A Reversible Audio Watermarking Scheme

Hyoung Joong KIM, Vasiliy SACHNEV, Ki Seob KIM
Department of Control and Instrumentation Engineering
Kangwon National University, Chunchon 200-701, Korea
E-mail: khj@kangwon.ac.kr, {vasys, kskim}@multimedia.kangwon.ac.kr

Abstract

A reversible audio watermarking algorithm is presented in this paper. This algorithm transforms the audio signal with the integer wavelet transform first in order to enhance the correlation between neighbor audio samples. Audio signal has low correlation between neighbor samples, which makes it difficult to apply difference expansion scheme. Second, a novel difference expansion scheme is used to embed more data by reducing the size of location map. Therefore, the difference expansion scheme used in this paper theoretically secures high embedding capacity under low perceptual distortion. Experiments show that this scheme can hide large number of information bits and keeps high perceptual quality.

Keywords: Reversible data hiding, integer wavelet transform, audio watermarking, data hiding.

I. INTRODUCTION

Recently, reversible data hiding schemes have got much attention. The main idea of reversible data hiding methods is allowing reconstruction of the original signal without any loss after extracting the hidden data. Reversible image watermarking algorithms are widely available. They embed data into images mostly quantized by 8-bit levels in case of grayscale images. Of course, color image watermarking is considered as an extension of the grayscale embedding. However, reversible audio watermarking algorithms are few in number. Basically audio is different from image. First, time-domain correlation between neighbor samples is low in audio. High correlation is important for compression and embedding data as well. Therefore, low correlation means that embedding data into audio signal is slightly more difficult than image. Second, sample values of audio signal mostly cluster around 0 which is a mean mostly, and most of their absolute values are less than 0.5, which is much favorable for bit shifting than image.

Reversible watermarking schemes are basically classified as follows:

1. *Companding and shifting scheme* [3, 4]: A sample value x can be shifted by n bit position as $(2^n)x$. For example, one bit shifting, $2x$, makes an even number whatever the original value x was. Thus, we can embed one bit of information, b, as $2x + b$. However, such bit shifting can cause overflow error. A compander, for example, $Q(\sqrt{x}/2)$,

can reduce the chance of overflow error because the value $2(Q(\sqrt{x}/2)) + b$, shifting after companding, is mostly smaller than x itself, where Q denotes a quantizer. However, companding error must be embedded as well because $4(Q(\sqrt{x}/2))^2$ equals hardly to x. If there is no companding error, the embedding capacity can be at least one bit per sample. Due to the second feature abovementioned, this technique is well fit for the audio watermarking.

2. *Difference shifting scheme* [2]: A difference of a pair of two sample values can be shifted to embed message. (Difference of a group of sample values can be used instead of a pair.) Not all pair can be shifted due to the possible overflow or underflow error. The location map marks which pairs have been shifted and which pairs have not. If the location map is not necessary to mark, the embedding capacity is at least one bit per two samples. Due to the first feature abovementioned, this technique is well fit for the image watermarking.

Of course, many kinds of novel reversible watermarking schemes are available. However, among them, in this paper, only two schemes abovementioned are considered. Companding and shifting scheme has been already available [3]. However, difference shifting scheme for audio watermarking has not been published so far. The difference expansion method, proposed by Tian [2], which is known as difference shifting scheme, can be used for embedding data into audio if underflow or overflow problem is solved. This overflow problem is serious since it causes perceptible distortion. In case of image and video, high correlation between neighbor samples produces small difference. Shifting small difference produces small perceptible distortion and reduces the chance of overflow and underflow problem. However, in case of audio, due to low correlation, difference between neighbor samples is relatively large, which may cause significant perceptible distortion and show high possibility of overflow or underflow problem.

In this paper, the difference expansion scheme is applied to the reversible audio watermarking. Correlation can be increased by applying any transform. Integer wavelet transform produces integer values, which makes the transform allow one-to-one mapping by producing integer output without causing any transform error. Transformed coefficients have higher correlation than original sample values. Thus, it is easy to apply difference expansion scheme in the transform domain. However, not all difference of pairs can be expanded. The location map marks which pairs have been expanded and which pairs have not. Decoder must have this location map to extract correct hidden data and recover original signal. To reduce the excessive overhead for the location map, the simplified location map is adopted. This paper shows that reversible audio watermarking is possible using difference expansion scheme. Moreover, this paper shows that the audio quality is very good after watermarking through experiments. Actual embedding capacity is also presented through experiments.

II. REVERSIBLE DIFFERENCE EXPANSION TRANSFORM

Digital audio signal $a(n)$ is a sequence of real numbers in the range of [-1;1]. The real numbers can be quantized into integer numbers represented by 16 bits ($2^{16} = 65536$ levels). The easiest quantization method is to quantize the real numbers into equivalent integer numbers as follows:

$$a_o(n) = \left\lfloor \frac{a(n)}{h} \right\rfloor, \tag{1}$$

where $a(n)$ is an original real audio sample value, h is a quantization step ($2/2^{16}$=3.0517×10^{-5}), $\lfloor x \rfloor$ is a floor operator, and $a_o(n)$ is an output integer audio sample value.

Reversible image watermarking schemes usually exploit high correlation between neighbor pixels in the time domain (or in the special domain). In case of audio samples the correlation between neighbor samples is not as high as images. So, it is difficult to use the difference expansion method in the time domain. In order to make the audio data more correlated, audio data in the time domain have to be transformed. The degree of correlation depends on the audio sequence. Due to the reversible watermarking, the integer wavelet transform is applied. Note that this transform eliminates ambiguity such that it maps the integer input to integer output by one-to-one manner.

Assume that we have two value pair (x, y), where $x, y \in Z$ in the transform domain and $0 \le x, y \le 65536$. We can define integer average value l and difference value h from the pair as follows:

$$l = \left\lfloor \frac{x+y}{2} \right\rfloor, \qquad h = x - y, \tag{2}$$

where the inverse transform of Equation (2) is given as follows:

$$x = l + \left\lfloor \frac{h+1}{2} \right\rfloor, \quad y = l - \left\lfloor \frac{h}{2} \right\rfloor \tag{3}$$

The reversible integer transforms in Equations (2) and (3) are also called the integer Haar wavelet transform or S transform. The reversible integer transforms allow a one-to-one mapping between (x, y) and (l, h). The new difference h' is computed by expanding the difference by a factor of 2 and embedding b as follows:

$$h' = 2 \cdot h + b \tag{4}$$

Note that the value $2h$ in Equation (4) produces even number regardless of whether h is odd or even. Thus, we have room to hide one bit of binary information b after expanding (or shifting) the difference by a factor of 2. For the expandable pairs, data embedding procedures are summarized as follows [2]:

$$x' = l + \left\lfloor \frac{h'+1}{2} \right\rfloor, \quad y' = l - \left\lfloor \frac{h'}{2} \right\rfloor \tag{5}$$

Audio data can be divided into two parts: perceptually sensitive part and non-sensitive part. The non-sensitive part of audio is better for embedding data, because the modified audio is relatively imperceptible. Embedding in the sensitive part must be careful and limited so that it is not perceptually noticeable. In most applications it is important that the watermark is undetectable to listeners. This ensures that the watermarked signal is not perceivably distorted and does not indicate the presence or location of watermarks. The perceptibility measure used in this paper is the SNR (signal-to-noise) expressed as follows:

$$SNR = 10 \cdot \log_{10} \left\{ \frac{\sum a^2(n)}{\sum (\tilde{a}(n) - a(n))^2} \right\}, \tag{6}$$

where $\tilde{a}(n)$ denotes the watermarked sample value.

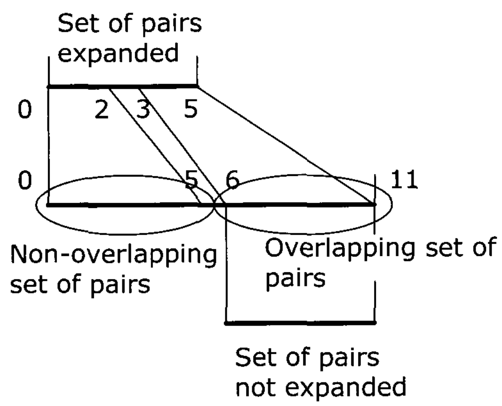

Figure 1. Set of ambiguous pairs (overlapping set of pairs) that requires the location map.

When we apply the difference expansion method, we need to mark which pair has been expanded and which pair has not. The location map maintains this information to allow correct decoding of embedded data and recovering original signal. The location map is a binary sequence for ambiguous sets where "1" denotes the expanded set and "0" for that has not been expanded. Assume that, for example, only the h values in between 0 and 5 will be expanded (see Figure 1). If the value is expanded, it produces the value in between 0 and 11 by Equation (4). After embedding, the values in between 0 and 5 are not ambiguous as long as they are expandable and changeable (refer to [1] for the definitions of expandability and changeability) because they have been expanded from the values in between 0 and 2 by Equation (4). However, the values in between 6 and 11 are ambiguous because the values have been expanded from the values in between 3 to 5, or they have not been expanded and they are as they have been. Among the ambiguous sets, the set S_o denotes the set of values not expanded, and the set S_e denotes the set of values expanded. Figure 1 shows the concept of this ambiguous set which occupies the overlapped part. One part comes from the expanded part and the other part comes from the original values not expanded.

Location map of Tian [2] has covered all pairs expanded set and all pairs of not expanded set regardless of their expandability and changeability. However, the simplified location map [1] is sufficient to cover only the ambiguous sets. Then, the location map size can be shrunk to the size of S_o plus the size of S_e. In the example, we consider the difference values to be expanded is in between 0 and 5. However, the difference value to be expanded can be decided to meet the desired embedding capacity. The location map can be compressed to maximize the embedding capacity [1, 2]. The detail of the difference expansion method can be found in [1, 2].

III. EXPERIMENTS

We apply the proposed method to five test sound tracks. For each track, we use 800,000 samples for the experiments. The audio signals used in the experiments are sampled at 44.1 kHz with 16 bits per sample.

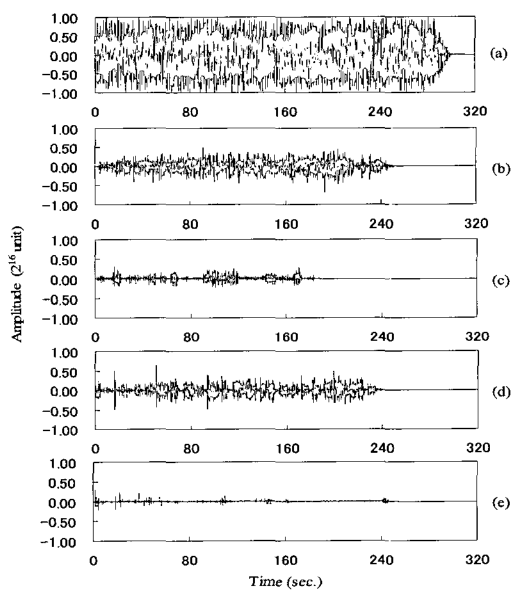

Figure 2. Five audio tracks for experiments.

As is mentioned above, the input audio signal is quantized, transformed using the integer wavelet transform, the differences of neighbor samples in the transform domain are expanded, and the simplified location is also embedded. We expand both low and high frequency coefficients first and only the high frequency coefficients second. Needless to say, both frequency coefficients allows more than double embedding capacity than high frequency coefficients only as is shown in Table 1. It shows that low frequency coefficients sometimes have better correlation than high frequency coefficients depending on audio tracks. SNR is over 30 dB for all experiments. In the experiments, the maximum embedding capacity is 37/80 bit per sample at 34 dB in the track 4. The track 4 is a performance by a traditional Japanese string instrument. Therefore, this audio track consists of monotone of the string sound mostly in the time domain, and consequently few coefficients in the frequency domain. Thus, its correlation in the frequency domain is relatively high. Figure 2 shows the audio clips for each track.

Track number	Both high and low frequencies		High frequency only	
	Payload [bits]	SNR [dB]	Payload [bits]	SNR [dB]
1	268753	32.0279	123453	36.0565
2	305333	36.4528	142041	41.7388
3	287596	31.5487	136257	38.4472
4	375670	34.0396	180319	41.7762
5	307378	30.1041	150535	35.9816

Table 1. Embedding capacity and SNR of the proposed watermarking scheme

IV. CONCLUSIONS

In this paper a reversible audio watermarking scheme is presented. This method is based on the integer wavelet transform and difference expansion scheme. This paper shows that the difference expansion scheme can be used for audio. This paper shows that embedding capacity and SNR can be controlled.

V. ACKNOWLEDGEMENTS

This work was in part supported by the BK21 Project, under the auspices of MOE, Korea.

REFERENCES

[1] H. J. Kim, and V. Sachnev, "A novel difference expansion transform for reversible data embedding," *Proc. WIAMIS*, Inchon, 2006.

[2] J. Tian, "Reversible watermarking by difference expansion," *Proc. Workshop Multimedia and Security: Authentication, Secrecy, and Steganalysis*, 19- 22 (2002)

[3] M. van der Veen, F. Bruekers, A. Leest, and S. Cavin, "High capacity reversible watermarking for audio," *Proc. SPIE*, **5020**, 1-11 (2003)

[4] G. Xuan, C. Yang, Y. Zhen, Y. Q. Shi, and Z. Ni, "Reversible data hiding using integer wavelet transform and companding technique," *Lecture Notes in Computer Science*, **3304**, 115-124, (also *Proc. IWDW*) (2004)

Biography

Prof. Hyoung Joong Kim

Hyoung Joong Kim received his B.S. in Electrical Engineering from the Seoul National University, Korea in 1978, and M.S. and Ph.D. degrees in Control and Instrumentation Engineering from theSeoul National University in 1986 and 1989, respectively. He joined the Department of Control and Instrumentation Engineering, Kangwon National University, Korea in 1989, where he is currently a Professor. He was a visiting scholar at the University of Southern California during 1992-1993. He was a Prime Investigator of the iPCTV project during 1997-2000, and has developed MHP, ATVEF, and DASE data broadcasting platforms with the leading-edge companies in Korea including the Samsung Electronics, LG Electronics, Daewoo Electronics, Korea Broadcasting Systems. During 2000-2005, he was the Prime Investigator of the iMS project, and has developed an interactive broadcasting system based on DASE and ACAP, personalized broadcasting system over TV-Anytime specification, and TVN for home networking solution over digital television. He was the founder of the International Workshop on Digital Watermarking (IWDW). He has served as a Technical Program Chair or Co-Chair and Organizing Chair of many international conferences including IWDW2002, IWAIT2004, IWDW2005, PCM2005, and WIAMIS2006. Currently, he is the Chairman of the Korea e-Learning Society, Editor-in-Chief of the Institute of Electronics Engineers of Korea, Korean Society of Broadcast Engineers and Korea Institute of Information and Telecommunication Facilities Engineering. He was one of the Guest Editors of the IEEE Transactions on Circuits and Systems for Video Technology in 2003 and EURASIP Journal of Applied Signal Processing in 2006. He was also the Head of the Media Service Research Center (MSRC-ITRC), one of the excellent research centers funded by the Ministry of Information and Communication, Korea during 2003-2005 and is Head of the U-Home Research Center, one of the excellent research centers funded by the Ministry of Education, Korea. He has contributed to the MPEG specifications including the digital item adaptation in MPEG-21, file format in MPEG-21 and MPEG-4, and symbolic music representation in MPEG-4. His interests include multimedia computing, multimedia security, semantic analysis, and e-learning applications.

Vasiliy SACHNEV

Vasiliy SACHNEV received his B.S and M.S. in Electrical Engineering from the Komsomolsk-na-Amure State Technical University, Russia in 2000 and 2004 respectively.
He was scholar of Gagarin and Russian Federation Government honors during 2000-2001 and 2001-2002 respectively. Now he is a PhD student of Kangwon Nation University, Korea. His interests include multimedia security, digital watermarking.

Ki Seob KIM

Ki Seob KIM received his B.S in Control and Instrumentation Engineering from the Samcheok National University and M.S. in Control and Instrumentation Engineering from the Kangwon National University. Now he is a PhD student of Kangwon Nation University, Korea. His interests include multimedia security, digital watermarking, MPEG-21.