Lossless Audio Coding using Integer DCT

Min Ho Kang, Sung Woo Lee, Se Hyoung Park, and Jaeho Shin
Dept. of Electronic Engineering, Dongguk University, Seoul,100-715, Korea.
Tel : +82-2-2260-3336 Fax : +82-2-2272-5667 E-mail: icewhere@dongguk.ac.kr

Abstract:

This paper proposes a novel algorithm for hybrid lossless audio coding, which employs integer discrete cosine transform. The proposed algorithm divides the input signal into frames of a proper length, decorrelates the framed data using the integer DCT and finally entropy-codes the frame data. In particular, the adaptive Golomb-Rice coding method used for the entropy coding selects an optimal option which gives the best compression efficiency. Since the proposed algorithm uses integer operations, it significantly improves the computation speed in comparison with an algorithm using real or floating-point operations. When the coding algorithm is implemented in hardware, the system complexity as well as the power consumption is remarkably reduced. Finally, because each frame is independently coded and is byte-aligned with respect to the frame header, it is convenient to move, search, and edit the coded data.

Keywords : Lossless Audio Coding, Integer DCT, Lifting Step, Adaptive Golomb-Rice Coding

1. INTRODUCTION

Lossless audio coding has been studied less than lossy audio coding. But the interest about lossless audio coding is increasing recently. In the field of audio coding, there are two types of scheme. One is lossy audio coding, and the other is lossless one. Examples of lossy coding are MPEG layer 1/2/3, AAC and AC-3. These are commonly used in many applications and have been improved. Especially, the Layer-3 of MPEG I is used popular in audio service through Internet since it has a high compression ratio(more than 1:12) and provides CD-quality sound. Thus, it is frequently used to store audio signals in disc. Meanwhile, lossless audio signal compression schemes haves been hardly studied compared to lossy compression. But, the demand of high quality and lossless compression of the original audio signals is gradually increased. Typical applications are It is useful in digital versatile disc(DVD) and audio signal database systems which require the nature of the original audio signal or the characteristics [1].

Lossy audio coding schemes are mostly using perceptual property of human hearing, but lossless coding is using prediction[1] or transform[2] method which reduce the redundancy of signal. Also, it must be decoded easier then lossy compression. General lossless data compression method like zip, pkzip are not adequate for audio signal because the properties of audio signal is different from text data properties. Therefore, audio signal is required its own lossless compression method [3].

There are several methods for lossless audio coding such as Shorten, LTAC, MusiCompress, and AudioPaK [1][2][3][5]. Shorten[1] proposed linear prediction method in time domain. OggSquish and DVD-audio are using IIR prediction filters[4][5]. But LTAC[2] is a transform coding using discrete cosine transform(DCT). It has the flexibility of audio quality and bit rate.

In the next section, we specify the main operations, i.e. we describe framing process, transform coding process, and entropy coding process. And introduction of integer DCT is involved. In the section III, the proposed lossless coding scheme is presented. In the section IV, we evaluates the performance from the experimental results. Finally, we make a conclusion in the section V.

2. LOSSLESS AUDIO CODING SYSTEM

2.1. The main operations of the system

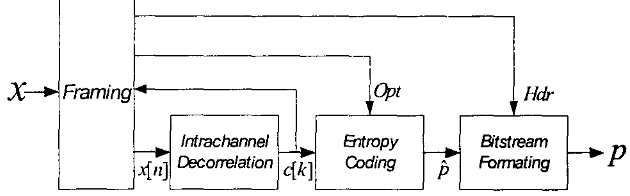

Fig. 1 The basic operations in most lossless compression algorithms definited by M. Hans

The basic operations in lossless compression algorithms are well defined in [1][3]. Fig. 1 is a block diagram of the operations involved in compressing a single audio channel. The first stage of the block diagram is the framing operation that provides an important and necessary property for most applications with digital audio. In this operation, digital audio signal is divided into independent frames of equal time duration. If this time scale is too short or long, it incurs overhead in the computation and transmission. And we can edit because of framing. Thus, it is important to determine the appropriate frame length.

After framing, input signals go through intrachannel decorrelation block followed by entropy coding. This intrachannel decorrelation is inevitable to achieve large compression ratio. There are two types to reduce data redundancy. One is prediction (analysis), and the other is transform coding. We will examine these two methods in the following subsections, and then we describe the entropy coding for residual signals and error signals that are generated from intrachannel decorrelation operation.

2.2. Integer DCT using lifting scheme

Since there exist errors for quantizing the transforming coefficients, it is impossible to use them for lossless compression. Therefore, we propose Integer DCT[9] without float-point multiplications or integer transforms. Integer DCT possess some features of their corresponding float-point transforms such as the decorrelation property, whereas their computational cost is less. Since integer DCT require only integer arithmetic(additions and possibly multiplications), their implementation is greatly simplified. In addition, if sufficient word length is used to represent the intermediate data, the round-off error can completely be eliminated.

Transforms like the DFT, DCT, or the MDCT can be formulated in terms of so-called Givens rotations. The lifting scheme can be applied to get an invertible integer approximation of each Givens rotation[6][7][8].

This matrix factored into three lifting matrix[9]. In this lifting matrix, $(\cos\alpha - 1)/\sin\alpha$ and $\sin\alpha$ are real numbers.

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} = \begin{pmatrix} 1 & \frac{\cos\alpha - 1}{\sin\alpha} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \sin\alpha & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{\cos\alpha - 1}{\sin\alpha} \\ 0 & 1 \end{pmatrix} \quad (1)$$

To approximate to these real numbers by using integer numbers, the notation RB(s) is used. RB(s) denotes a number that is of the form $\beta/2^{\lambda}$, where β and λ are integers, and approximates to the real number s.

Consequently, the Givens rotation is, approximated by RB(s).

2.3. Intrachannel decorrelation

Intrachannel decorrelation is generally composed of predictive model and lossy coding model. In this paper, we also propose lossy coding model proposed by M. Purat. Fig. 2 shows the lossless transform coding system using lossy coding model.

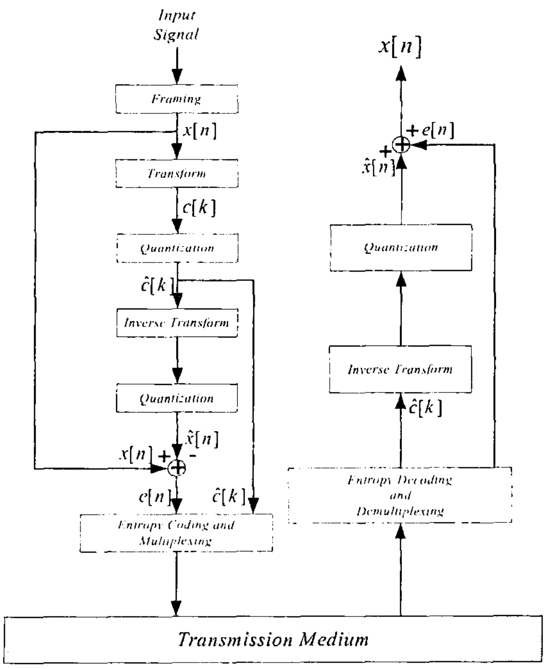

Fig. 2 Structure of lossless coding system using lossy coding model.

The transform and inverse transform represent an arbitrary orthonormal transform with block length N and its inverse, respectively, and the transform is discrete cosine transform(DCT) in [2]. The reason why DCT is used in lossless transform coding is energy compaction property of DCT, which permits a good reconstruction from a few quantized values of the coefficients. Many coefficients are very small or even zero, and moreover, they constitute an uncorrelated source when using a suitable transform. Next, we implement quantization. The result of the quantization is equivalent to integer truncation, so we cannot reconstruct the signal perfectly.

3. PROPOSED LOSSLESS CODING ALGORITHM

We propose integer DCT to remove correlation between the framed signals. Fig. 3 shows entire block diagram of proposed lossless coding algorithm. The decorrelated signal is coded by Adaptive Golomb-Rice coder to get optimal bit allocation.

The input signal x is divided to framed signal $x_l[n]$ by going through framing block. The length of this frame N is from 512 to 4096. If sampling frequency is 44.1kHz, the time interval is from 12ms to 93ms.

In this paper, we factored DCT_{IV} into lifting step[6].

The DCT_{IV} of length N

$$DCT_{IV}^{(N)} = \left(\sqrt{\frac{2}{N}} \cos \frac{(2k+1)(2l+1)\pi}{4N} \right)_{k,l=0,\dots,N-1} \quad (2)$$

can be decomposed into two DCT_{IV} of length N/2.

Define the $N \times N$ matrices L and M by

$$\begin{pmatrix} L_{k,k} & L_{k,N-1-k} \\ L_{N-1-k,k} & L_{N-1-k,N-1-k} \end{pmatrix} =$$

$$\begin{pmatrix} \cos\left(\frac{2k+1}{4N}\pi\right) & -\sin\left(\frac{2k+1}{4N}\pi\right) \\ -\sin\left(\frac{2k+1}{4N}\pi\right) & -\cos\left(\frac{2k+1}{4N}\pi\right) \end{pmatrix} \quad k = 0,\dots,N/2-1$$

$$L_{k,l} = 0 \qquad \text{else} \quad (3)$$

$$M = \frac{1}{\sqrt{2}} \begin{pmatrix} I_{N/2} & I_{N/2} \\ -I_{N/2} & I_{N/2} \end{pmatrix} \quad (4)$$

P and Q are $N \times N$ permutation matrices.

With these matrices the DCT_{IV} of length N can be decomposed into

$$DCT_{IV}^{(N)} = L \begin{pmatrix} DCT_{IV}^{(N/2)} & 0 \\ 0 & DCT_{IV}^{(N/2)} \end{pmatrix} MQP \quad (5)$$

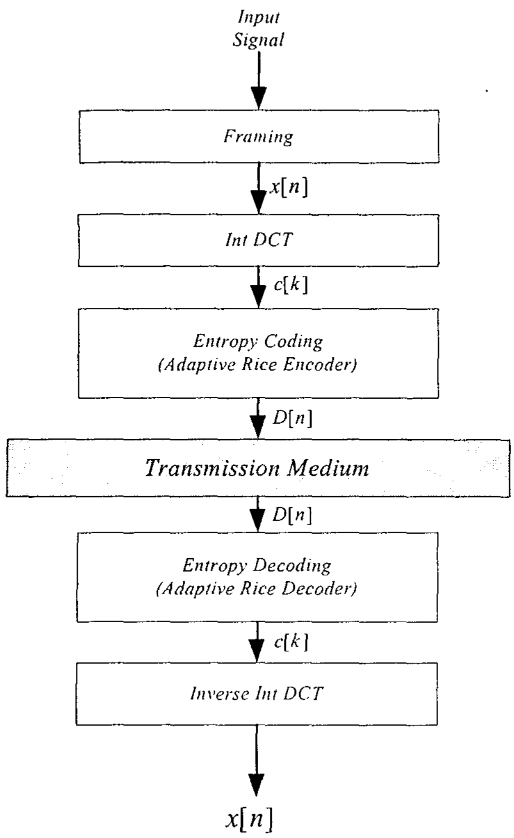

Fig. 3 The entire block diagram of proposed lossless coding algorithm.

In this paper, for entropy coding of decorrelated signal, we used Adaptive Golomb-Rice coding algorithm recommended by the Consultative Committee on Space Data Standards(**CCSDS**)[10]. In rice coding algorithm, blocks of source samples are encoded independently and side information does not need to be carried. So, it is possible to edit encoded data directly, and the structure is simple. There are four options in Adaptive Golomb-Rice coding. These options are Zero-block option, 2nd Extention option, FS(Fundamental Sequence) option, and Split option. The compression efficiency can be improved by selecting the option allocated minimal bit among these four options. Fig. 4 shows the Adaptive Golomb-Rice coding algorithm.

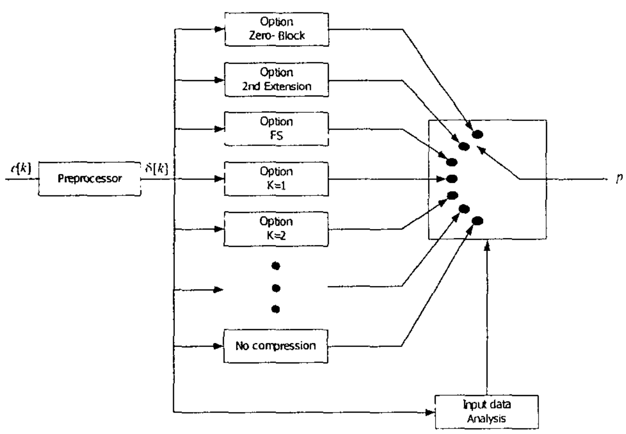

Fig. 4 Block diagram of Adaptive Golomb-Rice coding algorithm.

4. EXPERIMENTAL RESULTS

In this paper, we applied proposed algorithm to several audio CD(MAX). IDL1, IDL2, and IDL3 types are proposed method of compression. The frame sizes which we proposed are 2048 and 4096, and segment sizes are 16, 32, and 64. Table 1 shows the property of these three types.

Table 1. Parameters of IDL1, IDL2, and IDL3 types

Type	Frame Size	Segment Size
IDL1	4096	32
IDL2	4096	64
IDL3	2048	16

Table 2 shows the experimental results compared with Shorten, and LTAC. The compression ratio is given by

$$Ratio(\%) = \frac{Compressed\ file\ size\ in\ byte}{Original\ file\ size\ in\ byte} \times 100 \quad (6)$$

Table 2. Compression ratio of Shorten, LTAC, and proposed algorithms.

MAX CD	Shorten	LTAC	IDL1	IDL2	IDL3
Track1	81.53	77.21	77.47	77.19	78.71
Track2	67.17	66.28	68.99	68.61	69.61
Track3	77.59	76.85	77.39	76.97	78.16
Track4	75.48	71.74	72.60	72.19	74.08
Track5	60.89	57.31	61.11	60.83	62.55
Track6	73.78	70.84	71.63	71.29	72.57
Track7	68.19	64.29	68.19	67.86	69.32
Track8	73.42	71.65	73.75	73.34	74.46
Track9	68.14	64.44	67.99	67.65	69.71
Track10	74.02	68.15	69.21	68.78	71.06
Track11	74.89	69.98	72.60	72.22	73.87
Track12	70.34	67.71	69.94	69.50	71.29
Track13	62.82	59.73	62.04	61.68	63.78
Track14	63.46	59.92	62.07	61.74	63.83
Track15	65.35	62.49	66.68	66.40	68.18
Track16	70.21	65.18	69.03	68.83	70.38
Track17	65.06	61.43	63.92	63.75	65.95
Average	**69.70**	**66.31**	**69.09**	**68.75**	**70.44**

And Fig. 5 is the graph of result values. And these values are convenient to compare the compression ratio of table 2.

As you can see in Table 2 and Fig. 5, most of the compression ratios of proposes algorithms are from 68% to 69%, except for IDL3. These results shows that the compression ratio of proposed algorithm is better than that of Shorten. But, it has bad compression ratio compare to that of LTAC.

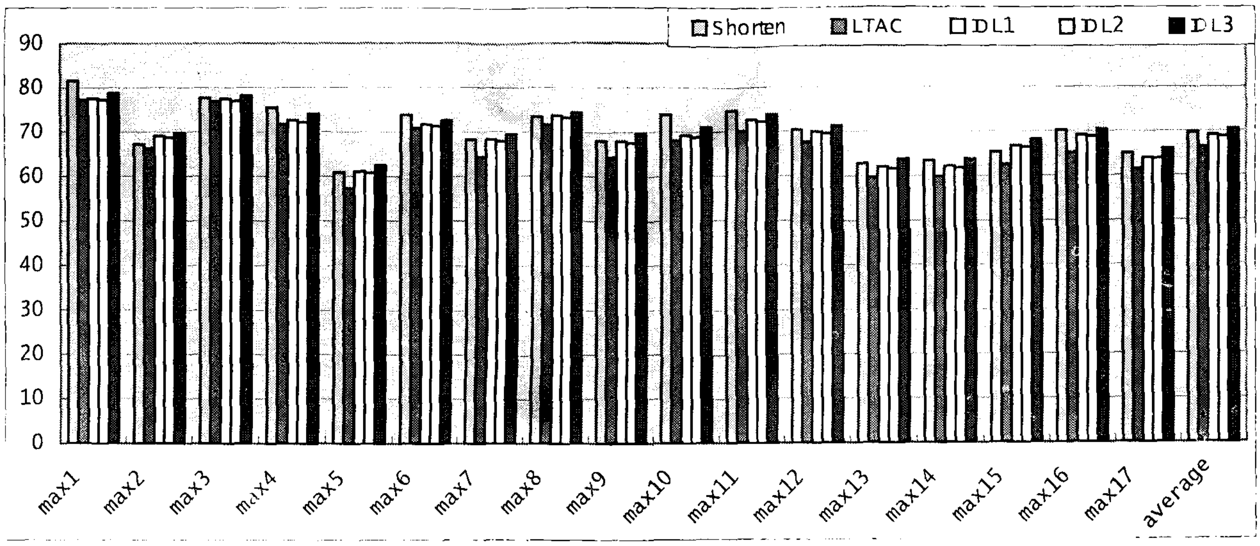

Fig. 5 Compression ratio of Shorten, LTAC, and proposed algorithms.

5. CONCLUSION

In this paper, we present a lossless audio coding algorithm which is based on integer discrete cosine transform. The simulation result show that the proposed algorithm have the compression ratio as good as conventional compression algorithm. Whereas the proposed algorithm uses integer operations, so it signify- cantly improves the computation speed in comparison with an algorithm using real or floating-point opera- tions. When the coding algorithm is implemented in hardware, the system complexity as well as the power consumption is remarkably reduced.

References

[1] T.Robinson, "SHORTEN: Simple lossless and near-lossless waveform compression", Cambridge Univ.Eng. Dept., Cambridge, UK, Tech.Rep.156, 1994.

[2] M.Purat, T.Liebchen, and P.Noll, "Losselss transform coding of audio signals," in Proc. 101nd AES Conv., Munich, Germany, 1997, preprint 4414.

[3] M.Hans, R,W. Schafer, "Lossless Compression of Digital Audio," IEEE Signal Preccssing Magazine, July, 2001.

[4] P.Craven, M.Law, and J.stuart, "Lossless compression using IIR prediction filters," in Proc.102nd AES Conv., Munich, Germany, 1997, preprint 4415.

[5] A.Bruekers, A.Oomen, and R.van der Vleuten, "Lossless coding for DVD audio, " in Proc. 101st AES Conv., Los Angeles, Ca, Nov.1996, preprint 4358.

[6] Geiger, R, Yokotani, Y, and Schuller, G, "Improved integer transforms for lossless audio coding", Signals, Systems & Computers, 2003 The Thrity-Seventh Asilomar Conference on, vol. 2, pp. 2119-2123, Nov. 9-12, 2003.

[7] Jie Liang, and Tran, T.D, "Approximating the DCT with the lifting scheme: systematic design and applications", Signals, Systems and Computers, 2000. Conference Record of the Thirty-Fourth Asilomar Conference on, vol. 1, pp. 192-196, 29 Oct.-1 Nov. 2000.

[8] Jie Liang, and Tran, T.D, "Fast multiplierless approximations of the DCT with the lifting scheme", IEEE Trans on Signal Processing, vol. 49, pp. 3032-3044, Dec. 2001.

[9] Yonghong Zeng, Lizhi Cheng, Guoan Bi, and Alex C. Kot, "Integer DCTs and Fast Algorithms", *IEEE Trans on Signal Processing*, vol. 49, November 2001.

[10] CCSDS 120.0-G-1: Lossless Data Compression. Green Book. Issue 1. 5/1997.