MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11-13, 2002, Nara-ken New Public Hall, Nara, Japan

13—9

A Watermarking Scheme to Multimedia Contents Using Simple-scrambler and Band Division.

Munetoshi Iwakiri *
Department of Computer Science
National Defense Academy

Shinji Ozawa †
Department of Information and Computer Science
Keio University

Kineo Matsui ‡
Department of Computer Science
National Defense Academy

Abstract

This paper presents a digital watermarking technique for multimedia contents. It is mainly composed of two steps: the first is to divide the frequency band into several sub-bands and the second is to embed a watermark into one of these bands by the spectrum spreading method using simple-scrambler. The proposed scheme makes a watermark robust against data compression, and it keeps the contents in high quality because we select a specified band among many sub-bands at random to embed the watermark. For example, our experimental results show that the quality of the watermarked contents is good in about 45dB and the watermark signals can be detected about 90% under the popular data compressions.

1. Introduction

A serious problem has arisen in copyright protection of the digital multimedia contents from the networked worlds, because the digital contents are easily copied out of any permission and widely distributed without the degradation. As one of the solution for this problem, a watermarking technique is introduced recently. This is to embed some copyright information into the digital contents so that any user may not perceive it in advance while the content is circulated, and is to activate the watermark if an illegal use would be detected in the network.

Watermarking techniques[1] proposed now are most applied for the contents such as image, video and audio, respectively. There are a few studies for multimedia scheme up to date, but severe conditions for the multimedia contents are imposed on the watermark in practical use. In this paper, we try to improve this scheme by using the sub-band division technique so that the embedded watermark becomes robust under the data compression.

First, a series of multimedia data are classified into two sets of component in high and low frequency ranges based on sub-band technique. This process is recurrently applied to each low band and it is divided into the specified sub-bands. Next, a band is chosen among them and a watermark is embedded into the band. Moreover, the spectrum components of watermark are spread onto the components in low frequency domain, and it generates only a bit of noise which we may not find.

2. Proposal Technique

This section describes mainly a watermarking technique to the digital audio based on the procedure of sub-band system with scrambing, and the technique is applied to the digital image watermarking.

2.1 Band Division

We first define several symbols in this subsection; let $l_0(t)$, $l_1(t)$ and $h_1(t)$ be sampled sequences of component L_0 in the original band, L_1 in the low frequency band and H_1 in high frequency band at time $t = 0, 1, 2, \cdots$, respectively, where $l_1(t)$ and $h_1(t)$ are written by $l_0(t)$ as

$$l_1(t_1) = \frac{l_0(2 \cdot t_1) + l_0(2 \cdot t_1 + 1)}{2} \quad (1)$$

$$h_1(t_0) = l_0(t_0) - l_1(\lfloor \frac{t_0}{2} \rfloor) \quad (2)$$

where $t_1 = t_0/2$. Equations (1) and (2) show that L_1 becomes a half of L_0, which is a low frequency

*Address: 1-10-20 Hashirimizu, Yokosuka-shi, Kanagawa 239-8686 Japan. E-mail: iwak@nda.ac.jp
†Address: 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522 Japan. E-mail: ozawa@ozawa.ics.keio.ac.jp
‡Address: 1-10-20 Hashirimizu, Yokosuka-shi, Kanagawa 239-8686 Japan. E-mail: matsui@nda.ac.jp

band, and H_1 consists of high-frequency components as well as those in L_0. Repeating these frequency divisions, we have the following sequences step by step as

$$\begin{aligned} L_0 &\rightarrow \{L_1, H_1\} \\ L_1 &\rightarrow \{L_2, H_2\} \\ &\vdots \\ L_{j-1} &\rightarrow \{L_j, H_j\} \end{aligned} \quad (3)$$

From the disjoint bands of L_1 and H_1, inversely, we have the original band L_0 by

$$l_0(t_0) = l_1(\lfloor \frac{t_0}{2} \rfloor) + h_1(t_0) \quad (4)$$

where $t_0 = 0, 1, 2, \cdots$. In the same way, we restore the upper bound L_{i-1} from both the sub-band L_i and H_i.

2.2 Scrambling

We first define several symbols in this subsection; let the scrambling table $F = \{f(g), 0 \le g \le N-1\}$ of length N be the random permutated numbers(e.g., $N = 10$, $F = \{1,6,2,3,4,0,9,8,7,5\}$) as a scrambling key. Then we have a scrambled signal $x(g)$ from input signal $s(f(g))$, i.e.,

$$x(g) \leftarrow s(f(g)),\ g = 0, 1, \cdots, N-1. \quad (5)$$

On the other hand, we have an inverted signal $s'(f(g))$ from scrambled signal $x(g)$, i.e.,

$$s'(f(g)) \leftarrow x(g),\ g = 0, 1, \cdots, N-1. \quad (6)$$

We write the scrambling as SCR and the inverse scrambling as ISCR, respectively. It is effective that SCR diffuses the spectrum of the signal.

2.3 Watermarking Method

A schematic diagram of our watermarking method is shown in Fig.1. First, a source data S is a sequence of signals $s(t)$ for $t = 0, 1, 2, \cdots$, where S is equal to L_0 in this paper, and it is divided into subbands L_j and $\{H_1, H_2, \cdots, H_j\}$. Next, we apply the scrambling scheme (i.e., SCR) to L_j so as to diffuse a watermark signal into a specified band.

The scrambled signal $x(t_j)$ is, next, converted in the frequency area by the modified discrete cosine transformation(i.e., MDCT)[2]. This transformation requires $2M$ samples at t_j and generates M coefficients. Let $X_i(k)$ be an MDCT coefficient at i-th frame.

Now, we define a random number r_i that specifies a frequency component to embed a watermarking bit b_i, and a positive real number p that controls

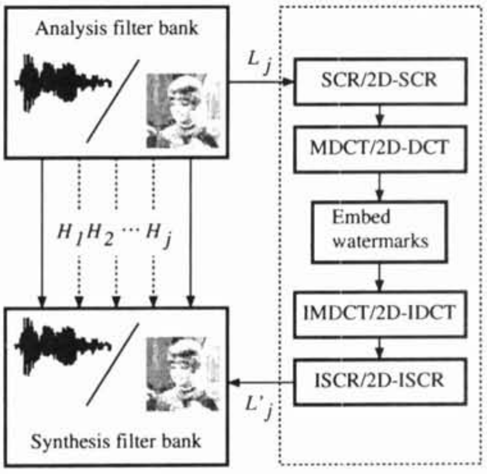

Figure 1: The processing block for embedding.

the watermarked level, where p is experimentally estimated under considering the sound quality and the robustness of watermark. Let R stand for a set of $\{r_i | i = 0, 1, ..., M-2\}$. It will be a key for restoring the embedded signals to the watermark. Next $X_i(r_i)$ are selected as shown in Fig.2 and it will be quantized by the absolute value of the scale parameters δ_i, it is caluculated as follows:

$$\delta_i = \frac{p}{M} \sum_{k=0}^{M-1} |X_i(k)|. \quad (7)$$

The adaptive control of the scale parameter can be resist the level modulation attack than the ordinary techniques using the fixed quantization. Then the following algorithm is applied:

Embedding algorithm

(a) To embed $b_i = 0$:

If the round number of $< |X_i(r_i)|/\delta_i >$ is even, then $X_i(r_i)$ remains as it is, otherwise $X_i(r_i)$ is exchanged with the $< |X_i(r_i)|/\delta_i >$ in even number, where the symbol $<$ (real number) $>$ means a round operation of the real number to the nearest integer.

(b) To embed $b_i = 1$:

If the round number of $< |X_i(r_i)|/\delta_i >$ is odd, then $X_i(r_i)$ remains as it is, otherwise $X_i(r_i)$ is exchanged with the $< |X_i(r_i)|/\delta_i >$ in odd number.

This completes the embedding process when the $|X_i(r_i)|$ is replaced with the corresponding absolute values of MDCT coefficients. Thus, let $X'_i(k)$ be the watermarked $X_i(k)$.

The watermarked componets $X'_i(k)$ is converted to the watermaked sampling sequences $x'(t_j)$ at time

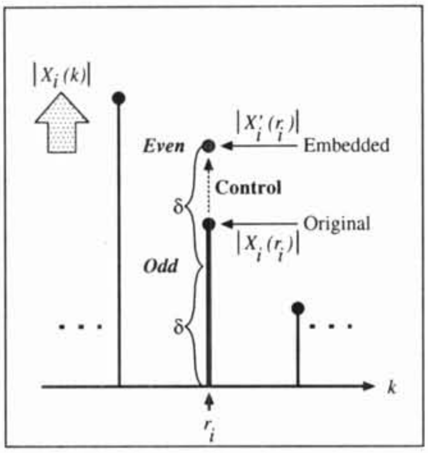

Figure 2: The embedding control($b_i = 0$).

t_j by the invert MDCT(i.e.,IMDCT)[2]. Since using ISCR, we get inverted scrambling sequence $x'(t_j)$, and it is composed of $l_j(t_j)$ with watermark signal.

This processing is applied to the component L_j. Then we have the watermarked audio data S' from L'_j and $\{H_1, H_2, \cdots, H_j\}$. Using the sampling rate B of original audio and embedding parameter (j, M), the watermarking bit rate β is calculated by

$$\beta = \frac{B}{M \cdot 2^j}. \tag{8}$$

For example, when $B = 44,100$[Hz], $j = 10$ and $M = 16$, we say that the embedded rate is $2.69 (= 44,100/(16 \times 2^{10}))$[bit/s].

2.4 Decoding Method

First, let us divide watermarked audio data S' into sub-bands L'_j and $\{H'_1, H'_2, \cdots, H'_j\}$. We get $l'_j(t_j)$ from L'_j and $x'(t_j)$ using SCR with scrambling key F. Then, the MDCT coefficients $X'_i(r_i)$ of the L'_j are checked using the scale parameters

$$\delta'_i = \frac{p}{M} \sum_{k=0}^{M-1} |X'_i(k)| \tag{9}$$

and we estimate the embedded signal by the following step ; i.e., if

$$< \frac{1}{\delta'_i} |X'_i(r_i)| > \quad : \quad \text{Even number}$$

then we have $b'_i = 0$, otherwise $b'_i = 1$.

Note that nobody can restore the watermark unless both the scale parameter p and the scrambling key F with r_i for decoding are identical to the one for scrambling, and R is legal for the scheme.

2.5 Applied to Image Watermarking

The difference between the digital audio and image is only the structure of the sampled data in this technique, and thus the above elementary watermarking schemes are easily applied to the digital image watermarking, that is, only the processing in a single dimension is expanded to two-dimensional(2-D) such as 2-D sub-banding, 2-D scrambing , 2-D discrete cosine transformation, and so on.

3. Experimental Results

We study the effect of watermark on digital multimedia contents(i.e., Audio and Image) quality and that of lossy data compression techniques such as MPEG1 audio LayerIII(i.e., MP3) and JPEG[2].

The digital data listed in Table 1 are used for our experiments. The experimental audio data(i.e., Classic, Jazz and Dance) are digitized at the rate of 44.1 kHz and are expressed in 16 bits per each sample. Ordinarily, these data are constituted in stereo-typed format, but we use only a single channel(right side) of stereo audios in order to simplify the discussion. On the other hand, the experimental image data(i.e., Girl, Moon and Facs) are in the standard image data base(i.e.,SIDBA).

3.1 Audio Quality and Robustness for the MP3 compression

We introduce the 32[ms] segmental signal-to-noise ratio[3], i.e., $SNR_{seg}(32[ms])$, corresponding to subjective evaluation, approximately. Since the watermarking bits are embedded to L_j component, it seems to be able to hide the watermark inaudible. The audio qualities SNR_{seg} to p are also shown in Fig.3 when $j = 10$ and $M = N = 16$(i.e.,β=2.69[bit/s]), and those are deteriorated as p is greater. For example, we have about 45[dB] in the audio level for $p = 0.8$, but we could not detect any noise on the audio.

Next, we check our scheme for the MP3, data compression method under the compression rate of

Table 1: The experimental data.

Name	Type	Samples	Physical size
Classic	Audio	440,832	10[sec.]
Jazz	Audio	440,832	10[sec.]
Dance	Audio	440,832	10[sec.]
Girl	Image	65,536	256×256[pixel]
Moon	Image	65,536	256×256[pixel]
Facs	Image	65,536	256×256[pixel]

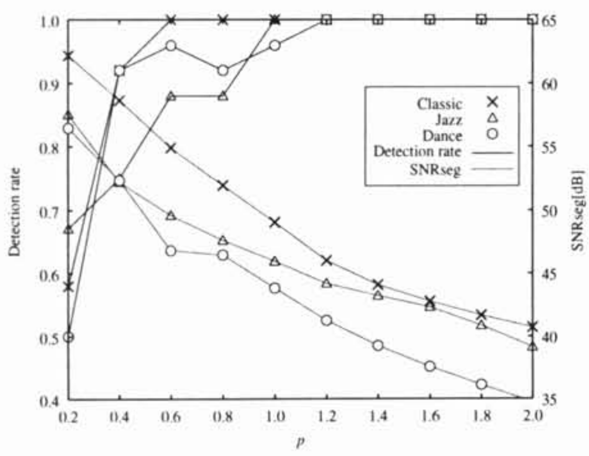

Figure 3: Detection rate with SNR_{seg} to p

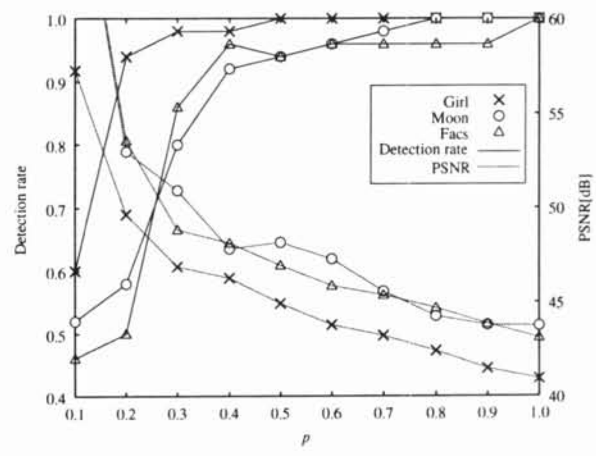

Figure 4: Detection rate with PSNR to p

1/10. A detection rate is defined with the rate of which the decoded bits in the watermark sequence coincide with the embedded ones. For example, when a single bit is not matched in the sequence of 25 bits, we say that the detection rate is $0.96 (= 24/25)$. The experimental results are shown in Fig.3, where the watermark signals are embedded for p into each sample under $j = 10$ and $M = N = 16$. Then, the samples are compressed by the MP3 algorithm. Our results show higher detection rate about 90[%] in $p \geq 1.0$, and the detection rate is usually satisfactory for copy-protection with watermarking.

3.2 Image Quality and Robustness for the JPEG compression

Since the watermarking bits are embedded to L_j component, it seems to be able to hide the watermark invisible. We introduce the Peak Signal-to-Noise Ratio, so called PSNR[1], corresponding to subjective evaluation, approximately. The image qualities PSNR to p are also shown in Fig.4 when both the size of the 2-D scrambling and DCT are equal to the sub-band component $j = 2$(i.e., L_2 : 64×64). Those PSNR are deteriorated as p is greater, where 50 bits of watermark are embedded into DCT coefficient selected at random. For example, we have a good picture in 45[dB] or more under $p \leq 0.4$, and thus no one can find any noise on the image.

Next, we check our scheme for JPEG, where the "`-quality`" is set to "20" for "`convert`" command in ImageMagick(ver.4.2.8) tools§ . The experimental results for p are shown in Fig.4, where the watermark signals are embedded into each sample under the above conditions. Then, the samples are compressed by the JPEG algorithm. Our results show higher detection rate about 90[%] in $p \geq 0.4$, and these would be satisfactory for the usual copy-protection system.

§ http://www.simplesystems.org/ImageMagick/
The "`-quality`" is selected from "0"(*worst*) to "100"(*best*) for the quality control of the decoded JPEG image.

4. Conclusion

This paper proposes a watermarking scheme to digital multimedia contents in high quality. It is robust for popular data compression because the watermarks are spread and concealed from illegal spectrum analysis. It is important in our scheme that we have to select a set of control parameters $\{j, p, M\}$ in optimal under preliminary tests. Increasing the detection rate of watermark will be required into practical use with the incorporation of error-correcting technique and adaptive control method. It will be available for detecting an illegal copy under the internet environment. Furthermore, we must study a protection scheme in this method against the band-pass filtering, resampling, and so on, and it is important to try to estimate the watermarked contents quality by introducing the Mean-Opinion-Score as subjective evaluation.

References

[1] Katzenbeisser,S., Petitcolas,F.A.P.: Information Hiding Techniques for Steganography and Digital Watermarking, Artech House, 2000.

[2] K.R.Rao, J.J.Hwang, "Techniques and Standards for Image," Video and Audio Coding, Prentice Hall, 1996.

[3] K.Ozawa, "The high efficiency sound coding techniques for digital movile-communication," Triceps Pub. Co., 1992(in japanese).