ACOUSTICAL LETTER

Reduction of distributed data size in audio content fingerprinting (CoFIP)

Kotaro Sonoda*, Ryouichi Nishimura† and Yôiti Suzuki‡

Research Institute of Electrical Communication/Graduate School of Information Sciences, Tohoku University, Katahira 2–1–1, Aoba-ku, Sendai, 980–8577 Japan

(*Received 11 January 2005, Accepted for publication 14 March 2005*)

Keywords: Watermark, Music signal, Time-spread echo
PACS number: 43.60.Dh [DOI: 10.1250/ast.26.362]

1. Introduction

Recently, the contents of digital multimedia are widely distributed through the Internet due to the progress of broadband networking technologies. As a result, management and protection of the copyright of such contents have become highly problematic. Fingerprinting, by which a unique fingerprint is stamped on each digital multimedia content, is a possible solution to such problems. However, this is difficult to realize in the case of a non-interactive delivery system, e.g., a broadcast system, because the broadcasted data are common among all users.

To solve this problem, "Content Fingerprinting (CoFIP)" has been proposed [1]. Though CoFIP was originally proposed for still images, it has been also implemented for audio signals [2]. In the original CoFIP algorithm, contents are distributed to users in the form of a package consisting of encrypted watermarked component objects (parts) and an unencrypted remaining background object (body). A user purchases a digital key to decrypt only one combination of the watermarked objects from the package to generate an individualized content by combining those objects and the background object.

In original CoFIP, a distributed package containing $N \cdot M$ watermarked objects, i.e., M different watermarks for each of N component objects, and a body of background objects can generate M^N fingerprints. For example, if one hundred fingerprinted contents are required, two component objects with 10 variations of watermark (realizing 10^2 fingerprints) or five component objects with three variations of watermark (realizing 3^5 fingerprints) are required. In such cases, the distributed package in the conventional audio CoFIP [2] becomes 21 or 15 times larger in size than the original.

Since this is quite large for practical use, this paper details a method to reduce the size of distributed package by downsizing each object.

2. Conventional method

The generating process of conventional audio CoFIP shown in Fig. 1 [2] is as follows:

Step-1

An original signal (L samples) is transformed into a representation on a time-frequency plane. In conventional implementation, a scalogram by Daubechies wavelet transformation is used. This scalogram has $I(=\log_2 L)$ scale levels. The number of coefficients on the i-th scale level is $L/2^i$. The set of coefficients on the i-th scale level is represented as

$$s_i = \left\{ s_{i,1}, \ldots, s_{i,\frac{L}{2^i}} \right\}, \quad 1 < i < I.$$

Step-2

Scale levels that have the M highest energy are selected for component objects. If the l_j-th scale level is selected for a component object, scalograms of that object and the background objects would consist of the following coefficients:

$$p_i = \begin{cases} \left\{ s_{i,1}, \ldots, s_{i,\frac{L}{2^i}} \right\} & i = l_j \ (j = 1, \ldots, M) \\ \{0, \ldots, 0\} & \text{otherwise,} \end{cases}$$

$$b_i = \begin{cases} \{0, \ldots, 0\} & i = l_j \ (j = 1, \ldots, M) \\ \left\{ s_{i,1}, \ldots, s_{i,\frac{L}{2^i}} \right\} & \text{otherwise,} \end{cases}$$

where p_i and b_i are the set of coefficients on the i-th scale level of the component object and the background object, respectively.

Step-3

These scalograms are inversely transformed to generate the signals of the component objects and the background object.

Step-4

Each component object signal is fingerprinted by using a watermarking method with various keys. In conventional implementation, the time-spread echo watermarking method [3] is applied with various PN sequences. It is realized by convolving the signal of a component object with a multi-echo kernel, $k(n)$, which is constructed by using a PN sequence as follows:

$$k(n) = \delta(n) + \alpha \cdot p(n - \Delta), \quad 0 < \alpha < 1,$$

where $p(n)$ is a PN sequence whose amplitude is originally ± 1, α is the amplitude of the PN sequence, and $\delta(n)$ is the Dirac delta function.

*e-mail: kotaro@ais.riec.tohoku.ac.jp
†e-mail: ryou@ais.riec.tohoku.ac.jp
‡e-mail: yoh@ais.riec.tohoku.ac.jp

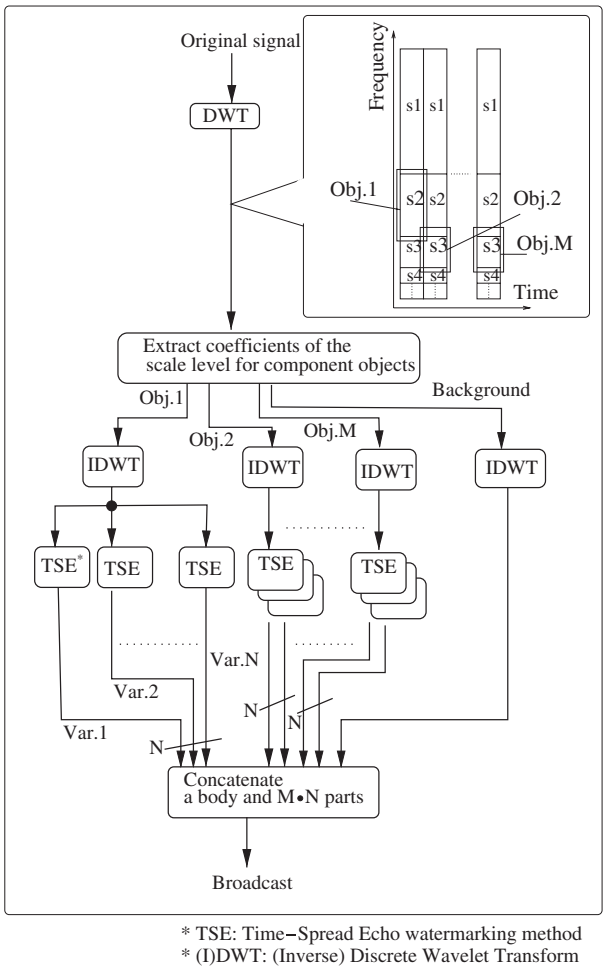

Fig. 1 Fingerprinting process in conventional audio CoFIP.

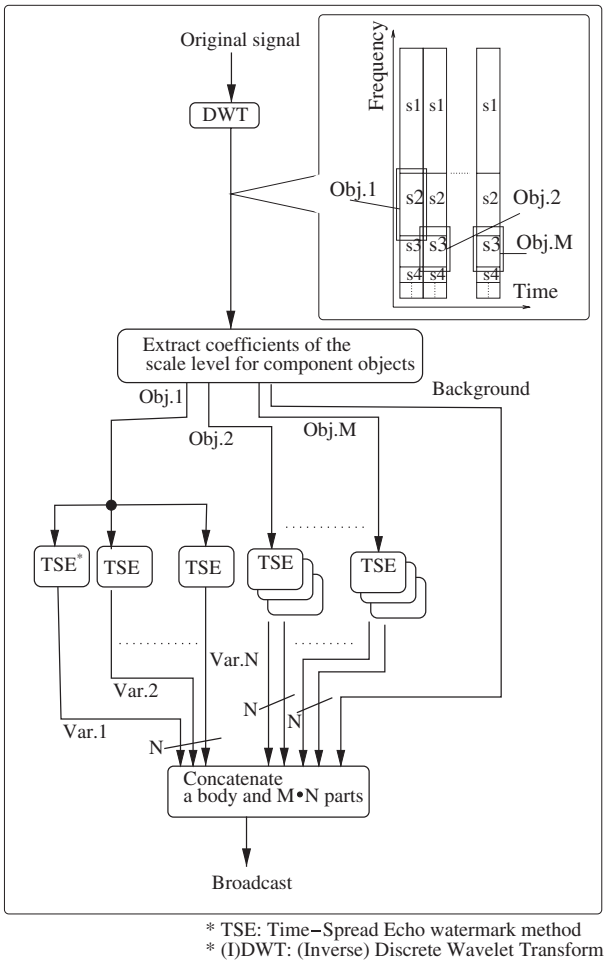

Fig. 2 Fingerprinting process in proposed method.

Step-5

These steps are repeated temporally while N component objects are being selected.

Step-6

These watermarked component object signals are packed together with the background body signal into the package which is to be distributed.

Accordingly, it should be noted that each object contained in a distributed package is of the same length of time as the original audio signal if the user is allowed to combine the watermarked objects with the background object. As a result, the size of distributed package increases $(M \cdot N + 1)$ times of original content.

3. Proposed method

We propose that watermarks be embedded into selected objects by directly applying the time-spread echo method to wavelet coefficients and composing the package data in the form of wavelet coefficients. In other words, the inverse wavelet transform to reconstruct the time domain signal in the third step of the above-mentioned steps is moved to the individualization process on user's side.

The proposed method is depicted in Fig. 2. The reduction ratio depends on which scale level of wavelet coefficients are used to embed watermarks. Because of the discrete wavelet transform, the data length of each scale level is reduced by 50% as the scale level increases. Hence, when the i-th scale level is used for a watermarked component object, the reduction ratio becomes $1/2^i$. The scale level employed is adaptively selected depending on the energy of each scale level of the original signal as in the conventional method. The chip rate of echo spreading is the same as the sample rate of the coefficients on the selected scale level. Therefore, when the i-th scale level is selected, the chip rate in the time domain becomes around $1/2^i$ times lower than that when the conventional method is used.

4. Experiment

In order to evaluate the performance of the proposed method, the size of the package to be distributed as well as the performance of watermark detection were numerically calculated using actual music signals. Conditions of the experiment are listed in Table 1. The length of the PN sequences were set so as to be a quarter of the length of the selected scale level. The amplitudes of the PN sequence were set to equalize the power of the echo kernel.

Figures 3 and 4 show the results of the experiment. The ordinate of Fig. 4 is the ratio of the signal (level of the cross-cepstrum between the cepstrum of watermarked signal and the PN sequence at a correct time lag) to noise (the second highest level of the cross-cepstrum at other time lags). The higher the

Table 1 Conditions to generate a distributor package.

Genre of original signal	classical and Japanese traditional (J-Traditional)
Format	RIFF wave, 44.1 kHz sampling, 16 bit, monaural
Wavelet	Daubechies 20th
Length of samples (L)	65,536
Variation of watermarks (M)	3
Number of partial objects (N)	1
In the Conventional method	
Scale level to be selected	The first to the third levels and one chosen from the forth to the seventh level
Length of PN sequence ($L_{\text{conv.}}$)	1,023
Amplitude of PN sequence ($\alpha_{\text{conv.}}$)	0.005
In the Proposed method	
Scale level to be selected	One chosen from the forth to the seventh level
Length of PN sequence ($L_{i,\text{prop.}}$)	$L/2^i/4-1$
Amplitude of PN sequence ($\alpha_{i,\text{prop.}}$)	$\alpha_{\text{conv.}} \cdot \sqrt{\frac{L_{\text{conv.}}}{L_{i,\text{prop.}}}}$ (To equalize the power of the echo kernel)

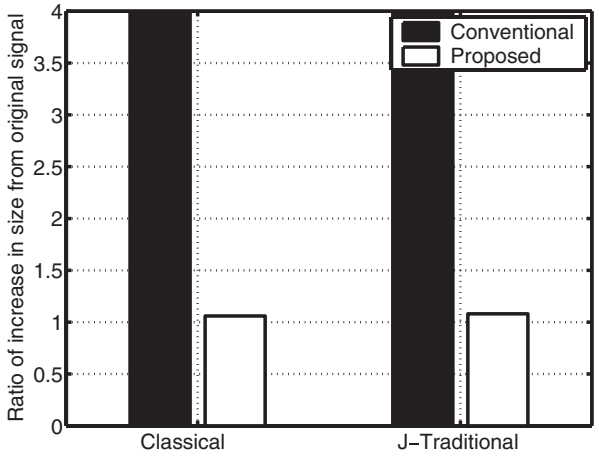

Fig. 3 Size of package to be distributed.

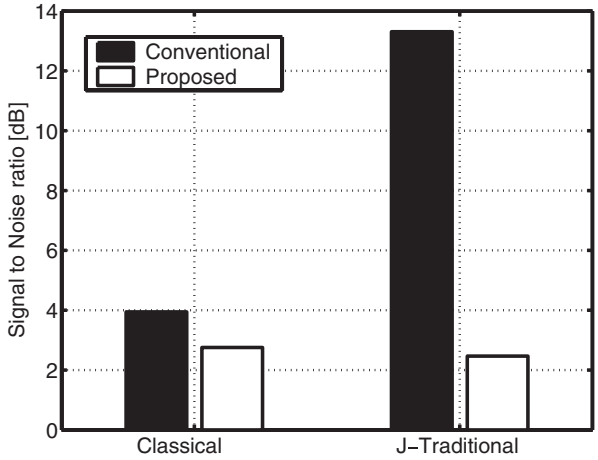

Fig. 4 Clarity of the detection of the embedded watermark.

value, the better the detection performance which can be expected.

As shown by Fig. 3, the size of package to be distributed can be 1.06 or 1.08 times larger than the original content for a piece of classical or Japanese traditional music, respectively, in the proposed method. These values are far less than that of the conventional method. Moreover, as shown by Fig. 4, the SN ratios for both pieces of music have positive values. Therefore, this reduction is realized without vital degradation in watermark detection.

5. Conclusion

A method to reduce the size of the distributed package in audio CoFIP was here in proposed. This method is realized by applying the time-spread echo method to the wavelet coefficients rather than to the time signal, as well as distributing the package in the form of wavelet coefficients. Applying this method to some actual music signals, it was possible to reduce the size of the package to be distributed more than half without significant deterioration of detection performance.

Acknowledgment

This study was partially supported by JST (Japan Science and Technology Agency) and Tohoku University 21st century COE (Center of Excellence) program.

References

[1] Y. Takahashi, T. Aoki and H. Yasuda, "Copyright Protection with CoFIP Scheme," *Forum on Information Technology 2002*, N-19, pp. 275–276 (2002).

[2] K. Sonoda, R. Nishimura, Y. Suzuki and T. Aoki, "Implementation of CoFIP for music signals," *18th Int. Congr. Acoust.*, pp. 1011–1014 (2004).

[3] B.-S. Ko, R. Nishimura and Y. Suzuki, "Time-spread echo method for digital audio watermarking using PN sequences," *IEEE Int. Conf. Acoust., Speech, and Signal Processing*, pp. 2001–2004 (2002).