

二维 DCT 在 MVP 上的并行实现研究*

沈国斌** 杨绿溪 茅一民

(东南大学无线电系 南京 210096)

【摘要】 DCT 是一种被广泛应用于图像压缩编码中的正交变换。DCT 的运算量极大,因而在一些高速或实时场合,能否快速实现 DCT 就成为一个关键因素。TM S320C80(MVP)作为新一代可编程的通用数字信号处理芯片,它支持 SIMD 和 MIMD 两种并行模式。文中描述了 DCT 的快速算法和 TM S320C80 的基本结构及特性,并详述了并行实现的策略,对两种并行模式的加速比作了比较,然后选用了一种高效算法,在 TMS320C80 上实现。并对结果进行了理论上的分析,证明能满足实时性的要求。

关键词 微处理机; 芯片; 多媒体数字处理器; 离散余弦变换; 并行处理

中图分类号 TP302 TN911.76

由于离散余弦变换(DCT)是最接近于统计最优变换—Karhunen-Loeve 变换(KLT)的正交变换,因此被广泛应用于运动或静止图像的压缩编码算法中。已有的各种成熟的压缩标准如 JPEG、MPEG-1、MPEG-2、H.261 以及 HDTV^[1~4] 等都无一例外地采用 DCT/IDCT。由于它能使能量集中于低频区,并可利用人眼的视觉特性,再结合其他的编码方法,可产生很好的压缩效果。

DCT 的性能是优越的,但是计算量极大。一个大小仅为 8×8 的图像子块进行二维 DCT(2D DCT)需进行 8192 次乘法和 3584 次加法操作。在 MPEG-1 解码过程中,计算 IDCT 的时间占了总时间的 38.7%^[5]。尽管各种各样的快速算法可以大幅降低运算量(可达 1~2 个数量级),但其计算量仍然很大。人们总是希望压缩和解压速度越快越好,许多应用场合如 Motion-JPEG, MPEG-1, 2 等要求做到实时解压。尽管近几年来处理器的速度越来越快,但仍然难以满足实时性的要求,而用由于工艺水平的现状和物理极限的存在,一味地从提高处理器速度方面来解决实时问题,显然是不明智的。一个比较好的方法是并行处理。所谓并行处理就是根据事件的同时性、并发性和流水特性,利用多台或等效多台处理机同时来进行处理^[6]。静止图像具有空域并发性,序列图像还具有时域并发性,这些并发特性都十分有利于并行处理。许多厂家已经生产专用的 VLSI 芯片,它们都利用了并行计算,达到了实时性的要求,如 C-CUBE CL550^[7]。但是它们都存在功能单一、通用性差的缺点。而各大公司也已相继推出功能十分强大的通用 DSP 芯片,如 TI 公司的 TM S320C8X 系列、AT&T 的 AVP 系列等。如何在通用 DSP 芯片上实现实时的压缩和解压就成为一个热门的课题。本文研究了如何在 MVP(Multimedia Video Processor)上实现 2D DCT。

1 TM S320C80 的结构和特征

作为 TI 公司 TMS320 数字处理芯片系列的最佳典型,TMS320C80(亦被称为多媒体视频处理器 MVP)是数字信号处理器的一个真正的突破。四个高性能的、并行的 32-bit ADSP 和一个 32-

1996 年 9 月 4 日收稿,1996 年 10 月 8 日修改定稿

* 国家自然科学基金资助,基金编号:69402003

** 男 24 岁 硕士

bit RISC的主处理器被集成在一块芯片上,可以做到 2BOPS(每秒千兆 RISC操作)和 400 MB/s 的数据速率

1.1 MVP的结构及说明^[8,9]

MVP是由一个主处理器(MP)、四个并行的高级数字信号处理器(ADSP)、传输控制器(TC)、视频控制器(VC)、片内存储器(On-chip RAM)以及交叉开关(Crossbar)所组成。其结构如图 1所示。

1.2 MVP所独有的全新的特征

1.2.1 MP的特征

MP是一个内部集成有浮点运算单元(FPU)的 RISC处理器。它还具有复杂的移位、合并逻辑以及硬件支持的比特检测功能

1.2.2 ADSP的特征

1) 64-bit的长指令:一条指令可以相当于 15条 RISC操作

2) 片内存储器的单周期存取:允许每个处理器在每个周期内可有两个与数据处理操作并发进行的 32-bit数据传输。

3) 三个操作数且可分裂的 32-bit的 ALU(算术逻辑单元):可被分裂为两个 6-bit或四个 8-bit的单元,以便并发计算。

4) 16×16 bit的可分裂的乘法器:可以分裂成两个能并发进行的 8×8 bit乘法单元。

5) 三个无辅助操作的硬件循环控制器

6) 产生地址用的专用加法器:亦可进行算术操作

7) 有条件的 ALU和数据传输操作:可以代替许多比较、分支操作。

1.2.3 TC的特征

TC是一个智能的 DMA控制器。它主要负责与外部存储系统的接口。它能区分不同优先级的数据传输请求;支持多种传输模式,如多维传输、表控传输等。

1.2.4 Crossbar的特征

Crossbar是联系处理器和内存的纽带,是全连通的。它的接通速率非常高,一次连接只需一个指令周期。Crossbar的连接还能保证通过优先级来解决竞争问题

1.2.5 On-chip RAM的特征

MVP的片内 RAM为 50 kb大小的共享存储器(SRAM),分为 25个 2 kb的模块。每个模块可充当高速指令缓存器、高速数据缓存器、数据或参数存储器。每个 ADSP的局部端口只能与 8 kb的数据、参数存储器相连。

1.2.6 MVP的特征

以上所述特征决定了 MVP是一种 MIMD的并行处理芯片。当然,也可用于实现 SIMD和 SPMD的算法

2 DCT算法及常用快速算法

二维 DCT(2D DCT)的定义为

$$F(u, v) = \frac{2}{N} C(u) C(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos [^c(2x+1)u/2N] \cos [^c(2y+1)v/2N] \quad (1)$$

二维 IDCT(2D IDCT)的定义为

$$f(x, y) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u) C(v) F(u, v) \cos [^c(2x+1)u/2N] \cos [^c(2y+1)v/2N] \quad (2)$$

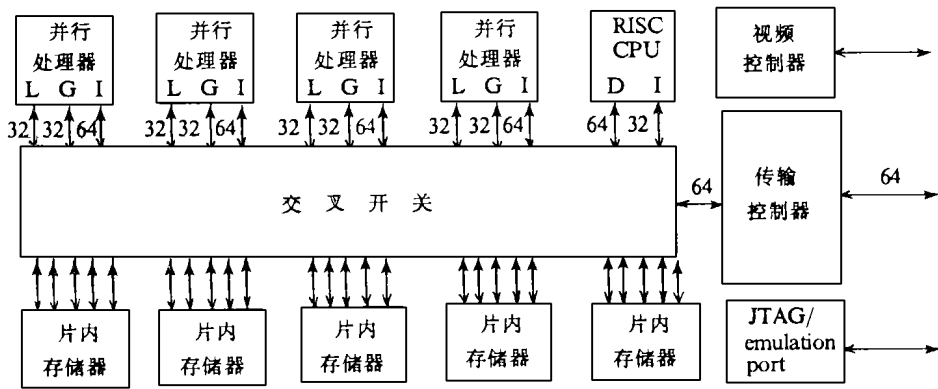


图 1 MVP 的组成结构

注: RISC CPU 是精简指令集中央处理单元

式中 (x, y) 是空域坐标; (u, v) 是变换域坐标 $C(u) = C(v) = \begin{cases} 1/\sqrt{2} & u = v = 0 \\ 1 & \text{其他} \end{cases}$
 它们的矩阵表示形式分别为 $F = CXC$ 和 $X = CFC$ (3)
 式中 F, X 分别为变换域和空域的矩阵, 通常 $N = 8$

$$C = [c(k, l)]_{N \times N} \quad c(k, l) = \frac{2}{N} \cos \left[\frac{(2k+1)lC}{2N} \right] \quad k, l = 0, 1, \dots, N-1$$

由于直接计算 DCT 的运算量极大 (见表 1), 无论对 2D DCT 还是 1D DCT 都要采用快速算法。对 2D DCT 而言, 快速算法主要分两大类, 即行列分解算法和矢量基算法。对于 MVP 之类的 DSP 芯片内部都有高速乘法器, 使得作一次乘法只需一个周期, 和作一次加法的时间相差无几; 加之行列分解算法又比较易于实现 (见表 1), 故仅采用行列分解算法。所谓行列分解算法就是 2D DCT 分成两个级连的 1D DCT (由式 (1) 易见其正确性), 即先进行行的 1D DCT, 然后将中间结果进行转置, 再进行列的 1D DCT。可见, 1D DCT 的快速算法对减少计算量有很大的贡献, 常用的 1D DCT 的快速算法^[10-13]及其所需的乘法、加法次数见表 2。

表 1 $N \times N$ 点 2D DCT 的运算量

计算方式	乘法次数	加法次数	难易程度
直接计算	N^4	N^4	极易
行列分解	$2N^3$	$2N^3$	易
快速行列分解	$N^2 \log_2^2 N$	$N^2 \log_2^2 N$	较易
矢量基	$3N^2 \log_2^2 N / 4$	$3N^2 \log_2^2 N$	难
		$- 2N(N-1)$	

表 2 8 点的 1D DCT 快速算法比较

快速算法	乘法次数	加法次数
Lee's	12	29
Chen's	16	26
Wang's	13	29
Arai's	5	29

注: 快速行列分解指分解后用快速算法实现

3 DCT 的并行实现方式及其到 MVP 上的映射

从式 (1)~(3) 易见 2D DCT 具有固有的空域并行性。对 M 个处理机而言, 可以通过简单的数据分布 (无需运用重叠技术, 因为只对 8×8 块操作), 以 SIMD 的方式做到近似 M 的加速比。之所以达不到 M 是因为数据通信时间也是总的处理时间的一部分。因 MVP 中有四个 ADSP, 为了便于比较, 所以以处理机数目 $M=4$ 为例来说明在一般并行机上进行处理的方案。设有四个 8×8 子块, 则存在以下两种可能:

- 1) 每个处理机处理一块; 显然加速比为 4

2) 多个处理机共同处理一块,则处理每块的时间减少,但处理器之间的通信次数增多且需重复多次。存在两种情况,以两个处理器共同处理一块为例,其他类同,如图 2 所示。

- (1) 在图 2a 中构成流水线,由于不同的处理机同时进行的是不同的操作,为 MIMD 方法
- (2) 在图 2b 中完全并发,由于不同的处理机同时进行的是相同的操作,为 SIMD 方法。

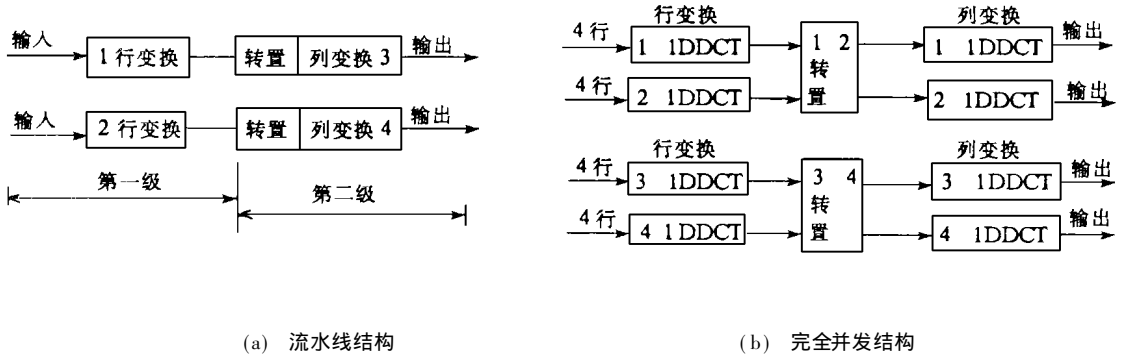


图 2 流水线和完全并发结构示意图 (图中 1, 2, 3, 4 表示四个 ADSP)

多处理机进行转置操作时,可以进行并行传送,因而传送次数相应减少。图 2b 的方法其加速比亦近似于 4 图 2a 的方法其加速比如下,显然介于 2 和 3 之间

$$\text{Speedup} = 4[8R_t + 8C_t + T] / [8R_t + \max(8C_t + T, 8R_t) + 8C_t + T] \approx 2(16R_t + T) / (12R_t + T) \quad (\because R_t \approx C_t)$$

其中 R_t 为一行行变换时间; C_t 为一列列变换时间; T 为转置时间

从实际应用的角度 (例如考虑到 DCT 的后续操作,因为 DCT 总是和别的算法一起使用的。)出发,我们选择方法一,因为加速比最大而且简单。即在用 MVP 实现时,每个 8×8 块由一个 ADSP 单独处理。

尽管 Arai's 算法所需运算量最小,但由于数据之间的相关性很强,不利于并行化。每个 ADSP 可以在一个指令周期内并行实现一次乘、一次可分裂的 ALU 操作以及两次内存操作,因此要选择这样一种快速算法:其运算量小,数据间相关性不很强,且乘加次数之比接近于 1:2,以充分发挥 MVP 的性能。显然 Lee's 算法是最优的。Lee's 算法的核心是如图 3 所示的两个蝶形结。

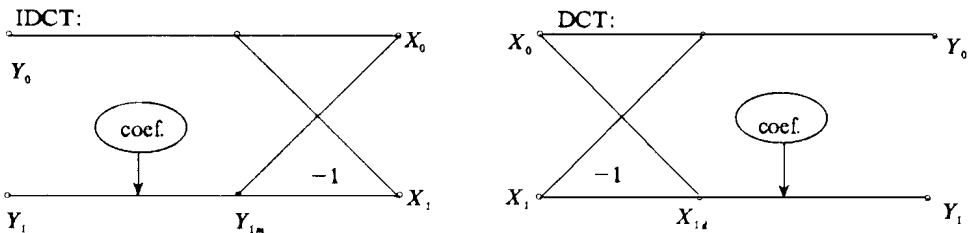


图 3 Lee's 算法

表示为

$$\begin{aligned} Y_m &= Y_1 * \text{coef} & x_0 &= Y_m + Y_0 & x_1 &= x_0 - Y_1m \\ x_1d &= x_0 - x_1 & Y_0 &= x_0 + x_1 & Y_1 &= x_1d * \text{coef} \end{aligned}$$

由此可见,实现一个蝶形结需一次乘和两次加操作。8 点的 DCT 需 12 个这样的蝶形结以及一些附加操作。

对目前的各种标准,包括医学图像压缩,采用 16 bit 来计算就可以满足精度要求。因为 MVP

的 ALU 和寄存器都是 32 位的, 为充分发挥其效率, 必须把两个 16 bit 的像素组合到 32 bit 的寄存器中。通常都是把相邻的两个像素进行组合, 这样就使得行变换和列变换要进行不同的操作。具体实现如下

; COL DCT/IDCT $d_3 = \text{mealu}(\text{LAB}; d_1 - d_2) \parallel [\text{par. multiply}]$

$d_4 = r(d_3 * d_7) \lll 3$

$\parallel d_5 = \text{mealu}(\text{LAB}; d_1 + d_2)$

$d_6 = r(d_4 * d_7) \lll 3 \parallel [\text{par. ealu0}]$

; ROW DCT/IDCT:

$d_2 = r(d_1 * d_7) \lll 3 \parallel [\text{par. ealu0}]$

$d_3 = \text{mealu}(\text{LAB}; d_2 - ((d_2 \parallel 16) \& \sim \% \% 16) - (d_2 \parallel 16) \& \% \% 16)$

$\parallel [\text{par. multiply}]$

其中 d_1, d_2, \dots, d_7 都是 MVP 中 32 位数据单元寄存器, $[\text{par.} \dots]$ 表示可以并发进行的操作。除去并行操作, 行变换每个蝶形结所需的净指令数为 1, 列变换的为 2, 但列变换每次进行的都是两列, 而行变换只一行。由于乘法是 16×16 位的, 故 MVP 的乘法器不能分裂开使用, 但其他特性的利用使得可以在保证精度的前提下连续进行高速乘法运算。32 位的 ALU 被分裂成两个 16 位的 ALU (一条 $\text{mealu}()$ 指令完成两个加法操作)。当然, 每条指令中都要有内存操作, 以为下面的指令准备好操作数。如此, 实现一个 8 点的行变换需 16 个指令周期, 两个 8 点的列变换需 29 个指令周期。不妨取平均数 16 个指令周期, 即 320 ns (MVP 的时钟频率为 50 MHz) 就可实现 8 点的 1D DCT。一个 2D DCT 要 16 个 1D DCT 实现, 故 8×8 点的 2D DCT 所需时间为 5120 ns 。设有 512×512 的图像, 共有 4096 个 8×8 子块, 则每个 ADSP 处理 1024 块, 对它进行 2D DCT 的时间为 $5120 \text{ ns} \times 1024 = 5.24 \text{ ms}$, 这个速度可以满足实时性的要求。

4 结 论

本文描述了 MVP 的结构和特性, DCT/DCT 的快速算法以 DCT 的并行特性, 详细讲述了 2D DCT 的并行策略, 并且就 MVP 的特性选择了一种最合适的算法, 详细描述了该算法的核心及其到 MVP 上的映射实现。分析结果表明如此实现的 DCT 速度很快, 完全可以用于实时场合。

参 考 文 献

- Wallace G K. Overview of the JPEG (ISO/CCITT) still image compression standard, image processing algorithms and techniques. Proceedings of the SPIE, 1990, 1244: 220-233
- CCITT. Recommendation H. 261-Videocoder for audiovisual services at p x 64 kbits/s. Geneva, 1990
- Coding of moving pictures and associated audio-for digital storage media at up to 1.5 Mbits/s. MPEG 1 Committee Draft of Standard 11172, Int'l Organization for Standardization, 1991
- Coding of moving pictures and associated audio. MPEG 2 Committee Draft of Standard 13818, Int'l Organization for Standardization, 1993
- Algorithmic and architectural enhancements for real-time MPEG-1 decoding on a general purpose RISC workstation. IEEE Trans on Circuits and Systems for Video Technology, 1995, 5
- 张德富. 并行处理技术. 南京: 南京大学出版社, 1993
- CL550 users manual, C-cube microsystems. Milpitas, CA, 1992
- TMS320C80(MVP) master processor user's guide. Texas Instruments Com
- TMS320C80(MVP) parallel processor user's guide. Texas Instruments Com

- 10 Lee B G. A new algorithm to compute the discrete cosine transform. IEEE Trans ASSP 1984, ASSP-32(6): 1243~1245
- 11 Chen W H, Harrison Smith C, Fralick S C. A fast computational algorithm for the discrete cosine transform. IEEE Trans Comm. 1977, COM-25(9): 1004~1009
- 12 Arai Y, Agui T, Nakajima M. A fast DCT-SQ scheme for images. Trans IEICE, 1988, E71(17): 1095~1097
- 13 Zhongde Wang. Fast algorithms for the discrete W transform and for the discrete fourier transform. IEEE Trans ASSP, 1984, ASSP-32(4): 803~816

Implementation of 2D DCT on MVP

Shen Guobin Yang Luxi Mao Yimin

(Southeast University Nanjing 210096)

Abstract DCT is an orthogonal transformation widely used in image compression applications. Because of its high computational cost, the operating speed of DCT will be a key factor in the case where real time processing are required. TMS320C80(MVP) is a recently developed general-purpose programmable digital signal processor, which can operate in both SIMD and MIMD concurrent modes. This paper presents the fast algorithms of DCT and the advantages of TMS320C80 architecture, and discusses the parallel strategy of 2D DCT on MVP. The speedups of SIMD and MIMD modes are compared. Finally, Lee's fast algorithm is chosen and mapped onto MVP with SIMD mode. The results show that the implementation can meet the requirements of real time applications.

Key words microprocessor; chip; multimedia video processor; discrete cosine transform; parallel processing

编辑 徐安玉 黄辛

· 科研成果介绍 ·

微带化 8 mm 波铁氧体隔离器与环行器

主研人员: 刘强 曾升权 陈巧生 钱汝彪

该 8 mm 波段微带隔离器和环行器选用落入 (Drop-in) 式结构, 其特点是使用商标为 Duroid 5880 的双面敷铜聚四氟乙烯薄板作器件的微带基片。

技术性能指标分别达到:

1) 落入式微带环行器:

插入损耗 < 1 dB 隔离 > 18 dB; $V_{SER} < 1.3\Omega$ 绝对带宽: 0.7 GHz

2) 落入式微带隔离器:

正向损耗 < 1 dB 反向损耗 > 20 dB; 绝对带宽: 0.5 GHz

与其他形式的毫米波微带环行器相比, 该器件所用铁氧体样品厚度较大, 退磁场小, 所需外磁场可大大减小, 恒磁体的尺寸也可大大减小。

该种结构的毫米波微带铁氧体器件在国内是首次研制成功, 与国外同类器件相比, 在隔离特性方面具有明显的优势, 属国内领先水平。

· 科 卜 ·