.

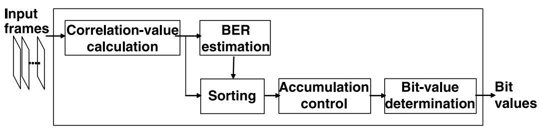


Fig. 4 Process of WM detection.

Our method can be applied to the various kinds of correlation-based WM schema described in Section 2.

3.2 Process Flow

The process flow for our proposed method of detecting WMs is outlined in **Fig. 5**. Step D3 represents the flows for statistical-value calculation and BER estimation. Step D4 represents the flow for sorting, and Steps D5 and D6 represent the flow for accumulation control. Step D7 represents the flow for bit-value determination.

Step D1: Do the following steps over $f_0 = 1, F+1, 2F+1, \ldots$.

Step D2: Input F watermarked frames, $\mathbf{y}'^{(f)}$s ($f = f_0, \ldots, f_0 + F - 1$), and divide FRM subregions, $\mathbf{y}'^{(f,r)}_k$s ($f = f_0, \ldots, f_0 + F - 1,\ r = 1, \ldots, R,\ k = 1, \ldots, M$).

Step D3: For each region, $\mathbf{y}'^{(f,r)} = \bigcup_k \mathbf{y}'^{(f,r)}_k$ ($f_0 \le f \le f_0 + F - 1, 1 \le r \le R$), calculate a set consisting of M statistical values, $\mathbf{v}^{(f,r)} = \{v^{(f,r)}_k \mid 1 \le k \le M\}$, and estimate BER $p(\mathbf{v}^{(f,r)})$ from set $\mathbf{v}^{(f,r)}$. The statistical value, $v^{(f,r)}_k$, of subregion $\mathbf{y}'^{(f,r)}_k$ is given by correlating WM pattern $\mathbf{m}_k$ with region $\mathbf{y}'^{(f,r)}_k$. That is,

$$v^{(f,r)}_k = \frac{1}{\lfloor N/(RM) \rfloor} \sum_{i=1}^{\lfloor N/(RM) \rfloor} m_{k,i} y'^{(f,r)}_{k,i} = \frac{1}{\lfloor N/(RM) \rfloor} \sum_i m_{k,i} y^{(f,r)}_{k,i} \pm \mu^{(f,r)}, \tag{11}$$

Doing the above process over FR regions ($f = f_0, \ldots, f_0 + F - 1, r = 1, \ldots, R$), we obtain FR sets, $\mathbf{v}^{(f_0,1)}, \ldots, \mathbf{v}^{(f_0+F-1,R)}$, and the corresponding FR BERs $p(\mathbf{v}^{(f_0,1)}), \ldots, p(\mathbf{v}^{(f_0+F-1,R)})$.

Step D4: Sort the FR sets, $\mathbf{v}^{(f_0,1)}, \ldots, \mathbf{v}^{(f_0+F-1,R)}$, by the correspond-

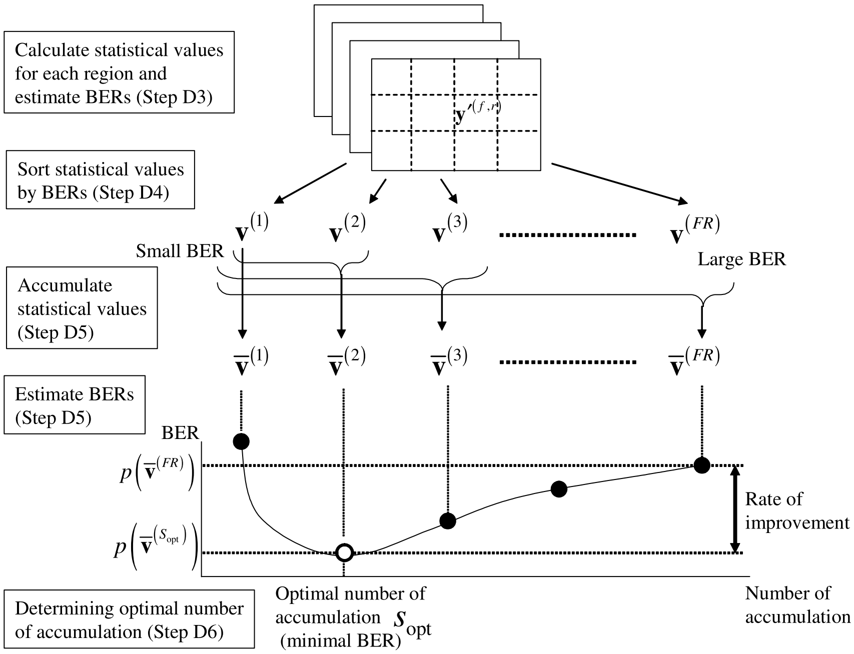


Fig. 5 Overview of statistically adaptive accumulation.

ing BERs and rename the suffixes of the sets and the BERs in ascending BER order. Thus, we obtain the FR sets, $\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(FR)}$, that satisfy $p(\mathbf{v}^{(1)}) \leq \ldots \leq p(\mathbf{v}^{(FR)})$.

Step D5: Generate the accumulated sets, $\bar{\mathbf{v}}^{(s)}$s $(s = 1, \ldots, FR)$, from the FR sets with $\bar{\mathbf{v}}^{(s)} = 1/s \sum_{i=1}^{s} \mathbf{v}^{(i)}$ and estimate the BERs, $p(\bar{\mathbf{v}}^{(s)})$ $(s = 1, \ldots, FR)$, from the FR accumulated sets.

Step D6: Select the set of statistical values with the smallest BER, $\bar{\mathbf{v}}^{(s_{\text{opt}})}$, where s_{opt} represents the optimal number of accumulations:

$$s_{\text{opt}} = \arg \min_{1 \leq s \leq FR} p(\bar{\mathbf{v}}^{(s)}). \tag{12}$$

Step D7: Determine M embedded bits, b_ks, by comparing $\bar{v}_k^{(s_{\text{opt}})}$ with a threshold value, T (> 0), as was the case with Step D4 in Section 2:

$$b_k = \begin{cases} 1 & \text{if } \bar{v}_k^{(s_{\text{opt}})} \geq T \\ 0 & \text{if } \bar{v}_k^{(s_{\text{opt}})} \leq -T \\ \text{“not detected”} & \text{if } -T < \bar{v}_k^{(s_{\text{opt}})} < T. \end{cases} \tag{13}$$

Note that set $\bar{\mathbf{v}}^{(FR)}$ in Step D5 is equal to set $\mathbf{v}$ in Step D4 in Section 2.1.2 and thus that $p(\bar{\mathbf{v}}^{(FR)})$ represents the BER with the proposed method.

3.3 Estimation of BER

We employed procedures in inferential statistics to estimate the BERs of the region, because its statistical values follow a normal distribution dependant on the WM signal and image noise. The basic method of estimating BER described in Section 3.1 was proposed by Echizen et al.[15]. This method can be used to estimate the BER from a watermarked still picture after image processing by using inferential statistics. We expanded it to estimating the BER of regions to implement our own method.

3.3.1 BER of Region

We estimated the BER of each watermarked region from the M statistical values, $v_k^{(f,r)}$ in formula (11), of the proposed WM detection method. As explained in Step D4 in Section 2.1.2, each statistical value, $v_k^{(f,r)}$, follows a normal distribution with mean $\pm\mu^{(f,r)}$ and variance $\sigma^{2(f,r)}$ if the number of $m_{k,i}y_{k,i}^{(f,r)}$s, $\lfloor N/(RM) \rfloor$ is sufficiently large:

$$v_k^{(f,r)} \sim \begin{cases} N(\mu^{(f,r)}, \sigma^{2(f,r)}) & \text{if } b_k = 1 \\ N(-\mu^{(f,r)}, \sigma^{2(f,r)}) & \text{if } b_k = 0. \end{cases} \tag{14}$$

The calculation of BER when $b_k = 1$ is embedded is explained by **Fig. 6**, where the gray area indicating the probability of erroneously detecting $b_k = 0$ is given by

$$p(\mathbf{v}^{(f,r)})_{\mathrm{BE}|b_k=1} = \int_{-\infty}^{-T} \phi(v;\mu^{(f,r)},\sigma^{2(f,r)})\,dv, \tag{15}$$

where $\phi(v;\mu^{(f,r)},\sigma^{2(f,r)})$ is the probability density function of $N(\mu^{(f,r)},\sigma^{2(f,r)})$. The probability of detecting $b_k = 1$ erroneously (when the embedded bit is 0) is correspondingly given by

$$p(\mathbf{v}^{(f,r)})_{\mathrm{BE}|b_k=0} = \int_{\infty}^{T} \phi(v;-\mu^{(f,r)},\sigma^{2(f,r)})\,dv. \tag{16}$$

From formulas (15) and (16), $p(\mathbf{v}^{(f,r)})_{BE|b_k=1} = p(\mathbf{v}^{(f,r)})_{BE|b_k=0}$. Thus, the BER of the region, $\mathbf{y}'^{(f,r)}$, for an arbitrary embedded bit is

$$p(\mathbf{v}^{(f,r)}) = \int_{-\infty}^{-T} \phi(v;\mu^{(f,r)},\sigma^{2(f,r)})\,dv. \tag{17}$$

As shown by formula (17), the mean $\mu^{(f,r)}$ and variance $\sigma^{2(f,r)}$ of a normal distribution should be used to obtain BER. There are, however, the following problems: (1) The information we obtain from the watermarked region $\mathbf{y}'^{(f,r)}$ is not the $\mu^{(f,r)}$ and $\sigma^{2(f,r)}$ but M statistical values $v_k^{(f,r)}$. (2) The statistical values are subject to change due to image processing, and the two normal distributions the values follow are getting closer to each other.

The $\mu^{(f,r)}$ and $\sigma^{2(f,r)}$ should thus be estimated from these statistical values that follow the mixture normal distribution.

3.3.2 EM Algorithm

The expectation-maximization (EM) algorithm is a representative maximum-likelihood method of estimating the statistical parameters of a probability distribution [16),17)]. In the case of a mixture normal distribution comprised of two normal distributions (i.e., $N(\mu^{(f,r)},\sigma^{2(f,r)})$ and $N(-\mu^{(f,r)},\sigma^{2(f,r)})$) that the statistical values $v_k^{(f,r)}$ follow, the EM algorithm can be used to estimate the probability $w_k^{(f,r)}$ that each $v_k^{(f,r)}$ follows $N(\mu^{(f,r)},\sigma^{2(f,r)})$, $\mu^{(f,r)}$, and $\sigma^{2(f,r)}$. The relation between $w_k^{(f,r)}$, $\mu^{(f,r)}$, and $\sigma^{2(f,r)}$ is given by

$$\mu^{(f,r)} = \frac{1}{M\alpha^{(f,r)}} \sum_k w_k^{(f,r)} v_k^{(f,r)}, \tag{18}$$

$$\sigma^{2(f,r)} = \frac{1}{M\alpha^{(f,r)}} \sum_k w_k^{(f,r)} v_k^{(f,r)2} - \mu^{(f,r)2}, \tag{19}$$

where $\alpha^{(f,r)} = 1/M\sum_k w_k^{(f,r)}$ is the weighting factor for $N(\mu^{(f,r)},\sigma^{2(f,r)})$ to the mixture normal distribution. These parameters are sequentially updated from initial values by iterative calculation and $\mu^{(f,r)}$ and $\sigma^{2(f,r)}$ are used as estimates when they are converged. See the appendix for the details on this calculation.

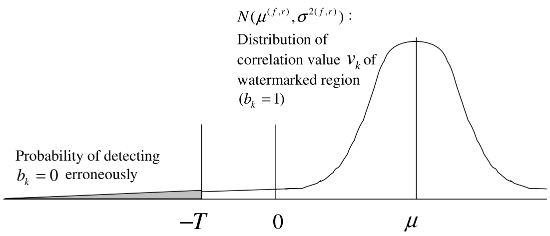


Fig. 6 Calculation of BER.

4. Experimental Evaluation

4.1 Survivability against MPEG-2 Encoding

The ability of the method we propose to detect WMs after MPEG-2 encoding with three different bit rates (3 M, 4 M, and 5 Mbps) was compared experimentally with that of the previous method by using the following standard motion pictures [18)] (450 frames of 720×480 pixels) having different properties (see **Fig. 7**):

(a) Walk through the Square ("Walk"): – People walking in a town square – not much movement.

(b) Whale Show ("Whale"): – Spraying whale with audience – a great deal of movement.

The average (over 450 frames) peak signal to noise ratios (PSNRs) at three different bit rates are listed in **Table 1**.

(a) Walk (b) Whale

Fig. 7 Evaluated pictures.

Table 1 PSNRs at three different bit rates with MPEG-2 encoding.

	3 Mbps	4 Mbps	5 Mbps
Walk	29.01	29.27	29.42
Whale	23.53	24.08	24.44

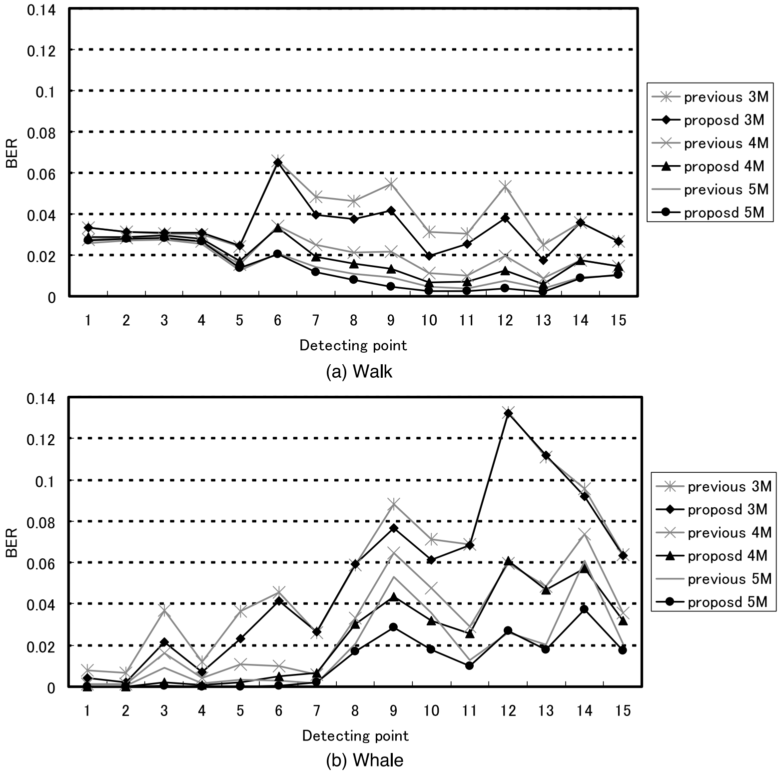


Fig. 8 Evaluation results.

4.1.1 Procedure

A WM pattern representing 256-bit information ($M = 256$) was generated by using a pseudo-random generator [19] and was embedded in each of four 360×240-pixel regions ($R = 4$) of every frame by using the procedures described in Section 2.1.2. After MPEG-2 encoding and decoding with three different bit rates (3, 4, 5 Mbps), 256-bit information was sequentially detected in 30-frame segments of the 450 frames of the watermarked pictures ($F = 30$, the number of detecting points was $450/30 = 15$) and the BERs measured using the proposed detection method described in Section 3.2 were compared with those measured using the previous detection method described in Section 2.1.2. The above procedure was done using 1000 different random WM patterns. Other parameters used in the evaluation were set as follows:

WM strength $\mu^{(f,r)}$: WM strength could be controlled for all pixels to minimize degradation in picture quality. We did not use such control, however, because the work reported in this paper focused on improving WM detection and also because such control might affect our evaluation of WM detection. We instead set the WM strength, $\mu^{(f,r)}$, of the formula (6), to be uniform for all pixels. The example strength we used was 3 for Walk and 4 for Whale. The corresponding PSNR in Walk was 38.6 and the PSNR in Whale was 36.1.

Threshold value T: To facilitate measuring of the BERs, we set the thresholds of formulas (10) and (13) to zeros ($T = 0$) without determining the results of "not detected."

4.1.2 Results

The average values (over 1000 WM patterns) of the measured BERs obtained using the proposed (black line) and the previous (gray line) methods are shown for each bit rate in **Fig. 8**, where the horizontal axis represents the detecting points from 1 to 15 and the vertical axis represents the average BERs. From Fig. 8, we can see that the BERs for the proposed method are better than or equal to those of the previous method at each bit rate. At all bit rates the average ratios of the proposed BER to the previous BER are 0.903 for Walk and 0.784 for Whale, and the effect of the proposed method

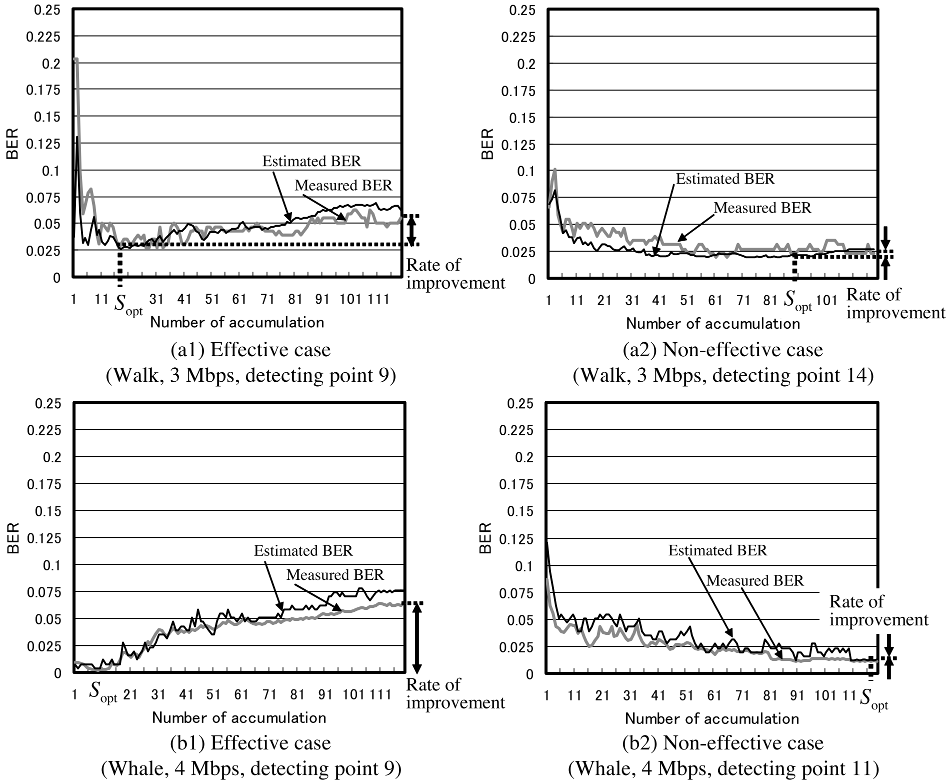


Fig. 9 Transitions in BERs.

is weaker for Walk than for Whale. The reason seems to be that sorting and accumulation control for the proposed method are less effective for Walk than for Whale because most regions of Walk were static and the BERs of their regions were less variable than those for Whale. The BERs estimated with the proposed method may actually be worse than the corresponding measured BERs. Therefore, for the proposed method to be practical, error-correcting codes [20] should be added to the embedded information so that it can be used determine whether the detected information was corrected or not. Moreover, since the new method contains the previous, its performance will always be better than that of the previous if error correction is applied to the series of information detected by each and correctly detected information is used. For both of the evaluated pictures, the proposed method thus yielded an average improvement of 15.7%.

From the plots of Fig. 8, we sampled two cases for each evaluated sample: effective and non-effective cases with the proposed WM detection. Example transitions in BERs estimated with the new method (corresponding to Fig. 5) are shown in **Fig. 9**, where the horizontal axis represents the number of accumulations (the order of accumulation is in ascending order of the estimated BERs) and the vertical axis represents the estimated (black line) and measured (gray line) BERs at the detecting points. From Fig. 9, we can see that the estimated BERs were roughly consistent with the experimentally measured ones—there were only small differences between them. These differences are probably because an insufficient number of pixels was allocated to each statistical value, $\lfloor N/(RM) \rfloor$, for it to follow a normal distribution, and the proposed method estimates the BER based on these values. There are three ways to increase $\lfloor N/(RM) \rfloor$: increase the number of pixels in the video frames (N), reduce the number of regions (R), and reduce the number of bit values (M). Because changing any of these parameter values might degrade performance, the change should be carefully made based on the target application.

In Figs. 9 (a1) (Walk at 3 Mbps, detecting point 9) and 9 (b1) (Whale at 4 Mbps, detecting point 9), the BER tends to increase with the number of accumulations and the rates of im-

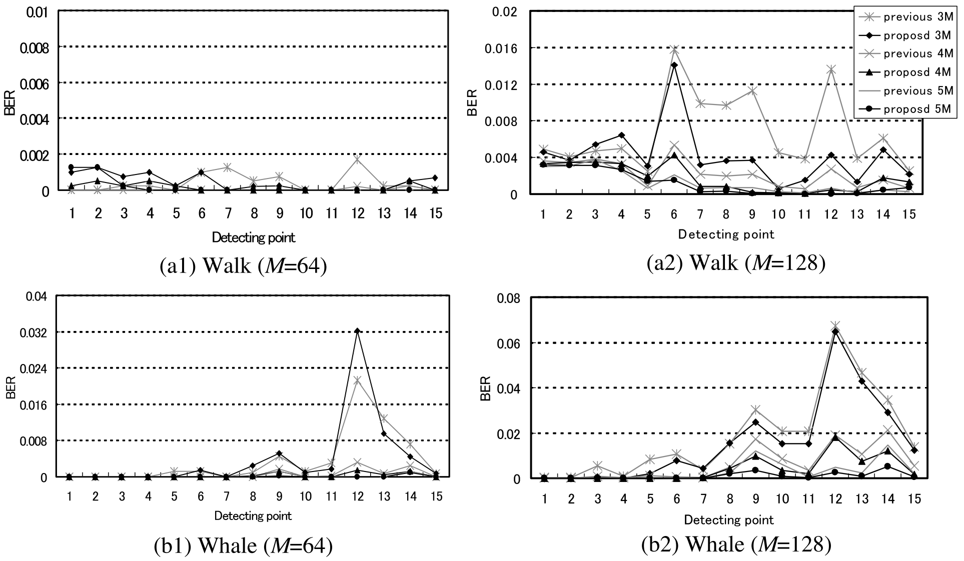


Fig. 10 Evaluation results with $M = 64$ and 128.

provement are about 2.5% for Walk and 6.5% for Whale. Similar trends were also found at all bit rates for the detecting points 7 through 13 for Walk, and 3 through 6, 9, 10, and 14 for Whale, where most parts of neighboring frames were relatively messy or moving compared to the other points. We can infer that the strengths of WMs in regions with varying picture properties were altered by MPEG-2 encoding and thus the BERs of the regions increased with the number of accumulations as a result of sorting with the proposed method.

In Figs. 9 (a2) (Walk at 3 Mbps, detecting point 14) and 9 (b2) (Whale at 4 Mbps, detecting point 11), on the other hand, the BER plot against the number of accumulations is flat or decreasing and the rate of improvement is nearly 0 for both samples. Similar trends were also found at all bit rates for the detecting points 1 through 6, 14, and 15 for Walk, and 1, 2, 7, 8, 11 through 13, and 15 for Whale, where most parts of the neighboring frames were static. We also can infer that MPEG-2 encoding did not vary the strengths of WMs in the regions and thus the BERs of the regions were not changed by sorting.

For all the MPEG-2 bit rates we evaluated, the proposed method could give lower or equal BERs dependant on the picture properties and yielded an average improvement of 15.7%. The proposed method can thus improve WM detection.

4.1.3 Effect of Number of Statistical Values on Detection

Because the new method estimates BER from M statistical values corresponding to M-bit values by using inferential statistics, its performance depends on M. We thus evaluated survivability against MPEG-2 encoding using two more M values (64 and 128) in addition to the 256 used in Section 4.1.2. **Figure 10** shows the average measured BERs obtained using the proposed and previous methods with $M = 64$ and 128 for both evaluated pictures (in the same manner as for $M = 256$ in Fig. 8).

From the plots in Fig. 10, we can see that, for $M = 128$ (Figs. 10 (a2) and (b2)), the BERs with the proposed method were better than or equal to those with the previous one, the same as for $M = 256$. For $M = 64$ (Figs. 10 (a1) and (b1)), on the other hand, the BERs with the proposed method were worse at some detecting points for both evaluated pictures. This means that the EM algorithm that estimates BERs from M statistical values was less reliable the smaller the M became.

Example transitions in estimated BERs at a detecting point for $M = 64$ (Whale, 3 Mbps, detecting point 12) are shown in **Fig. 11**. As the estimated BERs were not consistent with those measured, the proposed method misjudged the minimum BER. **Figure 12** has example histograms of M statistical values for $M = 64$ and 128 at the same detecting point. The horizontal axes represent the statistical values, and the vertical axes represent their frequencies. For $M = 128$ (Fig. 12 (b)), the histogram roughly describes a mixture normal distribution, i.e., a distribution comprising two normal distributions. For $M = 64$ (Fig. 12 (a)), the histogram tended to lose its shape because of the small number of samples. These results indicate that the proposed method is effective for $M = 128$ or more.

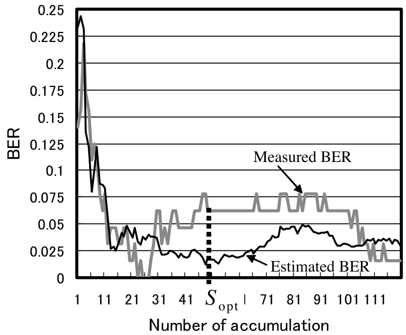
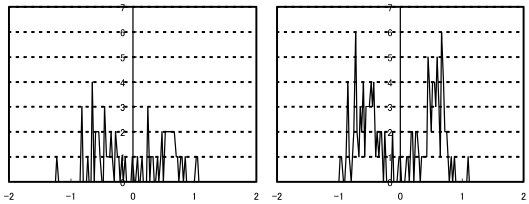


Fig. 11 Example transitions in estimated BERs at detecting point 12 for $M = 64$ (Whale, 3 Mbps).



(a) $M = 64$ (b) $M = 128$

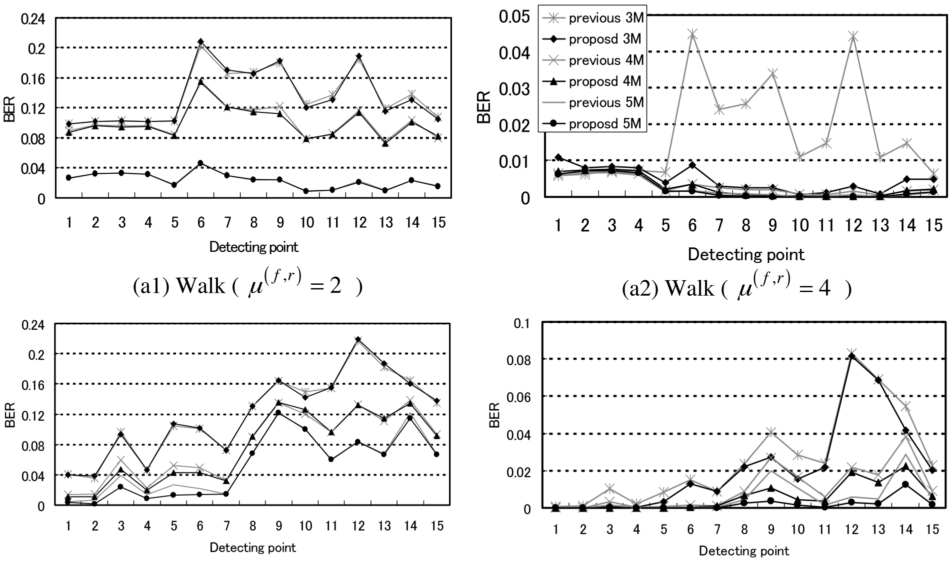
Fig. 12 Histograms for M statistical values.

4.1.4 Effect of WM Strength on Detection

To clarify what effect WM strength had on the ability of the proposed method to detect WMs, we evaluated its survivability against MPEG-2 encoding using two more WM strengths for all evaluated pictures, $\mu^{(f,r)} = 2$ and 4 for Walk and $\mu^{(f,r)} = 3$ and 5 for Whale, in addition to the $\mu^{(f,r)} = 3$ for Walk and 4 for Whale evaluated in Section 4.1.2.

Figure 13 shows the average measured BERs obtained using the proposed and previous methods for the above WM strengths (in the same manner as in Fig. 8), and **Table 2** shows the average ratios of the BERs obtained using the proposed and previous methods. We can see that the larger the $\mu^{(f,r)}$, the more effective the proposed method. This is probably because it is more effective when the WM signal of the region is strong against the corresponding noise and when the mixture normal distribution of the statistical values, from which the EM algorithm is estimated, has slight overlap between the two constitutive normal distributions. This is because the EM algorithm has trouble converging to the correct parameter values when the target mixed distribution has large overlaps between the constitutive distributions [21].

Future work should therefore focus on improving the precision of the inferential statistics so that all the BERs can be correctly estimated



(a1) Walk ($\mu^{(f,r)} = 2$) (a2) Walk ($\mu^{(f,r)} = 4$)

(b1) Whale ($\mu^{(f,r)} = 3$) (b2) Whale ($\mu^{(f,r)} = 5$)

Fig. 13 Evaluation results for different WM strengths.

even when $\mu^{(f,r)}$ is small.

4.2 Survivability against Video Transcoding

Video content is often re-encoded or transcoded when the content is used by other systems or devices (e.g., converted from NTSC format to VGA format or re-encoded from MPEG-2 format to MPEG-4 format). We thus evaluated survivability against such re-encoding and transcoding for a representative case—a watermarked video for broadcasting use is converted and re-encoded for distribution through the Internet. The evaluation was done in four steps.

Step 1: Embed 256-bit information ($M = 256$) into standard motion pictures (450 frames of 720×480 pixels) using the procedure described in Section 2.1.2.

Step 2: Encode and decode the watermarked pictures (Walk and Whale) using the MPEG-2 codec at three different bit rates (3, 4, 5 Mbps).

Step 3: Reduce the resolution of the watermarked pictures to that of the VGA format (640×480 pixels).

Step 4: Encode and decode the watermarked pictures in the VGA format using the MPEG-4 codec at a bit rate of 2 Mbps.

The 256-bit information was detected from the watermarked pictures after Steps 3 and 4 by using both the proposed and previous detection methods. The above procedure was done using 1000 different random WM patterns. Note that Steps 1 and 2 and the other parameters are the same as described in Section 4.1.1.

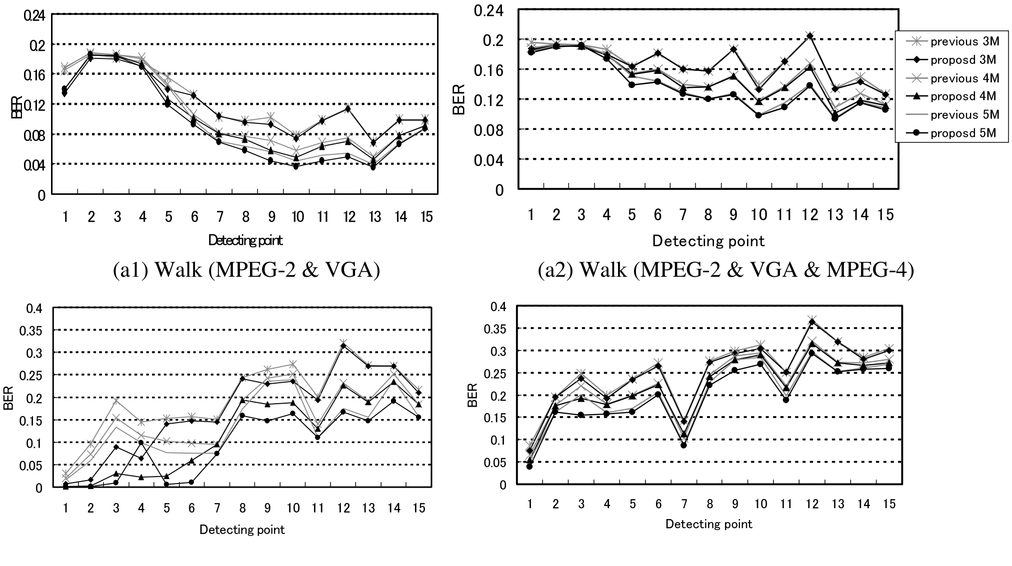
Figure 14 shows the average measured BERs obtained using the proposed and previous methods for the watermarked pictures after Step 3 (MPEG-2 encoding and VGA conversion) and after Step 4 (MPEG-2 encoding, VGA conversion, and MPEG-4 encoding), and **Table 3** lists the average ratios of the BERs obtained using the proposed and previous methods. The proposed method remained roughly effective in terms of survivability against MPEG-2 encoding and VGA conversion compared to only MPEG-2 encoding. However, its effectiveness was weakened by the additional MPEG-4 encoding. This is probably because the additional MPEG-4 encoding caused the WM signal to weaken against the corresponding noise, so the EM algorithm had trouble converging to the correct parameter values, resulting in a large overlap between the constitutive distributions of the statistical val-

Table 2 Average ratios of BERs obtained using proposed and previous methods for different WM strengths.

	$\mu = 2$	$\mu = 3$	$\mu = 4$
Walk	0.988	0.903*	0.731

	$\mu = 3$	$\mu = 4$	$\mu = 5$
Whale	0.967	0.784*	0.583

* Evaluation results in Section 4.1.2.



(a1) Walk (MPEG-2 & VGA) (a2) Walk (MPEG-2 & VGA & MPEG-4)

(b1) Whale (MPEG-2 & VGA) (b2) Whale (MPEG-2 & VGA & MPEG-4)

Fig. 14 Evaluation results for video transcoding.

Table 3 Average ratios of proposed BER to previous BER for video transcoding.

	MPEG-2	MPEG-2 & VGA	MPEG-2 & VGA & MPEG-4
Walk	0.903*	0.938	0.983
Whale	0.784*	0.776	0.970

∗ Evaluation results in Section 4.1.2.

ues. As in the case described in Section 4.1.4, future work should focus on improving the precision of inferential statistics.

5. Conclusion

Improving WM survivability is an essential requirement in research on video watermarking. Redundant coding is a basic method that can prevent errors by accumulating frames or regions. Accumulation, however, could cause the strength of WMs to decrease so much that the embedded bits may not be detected reliably. This paper proposed a method of detecting WMs using statistically adaptive accumulation that can prevent the strength of WMs from decreasing due to accumulation. Our method was used to estimate the BER of region using the expectation-maximization algorithm utilized in inferential statistics and it accumulated the subset of regions based on the estimated BERs so that the accumulated region had a minimal BER. Experimental evaluations using actual motion pictures revealed that our new method can improve WM detection after MPEG-2 encoding by an average of 15.7% and can be widely used in correlation-based watermarking. Future work should focus on improving the precision of inferential statistics so that the BER can be correctly estimated under any conditions.

References

1) Swanson, M., Kobayashi, M. and Tewfik, A.: Multimedia data-embedding and watermarking technologies, *Proc. IEEE*, Vol.86, No.6, pp.1064–1087 (1998).
2) Bloom, J., Cox, I., Kalker, T., Linnartz, J., Miller, M. and Traw, C.: Copy protection for DVD video, *Proc.IEEE*, Vol.87, No.7, pp.1267–1276 (1999).
3) Lin, E. and Delp, E.: Spatial Synchronization Using Watermark Key Structure, Security and Watermarking of Multimedia Contents, *Proc. SPIE*, Vol.5306, pp.536–547 (2004).
4) Kalker, T., Depovere, G., Haitsma, J. and Maes, M.: Video watermarking system for broadcast monitoring, Security and Watermarking of Multimedia Contents, *Proc. SPIE*, Vol.3657, pp.103–112 (1999).
5) Kusanagi, A. and Imai, H.: An Image Correction Scheme for Video Watermarking Extraction, *IEICE Trans. Fundamentals*, Vol.E84-A, No.1, pp.273–280 (2001).
6) Cox, I., Kilian, J., Leighton, T. and Shamoon, T.: Secure Spread Spectrum Watermarking for Multimedia, *IEEE Trans. Image Processing*, Vol.6, No.12, pp.1673–1687 (1997).
7) Hartung, F., Su, J. and Girod, B.: Spread spectrum watermarking: Malicious attacks and counterattacks, Security and Watermarking of Multimedia Contents, *Proc. SPIE*, Vol.3657, pp.147–158 (1999).
8) Hsu, C. and Wu, J.: Hidden Digital Watermarks in Images, *IEEE Trans. Image Processing*, Vol.8, No.1, pp.58–68 (1999).
9) Pereira, S. and Pun, T.: Robust template matching for affine resistant image watermarks, *IEEE Trans. Image Processing*, Vol.9, No.6, pp.1123–1129 (2000).
10) Seo, Y., Choi, S., Park, S. and Kim, D.: A Digital Watermarking Algorithm Using Correlation of the Tree Structure of DWT Coefficients, *IEICE Trans. Fundamentals*, Vol.E87-A, No.6, pp.1347–1354 (2004).
11) Liang, T. and Rodriguez, J.: Robust Watermarking Using Robust Coefficients, Security and Watermarking of Multimedia Contents II, *Proc. SPIE*, Vol.3971, pp.356–335 (2000).
12) Delaigle, J., Vleeschouwer, C., Goffin, F., Macq, B. and Quisquater, J.: Low cost watermarking based on a human visual model, *Proc. Eur. Conf. Multimedia Applications, Services Techniques*, pp.153–167 (1997).
13) Echizen, I., Yoshiura, H., Arai, T., Kimura, H. and Takeuchi, T.: General Quality Maintenance Module for Motion Picture Watermarking, *IEEE Trans. Consumer Electronics*, Vol.45, No.4, pp.1150–1158 (1999).
14) Bender, W., Gruhl, D. and Morimoto, N.: Techniques for data hiding, *Proc. SPIE*, Vol.2020, pp.2420–2440 (1995).
15) Echizen, I., Yoshiura, H., Anzai, K. and Sasaki, R.: Estimating the Bit-error-rate in Digital Watermarking Using Inferential Statistics, *J. IPS Japan*, Vol.42, No.8, pp.2006–2016 (2001) (in Japanese).
16) Redner, R. and Walker, H.: Mixture densities, maximum likelihood and the EM algorithm, *SIAM Review*, Vol.26, pp.195–239 (1984).
17) Xu, L. and Jordan, M.: On Convergence properties of the EM algorithm for Gaussian mixtures, *Neural Computation*, Vol.8, pp.129–151 (1996).
18) The Institute of Image Information and Television Engineers: Evaluation video sample

(standard definition).
19) Matsumoto, M. and Nishimura, T.: Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator, *ACM Trans. Modeling and Computer Simulation*, Vol.8, No.1, pp.3–30 (1998).
20) Wicker, S.: *Error control systems for digital communication and storage*, Englewood Cliffs: Prentice Hall (1995).
21) Akaho, S.: The EM Algorithm for multiple object recognition, *Proc. IEEE Intl. Conf. Neural Networks*, pp.2426–2431 (1995).

Appendix

A.1 Process flow for EM algorithm

The $\mu^{(f,r)}$ and $\sigma^{2(f,r)}$ are estimated from M statistical values $v_k^{(f,r)}$s in the following steps:

Step 1: Set the initial values of parameters $\alpha^{(f,r)}$, $\mu^{(f,r)}$, and $\sigma^{2(f,r)}$ to $\alpha^{(f,r,0)}$, $\mu^{(f,r,0)}$, and $\sigma^{2(f,r,0)}$.

Step 2: Do Step 3 through Step 5 over $t = 1, 2, \ldots$.

Step 3: For each k, calculate $w_k^{(f,r,t)}$ from $v_k^{(f,r)}$:

$$w_k^{(f,r,t)} = \frac{\alpha^{(f,r,t)}\phi(v_k^{(f,r)}; \mu^{(f,r,t)}, \sigma^{2(f,r,t)})}{g(v_k^{(f,r)}; \alpha^{(f,r,t)}, \mu^{(f,r,t)}, \sigma^{2(f,r,t)})}, \tag{20}$$

where $g(v; \alpha, \mu, \sigma^2)$ is the probability density function of the mixture normal distribution, i.e.,

$$g(v; \alpha, \mu, \sigma^2) = \alpha\phi(v; \alpha, \mu, \sigma^2) + (1-\alpha)\phi(v; \alpha, -\mu, \sigma^2). \tag{21}$$

Step 4: Update $\alpha^{(f,r,t)}$, $\mu^{(f,r,t)}$, and $\sigma^{2(f,r,t)}$ with the following formulas:

$$\alpha^{(f,r,t+1)} = \frac{1}{M}\sum_k w_k^{(f,r,t)}, \tag{22}$$

$$\mu^{(f,r,t+1)} = \frac{1}{M\alpha^{(f,r,t+1)}}\sum_k w_k^{(f,r,t)} v_k^{(f,r)}, \tag{23}$$

$$\sigma^{2(f,r,t+1)} = \frac{\sum_k w_k^{(f,r,t)} v_k^{(f,r)2}}{M\alpha^{(f,r,t+1)}} - \mu^{(f,r,t+1)2}. \tag{24}$$

Step 5: Stop the process and set

$$\mu^{(f,r)} = \mu^{(f,r,t+1)}, \tag{25}$$

$$\sigma^{2(f,r)} = \sigma^{2(f,r,t+1)}, \tag{26}$$

if the parameters satisfy the following conditions:

$$|\alpha^{(f,r,t+1)} - \alpha^{(f,r,t)}| < \delta, \tag{27}$$

$$|\mu^{(f,r,t+1)} - \mu^{(f,r,t)}| < \delta, \tag{28}$$

$$|\sigma^{2(f,r,t+1)} - \sigma^{2(f,r,t)}| < \delta. \tag{29}$$

The bit error rate of the region, $\mathbf{y}'^{(f,r)}$, $p^{(f,r)}$, is calculated from $\mu^{(f,r)}$ and $\sigma^{2(f,r)}$ using formula (17).

(Received November 30, 2005)
(Accepted June 1, 2006)
(Online version of this article can be found in the IPSJ Digital Courier, Vol.2, pp.537–550.)

Isao Echizen received his B.S., M.S., and D.E. from the Tokyo Institute of Technology, Japan, in 1995, 1997, and 2003. He joined the Systems Development Laboratory of Hitachi, Ltd., Yokohama, Japan, in 1997 and continued research on information security, copyright protection technologies, and multimedia application systems. He received the Best Paper Award from IPSJ in 2005. He is a member of the IEICE, ITE, and IEEE.

Yasuhiro Fujii received his M.S. and Ph.D. from the Department of Physics of the University of Tokyo, Japan, in 1996 and 2001. In 2001 he joined the Systems Development Laboratory of Hitachi, Ltd., Yokohama, Japan. His research is on image processing and information security. He received the Best Paper Award from IPSJ in 2005. He is a member of the IEICE.

Takaaki Yamada received his B.S. from Kyoto University, Japan, in 1988. He joined Hitachi, Ltd. in 1988 and has been engaged in R&D dedicated to multimedia application systems and copyright protection systems. He is now a Senior Researcher at the Systems Development Laboratory, Yokohama, Japan. He received the Best Paper Award from IPSJ in 2005. He is a member of the IEEE computer society.

Satoru Tezuka received his B.S. and Ph.D. from Keio University, Japan, in 1984 and 2000. He joined Hitachi, Ltd. and has been engaged in research on information security, public key infrastructure, and copyright protection. He is now a Senior Research Manager at the Systems Development Laboratory, Yokohama, Japan. He received the Best Paper Award from IPSJ in 2005.

Hiroshi Yoshiura received his B.S. and D.Sc. from the University of Tokyo, Japan, in 1981 and 1997. He joined Hitachi, Ltd. in 1981 and until March 2003 was a Senior Research Engineer in the company's Systems Development Laboratory. He is currently a Professor in the Department of Human Communications, the Faculty of Electro-Communications, the University of Electro-Communications. He has been engaged in research on information security and copyright protection technologies and received the Young Researcher Encouragement Award from IPSJ in 1990, the Best Paper Award from IPSJ in 2005, and the Industrial Technology Award from ISCI in 2005. He is a member of the IEICE, ISCI, JSAI, and IEEE.