A Fast Algorithm of Integer MDCT for Lossless Audio Coding

HUANG HaiBin, RAHARDJA Susanto, YU Rongshan, LIN Xiao
Agency for Science, Technology and Research (A*STAR)
Institute for Infocomm Research (I^{2}R), Singapore
{hhuang, rsusanto, rsyu, linxiao@i2r.a-star.edu.sg}

Abstract

In this paper, a new fast algorithm to implement integer modified Discrete Cosine Transform (IntMDCT) is proposed. It is shown that the total rounding operations required for this algorithm is only 2.5*N*. As a result, its approximation error is far less than that of the directly converted integer transforms. At the same time, the complexity is greatly reduced which results in improved performance of a lossless audio coding system employing the IntMDCT in terms of compression ratio and complexity costs.

1. Introduction

Reversible transforms that map integers to integers are important tools for lossless coding schemes. For integer inputs, such a transform generates integer outputs that approximate its floating-point prototype transform outputs. The original integer inputs can be completely recovered by passing the integer outputs from the forward transform through the inverse transform.

Recently, integer transform has attracted more and more attention for both image and audio coding [1-6]. Fast and efficient algorithms for high order N (N>16) on integer transform still poses a challenge in research. In audio coding, a modified discrete cosine transform (MDCT) is usually used. It is known that the MDCT can provide critical sampling with good frequency selectivity. With that, it is employed in MPEG-4 advanced audio coding (AAC) system [7]. Very recently, MPEG issued a call for proposal on lossless audio coding for its extension 3. Integer MDCT (IntMDCT) is one of the key components within this possible new coding scheme [8]. It is then critical that a fast IntMDCT with low rounding error be developed and exploited to meet the industry requirements.

The IntMDCT can be derived from its prototype – the modified discrete cosines transform type-IV. In his book [9], Malvar gave an efficient realization of MDCT by cascading a bank of Givens rotations with a DCT-IV block. It is well known that Givens rotation can be factorized into three lifting steps for mapping integers to integers [3]. With that, the realization of IntMDCT mainly relies on an efficient implementation of integer DCT-IV.

Integer transforms can be directly converted from their prototypes by replacing each Givens rotation with three lifting steps. For each lifting step there is one rounding operation, the total rounding number of an integer transform is three times the Givens rotation number of the prototype transform. For discrete trigonometric transforms, the number of Givens rotations needed is usually $N\log_2 N$, where N is the block size. Accordingly, the total rounding number is also $N\log_2 N$ for the family of directly converted integer transforms. Because of the rounding, an integer transform can only approximate its floating-point prototype. Obviously, the approximation error increases with the number of rounding operations.

In this paper, a novel fast algorithm for realizing reversible integer DCT-IV is proposed. The fast algorithm has immediate application in improving the performance of a lossless audio coding system. The total rounding number of this algorithm is only 2.5N while the approximation error is shown to be far less than that of the directly converted integer transforms.

As a result, the lossless compression ratio of a lossless audio coding system employing the fast algorithm improves and at same time the computational complexity is greatly reduced. This paper is organized as follows. Section 2 describes the new algorithm. Section 3 presents discussions and numerical results on performance comparisons. Finally, in Section 4 a conclusion on this work is discussed.

2. Integer Type-IV DCT

The discrete cosine transform type IV (DCT-IV) of an N-point real input sequence $x(n)$ is defined as follows [2]:

$$\mathbf{y} = \mathbf{C}_N^{IV} \cdot \mathbf{x} \qquad \textbf{2-1}$$

$$\mathbf{x} = \{x(n)\}_{n=0,1,\cdots,N-1}, \quad \mathbf{y} = \{y(m)\}_{m=0,1,\cdots,N-1}$$

where $\mathbf{C}_N^{IV}$ is the transform matrix defined as

$$\mathbf{C}_N^{IV} = \sqrt{\frac{2}{N}}\left[\cos\left(\frac{(m+1/2)(n+1/2)\pi}{N}\right)\right] \qquad \textbf{2-2}$$

$m = 0, 1, \ldots, N$-1 and $n = 0, 1, \ldots, N$-1

$\mathbf{C}_N^{IV}$ can be decomposed into the following six matrices:

$$\mathbf{C}_N^{IV} = \mathbf{R}_1 \cdot \mathbf{R}_2 \cdot \mathbf{S} \cdot \mathbf{T} \cdot_1 \mathbf{T}_2 \cdot \mathbf{P}_{eo} \qquad \textbf{2-3}$$

where $\mathbf{P}_{eo}$ is the even-odd permutation matrix, $\mathbf{R}_1$, $\mathbf{R}_2$, $\mathbf{T}_1$, $\mathbf{T}_2$, and $\mathbf{S}$ are

$$\mathbf{R}_1 = \begin{bmatrix} \mathbf{I}_{N/2} & \\ \mathbf{H}_1 & \mathbf{I}_{N/2} \end{bmatrix} \qquad \textbf{2-4}$$

$$\mathbf{R}_2 = \begin{bmatrix} \mathbf{I}_{N/2} & \mathbf{H}_2 \\ & \mathbf{I}_{N/2} \end{bmatrix} \qquad \textbf{2-5}$$

$$\mathbf{T}_1 = \begin{bmatrix} -\mathbf{D}_{N/2} & \mathbf{K}_2 \\ & \mathbf{I}_{N/2} \end{bmatrix} \qquad \textbf{2-6}$$

$$\mathbf{T}_2 = \begin{bmatrix} \mathbf{I}_{N/2} & \\ \mathbf{K}_3 & \mathbf{I}_{N/2} \end{bmatrix} \qquad \textbf{2-7}$$

$$\mathbf{S} = \begin{bmatrix} \mathbf{I}_{N/2} & \\ \mathbf{H}_3 + \mathbf{K}_1 & \mathbf{I}_{N/2} \end{bmatrix} \qquad \textbf{2-8}$$

$$\mathbf{K}_1 = -\left(\mathbf{C}_{N/2}^{IV} \cdot \mathbf{D}_{N/2} + \sqrt{2}\mathbf{I}_{N/2}\right) \cdot \mathbf{C}_{N/2}^{IV}$$

$$\mathbf{K}_2 = \frac{\mathbf{C}_{N/2}^{IV}}{\sqrt{2}},$$

$$\mathbf{K}_3 = \sqrt{2}\mathbf{C}_{N/2}^{IV} \cdot \mathbf{D}_{N/2} + \mathbf{I}_{N/2}$$

respectively and $\mathbf{K}_1$, $\mathbf{K}_2$, $\mathbf{K}_3$, $\mathbf{H}_1$, $\mathbf{H}_2$, $\mathbf{H}_3$, $\mathbf{D}$ are

$$\mathbf{H}_1 = \begin{bmatrix} & & & -\tan\frac{(N-1)\pi}{8N} \\ & & \iddots & \\ & -\tan\frac{3\pi}{8N} & & \\ -\tan\frac{\pi}{8N} & & & \end{bmatrix}$$

$$\mathbf{H}_2 = \begin{bmatrix} & & & \sin\frac{\pi}{4N} \\ & & \sin\frac{3\pi}{4N} & \\ & \iddots & & \\ \sin\frac{(N-1)\pi}{4N} & & & \end{bmatrix}$$

$\mathbf{H}_3 = \mathbf{H}_1$

$$\mathbf{D} = \begin{bmatrix} 1 & & & & \\ & -1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & -1 \end{bmatrix}$$

respectively. Equation **2-3** indicates that integer DCT-IV transform can be realized by five lifting stages. As these lifting stages have identical structures, only the transform equation for one stage is discussed here. It is trivial and straightforward to write down the transform equations for all the other stages.

Let $\mathbf{x}$ and $\mathbf{y}$ be the input and output $N \times 1$ integer vectors for the first transform stage, we have

$$\mathbf{y} = \mathbf{T}_2 \cdot \mathbf{x} \qquad \textbf{2-9}$$

Vectors $\mathbf{x}$ and $\mathbf{y}$ are divided into two halves, and the equation can be rewritten as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{N/2} & \\ \mathbf{K}_3 & \mathbf{I}_{N/2} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}$$

Thus, the forward integer transform is

$$\mathbf{y}_1 = \mathbf{x}_1$$

$$\mathbf{y}_2 = \lfloor \mathbf{K}_3 \cdot \mathbf{x}_1 \rfloor + \mathbf{x}_2 \qquad \textbf{2-10}$$

and the inverse integer transform is

$$\mathbf{x}_1 = \mathbf{y}_1$$

$$\mathbf{x}_2 = \mathbf{y}_2 - \lfloor \mathbf{K}_3 \cdot \mathbf{x}_1 \rfloor \qquad \textbf{2-11}$$

where $\lfloor * \rfloor$ denotes rounding operation.

3. Discussion and Experimental Results

From Equations **2-10** and **2-11**, there are $N/2$ rounding operations in each lifting stage. Therefore, from Equation **2-3**, the total number of rounding operations for the proposed integer DCT-IV algorithm is five times $N/2$, which is $2.5N$. From Equation **2-3**, it can be seen that the majority of computational power is on the four $N/2$ point DCT-IV subroutines $\mathbf{C}_{N/2}^{IV}$, when N is a large value, e.g. N = 1024. It can be roughly estimated that the complexity of the proposed integer DCT-IV to be twice that of the floating-point DCT-IV.

The proposed algorithm in this paper is compared with the IntMDCT implementation used in the MPEG-4 lossless audio codec [10] which was adopted as Reference Model 0 (RM0) in July 2003. Fig 1 shows the block diagram of performance evaluation of the transforms. Mean square error (MSE) is computed and the MPEG-4 lossless audio coding evaluation procedure is adopted where the results are compared with the direct implementation method used in the RM0. The MSEs for the forward and inverse IntMDCTs are defined as:

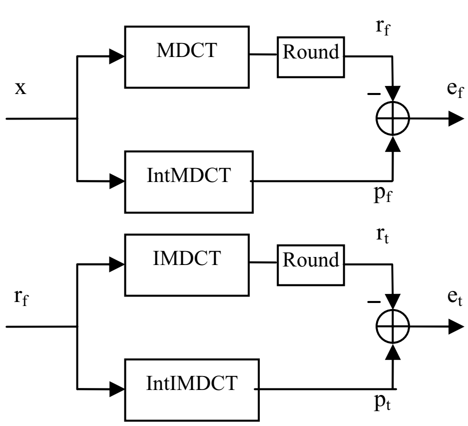

Fig 1. Evaluation test for IntMDCT and IntIMDCT

$$MSE = \frac{1}{L}\sum_{j=0}^{L-1}\frac{1}{1024}\sum_{i=0}^{1023} e_i^2$$

Table 1 shows a comparison of the proposed IntMDCT algorithm with the algorithm used in MPEG-4 RM0 and the new algorithm is integrated into the RM0 with the compression ratio performance compared with the original.

Table 1. Comparison of proposed IntMDCT algorithm with MPEG-4 reference model 0

	RM0 (Direct Implementation)	Proposed Algorithm	Compression ratio improvement
(IntMDCT) ***MSE**	1.789	0.547	0.87%
(IntIMDCT) ***MSE**	1.774	0.545	
(IntMDCT) ****MSE**	1.785	0.546	0.42%
(IntIMDCT) ****MSE**	1.766	0.546	
Complexity (instructions)	47,616	34,816	

*48kHz/16 bit test set; **96kHz/24 bit test set

In Table 1, it is shown that the new algorithm demonstrates superior performance compared with the algorithm employed in RM0. The RM0 of MPEG-4 lossless audio coding system is shown in Fig 2, which is the Advanced Audio Zip (AAZ) technology adopted by MPEG recently [10]. In the encoder, the input audio frames are transformed losslessly by using the integer MDCT (IntMDCT) to generate the IntMDCT coefficients. These coefficients are scaled, quantized and coded with perceptually weighted so that the noise introduced is best masked by the masking threshold of the human auditory system in the core AAC encoder. The resultant core layer bit-stream thus constitutes the minimum rate/quality unit of the final lossless bit-stream. For optimal coding efficiency, an error-mapping procedure is employed to remove the information that has been coded in the core layer from the IntMDCT coefficients before being noiselessly coded by bit-plane Golomb code (BPGC) [11] to form lossless bit stream. In the decoder, a reverse process is shown in the lower part of Fig. 2. According to MPEG audio test requirement, 450 seconds with 15 different types of music files are used for each 48kHz/16-bit and 96kHz/24-bit test sets. It can be seen that the total compression ratios for both sequences are improved, and this is achieved without introducing any amendment from the original entropy coder and simultaneously the complexity cost has been reduced for more than 26%.

4. Conclusion

In this paper, a new fast algorithm for realizing reversible integer type-IV DCT is proposed. This algorithm requires only 2.5*N* rounding operations for every block of *N* input samples. As a result, the approximation error is greatly reduced. The proposed algorithm is low in computational complexity and modular in structure. This algorithm is integrated and evaluated with MPEG-4 audio extension 3 lossless audio compression reference model 0. From the results it is clear that the new algorithm contributes to an overall improvement of compression ratio as well as lesser computational complexity.

References

[1] K.Komatsu, K.Sezaki, "Design of Lossless LOT and its performance evaluation", IEEE ICASSP2000, pp2119-2122

[2] S.Oraintara, Y.Chen, T.Nguyen, "Integer Fast Fourier Transform (INTFFT)", Proc. IEEE ICASSP2001

[3] R. Geiger, T. Sporer, J. Koller, K. Brandenburg, "Audio Coding based on Integer Transforms" *AES 111th Convention*, New York, USA, Sept. 2001

[4] R. Geiger, J. Herre, J. Koller and K. Brandenburg, "INTMDCT – a link between perceptual and lossless audio coding", Proc. IEEE ICASSP2002, pp1813-1816

[5] Jin Li, "A Progressive to Lossless Embedded Audio Coder (PLEAC) with Reversible Modulated Lapped Transform", Proc. IEEE ICME2003, ppIII-221—III-224 (reprint from ICASSP2003)

[6] Jia Wang, Jun Sun and Songyu Yu, "1-D and 2-D transforms from integer to integer", Proc. IEEE ICASSP2003, ppII-549 – II552

[7] Information Technology – Coding of Audiovisual Objects, Part 3. Audio, Subpart 4 Time/Frequency Coding, ISO/JTC 1/SC 29/WG11, 1998

[8] "Final Call for Proposals on MPEG-4 Lossless Audio Coding", ISO/IEC JTC 1/SC 29/WG 11 W5280, Shanghai, Oct. 2002

[9] H. S. Malvar, "Signal Processing with Lapped Transforms" *Artech House*, 1992

[10] R.Yu, X. Lin and S. Rahardja, "Advanced Audio Zip – A Scalable Perceptual and Lossless Audio Codec", ISO/IEC JTC1/SC29/WG11 (m9134) MPEG2002 Awaji, Dec 2002.

[11] R. Yu, C.C. Ko, S. Rahardja, and X. Lin, "Bit-Plane Golomb Coding for Sources with Laplacian Distributions," Proc. IEEE ICASSP 2003.

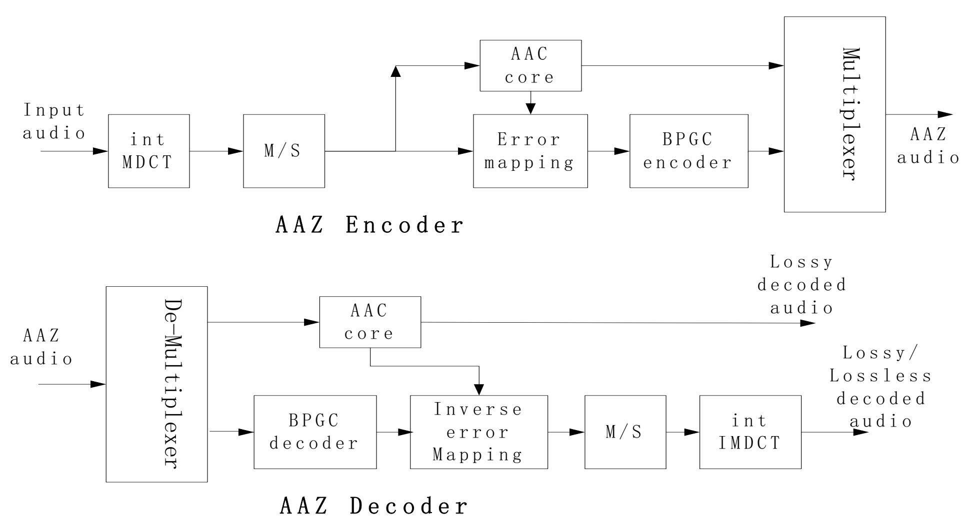

Fig 2. AAZ lossless audio encoder/decoder system diagram