REAL-TIME RECURSIVE TWO-DIMENSIONAL DCT FOR HDTV SYSTEMS †

C.T. Chiu and K.J. Ray Liu
Electrical Engineering Department and Systems Research Center,
University of Maryland, College Park, MD 20742 USA

ABSTRACT

The two-dimensional discrete cosine transform (2-D DCT) has been widely recognized as the most effective technique in image data compression. In this paper, we propose a new algorithm to compute the 2-D DCT from a frame-recursive point of view. Based on this approach, a real-time parallel lattice structure for the 2-D DCT is developed. The system is fully-pipelined with throughput rate N clock cycles for an $N \times N$ successive input data frame. This is the fastest pipelined structure known so far. Moreover, the 2-D DCT architecture is modular, regular, and requires only two 1-D DCT blocks which can be extended directly from the 1-D DCT array. We also propose a parallel 2-D DCT architecture and a new scanning pattern for the HDTV system to achieve higher performance.

1. INTRODUCTION

In recent years, much research has been focus on image data compression, especially for the application of the next generation TV, "HDTV". To make HDTV systems practical, bit rate reduction and data compression are indispensable [5]. The DCT coding approach has obtained most attention due to its superior energy compaction property and much simpler computations than the optimal Karhunen-Loeve transform (KLT). To satisfy the high speed video transmission system, fast and efficient algorithm to implement the 2-D DCT with simple hardware is strongly demanded [1, 2, 4]. The irregularity, global communication, and transposition delay of the existing 2-D DCT architectures have severe impacts on high speed video signal processing systems. Here we propose a new real-time recursive 2-D DCT architecture which requires only two 1-D DCT arrays and no transposition is required.

†The work is partially supported by ECD-8803012.

2. FRAME RECURSIVE 2-D DCT ARCHITECTURE

A new algorithm for the 2-D DCT by employing the frame-recursive concept [3] on successive input frames is presented. We adopt the frame-recursive approach since in digital signal transmission data arrive seriesly. Such approach can obtain the 2-D DCT in real-time recursively. Based on this method, a parallel and fully-pipelined 2-D DCT lattice structure which can dually generate the 2-D DCT and discrete sine-cosine transform (DSCT) is developed. The 2-D DCT $\{X_c(k,l,t) : k,l = 0,1,...,N-1.\}$ and 2-D discrete sine-cosine transform (DSCT) $\{X_{sc}(k,l,t) : k = 1,2,...,N; l = 0,1,...,N-1.\}$ of an $N \times N$ 2-D data sequence $\{x(m,n) : m = 0,1,2,...; n = 0,1,...,N-1.\}$ is defined as

$$X_c(k,l,t) = \frac{4}{N^2} C(k)C(l) \sum_{m=t}^{t+N-1} \sum_{n=0}^{N-1} x(m,n) \cdot \cos\left[\frac{\pi[2(m-t)+1]k}{2N}\right] \cos\left[\frac{\pi(2n+1)l}{2N}\right] \quad (1)$$

and

$$X_{sc}(k,l,t) = \frac{4}{N^2} C(k)C(l) \sum_{m=t}^{N+t-1} \sum_{n=0}^{N-1} x(m,n) \cdot \sin\left[\frac{\pi[2(m-t)+1]k}{2N}\right] \cos\left[\frac{\pi(2n+1)l}{2N}\right] \quad (2)$$

where

$$C(k) = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } k = 0 \text{ and } k = N, \\ 1 & \text{otherwise.} \end{cases}$$

In the following, we call $X_c(k,l,t)$ and $X_{sc}(k,l,t)$ the the t'th frame 2-D DCT and 2-D DSCT of an $N \times N$ 2-D data frame $x(m,n)$. The recursive relations for the $(t+1)$'th frame transformed data

$X_c(k,l,t+1)$ and $X_{sc}(k,l,t+1)$ as well as the t'th frame transformed data $X_c(k,l,t)$ and $X_{sc}(k,l,t)$ are given by

$$X_c(k,l,t+1) = \overline{X}_c \cos\left(\frac{\pi k}{N}\right) + \overline{X}_{sc} \sin\left(\frac{\pi k}{N}\right), \quad (3)$$

and

$$X_{sc}(k,l,t+1) = \overline{X}_{sc} \cos\left(\frac{\pi k}{N}\right) - \overline{X}_c \sin\left(\frac{\pi k}{N}\right), \quad (4)$$

where

$$\overline{X}_c = \frac{4}{N^2} C(k)C(l) \sum_{m=t+1}^{t+N} \sum_{n=0}^{N-1} x(m,n) \cdot \cos\left[\frac{\pi[2(m-t)+1]k}{2N}\right] \cos\left[\frac{\pi(2n+1)l}{2N}\right] \quad (5)$$

and

$$\overline{X}_{sc} = \frac{4}{N^2} C(k)C(l) \sum_{m=t+1}^{t+N} \sum_{n=0}^{N-1} x(m,n) \cdot \sin\left[\frac{\pi[2(m-t)+1]k}{2N}\right] \cos\left[\frac{\pi(2n+1)l}{2N}\right] \quad (6)$$

The relations between $\overline{X}_c$ and $\overline{X}_{sc}$ and the previous transformed data $X_c(k,l,t)$ and $X_{sc}(k,l,t)$ are

$$\overline{X}_c = X_c(k,l,t) + \delta_c(k,l,t) \frac{2}{N} C(k) \cos\left(\frac{\pi k}{2N}\right), \quad (7)$$

and

$$\overline{X}_{sc} = X_{sc}(k,l,t) + \delta_c(k,l,t) \frac{2}{N} C(k) \sin\left(\frac{\pi k}{2N}\right). \quad (8)$$

And the intermediate values $\delta_c(k,l)$ are

$$\delta_c(k,l) = \frac{2}{N} C(l) \sum_{n=0}^{N-1} [(-1)^k x(N,n) - x(0,n)] \cdot \cos\left[\frac{\pi(2n+1)l}{2N}\right]. \quad (9)$$

The relation between $X_c(k,l,t+1)$ and $X_c(k,l,t)$ is realized by lattice array II with lattice module shown in Fig. 1. It is noted that $\delta_c(k,l)$ in (9) is the 1-D DCT of the data vector which is the difference between the parity of the $(t+N)$'th row and t'th row of the 2-D input sequence. It can be shown that $\delta_c(k,l)$ can be generated time recursively by the lattice array I whose lattice module is plotted in Fig. 2. The fully-pipelined lattice structure for the 2-D DCT and DSCT is shown in Fig. 3 which includes one LAI, one $LAII$, and two circular shift arrays and shift register arrays.

The circular shift array in the middle of the system is an $N \times 1$ shift register array. This special shift register array loads an $N \times 1$ data vector from the LAI every N clock cycles, then it shifts the data circularly and sends the data to the $LAII$ every clock cycle. There are three inputs in $LAII$, δ_c, $X_c(k,l,t)$ and $X_s(k,l,t)$, where the δ_c comes from the circular shift array, and $X_c(k,l,t)$ and $X_s(k,l,t)$ from the shift register arrays located behind the $LAII$. We divide the $LAII$ into two groups: the $LAII_{even}$ and $LAII_{odd}$. Each includes $N/2$ lattice modules as shown in Fig. 3. The $LAII_{even}$ contains only those lattice modules for even transformed components k, while $LAII_{odd}$ contains only the odd lattice modules. The shift register array contains $2N \times N$ registers which are used to delay data for N clock cycles.

We will show how to apply the frame-recursive concept to obtain the 2-D DCT. Our approach is to send the input sequence $x(m,n)$ row by row directly into the LAI. It takes N clock cycles for the LAI to complete the 1-D DCT of one row input vector, then the array sends the 1-D DCT data in parallel to the $CSMII$ as shown in Fig. 3. The circular shift matrix II ($CSMII$) is an $(N+1) \times N$ sequential shift register. At the output of the $CSMII$, the 1-D transformed data of the $(t+N)$'th row and t'th row are added together according to (9) depending on the sign of the k components (see Fig.3). Then the results are sent to CSAs. The upper CSA translates the intermediate value $\delta_c(k,l)$ to the lattice array II_{even}, as do the lower CSA except that the signs of the output of the CSA are changed before being sent to the lattice array II_{odd}. Since $LAII_{even}$ and $LAII_{odd}$ have only $N/2$ modules, every δ_c is floating for $N/2$ clock cycles. It is noted that a specific 2-D transform data $X_c(k,l,t+1)$ and $X_{sc}(k,l,t+1)$ are updated recursively every N clock cycles from $X_c(k,l,t)$ and $X_{sc}(k,l,t)$. Therefore the outputs of the $LAII$ are sent into the shift register array (SRA) where data are delayed by N clock cycles. Each SRA contains $N/2$ shift registers each with length N. The data in the rightest registers are sent back as the $X_c(k,l,t)$ and $X_{sc}(k,l,t)$ of $LAII$. At the N^2 clock cycle, the 2-D DCT and DSCT of the 0'th frame are available. After this, the 2-D transformed data of successive frames can be obtained every N clock cycles.

There are many interesting results in this structure. First, the lattice array can be viewed as a filter bank. It is because every lattice module itself is an independent digital filter with different frequency components $l = 0, 1, ..., N-1$. Moreover, all the lattice modules in this architecture have the same structure which is regular, modular, and without global communication. Second, the system requires only 2 1-D DCT arrays and is fully pipelined with throughput rate N clock cycle for frame-recursive approach. A comparison with existing algorithms is given in Table 1.

3. APPLICATION TO HDTV SYSTEMS

Most of the 2-D DCT implementations in HDTV systems are based on the row-column decompositions methods [5, 6]. Although fast algorithms exist for the 1-D DCT, the second 1-D DCT cannot start until all the first 1-D DCTs are completed. To speed up the operations, one method is to execute the first 1-D DCT in parallel. For the 8×8 case, there are 8 1-D DCT blocks to perform the first transform simultaneously. Assuming that each signal is 10-bit long, in order to to satisfy the precision, then the total number of bits required in the input is 640 bits, which is not practical in the circuit realizations. From this point of view, our serial input 2-D DCT system is more practical in hardware implementations. Moreover, if the speed of the circuit components, such as the ROM and adder, is high enough, our 2-D DCT system can be executed as fast as the sample clocking rate.

Although our 2-D DCT implementations are effective, transforming a video frame of 1080×1920 still requires intensive computations. Therefore, we designed a 2-D DCT architecture suitable for the HDTV system to achieve higher performance. The block diagram of the 2-D DCT encoder is shown in Fig. 4, where five 2-D DCT chips are included. Five chips were used because the ratio of pixel numbers per line for luminance signal Y and color difference signals U and V is 4:2:2. As the sampling frequency of HDTV is very high, the pixels of Y are divided into four groups, in order to carry out DCT in parallel. Additionally, the color difference signal Y and U are switched alternatively to another DCT coder. The scanning processor shown in Fig. 4 is used to divide the signal into four luminance components and one color difference component. The outputs of the 2-D DCT transformed data are sent to the entropy encoder in parallel or through multiplexers.

Since the transform block size is 8×8, we divided the frame into 135×240 blocks and 240 channels as shown in Fig. 5. The 2-D DCT are executed on each channel whose scanning pattern is shown in Fig. 5. This scanning pattern reflects the fact that our system is based on row by row scanning order and is fully pipelined. Thus, such a scanning method would maximize the system throughput.

4. CONCLUSIONS

In this paper, we propose a new 2-D DCT algorithm based on a frame-recursive approach. The resulting 2-D DCT architecture can be obtained by using only two 1-D DCT arrays, at the same time, the transposition procedure is eliminated. It, therefore, does not have the drawback of the row-column decomposition method in which a transposition is needed between the first and the second 1-D DCT. The parallel 2-D DCT architecture and the scanning pattern proposed in Section 3 can process the video data in real time and eliminate the waiting time in the DCT codings so that the system performance can be maximized. Consequencely, our real-time parallel and fully-pipelined 2D-DCT structure is very attractive in high speed transmission systems.

References

[1] W. Ma, " 2-D DCT systolic array implementation," Electronics Letters, Vol. 27, No. 3, pp. 201-202, 31st Jan. 1991.

[2] P. Duhamel and C. Guillemot, "Polynomial Transform computation of the 2-D DCT," IEEE ICASSP Proc., pp. 1515-1518, March. 1990.

[3] K. J. R. Liu, and C. T. Chiu, "Unified Parallel Lattice Structures for Time-Recursive Discrete Cosine/Sine/Hartley Transforms," submitted to IEEE Trans. Acoust., Speech, Signal processing.

[4] B. Silkstrom, *et al.*, "A high speed 2-D Discrete Cosine-Transform," Integration, VLSI journal 5, pp. 159-163, 1987.

[5] K. Kinoshita, T. Nakahashi, and . Eto, "130 M bit/s (H4 rate) HDTV Codec based on the DCT algorithms," Electronics Letters, Vol. 26, No. 16, pp. 1245-1246, 2nd Aug. 1990.

[6] S. Cucchi, and F. Molo, "DCT-based Television Codec for DS3 digital Transmission," SMPTE Journal, pp. 640-646, Sep. 1989.

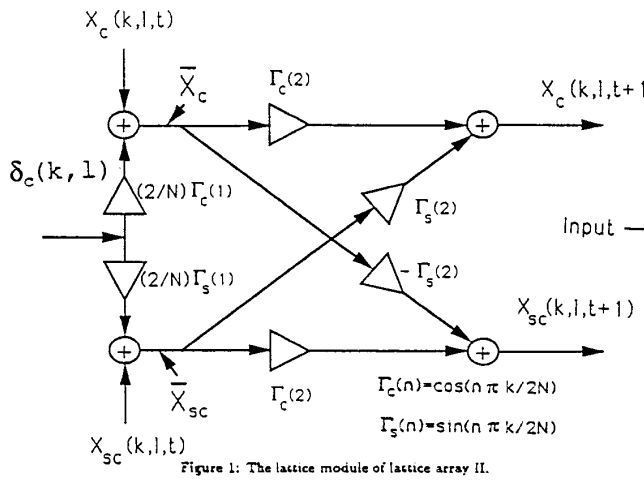

Figure 1: The lattice module of lattice array II.

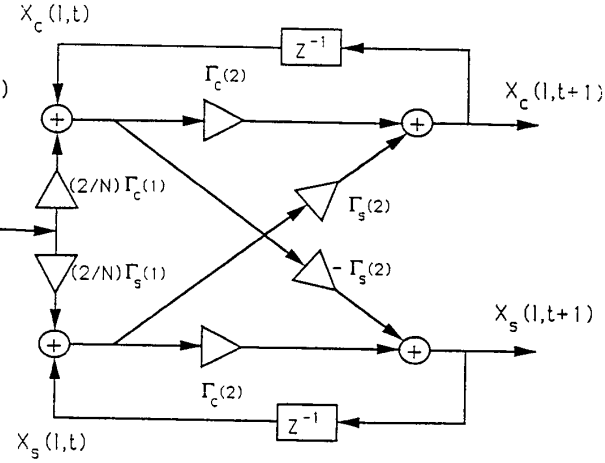

Figure 2: The lattice module of lattice array I.

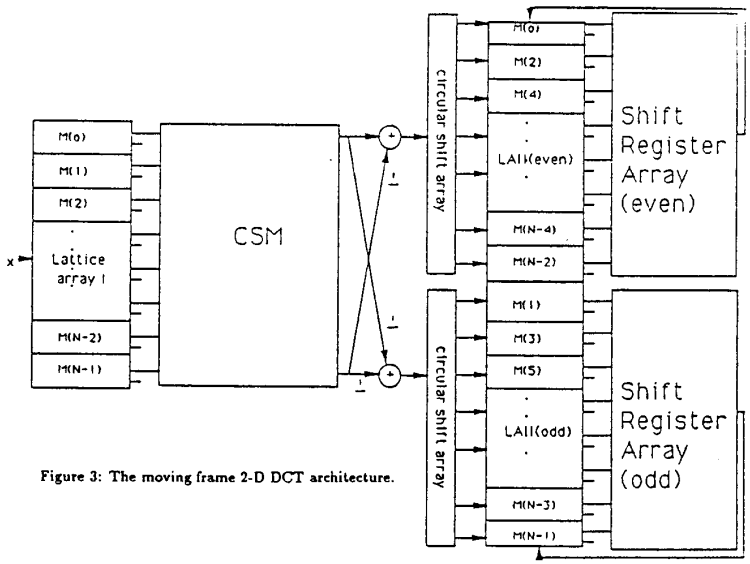

Figure 3: The moving frame 2-D DCT architecture.

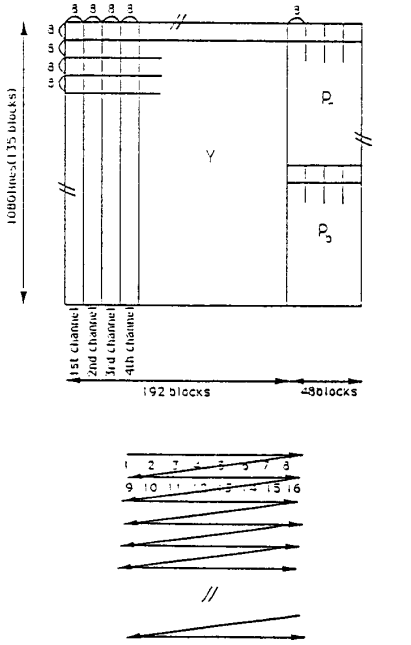

Figure 5: Block construction of a video frame and proposed scanning pattern.

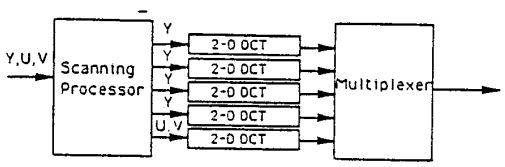

Figure 4: The block diagram of the DCT encoder.

	row-column method based on Chen in[2]	Duhamel[2] *et. al.*	Ma [1]	Liu-Chiu2D
No. of multipliers	$2N^2ln(N)$ $-6N^2/2+8N$	N^2 $+2N+2$	$4N(N+1)$	$8N$
Throughput	$N+$ transposition	$2N$	$2N+1$	N
Limitation on transform size N	power of 2	power of 2	no	no
Communication	global	global	local	local
I/O operation	*PIPO*	*PIPO*	*SIPO*	*SIPO*
Approach of of algorithm	indirect	direct	indirect	direct

Table 1: Comparisons of different 2-D DCT algorithms.