A NEW STRUCTURE OF INTEGER DCT LEAST SENSITIVE TO FINITE WORD LENGTH EXPRESSION OF MULTIPLIERS

Masahiro Iwahashi, Dang Khoa Dang, Masahiro Ohnishi, Somchart Chokchaitam*
Nagaoka University of Technology, Nagaoka, Japan, *Thammasat University, Bangkok, Thailand

ABSTRACT

While conventional DCT outputs real numbers and it is not suitable for lossless coding of images, the integer DCT (Int-DCT) outputs integer numbers and can serve as a lossless coding and lossy coding compatible to the conventional DCT. Recently, our group has proposed an optimum assignment method of word length of multipliers in the Int-DCT assuming "colored" spectrum of input signals to reduce total hardware complexity of the circuit. However, it has been ignored to include variations of the integer rotation transforms. This report theoretically and experimentally investigates various rotation transforms and finds the best combination to construct the Int-DCT least sensitive to finite word length expression of multiplier coefficients. As a result, improvement up to 1.8 [dB] to the existing one was confirmed for a colored signal.

1. INTRODUCTION

The discrete cosine transform (DCT) has found wide applications especially in compression (lossy coding) of image data [1]. It has become a core of many international standards such as JPEG and MPEG. However, the conventional DCT is not suitable for lossless coding of images since it essentially outputs real numbers (long length integers). As a result, it is well known that the lifting wavelet transform replaces the DCT for lossless image coding [2].

On the other hand, various types of the integer DCT (Int-DCT) that outputs integer numbers (short length integers) have been proposed [3-6]. The Int-DCT is composed of lifting steps accompanied by the rounding operations [2,7]. It attracts many researchers' attention since it serves as a lossless coding and also lossy coding compatible to the conventional DCT.

Recently, our group has reported how to assign word length to each of multipliers in the Int-DCT to reduce total hardware complexity of the circuit [8]. Since the image compression assumes "colored" spectrum of the signal, the report considered the input signal's "color-ness" to attain the best signal quality (PSNR) under the shortest average word length of the multipliers. Robustness of the assignment has also theoretically analyzed [9]. However, it has been ignored to include variations of the integer rotation transform (IR) in the Int-DCT.

In this report, we include "four" types of IR illustrated in **figure 1** into investigation. For example, the Int-DCT shown in **figure 2** and **figure 3** include "five" IRs. Therefore, we have 4^5 combinations to construct an Int-DCT. We indicate experimentally and theoretically that there is no difference when the input image signal is not colored, however, the best combination is superior to the existing one by up to 1.8 [dB] for colored signal such as image signals.

2. ROTATION TRANSFORMS FOR INTEGER DCT

The IR converts an input integer vector **X** into the output vector **Y** by **Y**=**TX** where

$$\mathbf{T} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} , \qquad (1)$$

$$\mathbf{X} = \begin{pmatrix} x(0) & x(1) \end{pmatrix}^t , \quad \mathbf{Y} = \begin{pmatrix} y(0) & y(1) \end{pmatrix}^t .$$

While the output signal above is not integers in general, the rounding operation at each lifting step in **figure 1** truncates a signal value into an integer [3,4].

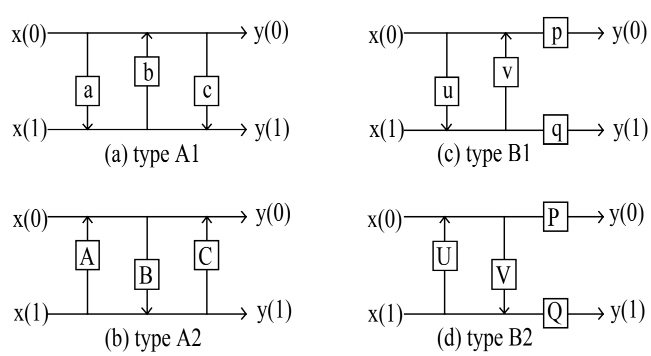

Fig.1 Variations of the integer rotation transform (IR).

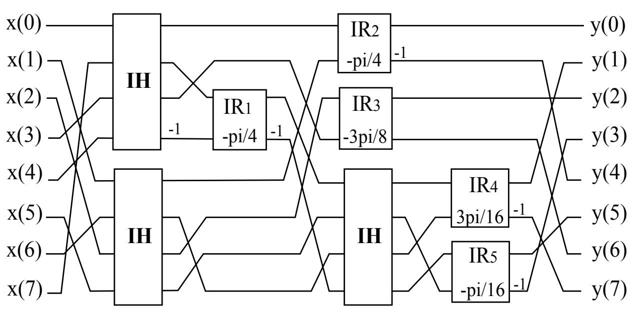

Fig.2 Fukuma's Int-DCT. **IH** is the Int- Hadanard transform [3,8].

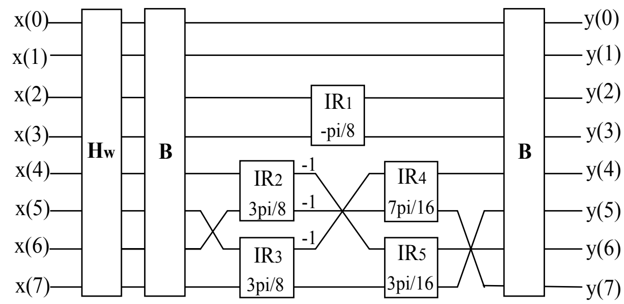

Fig.3 Soontorn's Int-DCT. $\mathbf{H}_w$ and **B** denote Int-Hadanard transform and permutation respectively [4].

When the IR is implemented in the structures in **figure 1**, the equation (1) is factorized into equations below. Note that the rounding operations are ignored to focus our discussion on "finite word length expression error" of the multipliers.

type A1

$$\mathbf{T} = \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix} \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix} \tag{2}$$

type A2

$$\mathbf{T} = \begin{pmatrix} 1 & C \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ B & 1 \end{pmatrix} \begin{pmatrix} 1 & A \\ 0 & 1 \end{pmatrix} \tag{3}$$

type B1

$$\mathbf{T} = \begin{pmatrix} p & 0 \\ 0 & q \end{pmatrix} \begin{pmatrix} 1 & v \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix} \tag{4}$$

type B2

$$\mathbf{T} = \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix} \begin{pmatrix} 1 & 0 \\ V & 1 \end{pmatrix} \begin{pmatrix} 1 & U \\ 0 & 1 \end{pmatrix} \tag{5}$$

The multiplier coefficients to be truncated in this report are given by comparing equation (1) and equations (2)-(5) as follows.

type A1

$$a = c = (1 - \cos\theta)\sin^{-1}\theta, \qquad b = -\sin\theta \tag{6}$$

type A2

$$A = C = -(1 - \cos\theta)\sin^{-1}\theta, \qquad B = \sin\theta \tag{7}$$

type B1

$$u = \cos^{-1}\theta \sin\theta, \qquad v = -\cos\theta\sin\theta$$

$$p = \cos^{-1}\theta, \qquad q = \cos\theta \tag{8}$$

type B2

$$U = -\cos^{-1}\theta\sin\theta, \qquad V = \cos\theta\sin\theta$$

$$P = \cos\theta, \qquad Q = \cos^{-1}\theta \tag{9}$$

3. SENSITIVITY FOR COLORED SIGNALS

When a multiplier coefficient h in either of the forward transform **T** or the backward transform $\mathbf{T}^{-1}$ is truncated into $\hat{h}$, error signal is observed in the decoded signal. Namely, the sensitivity of the coefficient to finite word length expression is experimentally evaluated by

$$S_h = \frac{\Delta\sigma_{err}}{\Delta h} \tag{10}$$

where

$\Delta h = \hat{h} - h$,

$h \in \{ a, b, c, A, B, C, p, q, u, v, P, Q, U, V \}$.

In the equation above, $\Delta\sigma_{err}$ denotes square root of variance of the error and it is theoretically calculated, when the backward transform contains $\hat{h}$, by

$$\Delta\sigma_{err} = \Delta h \sqrt{\frac{1}{N} trace\left[\frac{\partial \mathbf{T}^{-1}}{\partial h} \mathbf{T} \cdot \varepsilon\left[\mathbf{X}\mathbf{X}^t\right] \cdot \mathbf{T}^t \frac{\partial \left(\mathbf{T}^{-1}\right)^t}{\partial h} \right]} \tag{11}$$

where N and $\varepsilon[\cdot]$ denote size of the square matrix **T** and ensemble average respectively [9].

Especially for AR(1) model with correlation coefficient ρ as an example of colored signal, the sensitivity is theoretically evaluated by

$$S_h = \sigma_x \sqrt{\frac{1}{N} trace\left[\frac{\partial \mathbf{T}^t}{\partial h} \mathbf{T} \cdot R_{xx}(\rho) \cdot \mathbf{T}^t \frac{\partial \mathbf{T}}{\partial h} \right]} \tag{12}$$

where $R_{xx}(\rho)$ and $\sigma_x{}^2$ denote covariance and variance of the input signal respectively. As a result, it is explicitly described that the sensitivity is a function of correlation ρ of the input signal.

According to the discussion above, sensitivities of each multiplier in the IR variations in **figure 1** are calculated as below.

type A1

$$S^2{}_a = \sigma_x^2 / 2$$
$$S^2{}_b = cos^{-4}(\theta/2)(1+\rho sin\theta)\sigma_x^2 / 2 \qquad (13)$$
$$S^2{}_c = (1-\rho sin2\theta)\sigma_x^2 / 2$$

type A2

$$S^2{}_A = \sigma_x^2 / 2$$
$$S^2{}_B = cos^{-4}(\theta/2)(1-\rho sin\theta)\sigma_x^2 / 2 \qquad (14)$$
$$S^2{}_C = (1+\rho sin2\theta)\sigma_x^2 / 2$$

type B1

$$S^2{}_u = \sigma_x^2 / 2$$
$$S^2{}_v = cos^{-4}(\theta/2)(1+\rho sin2\theta)\sigma_x^2 / 2$$
$$S^2{}_p = cos^{2}(\theta/2)(1-\rho sin2\theta)\sigma_x^2 / 2 \qquad (15)$$
$$S^2{}_q = cos^{-2}(\theta/2)(1+\rho sin2\theta)\sigma_x^2 / 2$$

type B2

$$S^2{}_U = \sigma_x^2 / 2$$
$$S^2{}_V = cos^{-4}(\theta/2)(1-\rho sin2\theta)\sigma_x^2 / 2$$
$$S^2{}_P = cos^{-2}(\theta/2)(1-\rho sin2\theta)\sigma_x^2 / 2 \qquad (16)$$
$$S^2{}_Q = cos^{2}(\theta/2)(1+\rho sin2\theta)\sigma_x^2 / 2$$

The geometric mean of the sensitivity for type A1 and type A2 are illustrated in **figure 4** and **figure 5**. It is confirmed that the result of the experiment was in beautiful agreement with the value that had been obtained by theoretical calculation. In addition, it became clear that there is no difference between the variations when ρ =0. On the contrary, when ρ is not zero, it is expected that we can reduce sensitivity of the Int-DCT by choosing an appropriate combination of the variations of the IR in the Int-DCT.

Note that this report focuses on type A1 and type A2 because they require only three multipliers compare to four multipliers required for type B1 and type B2. In addition, type B1 and B2 contains multiplier coefficient value greater than unity as shown in equations (8) and (9). They degrade performance of lossless coding.

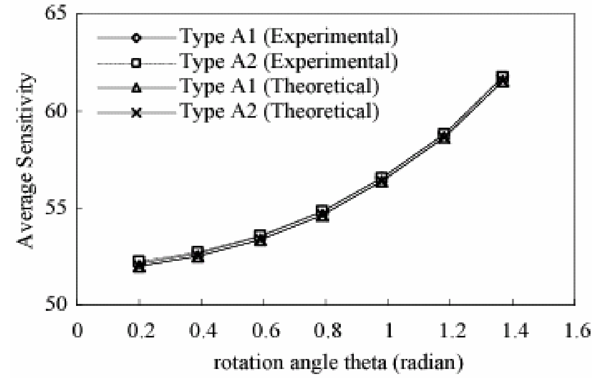

Fig.4 Average sensitivity for AR(1) with ρ =0.

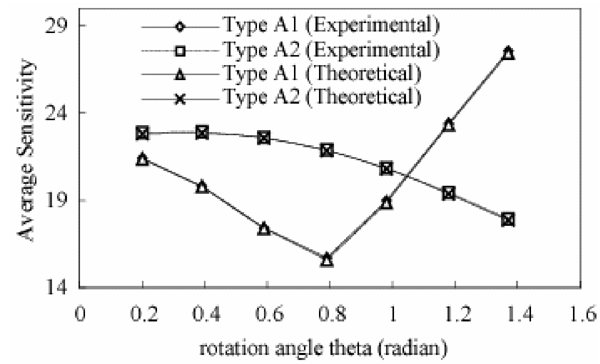

Fig.5 Average sensitivity for AR(1) with ρ =0.95.

4. THE LEAST SENSITIVE INTEGER DCT

As previously described, the Int-DCT shown in **figure 2** and **figure 3** include "five" IRi, $i \in \{1,2,\cdots,5\}$ and we have "two" types Ak, $k \in \{1,2\}$ for each IR. Therefore, we have 2^5 =32 candidates for the least sensitive Int-DCT. Defining N as a combination number by

$$N = \sum_{i=1}^{5} IRi(Ak) \times 2^{i-1} + 1$$
$$IRi(Ak) = \begin{cases} 0, & k = 1 \\ 1, & k = 2 \end{cases},$$

we evaluated all the possible combinations and investigated their average sensitivity and variance of the error signal due to truncation of the multipliers.

Table 1 Results for Fukuma's Int-DCT.

N	IR5	IR4	IR3	IR2	IR1	Average Sensitivity	PSNR	Remark
12	A1	A2	A1	A2	A2	4.78	**50.55**	Best
1	A1	A1	A1	A1	A1	5.54	**48.77**	Existing
22	A2	A1	A2	A1	A2	5.77	47.05	Worst

Table 2 Results for Soontorn's Int-DCT.

N	IR5	IR4	IR3	IR2	IR1	Average Sensitivity	PSNR	Remark
14	A1	A2	A2	A1	A2	3.8	**53.2**	Best
32	A2	A2	A2	A2	A2	3.89	**52.4**	Existing
19	A2	A1	A1	A2	A1	4.09	49.27	Worst

Table 1 and **table 2** summarize results for Fukuma's Int-DCT and Soontorn's Int-DCT respectively. We applied the optimum word length allocation method [8,9] to all the 32 combinations. AR(1) with ρ =0.95 is used for the analysis as a colored signal. Existing methods employ either type A1 or A2 only, namely indicated as "N=1" or "N=32".

In case of Fukuma's Int-DCT, the best combination (N=12) improves PSNR by 1.8 dB compare to the existing combination (N=1). Improvement of 0.8 dB in PSNR is also confirmed for Soontorn's Int-DCT. **Figure 6** shows a comparison of PSNR for 32 combinations when average word length $\overline{W} = 3$. It indicates that PSNR of Soontorn's Int-DCT is better than that of Fukuma's under the same average word length.

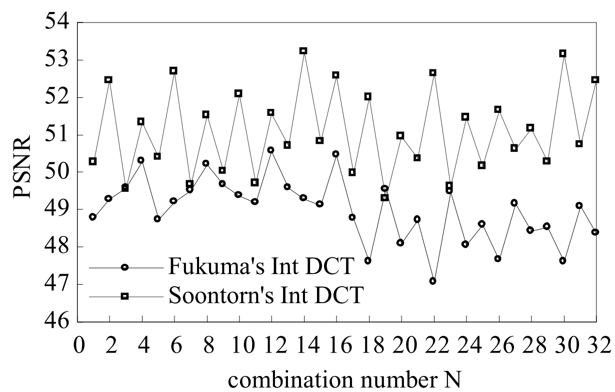

Fig 6 PSNR versus 32 combinations of two Int-DCTs.

5. CONCLUSIONS

We have theoretically and experimentally investigated sensitivity of variations of the integer rotation transform (IR) in the integer DCT (Int-DCT). It is explicitly described that the sensitivity depends on input signal's "color-ness" (correlation among pixels) and therefore we have utilized it to reduce word length of multipliers so that circuit scale of the Int-DCT can be reduced. As a result, the best combination of the IR types that gives us the best Int-DCT, namely the least sensitive to truncation of multiplier values, was found. Effectiveness of the finding was confirmed to be 0.8 to 1.8 [dB] in PSNR.

Analysis in this report has been limited to the matrix with size of two. It should be expanded to size of four or eight for further improvement in the future.

6. REFERENCES

[1] K.R.Rao and Yip, "Discrete Cosine Transform, Algorithm, Advantage and Applications". New York: Academic, 1990.

[2] D.S.Taubman, M.W.Marcellin, "JPEG 2000 - Image compression fundamentals, standards and practice", Kluwer Academic Publishers, 2002.

[3] S.Fukuma, K.Ohyama, M.Iwahashi, N.Kanbayashi, "Lossless 8-point Fast Discrete Cosine Transform using Lossless Hadamard Transform", IEICE Technical Report, DSP99-103, pp.37-44, Oct. 1999.

[4] Y.J.Chen, Soontorn Oraintara, T.Nguyen, "Integer Discrete Cosine Transform (Int DCT)", the 2nd International Conference on Information, Communications and Signal Processing, Singapore, Dec. 1999.

[5] W.Philips, "The lossless DCT for combined lossy / lossless image coding", IEEE ICIP 98, vol.3, pp.871-875, Oct 1998.

[6] S.Chokchaitam, M.Iwahashi, S.Jitapanakul, "A New Lossless-DCT for Unified Lossless / Lossy Image Coding", IEEE MWS-CAS 2004, July 2004.

[7] W.Sweldens, "The Lifting scheme: A construction of second generation wavelets", SIAMJ. Math. Anal, vol. 29, no. 2, pp. 511-546,1998.

[8] Iwahashi, O.Nishida, S.Chokchaitam, N.Kanbayashi, "Optimum Word Length Allocation for Multipliers of Integer DCT", IEEE ISCAS 2003, vol.2, pp.400-403, May 2003.

[9] M.Iwahashi, K.Nakagawa, S.Chokchaitam, Y.Tonomura, "Theoretical Analysis on Optimum Word Length Assignment for Integer DCT", IEEE ICIP '04, WA-L4, pp.2507-2510, (2004.10)