Paper

Improved Video Watermarking for MPEG Encoders Using Motion Analysis

(動き検出を利用したMPEGエンコーダ向け動画用電子透かし)

Isao Echizen† (member), Hiroshi Yoshiura††, Kousuke Anzai†, Jun'ichi Taguchi†, Kazunori Nakano† and Satoru Tezuka†

Abstract Maintaining picture quality is an essential requirement for video watermarking, and this paper shows that watermarks are less perceptible when they are embedded where picture contents are moving. It also describes a new watermarking method that uses motion detection to preferentially allocate watermarks to macro-blocks containing motion. As well it presents the results of experimental evaluations which demonstrate that this method significantly reduces degradation in picture quality. Further it shows that, for the same picture quality after MPEG encoding and decoding, 30% more watermarks can be embedded than with the previous method. This new method would be most effective when implemented in MPEG encoders, which already have the necessary motion analysis functions; it can also be used in content-distribution systems.

Key words: Video watermarking, Picture quality, MPEG encoders, Block-matching techniques

1. Introduction

Digital watermarking can be used to embed copyright and copy control information in motion pictures and will therefore be used in DVD players and recorders as well as in digital broadcasting equipment such as set-top boxes. Because watermarking must not degrade picture quality, the properties of pictures have been taken into consideration when determining the allocation of watermarks[1)~4)]. However, even when watermarks are preferentially embedded in messy areas rather than plain areas where they are easier to perceive, picture quality is often degraded because the conventional watermarking methods consider only the properties of each frame (still picture) and neglect inter-frame properties (i.e., motion). The method described in this paper allocates watermarks by using motion analysis and would be most effective when used in MPEG encoders, since they already have the functions needed for this analysis.

Received November 11, 2002; Revised October 17, 2003; Accepted November 28, 2003

† Systems Development Laboratory, Hitachi,Ltd.
(890, Kashimada, Saiwai-ku, Kawasaki, 212-8567 Japan)
†† Faculty of Electro-Communication, The University of Electro-Communications.
(1-5-1, Chofugaoka, Chofu, 182-8585, Japan)

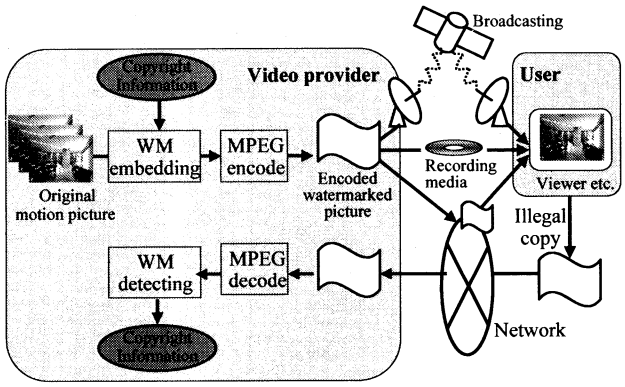


Fig. 1 One use of video watermarking.

2. Previous Methods for Maintaining Picture Quality

2.1 Typical use of video watermarking

As shown in **Fig. 1**, watermarks (WMs) representing copyright information are embedded into video pictures, and the watermarked pictures are compressed using an MPEG encoder. The video provider broadcasts the encoded pictures or distributes them over a network or on recording media such a DVD-ROM. When the encoded watermarked pictures are received, they are decoded, and the user can watch them on a TV set or a personal computer. If the user illegally copies and redistributes the pictures (for example, by placing them on a web page), the illegal copying and distribution can

be detected by detecting the copyright information embedded in the pictures. The video provider could even identify the illegal copier if the identifier of the user had been embedded in the pictures.

2.2 Requirements for video watermarking

Methods of video watermarking can be classified into two types: those changing the pixel values and those changing the frequency coefficients. This paper is concerned with the former type and focuses on representative methods of that type: those changing the luminance values.

Video watermarks should not degrade picture quality. That is, they should not interfere with the user's enjoyment. They should nonetheless be robust enough to be reliably detected after common image processing procedures such as MPEG, and they should never be found in unmarked pictures.

2.3 Definition of terminology

We first define some terminology used in this paper.

Watermark Embedding: Changing pixel values in order to embed information into images. The luminance set of the t th frame, which consists of $n \times m$ pixels, is

$$\mathbf{Y}^{(t)} = \{Y_{i,j}^{(t)} \mid 1 \leqq i \leqq n\,,\ 1 \leqq j \leqq m\},$$

so embedding operation E giving luminance set $\mathbf{Y}'^{(t)}$ of a watermarked frame is

$$E : \mathbf{Y}^{(t)} \rightarrow \mathbf{Y}'^{(t)}.$$

Watermark Strength: Luminance change due to watermark embedding. Watermark strength $f_{i,j}^{(t)}$ of pixels (i,j) of the t th frame is given by

$$f_{i,j}^{(t)} = |Y_{i,j}'^{(t)} - Y_{i,j}^{(t)}|.$$

Embedding Quantity: Total amount of luminance change distributed among pixels in a frame. The relation between the embedding quantity Q and the watermark strength $f_{i,j}^{(t)}$ is

$$Q = \sum_{ij} f_{i,j}^{(t)}.$$

2.4 Problem with previous methods for maintaining picture quality

Previous methods for maintaining picture quality can be classified into two types:

(a) Those that reduce the quantity of WMs embedded in pictures by enabling even a small amount of WMs to be robust enough to be precisely detected[5)~9)].

(b) Those that embed WMs where they are less perceptible[1)~4)].

One example of type (a) is an improvement[9)] in the patchwork algorithm reported by Bender and coworkers[10)], which embeds WMs by giving a statistically meaningful difference between the average luminance of pixels (a_i, b_i), $1/N \sum_{i=1}^{N} Y_{a_i,b_i}^{(t)}$, and the average luminance of pixels (c_i, d_i), $1/N \sum_{i=1}^{N} Y_{c_i,d_i}^{(t)}$, where the locations of pixels (a_i, b_i) and (c_i, d_i) are randomly chosen in each frame without overlapping. Small luminance changes in those pixels were shown to result in a statistically meaningful difference if (a_i, b_i) and (c_i, d_i) were neighboring for each i[9)]. The methods of type (a) reduce the luminance change while maintaining the robustness of the WMs and thus maintain picture quality.

The methods of type (b), on the other hand, maintain picture quality directly and are usually more effective. Most of them use "masking effects" found in the human eye system[1)~3)]. Because picture content masks WMs, these methods embed WMs strongly in messy areas of pictures and weakly in plain areas. The methods select as "messy areas" regions with large luminance differences and many high-frequency elements. The quality of motion pictures is still degraded, however, because these methods consider only the properties of each frame (still picture) and neglect inter-frame properties (i.e., motion) when selecting areas where WMs would be less perceptible.

3. Improved Video Watermarking for MPEG Encoders

3.1 New criterion for measuring WM imperceptibility

In our proposed method, the strength of embedded WMs is adjusted according to an "estimation of WM imperceptibility." That is, WMs are embedded more strongly in areas where they are estimated to be relatively imperceptible. This section describes our new criterion for estimating WM imperceptibility that takes motion into account.

(1) Analysis of WM imperceptibility

Analyzing the relation between WM imperceptibility (WMIP) and the motion in motion pictures, we observed the following:

・WMIP depends on the velocity of object movement between frames, as shown in **Fig. 2**(a), and the larger the velocity is, the less perceptible the WMs are.

・WMIP also depends on the degree of object deformation between frames as shown in Fig. 2(b), and the more the deformation is, the more imperceptible the WMs are.

These observations imply that WMIP depends not only on the properties of the still pictures making up the sequence of images constituting a motion picture,

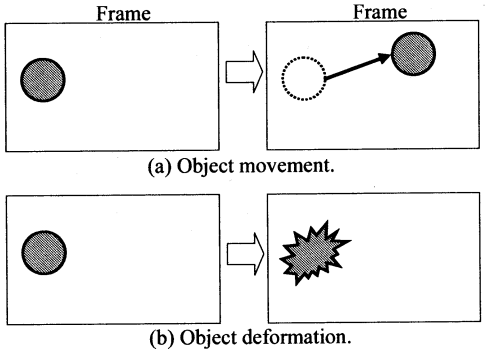


Fig. 2 Relation between WM imperceptibility and motion picture properties.

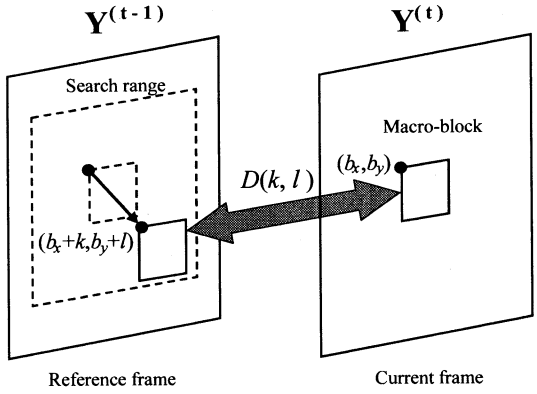


Fig. 3 Block-matching method.

but also on the motions of the objects: their velocities and their degrees of deformation.

In the following we describe the block-matching techniques widely used for motion estimation in MPEG encoders and present a new criterion for measuring WMIP that uses these techniques.

(2) **Block-matching techniques**

Block-matching techniques[11] are based on between-frame matching of MPEG macro-blocks (16 × 16-pixel blocks).

Step 1: The sum of absolute difference $D(k,l)$ between the macro-block of current frame $\mathbf{Y}^{(t)}$ and that of reference frame $\mathbf{Y}^{(t-1)}$, which is displaced by (k,l) from the block of $\mathbf{Y}^{(t)}$, is calculated using the following formula (see **Fig. 3**):

$$D(k,l) = \sum_{i,j=0}^{15} |Y^{(t)}_{b_x+i,b_y+j} - Y^{(t-1)}_{b_x+k+i,b_y+l+j}|,$$

where (b_x, b_y) is the pixel location representing the starting point of the macro-block.

Step 2: The $D(k,l)$s are calculated within search range k,l. A range of $-15 \leqq k,l \leqq 15$ is typically searched[11], and 31 × 31 $D(k,l)$s are generated.

Step 3: The motion vector $\mathbf{v}$ is identified as the vector (k_0,l_0) for the minimum of 31 × 31 $D(k,l)$s. That is, according to the formulas

$$d = D(k_0,l_0) = \min_{k,l} D(k,l),$$

where d is the inter-frame deformation of the object, and $\mathbf{v} = (k_0,l_0)$ is the motion vector representing the velocity of the movement of the object. If more than one (k_0,l_0) gives the same minimum value, $D(k_0,l_0)$, the shortest (k_0,l_0) vector is chosen as motion vector $\mathbf{v}$. The motion vector is thus $\mathbf{v} = (0,0)$ if all $D(k,l)$s are 0.

(3) **Criterion using the motion information**

We first define the following terminology.

WMIP $R_{i,j}$ of a pixel in a motion picture: A measure representing the degree of imperceptibility of a luminance change at a pixel (i,j). If, for example, the WMIP of pixel (1,0) is larger than that of pixel (2,0), a change in luminance at pixel (1,0) is less perceptible than a change at pixel (2,0).

As described in Section 3.1 (1), the $R_{i,j}$ of a pixel in a motion picture depends on two kinds of motion information as well as on the still picture properties. Thus, $R_{i,j}$ is a function of d, $\mathbf{v}$, and $S_{i,j}$:

$$R_{i,j} = F(d, \mathbf{v}, S_{i,j}), \tag{1}$$

where $S_{i,j}$ is the pixel's WMIP based only on the still picture properties. We next explain how WMIP $S_{i,j}$ and function F are determined.

(a) **Determination of WMIP $S_{i,j}$**

Various methods for determining $S_{i,j}$ have been proposed. We use the one reported by Echizen and coworkers[4]. This method uses luminance set $D_{i,j,\theta}$ on the line at angle θ centered on pixel (i,j), as shown in **Fig. 4**:

$$S_{i,j} = \min_{\theta} \sqrt{\sigma^2(D_{i,j,\theta})}, \tag{2}$$

where $\sigma^2(\bullet)$ is an operation that calculates the variance of elements in a set. They clarified the effect of the number of pixels in $D_{i,j,\theta}$ on WMIP $S_{i,j}$ and the effect of the resolution of angular orientation θ on WMIP $S_{i,j}$.

- If the number of pixels in $D_{i,j,\theta}$ is too large (i.e., the line segment included in $D_{i,j,\theta}$ is too long), $S_{i,j}$ is affected by picture areas distant from the target pixel (i,j) and thus does not adequately represent the local imperceptibility at pixel (i,j). For example, **Fig. 5** (a) and Fig. 5 (b) respectively show a sample image and magnified portions around the contour of the sample image. As shown in Fig. 5 (b), a watermark is perceptible at the target pixel (i_1,j_1) in a smooth area distant from the contour and thus S_{i_1,j_1} should be small. If the number of pixels in $D_{i_1,j_1,\theta}$ is too large, however,

$D_{i_1,j_1,\theta}$ lies across the contour and thus has luminance values whose variance is large. Consequently, S_{i_1,j_1} is affected by the distant contour and is misjudged to be large.

・If, on the other hand, the number of pixels in $D_{i,j,\theta}$ is too small, $D_{i,j,\theta}$ does not illustrate the image structure around the target pixel but instead reflects only the strong correlation of neighboring luminance values. Consequently, $S_{i,j}$ will be underestimated. For example, Fig. 5 (c) shows a magnified part of a messy area. A watermark is imperceptible at the target pixel (i_2, j_2) in this part and S_{i_2,j_2} should be large. If the number of pixels in $D_{i_2,j_2,\theta}$ is too small, however, $D_{i_2,j_2,\theta}$ does not characterize the messiness of this part and is calculated only on the neighboring pixels with luminance values that are almost the same. Consequently, S_{i_2,j_2} is misjudged to be small.

・The larger the resolution of angular orientation θ, the more precisely $S_{i,j}$ can be estimated.

Based on these analyses and experimental evaluations of image quality, they concluded that image quality is best when the number of pixels in $D_{i,j,\theta}$ is five and the angular orientation is eight, which corresponds to the finest angular resolution in the five-pixel case. We thus use these values here.

Real-time processing is important for video watermarking, but it may be difficult with the proposed method because the proposed method uses intensive processes such as calculating squares and square roots. We therefore estimated the possibility of real-time processing by using a media processor[12)13)] that has been used for various MPEG encoders. This processor supports two single-instruction multiple data (SIMD) operations that perform parallel processing of multimedia data at up to 1800 MIPS performance.

The possibility of real-time processing depends on the number of pixels in $D_{i,j,\theta}$ and the number of orientation θs because the amount of computation required for calculating $S_{i,j}$ is proportional to both value of these numbers. We therefore use as values those determined by Echizen and coworkers (i.e., five and eight, respectively). We assume that $S_{i,j}$ has 22-bit precision and that the input pictures have the standard definition (i.e., 8-bit luminance component, 720 × 480-pixel frame size, and 30 frames per second). Real-time processing of the proposed method with these parameters is estimated to require 3500 MIPS performance, which can be achieved by the parallel use of two media processors each capable of 1800 MIPS. Real-time processing

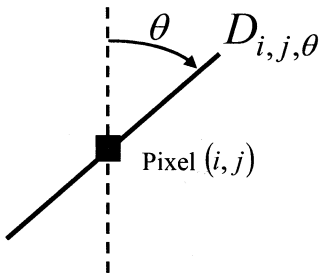
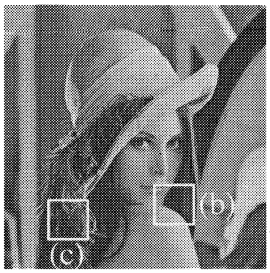
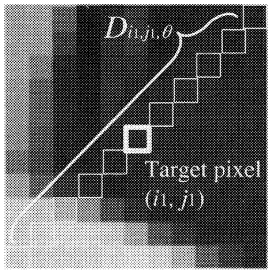


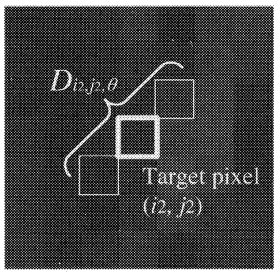
Fig. 4 Definition of $D_{i,j,\theta}$.



(a) Sample image.



(b) Too large.



(c) Too small.

Fig. 5 $D_{i,j,\theta}$ of inadequate size.

using a single media processor should be attainable in the near future because of the continuing improvements in processor speeds. The proposed method can thus be processed in real time on the target equipment, such as on MPEG encoders.

(b) Determination of function F

There has been little research on the determination of F, so there is still no clue for determining F optimally. We therefore used as the simplest candidate a two-step function of motion vector $\mathbf{v}$ and deformation quantity d. The suitability of this function was demonstrated by the results of the experimental evaluations given in Section 4.

To establish the two-step function specifically, we classified the macro-blocks in each frame into two types based on the motion estimation:

Static areas: Macro blocks in which the length of motion vector $|\mathbf{v}|$ is less than threshold value $T_{\mathbf{v}}$ and deformation quantity d is less than threshold value T_d: $d < T_d$ and $|\mathbf{v}| < T_{\mathbf{v}}$. Objects in these

areas are static.

Motion areas: Macro blocks in which $|\mathbf{v}|$ is not less than $T_{\mathbf{v}}$ or d is not less than T_d: i.e., $d \geq T_d$ or $|\mathbf{v}| \geq T_{\mathbf{v}}$. Objects in these areas are to be moving or deformed.

Based on the above classification, the WMIP of pixel $R_{i,j}$ is given by a two-step function:

$$R_{i,j} = F(d, \mathbf{v}, S_{i,j}) = \begin{cases} S_{i,j} & d < T_d, |\mathbf{v}| < T_{\mathbf{v}} \\ \alpha S_{i,j} & \text{otherwise} \end{cases}, \quad (3)$$

where coefficient α is greater than 1. $R_{i,j}$ is small in static areas and large in motion areas. Thus, this criterion is consistent with the results of our preliminary analysis and can be considered appropriate.

(4) **Setting $T_{\mathbf{v}}, T_d$, and α**

Threshold values $T_{\mathbf{v}}$, T_d and coefficient α in formula (3) are set as follows:

Step 1: The values of $T_{\mathbf{v}}$ and T_d are set according to the results of a subjective evaluation using standard video samples[14].

Step 2: The value of α is also set subjectively using the $T_{\mathbf{v}}$ and T_d values determined in Step 1.

Setting $T_{\mathbf{v}}$ and T_d: First, $T_{\mathbf{v}}$ is determined from the motion vectors measured, after MPEG compression, in areas where WMs are easier to perceive and in areas where they are harder to perceive (WM perceptibility is judged using MPEG-encoded standard pictures). Next, T_d is determined using a similar procedure.

Step 1: WMs are embedded by considering only the properties of the still pictures (embedding quantity Q is distributed among the pixels in each frame on the basis of only $S_{i,j}$, and the luminance of each pixel in each frame is changed based on the distributed quantity).

Step 2: The watermarked pictures are encoded using MPEG2 with four different bit rates (2, 4, 6, and 8 Mbps).

Step 3: After decoding, the pictures are evaluated subjectively to identify evaluators for discriminating areas in which WMs are easier to perceive from those in which they are harder to perceive.

Step 4: The motion vectors and the deformation quantities are measured in these areas to determine $T_{\mathbf{v}}$ and T_d.

Carrying out this procedure with five embedding quantities produced threshold values of $T_{\mathbf{v}} = 6$ and $T_d = 5000$.

Setting α: The value of α also needs to be set appropriately because if α were too small there would be no difference between the previous and proposed methods and if it were too large, the picture quality would be degraded in motion areas.

We therefore used several values of α when making a watermarked picture and evaluated the picture quality after MPEG2 compression using the following procedure:

Step 1: WMs are embedded into a standard video sample (Walk through the Square[14]) by using several values of α and the above threshold values of $T_{\mathbf{v}} = 6$ and $T_d = 5000$.

Step 2: These pictures watermarked with different values of α are encoded using MPEG2 at four different bit rates (2, 4, 6, and 8 Mbps).

Step 3: The decoded pictures are evaluated subjectively to identify evaluators for determining picture quality (details of the subjective evaluation procedure are described in Section 4.1).

Carrying out this procedure with five embedding quantity produced the values of α at the four different MPEG bit rates.

The results for 8 Mbps are shown in **Fig. 6**; the horizontal axis represents the average embedding quantity per pixel, q, and the vertical axis represents the level of picture quality. Pictures watermarked with three different values of α ($\alpha = 1.5, 2.5, 3.5$) were evaluated, and the picture watermarked with $\alpha = 2.5$ was judged to have the best quality. We thus set the value of α for 8 Mbps to 2.5.

The same evaluation was done at 2, 4, and 6 Mbps. The results, listed in **Table 1**, show that the value of α depends inversely on the bit rate. We used these values in the detailed evaluation reported in Section 4.

3.2 WM embedding system

(1) **Overall structure**

The proposed method would be most effective when implemented in MPEG encoders because they already have the necessary motion analysis functions. **Fig. 7** shows the overall system structure of the proposed method implemented in an MPEG encoder. As shown in Fig. 7(a), the watermarking process embeds copyright information into each frame by using the motion information from the motion estimation of the MPEG encoder. Then watermarked frames are encoded by the

Table 1 Values of α.

MPEG2 bit rate (Mbps)	α
2	3.5
4	3.0
6	2.8
8	2.5

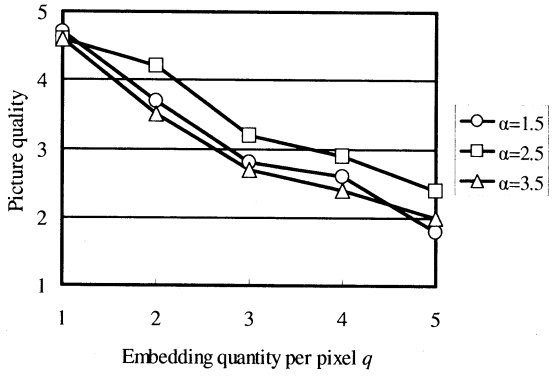


Fig. 6 Evaluated quality of watermarked pictures (8 Mbps).

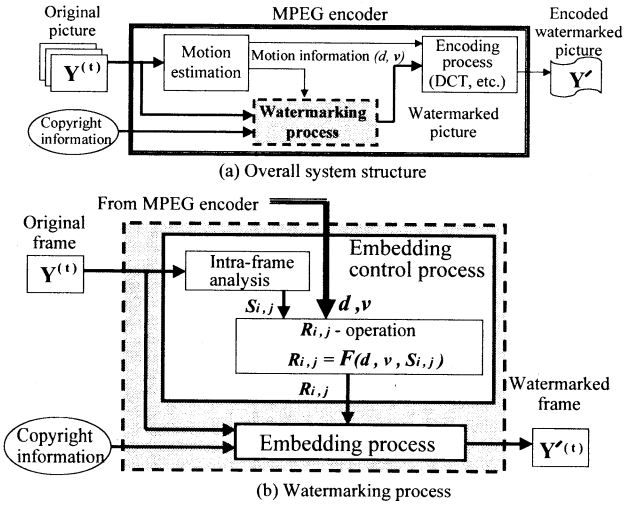


Fig. 7 Watermark embedding system using the proposed method.

encoding process.

Fig. 7(b) shows the structure of the watermarking process, which consists of an embedding process and an embedding control process:

Embedding process: Embeds into each frame WMs representing the copyright information. The WM strength is determined using the $R_{i,j}$ calculated in the embedding control process.

Embedding control process: Calculates $R_{i,j}$ (WMIP of each pixel) of the original frames. This process consists of two subprocesses:

(a) Intra-frame analysis: Calculates $S_{i,j}$ for each frame by using formula (2).

(b) $R_{i,j}$-operation: Calculates $R_{i,j}$ from $S_{i,j}$, $\mathbf{v}$, and d.

The embedding control process can be used in various WM embedding systems because it is combined with the embedding processes only though $R_{i,j}$, which can be calculated independently of the embedding process.

(2) Details of watermarking process

The process flow of the 1-bit-WM embedding is described below. Steps 2 and 3 represent the flow of the embedding control process and Steps 4 and 5 represent that of the embedding process. For multiple-bit embedding, each frame is divided into regions, and the 1-bit embedding process is applied to each region. See appendix for the corresponding WM detection.

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

Step 2: Calculate $S_{i,j}$ from luminance set $\mathbf{Y}^{(t)}$ of the input frame.

Step 3: Calculate $R_{i,j}$ from $\mathbf{v}$, d (calculated by motion estimation of the encoder) and $S_{i,j}$.

Step 4: Determine WM strength $f_{i,j}^{(t)}$ at pixel (i,j) of frame $\mathbf{Y}^{(t)}$:

$$f_{i,j}^{(t)} = \frac{R_{i,j}}{\sum_{ij} R_{i,j}} Q, \tag{4}$$

where Q is the specific WM embedding quantity for frame $\mathbf{Y}^{(t)}$; Q is set by the system designer and controlled by the users in accordance with the intended use. This control means that the Q for a frame is distributed among its pixels in proportion to $R_{i,j}$.

Step 5: Generate watermarked frame $\mathbf{Y}'^{(t)}$ by adding WM pattern $\mathbf{M}$ $(= \{M_{i,j} = +1 \text{or} -1 \mid 1 \leqq i \leqq n, 1 \leqq j \leqq m\}$ comprising a pseudo random array of ± 1s) to original frame $\mathbf{Y}^{(t)}$ depending on embedding bit b:

$$Y_{i,j}'^{(t)} = \begin{cases} Y_{i,j}^{(t)} + f_{i,j}^{(t)} M_{i,j} & \text{if } b = 1 \\ Y_{i,j}^{(t)} - f_{i,j}^{(t)} M_{i,j} & \text{if } b = 0 \end{cases}. \tag{5}$$

4. Experimental Evaluation

4.1 Subjective evaluation of watermarked picture quality after MPEG encoding and decoding (i.e., picture quality that users actually see)

(1) Procedure

To evaluate the MPEG quality, we used 30 standard motion pictures[14], each of which had 450 frames of 720×480 pixels. After the previous and proposed methods were used to embed WMs into luminance set $\mathbf{Y}^{(t)}$ of the motion picture, the visual qualities of these two watermarked pictures were compared by standard subjec-

Table 2 Level of disturbance and rating scale.

Disturbance	Points
Imperceptible	5
Perceptible but not annoying	4
Slightly annoying	3
Annoying	2
Very annoying	1

tive evaluation[15] after MPEG2 encoding and decoding using a hardware codec.

Proposed method: WMs were embedded using the procedures described in Section 3.2 with embedding bit $b = 1$.

Previous method: The watermarking procedures were the same as those of the proposed method except that $S_{i,j}$ instead of $R_{i,j}$ was used in Step 4. The embedding quantity for each frame was set the same as that with the proposed method. The difference between $R_{i,j}$ and $S_{i,j}$ results in different distributions of WMs among the pixels.

The 30 sample pictures were categorized into five types according to the degree of object movement between frames and the degree of object deformation between frames (see **Fig. 8**).

(a) Walk through the Square (Walk): People walking in a town square: medium movement and medium deformation.

(b) Flamingos: Pan-scanned scene of flamingos: much movement and little deformation.

(c) Whale Show (Whale): Spraying whale with audience: little movement and much deformation.

(d) Entrance Hall (Hall): Three people talking in an entrance hall: little movement and little deformation.

(e) Rustling Leaves (Leaves): Fluttering leaves: much movement and much deformation.

For the proposed method, we used the threshold values ($T_v = 6$, $T_d = 5000$) determined in Section 3.1 (4) and the values of α listed in **Table 1**. The pictures were subjectively evaluated using the procedure described in Recommendation ITU-R BT.500-7[15]. The original and watermarked pictures after MPEG encoding and decoding were shown simultaneously to ten evaluators, who rated the picture quality according to the scale listed in **Table 2**. For each picture, the average of the ten scores was used as the quality level.

(2) **Results**

Evaluation results are shown in **Fig. 9**, in which the horizontal axis represents averaged embedding quantity per pixel q:

$$q = \frac{1}{720 \times 480} Q,$$

and the vertical axis represents the level of picture quality. From Fig. 9 we can see the following:

- Fig. 9(a1)-(a4) show the results for Walk (the standard) at the four different bit rates. The degradation in picture quality was significantly reduced at 4, 6, and 8 Mbps (Fig. 9(a2), (a3), and (a4)) because the proposed method preferentially allocates WMs to the moving areas where they should be less perceptible. This is illustrated in **Fig. 10**, which shows the strength of watermarking in a frame. The proposed method was not effective at 2 Mbps (Fig. 9(a1)) because the noise with low-bit-rate MPEG dominates the WM noise, preventing the WM noise from being recognized. In the standard case, the proposed method can embed 1.3 times as many WMs at the same picture quality.
- Fig. 9(b) and (c) show that the proposed method also significantly reduces degradation in picture quality (only the results for 8 Mbps are shown due to space limitations).
- The proposed method was not as effective for the pictures in which there was little movement and deformation (Fig. 9(d)) because most parts of the frames were static, so it is difficult to preferentially allocate WMs to parts with motion.
- For the same reason, the proposed method was not effective for the picture with much movement and deformation (Fig. 9(e)) (i.e., there was little basis for preferential allocation because most parts of its frames were "busy").
- For all the pictures evaluated, the quality of the proposed method was equal to or better than that of the previous method. This shows the effectiveness of the proposed method.

4.2 Estimation of WM detection rate after MPEG encoding and decoding

This section describes the results of evaluating the proposed method with respect to improvement in WM detection by comparing the WM detection ratios of the proposed method and the previous method. The ratios were measured at the same picture quality after MPEG encoding.

(1) **Procedure**

We evaluated the detection rate using the standard picture (Walk-through-the-Square).

Step 1: Select the level of picture quality, L, at which the evaluation is to be done (for example, $L = 4$, "Perceptible but not annoying"). Obtain aver-

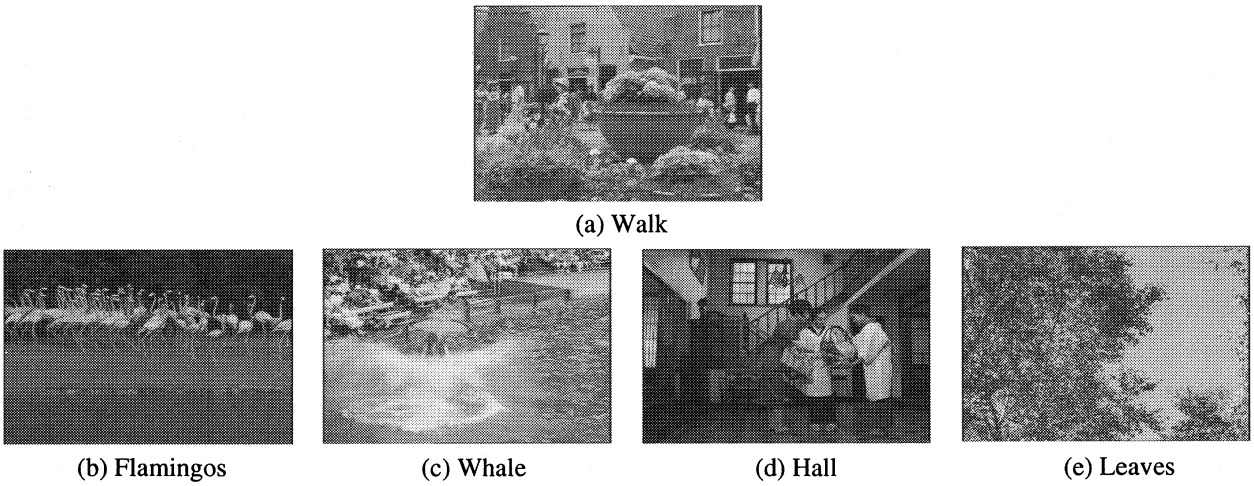


Fig. 8 Evaluated pictures.

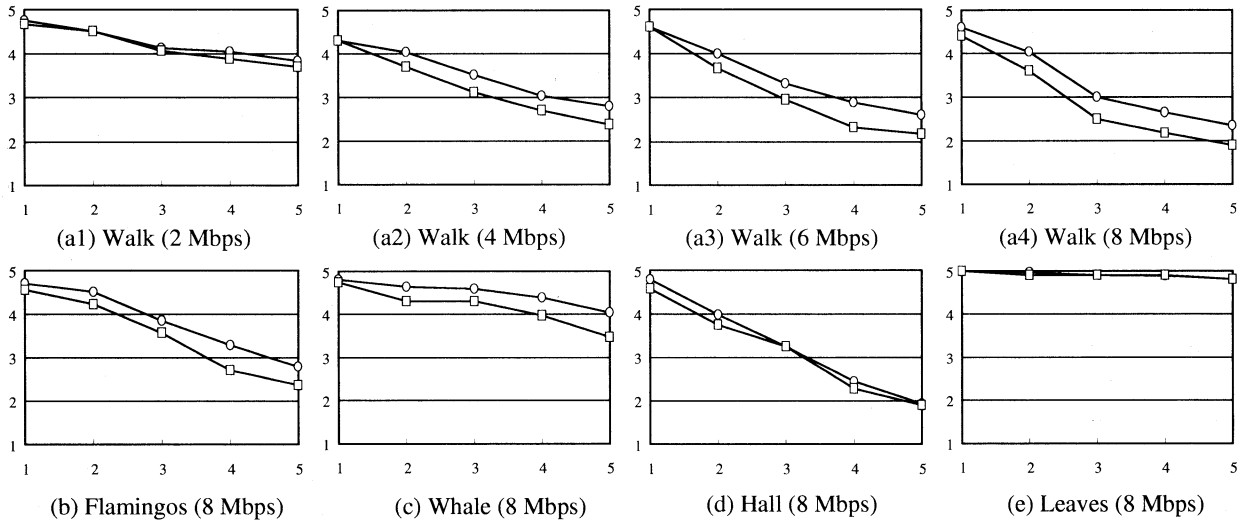


Fig. 9 Quality of watermarked pictures (circles and squares indicate proposed and previous methods).

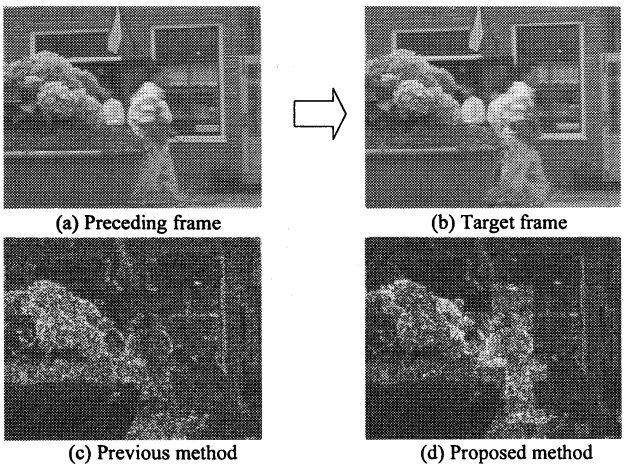


Fig. 10 Strength of watermarking (white indicates strong watermarking).

age embedding quantity $q^{(L)}_{\text{proposed}}$ for the proposed method that corresponds to L by referring to the relationship between q and picture quality shown in Fig. 9. Similarly obtain average embedding quantity $q^{(L)}_{\text{previous}}$ for the previous method. For instance, for Fig. 9(a4), $q^{(4)}_{\text{proposed}}$ and $q^{(4)}_{\text{previous}}$ are, respectively, 2.0 and 1.5.

Step 2: Embed 64-bit information into each of the 450 frames of the picture using the proposed and previous methods so that the average embedding quantities are, respectively, $q^{(4)}_{\text{proposed}}$ and $q^{(4)}_{\text{previous}}$ (resulting in two watermarked pictures with the same quality).

Step 3: Encode and decode the watermarked pictures using an MPEG2 hardware codec.

Step 4: Using a process based on WM detection described in the appendix, detect the 64-bit information in each frame of the two watermarked pictures.

Step 5: Count the number of frames in which the

Table 3 Percentage of correct detection.

	2 Mbps	4 Mbps	6 Mbps	8 Mbps
Proposed	36.0	51.6	57.6	76.7
Previous	34.2	37.6	49.1	55.3

information was correctly detected.

This procedure was performed for the four different MPEG2 bit rates (i.e. 2, 4, 6, and 8 Mbps).

(2) **Results**

Evaluation results are summarized in **Table 3**, which lists the percentages of the 450 frames from which information was detected correctly. At all MPEG2 bit rates, the proposed method gave better detection ratios *. Thus the proposed method can improve the WM detection.

4.3 Application

As mentioned above, the proposed method would be most effective when implemented in MPEG encoders. Video recorders, such as personal video recorders and DVD video recorders, are widely used in the field of A/V consumer products. These recorders, however, can be used to illegally copy, redistribute, and play video. Implementation of the proposed method in video recorders and players would help prevent such illegal activities by (1) embedding authentication information in video and (2) detecting the embedded authentication information and using it to control playback.

Fig. 11 shows the structure of a video recorder in which the proposed method is implemented. The recorder encodes signals output from a TV tuner, VCR, DVD, or camcorder and stores the encoded data on a recording medium, such as an HDD, DVD-RAM/R, or memory card. WMs representing authentication information are embedded into the luminance signal (Y) of the video during MPEG encoding, and the watermarked MPEG bit stream is recorded on the recording media. When the recorded video is played on a player, which could be the same device as the recorder if it includes a playback function, the player detects the WMs from the MPEG-decoded signal (Y) of the video. It then controls the playback based on the authentication information it detected. For example, the recorder's ID or time code might be used as the authentication information. The player could then stop playback of video that was recorded by another recorder and illegally copied, or it could stop playback when the viewing term expires. Similar application is possible in content-distribution systems for video via network such as content servers and client PCs.

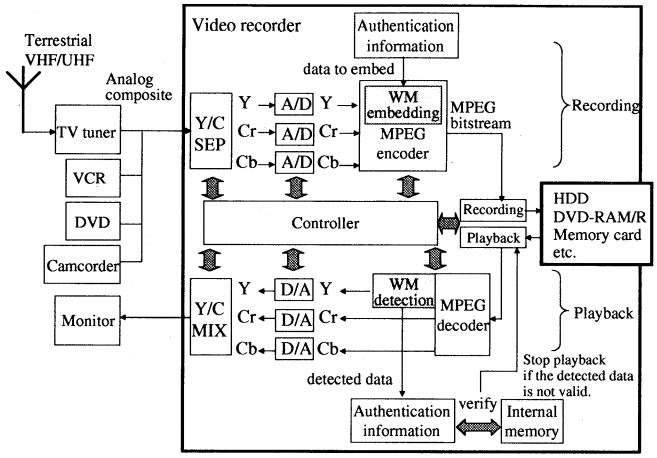


Fig. 11 System structure of video recorder using proposed method.

5. Conclusion

Maintaining picture quality is an essential requirement for video watermarking. This paper showed that watermarks are less perceptible where picture contents are moving than where they are static. It then described a new watermarking method that uses motion detection to preferentially allocate watermarks to moving parts of pictures. Also described were a motion-based criterion for estimating watermark imperceptibility, a method for controlling watermark allocation, and an implementation of the proposed method in an MPEG2 encoder. Experimental evaluations demonstrated that degradation in picture quality was significantly reduced with the proposed method and that the proposed method can, for a given picture quality, embed 30% more watermarks than the previous method. The proposed method would be most effective when implemented in MPEG encoders because they already have the motion analysis functions needed. Future work will focus on improving the criterion for watermark imperceptibility and evaluate watermark survivability when image processing is other than MPEG2 (i.e., MPEG4, resolution change).

We thank Dr. Nagaaki Ohyama, Dr. Toshiya Itoh, Dr. Hiroshi Nagahashi, Dr. Takao Kobayashi, and Dr. Masahiro Yamaguchi of the Tokyo Institute of Technology for their insightful comments, suggestions, and constant support. We are also grateful to Mr. Yukio Fujii, Mr. Yoshifumi Fujikawa, Mr. Koichi Terada, and Mr. Takayoshi Yahata of Hitachi, Ltd. for their de-

* These detection ratios are too low for practical use but they can be improved sufficiently by using a better coding technique, such as redundant coding or error-correction coding.

tailed comments.

〔References〕

1) Swanson et al.: Multimedia data-embedding and watermarking technologies, *Proc. IEEE*, Vol. 86, No. 6, pp. 1064–1087 (1998).
2) Delaigle et al.: Watermarking algorithm based on a human visual model, *Signal Processing*, Vol. 66, pp. 319–335 (1998).
3) De Vleeschouwer et al.: Invisibility and application functionalities in perceptual watermarking–An overview, *Proc. IEEE*, Vol. 90, No. 1, pp. 64–77 (2002).
4) Echizen et al.: General quality maintenance module for motion picture watermarking, *IEEE Trans. Consumer Electronics*, Vol. 45, No. 4, pp. 1150–1158 (1999).
5) Echizen et al.: Estimating the Bit-error-rate in Digital Watermarking Using Inferential Statistics, *IPSJ Journal*, Vol. 42, No. 8, pp. 2006–2016 (2001).
6) Hartung et al.: Digital watermarking of raw and compressed video, *Proc. SPIE*, Vol. 2952, pp. 205–213 (1996).
7) Kundur et al.: Digital watermarking using multiresolution wavelet decomposition, *Intl. Conf. Acoustics, Speech and Signal Processing*, Vol. 5, pp. 2969–2972 (1998).
8) Sakai et al.: Evaluation of digital watermarking based on frequency transform, *Technical Report of IEICE*, ISEC 97–22, pp. 87–96 (1997).
9) Kobayashi et al.: DataHiding based on Neighbor Pixels Statistics, *Proc. IPSJ 56th Annual Conf.*, 1V–03 (1998).
10) Bender et al.: Techniques for data hiding, *Proc. SPIE*, Vol. 2020, pp. 2420–2440 (1995).
11) Dufaux et al.: Motion estimation techniques for digital TV: a review and new contribution, *Proc. IEEE*, Vol. 83, No. 6, pp. 858–876 (1998).
12) Kojima et al.: A Vision and Architecture of Media-processor (MAP), *Proc. IPSJ 59th Annual Conf.*, 1C–01 (1999).
13) Ishiguro et al.: MPEG-4 Simple Profile Video Encoder on MAPCA, *Proc. 2001 ITE Annual Convention*, 8–2 (2001).
14) The Institute of Image Information and Television Engineers, Evaluation video sample (standard definition).
15) Rec. ITU-R BT.500-7, Methodology for the subjective assessment of the quality of television pictures (1995).

〈Appendix〉

The process flow of the 1-bit-WM detection associated with the WM embedding described in Section 3.2 comprises the following steps:

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

Step 2: Calculate bit-decision value v by correlating WM pattern $\mathbf{M}$ with watermarked image $\mathbf{Y}'^{(t)}$. That is,

$$\begin{aligned} v &= \frac{1}{mn}\sum_{ij} M_{i,j} Y'^{(t)}_{i,j} \\ &= \frac{1}{mn}\sum_{ij} M_{i,j}(Y^{(t)}_{i,j} \pm f^{(t)}_{i,j} M_{i,j}) \\ &= \frac{1}{mn}\sum_{ij} M_{i,j} Y^{(t)}_{i,j} \pm \frac{1}{mn}\sum_{ij} f^{(t)}_{i,j}, \end{aligned}$$

where $M_{i,j}$ are random arrays of ± 1. Thus, v is expressed by

$$v \simeq \pm \frac{1}{mn}\sum_{ij} f^{(t)}_{i,j}.$$

Step 3: Determine bit-value b by comparing v with threshold value $T(>0)$:

$$b = \begin{cases} 1 & \text{if } v \geqq T \\ 0 & \text{if } v \leqq -T \\ \text{“not detected”} & \text{if } -T < v < T \end{cases}.$$



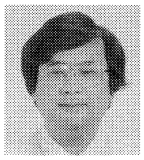
Isao Echizen received B.S., M.S., and Ph.D. degrees from the Tokyo Institute of Technology, Japan, in 1995, 1997, and 2003. In 1997 he joined the Systems Development Laboratory of Hitachi, Ltd., Yokohama, Japan, and continued research on information security and copyright protection, especially digital watermarking.



Hiroshi Yoshiura received B.S. and Ph.D. degrees from the University of Tokyo, Japan, in 1981 and 1997. In 1981 he joined Hitachi, Ltd. and was a Senior Research Engineer in the company's Systems Development Laboratory until March 2003. He is currently an associate professor in the Department of Human Communications, the Faculty of Electro-Communications, the University of Electro-Communications. His research interests include network security and copyright management technologies.



Kousuke Anzai received B.S. and M.S. degrees from Tohoku University, Japan, in 1996 and 1998. In 1998 he joined the Systems Development Laboratory of Hitachi, Ltd., Yokohama, Japan, and continued research on digital watermarking.



Jun'ichi Taguchi received B.S. and M.S. degrees from Kyushu University, Japan, in 1984 and 1986. He received a Ph.D. degree from Tohoku University, Japan, in 1998. In 1986 he joined the Systems Development Laboratory of Hitachi, Ltd., Kawasaki and Yokohama, Japan, and continued research on magnetic resonance imaging, remote sensing, signal processing, and digital watermarking as a Senior Researcher.



Kazunori Nakano received B.S. and M.S. degrees from Yamagata University, Japan, in 1990 and 1992. In 1992 he joined the Systems Development Laboratory of Hitachi, Ltd., Yokohama, Japan, and continued research on image processing techniques for remote sensing and digital watermarking.



Satoru Tezuka received B.S. and Ph.D. degrees from Keio University, Japan, in 1984 and 2000. In 1984 he joined Hitachi, Ltd. and has been engaged in research of information security, public key infrastructure, and copyright protection. He is now a Senior Research Manager in the Systems Development Laboratory, Yokohama, Japan.