An Efficient Architecture of Transform & Quantization Module in MPEG-4 Video Codec

Kibum suh[1] and In-Kuen Song[2]
[1]Dept. of Electronic and Communication Eng. Woosong Univ.
17-2, Jayang-Dong Dong-Ku Deajeon, Korea.300-718
Tel +82-42-630-9705 Fax: +82-42-630-9719 Email kbsuh@lion.woosong.ac.kr
[2]Dept. of Electronic and Communication Eng. Woosong Univ.
17-2, Jayang-Dong Dong-Ku Deajeon, Korea.300-718
Tel +82-42-630-9701 Fax: +82-42-630-9719 Email iksong@lion.woosong.ac.kr

ABSTRACT: In this paper, a VLSI architecture for transform and quantization module, which consists of 2D-DCT, quantization, AC/DC prediction block, scan conversion, inverse quantization and 2D-IDCT, is presented.
The architecture of the module is designed to handle a macroblock data within 1064 cycles and suitable for MPEG-4 video codec handling CIF image formats. Only single 1-D DCT/IDCT cores are used for the design instead of 2-D DCT/IDCT, respectively. 1-bit serial distributed arithmetic architecture is adopted for 1-D DCT/IDCT to reduce the hardware area in this architecture. As the result, the maximum utilization of hardware can be achieved, and power consumption can be minimized. The proposed design is operated on 27MHz clock. The experimental results show that the accuracy of DCT and IDCT meet the IEEE specification.

1. INTRODUCTION

Among various transform techniques for image compression, the discrete cosine transform (DCT) is the most popular and effective one in practical applications because it gives an almost optimal performance and can be implemented at an acceptable cost. It has been commonly adopted in the various standards for image compression such as H.263, JPEG (Joint Photographic Expert Group), MPEG (Moving Picture Experts Group) and HDTV. In the application requiring low power consumption like mobile electronics equipment, the DCT processor which consumes less power and has smaller area becomes a key component.

Because two of 2D DCT/IDCT core are required in such an application as the hybrid coding loop of the MPEG-4 codec, the area of the codec will be increased.

Furthermore MPEG-4 standard[1] has different coding methodologies compared with MPEG-2/1. For example, DC/AC prediction is different from MPEG-2/1. In the designing of MPEG-4 codec., the problem of AC/DC prediction block area is serious.

Therefore, this paper describes the architecture of the DCTQ module maximizing the utilization of the hardware resources.

Firstly, we propose the architecture for DCT and IDCT. Secondly, we propose the architecture for AC/DC prediction block.

The contents of the paper are as follows. The components in the Transform and Quantization module, its architecture are described in sections 2, and 3, respectively. The simulation results and conclusion is described in section 4 and 5, respectively.

2. THE DATAFLOW IN THE MPEG-4 CODEC

The data flow of MPEG-4 has different paths in the Encoder and decoder shown in figure 1. In the encoder path, after 2D FDCT, quantization is performed on transformed coefficient and DPCM is performed on quantized coefficient with the AC/DC predicted value. After determination of prediction direction by AC/DC prediction block, scan module store the AC/DC prediction results in the QCO(Quantized Coefficient) buffer with their scan order.

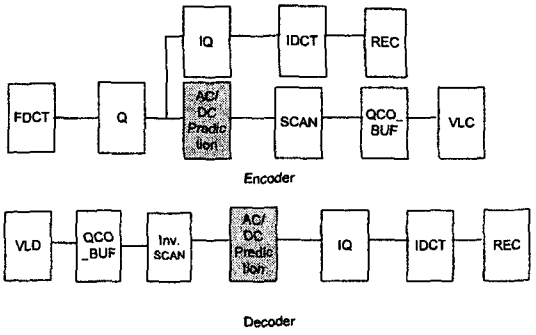

Figure 1. The dataflow of MPEG-4 codec

VLC module generates the bitstream from the QCO buffer. To reconstruct the frame memory for motion estimation, IQ and 2D-IDCT module are required.

In the decoder path, results from VLD are stored in the QCO buffer. After AC/DC prediction module determines the prediction direction, inverse scan module reads the data from QCO-buffer with its inverse scan order and transfer to AC/DC prediction block.

The results of AC/DC prediction are inverse-quantized and inverse DCTed to reconstruct the frame memory. AC/DC prediction is only used for intra macroblock of MPEG-4 codec, not used for short video header mode (H.263 mode)

For our implementation of MPEG-4 codec, our main purpose is to design a DCTQ module capable of 30 frame CIF processing for encoder and decoder with minimum hardware resources.

For hardware size reduction, as DCT/IDCT has orthogonal transform properties, 1-D DCT and 1D-IDCT can be commonly utilized, and substituted to 2-D DCT and IDCT respectively. Actually, when 2-D DCT is designed using row-column decomposition method, it is made up of two 1-D DCT processors and intermediate transposition memory. As the time of encoding and decoding are separated by frame base, all modules in figure 1 are shared for encoding and decoding

3. HW ARCHITECTURE AND TIMING

3.1. 2D-DCT/IDCT

The architecture of 2D IDCT core is illustrated in Figure 2. A number of fast DCT algorithms and implementation methods have been developed to reduce the number of multiplications and hardware size[4,5,7].Chen's algorithm[6] is employed in the construction of 1D DCT and IDCT. It also follows distributed arithmetic architecture in [3].

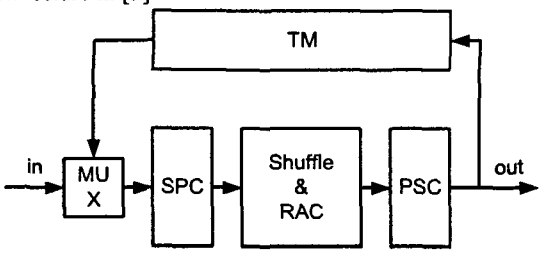

Figure 2. Proposed 2-D IDCT core architecture

To reduce the hardware area of DCT/IDCT, 1bit serial distributed arithmetic method is used.

For every 16 clock, external input is applied to 2D-DCT block during first 8 clocks and internal input from transposition memory(TM) is applied during last 8 clocks.

Both FDCT and IDCT have 3-step pipeline architecture. First step is serial to parallel converting process(SPC). Second step is shuffle and RAC(ROM and accumulation) and last step is parallel to serial converting process(PSC).

Table 1 shows the operating cycle of 2D DCT. In the table 1, 0(0.*) means processing block index 0 and 0^{th} 1D-DCT in row dimension.

912 (=114x8) clock cycles are needed to process one macroblock. In the table 1, the input from transposition RAM is marked with underscore line. After the 152(=19*8) cycles of initialization, FDCT block outputs the coefficient for every alternate 8 cycle.

Table 1. The operation cycle of 2D IDCT

CLK/8	SPC	RAC	PSC	CLK/8	SPC	RAC	PSC
1	0(0,*)	NOP	NOP	22	0(*,2)	1(2,*)	0(*,1)
2	NOP	0(0,*)	NOP	23	1(3,*)	0(*,2)	1(2,*)
3	0(1,*)	NOP	0(0,*)	24	0(*,3)	1(3,*)	0(*,2)
4	NOP	0(1,*)	NOP	25	1(4,*)	0(*,3)	1(3,*)
5	0(2,*)	NOP	0(1,*)	26	0(*,4)	1(4,*)	0(*,3)
6	NOP	0(2,*)	NOP	27	1(5,*)	0(*,4)	1(4,*)
7	0(3,*)	NOP	0(2,*)	28	0(*,5)	1(5,*)	0(*,4)
8	NOP	0(3,*)	NOP	29	1(6,*)	0(*,5)	1(5,*)
9	0(4,*)	NOP	0(3,*)	30	0(*,6)	1(6,*)	0(*,5)
10	NOP	0(4,*)	NOP	31	1(7,*)	0(*,6)	1(6,*)
11	0(5,*)	NOP	0(4,*)	32	0(*,7)	1(7,*)	0(*,6)
12	NOP	0(5,*)	NOP	33	2(0,*)	0(*,7)	1(7,*)
13	0(6,*)	NOP	0(5,*)	34	1(*,0)	2(0,*)	0(*,7)
14	NOP	0(6,*)	NOP	35	2(1,*)	1(*,0)	2(0,*)
15	0(7,*)	NOP	0(6,*)	36	1(*,1)	2(1,*)	1(*,0)
16	NOP	0(7,*)	NOP	...	.	.	.
17	1(0,*)	NOP	0(7,*)	....	...	...	...
18	0(*,0)	1(0,*)	NOP	111	NOP	5(*,6)	NOP
19	1(1,*)	0(*,0)	1(0,*)	112	5(*,7)	NOP	5(*,6)
20	0(*,1)	1(1,*)	0(*,0)	113	NOP	5(*,7)	NOP
21	1(2,*)	0(*,1)	1(1,*)	114	NOP	NOP	5(*,7)

Figure 3 shows the timing diagram of Transform and quantization module.

3.2. The architecture of AC/DC Prediction Block.

To perform the AC/DC prediction, the large memory for storing prediction data is needed. Because the memory to store DC value and AC value of the slice is so large.

Figure 4 shows the position of prediction value for block. For each block, to determine the direction of prediction DC coefficient of 3 adjacent blocks has to be read.

In the figure, for the block index 1 of row R1, left top block, left block and top block is 2,0,3, respectively. For the block index 3 of row R2 , left top block , left block and top block is 0,1,2, respectively. The macroblock consists of four 4 block such as 0,1,2,3 shown in the figure 5.

Therefore, for the processing of one macroblock, the DC value of R0 is required to predict the direction of block in R1 and the DC value of R1 is required to predict the direction of block in R2.

So the 2 line memory is needed for macroblock processing, but in the proposed architecture, one line memory is required.

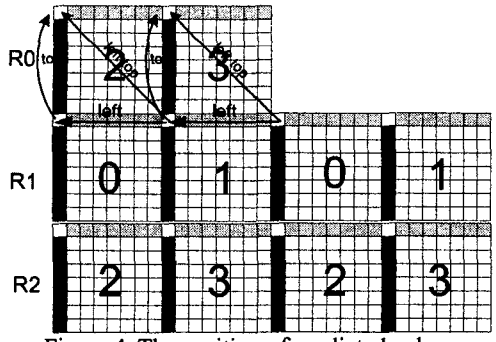

Figure 4. The position of predicted value

Figure 5 shows the structure of memory for prediction. The horizontal memory stores the DC and horizontal AC coefficient, and vertical memory stores vertical AC coefficient. LT_DC_value memory stores the left top value to be predicted. Its value is updated with top value of each block when a block read top value and is read according to the table 2.

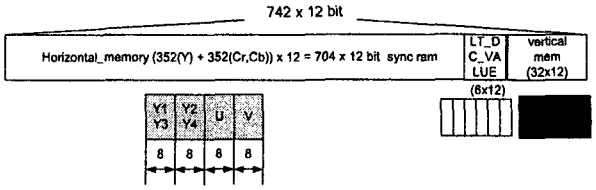

Figure 5. The structure of prediction memory

If current block index is 0, value of Lt is read from index 1 of LT_DC_VALUE memory and top value of block index 0 is stored in the LT_DCT_VALUE memory with index 0. That's why top value of block 1 represents LT value of block 0(see figure 4). Therefore, by copying the DC component of top to LT_DC_VALUE memory, we need only one row for prediction RAM.

Table 2. Storing left top value and reading for each block index

Block index	0	1	2	3	4	5
Storing lt_value	0	1	2	3	4	5
Lt memory read	1	0	3	2	4	5

Figure 6 shows the memory access of AC/DC prediction. In the figure, shaded index is reading and storing point for prediction RAM. Lightly shaded points represent the horizontal AC coeffieicient access, and thickly shaded block represent the vertical AC coefficient access.

To determine the prediction direction, we have to compare the horizontal gradient of DC value with vertical gradient. The determination of prediction is to be done before the prediction input. So reading the prediction data from memory starts 2 clocks before prediction input. After reading of LT value(B) from LT_DC_VALUE, L value(A) and T value(C), the gradient value is calculated and the direction of prediction is determined. When top value is read from horizontal memory, the value of top is stored into LT_DC_VALUE RAM area with the index shown in the TABLE 2. In the figure 6, LT Save is done immediately after the reading top value.

4. SIMULATION AND RESULTS

The 10000 block generated by a random function is applied to the hardware model of DCT/IDCT for the verification. The result of this is summarized in Table 3. The internal bit width is 16 bit, and it is sufficient to meet the IEEE specification [2].

Hardware complexities are denoted in table 4. Here, total gate count by equivalent 2-input NAND gate is 42,180 gates except RAM. The designed library is Hyundai 0.35 um CMOS standard cell TLM process.

Table 3. Accuracy simulation results for the proposed architecture

TEST	IEEE Spec.	Proposed
Pixel Peak Error	1	1
Peak pixel Mean Square Error	0.06	0.01310
Overall Mean Square Error	0.02	0.01020
Peak pixel Mean Error	0.015	0.009
Overall Mean Error	0.0015	0.0005

Table 4. Area reports of the proposed DCTQ

Components names		Area
2D-FDCT	Gate Count	7005
	RAM	64 x 16(bit)
2D-IDCT	Gate Count	8091
	RAM	64 x16(bit)
Q/IQ	Gate Count	3514
AC/DC Prediction Block	Gate Count	17939
	RAM	742x12(bit)
Scan Logic	Gate Count	2841
Quantized Coeff. Buffer	RAM (QCO_BUF)	384x12(bit)
AMBA interface	Gate Count	2790
Total gate		42180
Total RAM size		15536 bit

5. CONCLUSION

In this paper, the simplified architecture for transform & quantization module in MPEG-4 codec is proposed. As the architecture can be placed in the regular fashion, it is appropriate to be implemented with commercial ASIC technologies. It uses the structure and characteristics of MPEG-4 video codec, so the utilization of hardware resource is enhanced effectively, and power consumption is reduced. Therefore, it can be applied to the portable multimedia terminals for wireless multimedia services. The proposed design is implemented in MOVA ASIC[8].

6. REFERENCES

[1] ISO/IEC 14496-2 Final Draft of International Standard

[2]"IEEE standard Specifications for the Implementation of 8x8 inverse discrete cosine transforms," IEEE standard, pp.1180-1190, March 1991.

[3] S. A. White, "Applications of Distributed Arithmetic to Digital Signal Processing:A Tutorial Review," IEEE ASSP Magazine, July 1989.

[4] M. T. Sun, T. C. Chen, and A. M. Gottlieb, "VLSI Implementation of a 16x16 Discrete Cosine Transform Chip", IEEE Trans. Circuits and Syst., pp. 610-617, April 1989.

[5] A. Madisetti et al., "A 100 MHz 2-D DCT/IDCT Processor for HDTV Applications", IEEE Trans. CAS for Video Tech., Vol.5,No.2, pp. 158-165, APRIL 1995.

[6] W.H. Chen, C.H.Smith, and S.C. Fralick, "A fast computational algorithm for the discrete cosine transform" IEEE trans. Comm. Vol. Com-25,no-9,pp1004-1009, Sept.1977

[7] S. M. Park and et al., "A Single-Chip Video/Audio Codec for Low Bit Rate Application", ETRI J., pp. 20-29 vol. 22, no. 1, Mar. 2000.

[8] Juhyun Park, Koo bontae, Kim sung min, Ikgyun kim, hanjin cho, "MPEG-4 Video Codec for Mobile Multimedia Applications", ICCE 2001,June, 2001.

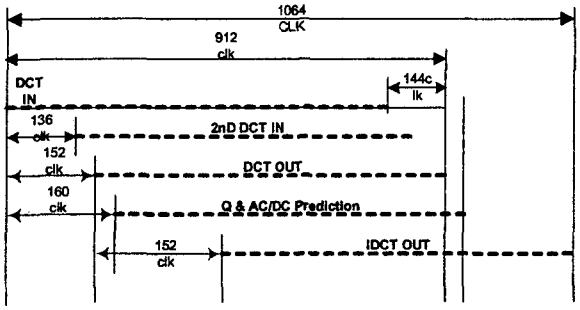

Figure 3. Timing of the proposed architecture

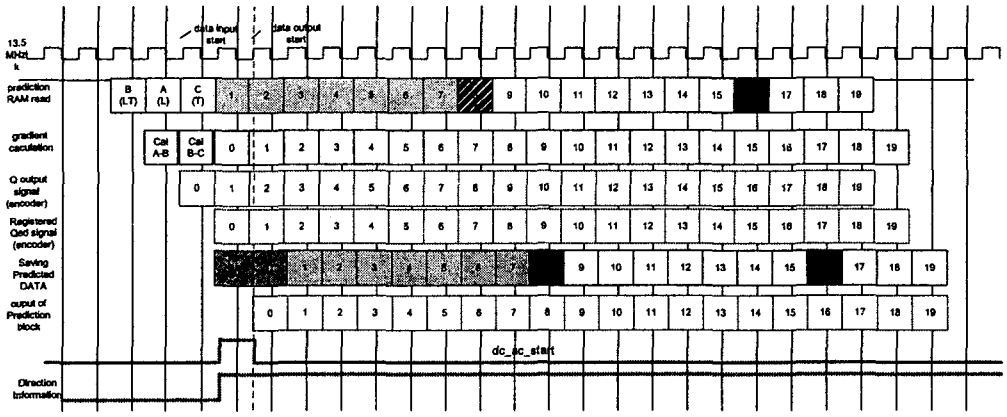

Figure 6.The memory access and operation of AC/DC prediction