

基于 DM642 的视频编码 Cache 优化策略

Cache optimization strategy of video coding based on DM642

(清华大学)王熹微 唐 昆 崔慧娟

Wang, Xiwei Tang, Kun Cui, Huijuan

摘要: TMS320DM642 DSP 是 TI 公司新近推出的一款高性能数字多媒体处理器,它的两级高速缓存(Cache)结构为高复杂度视频编码算法的高效率实现提供了有力的保证。综合考虑视频编码算法特点和 Cache 结构特征实现的算法,能够满足系统整体需求和指标要求。本文首先简要介绍 DM642 的 Cache 结构特点,然后针对视频编码算法的高复杂度,提出一种 Cache 优化策略,能够很好的解决应用系统中的实时实现问题。

关键词: DM642; Cache; 视频编码; 优化策略

中图分类号: TN91

文献标识码: A

文章编号: 1008-0570(2005)9-2-0084-03

Abstract: TMS320DM642 is one high performance digital media processor recommended by TI Corp. Its two level cache structure could help to realize the high-complexity video coding algorithm efficiently. The algorithm, which takes the feather of the video coding and the two-level cache structure into account, would meet the requirement of the real-time application. In this paper, the DM642 cache structure is introduced first, and then according to the high complexity of video coding algorithm, a cache optimization strategy is proposed to realize the real-time video application system.

Key words: DM642; Cache; Video coding; Optimization strategy

随着多媒体业务的飞速发展,低速率高质量的活动图像编码成为当前国内外研究的热点之一。由国际电联和 MPEG 标准组织 ISO 共同发布的 H.264/AVC 视频压缩标准,以技术的显著优势和多个标准组织的支持,迅速应用到视频服务、媒体制作发行、固定及移动运营网络、设备终端制造等多个领域。

但 H.264 在大幅度提高编码效率的同时,带来了运算复杂度猛增的新问题。许多针对提高图像质量,加快编解码速度方法被提出并实现, H.264 编码器在 DSP 上的应用也日臻成熟。在目前的 DSP 平台上实时实现大尺寸、高分辨率的视频编解码仍旧是大家关注的焦点。如何合理利用 DM642 平台的优势高效率的实现 H.264 视频编码算法成为急需解决的问题。

1 视频编码算法的高复杂度

H.264 包括视频编码层 (VCL) 和网络抽象层 (NAL), VCL 研究基于运动补偿、变换编码、熵编码技

王熹微:博士

基金项目:国家自然科学基金项目(60272020)

术来提高视频编码效率, NAL 用于支持 H.264 的网络友好,将 VCL 编码数据组合到现有网络中以提供更高的灵活性,目的是基于兼容当前视频编码的流行标准 H.263 和 MPEG-4 技术,找到一种新标准用于取代现有任何实现高质量视频的视频编码标准。

它以运动补偿和变换编码为框架,利用多参考预测帧提高编码效率,1/4 像素搜索提高预测精度,基于 4x4 块编码、多种帧间预测模式提高峰值信噪比。其中,七种不同块尺寸大小的帧间预测技术可以节省 15% 的码率,1/4 像素精度估值相对于整像素估值可以节省 20% 的码率,采用五帧参考帧预测可节省 5~10% 的码率。这带来了运算复杂度猛增的新问题。

2 TMS320DM642 DSP 及 Cache 结构

2.1 DM642 平台简介

DM642 是德州仪器(TI)公司最新推出的一款专门面向多媒体处理领域应用的处理器,是构成多媒体通信系统的良好平台。它丰富的外围接口使得它近乎是一个多媒体嵌入式系统的单芯片硬件平台;它的完全可编程性,又可以使得它能够兼容正在发展的各种多媒体信号处理标准,构成通用的软件平台。

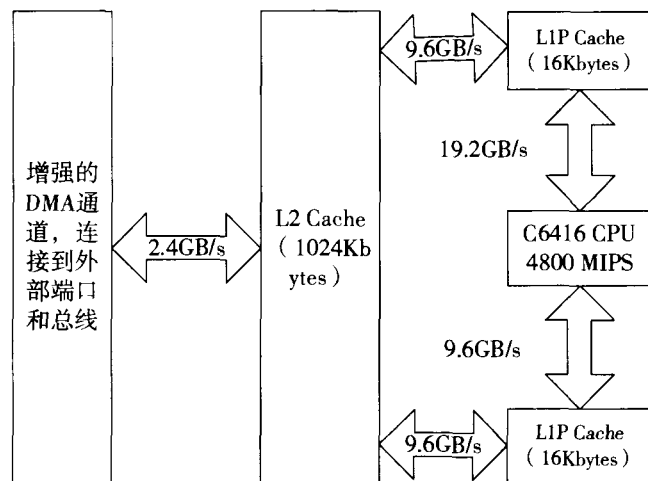


图 1 两级 CACHE 结构框图

DM642 采用的是 C64x DSP 核, 时钟频率高达 600MHz, 提供 8 个并行处理单元, 处理能力高达 4800MIPS, 拥有两级 Cache 结构, 适用于 V2oIP、视频点播 (VOD)、多通道数字录像应用, 以及高质量的视频编码、解码解决方案。为满足视频和图像处理的需要,

还集成了三个高精度可配置的视频接口、面向音频应用的串行端口 McASP、10Mb/s 或 100Mb/s 模式自适应的以太网口(EMAC)等外设。其改进的数据路径结构能够提供一个时钟周期内两个 64-bit 的读取/存储,在图像算法处理方面能力得到了很大提高,相对于 300MHz 主频的 C62x DSP 能够最多有 19 倍的性能改善。

表 1 L1P、L1D、L2 参数比较

	L1P	L1D	L2
组织形式	直接映像	2路组合	4路组合
行大小	32Byte	64字节	128 Bytes
Set 数量	512个(512行)	128个	
读 miss 指派	向 L2 请求一行(CPU 阻塞<=8周期)	在 L1D 中指派一行	数据取自 EDMA, 存储在 L2 中,并送往 L1
读 hit	数据直接由 L1P 中读出	数据直接由 L1D 中读出	由 L2 读数据
写 miss 指派	不支持 L1P 写	不在 L1D 中指派, 数据送给 L2	先将数据取自 EDMA, 写入 LRU
写 hit	不支持 L1P 写	写入 L1D 中命中的地方	写入命中的 L2 单元
CPU 存取时间	1个周期(1cycle)	1个周期(1cycle)	
代换策略		2路最早使用原则	4路最早使用原则

2.2 DM642 的 cache 结构

DM642 的高性能还得益于 CPU 的两级高速缓存结构,其结构框图如图 1 所示。

CPU 和一级程序高速缓存(L1P)及一级数据高速缓存(L1D)直接连接,L1P 和 L1D 分别为 16KByte,第二级缓存(L2)有 256KByte,程序空间和数据空间是共享的,可以设置成存储单元(memory),高速缓存(cache),或者是这两者的结合。表 1 列出的是 L1P、L1D 和 L2 各个性能参数和使用方式的对比。

L2 一部分配置为 SRAM 时,映像入 C64x 寻址空间,CPU 可以直接进行存取,L2 一部分配置为 cache,大小可为 32KB、64KB、128KB、256KB 中的一种,任何被配置为 Cache 的 L2 SRAM 不处于存储器映像中,不能被存取。

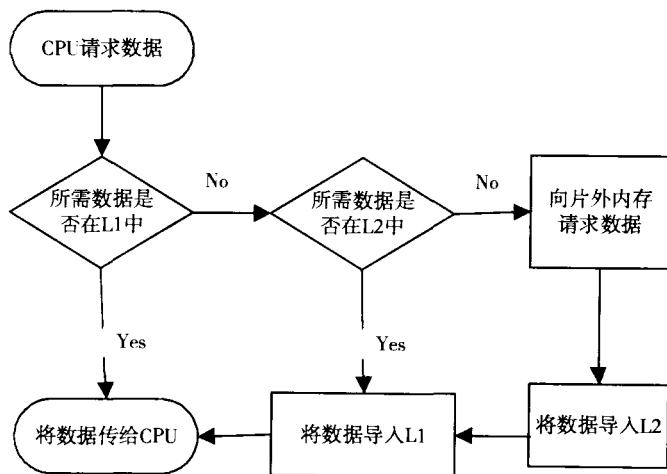


图 2 两级缓存的访问原理

两级 Cache 的工作原理如图 2 所示。CPU 只对 L1 的数据进行访问,程序代码和数据必须经过存储器到

L2, L2 到 L1 的逐级搬移才能被 CPU 访问。在 CPU 处理片内的数据时可以通过 EDMA 把片外的数据倒入片内,并行工作以提高效率。由于内部存储器的工作频率与 DSP 内部时钟同频,而远远高于片外存储器的频率,这就解决了 DSP 外部时钟频率小于内部时钟频率的问题。实验表明,合理利用两级缓存配合低工作频率的外部存储器,系统的效率能够达到全部使用高工作频率的内部存储器的 80%到 90%。

3 Cache 优化策略及优化结果

Cache 的性能主要依赖于 cache 行的重复使用情况。一般性应用中涉及大量不可预测的内存访问时(比如控制流程等)使用 L2 cache;在实时信号处理系统中必须要使用 L2 SRAM。具体的优化方法,包括应用层优化和程序层优化两方面。

3.1 应用层优化

3.1.1 合理配置 L2 的 Cache 和 SRAM 的大小

配置的总原则是将尽量多的关键数据分配在片内,Cache 越大越好,对于不同的应用需要用不同的配置。最优配置需要在开发中根据经验和实际的测试结果进行选择。

在我们的实时系统中把 cache 配置成 128KByte,剩余的 128KByte 作为 SRAM 使用。从表 2 可以看到,将 L2 的 Cache 设为 128KByte 时,系统性能最佳。

表 2 不同 L2 配置方式对系统性能影响

配置的 Cache 大小 (KByte)	CPU Load (%)	帧率(帧/秒)
128	76	25
64	77	20
32	86	13
0	86	13

3.1.2 信号处理代码与通用处理代码的分别对待

通用目的处理程序一般而言是顺序执行的数据流和条件分支,没有太多的并行性,执行时依赖于不同的模式和条件,数据内存访问具有更大的随机性,对代码空间的访问更多的是带分支的线性流程。L2 SRAM 不足以大到存放系统所有的数据和程序,因此把通用用途的代码和相关数据放在片外并将该存储空间设置成允许 Cache 方式。这种方式将把 L2 SRAM 留给对性能要求更好的信号处理部分的代码。

3.2 程序层优化

Cache 在程序层的优化,关注的是数据、代码在内存中的分配方式和函数的调用方式。优化的目的是尽可能减少 cache miss 的次数和 CPU 停止周期的次数。

3.2.1 选择合适的数据类型以减少对内存带宽的需求

在视频编码系统中的变换编码部分,H.264 中采用的是 4x4 整型变换。由于视频编码的输入是 8-bit

技术创新

精度的数,依据整型变换变化矩阵的特点很容易分析出在变换过程中的数,用 16 位来表示即可,不必要定义为 32 位。

对于代码段:for(i=0; i<n; i++) r[i]=x[i]-c; 表 3 给出了数组 x[] 定义成 int 型和 short 型对应参数比较。在改用 short 型定义 x[] 后,该段代码的执行速度得到了 2.3 倍的提高,相应的 cache miss 次数减半。

表 3 不同数据类型对实现性能和 cache 效率的影响

	x[] 定义为整型 (周期)	x[] 定义为 short 型 (周期)
执行周期	3117	1071
L1D 停止周期	2063	1152
L1P 停止周期	17	22
总执行周期	5197	2248

3.2.2 对处理链进行优化

对于视频编码应用来说,数据流是顺序的,即前一个函数的输出将作为下一个函数的输入。函数 1 当前的输入在 L1D 中,输出数据将被存放在低一级的存储空间(L2 或者外部存储器),函数 2 在读取数据时会发生 miss。为了减少该类 miss,可将函数 1 的输出写入 L1D,则该部分数据可以直接重新访问而不会发生 CPU 停止。这种优化的理念可以用图 3 来表示。

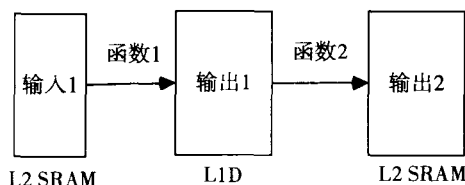


图 3 两个函数的处理链

3.2.3 避免 L1P 因冲突造成的读 miss

当读取内存地址映射到相同 cache 单元却不在 cache 中的代码时,会导致这类读 miss 的产生。编译器和连接器不会考虑到 cache 冲突问题,必须手动优化这部分代码。通过下面的例子来说明这类读 miss 的几种情况和相应的解决方法

```

For(i=0; i<N; i++) { func1(in[i], tmp); func2 (tmp, out[i]); }
  
```

PCI 总线 IEEE-488 接口卡及系列软件

西安交通大学在推广 ISA 总线 IEEE-488 接口卡 (GPIB-PC1、GPIB-PC2、GPIB-PC3) 基础上,最近又研制成功 PCI 总线的 IEEE-488 接口卡 (GPIB-PC4)。GPIB-PC4 卡符合国际标准,适用于 PCI 总线微机,通信速度相对 ISA 总线卡大大提高。与其配套的驱动软件支持 Windows 环境下 VB、VC 和 DELPHI 语言编程,也可在 LabVIEW 环境下使用。编程使用方法与 ISA 总线卡完全相同,使用方便。该卡性价比优于国外同类产品,购买方式为款到寄货,免费保修,负责解答各种技术问题,有意者来信来电联系。

通信地址:西安交通大学自动控制系 邮编:710049 联系人:吴彩玲
电话:029-82668665-263, 82668665-163, 82660338
手机:13152188561 传真:029-82660338, 029-83237910

1、循环里两个函数映射到 L1P 中的地址空间有重叠的情况:

解决方法:调整函数在链接时候的顺序,保证一个循环中的两个函数 func1 和 func2 映射到 L1P 中的地址空间不重叠。这个可以通过将 func1 和 func2 在内存中连续存放来实现。

2、循环里的所有函数的代码大小比 L1P 大的情况:

解决方法:循环拆分。循环被拆分成如下形式,

```

For(i=0; i<N; i++) { func1(in[i], tmp[i]); }
  
```

```

For(i=0; i<N; i++) { func2(tmp[i], out[i]); }
  
```

这种改进方式会增加用于存储输出数据的临时缓冲区的大小。

在我们的视频编码系统中,对比例缩放归一化部分的代码进行这类优化,将该部分代码分割成一系列的核心循环以避免 L1P 因冲突造成的读 miss,对该部分的数据依照 3.2.2 中提及的方式优化,经测试可知,优化之后的程序,cache 的使用效率能达到 90%。

4 结论

Cache 工作在 CPU 时钟频率上,静态 L1 高命中率和灵活的 L2 访问机制提供了低成本高性能的使用。依据 DM642 两级 cache 特征,对高复杂度的视频编码算法 H.264 进行 cache 使用的优化,取得了非常显著的效果。

参考文献

[1]TMS320DM642 Video/Imaging Fixed-Point Digital Signal Processor Data Manual. Texas Instruments, sprs200B, May 2003

[2]TMS320C6000 DSP Cache User's Guide. Texas Instruments, spru656A, May 2003

[3]Cache Analysis User's Guide. Texas Instruments, spru575A, January 2003

作者简介:王熹微 (1977.11-),汉族,女,四川人,信息与通信系统在读博士,主要研究方向多媒体通信视频编码算法研究。E-mail: wangxiwei00@mails.tsinghua.edu.cn。唐昆(1945-),汉族,男,清华大学电子工程系博士生导师,Email: tangkun@mail.tsinghua.edu.cn。

Author brief introduction: Wang Xiwei, Female, Born in 1977, Phd. Candidate in DEE of Tsinghua Univ., focus on multimedia communication video coding algorithm research; Tang, Kun, Male, Born in 1945, Phd. Advisor in DEE of Tsinghua Univ. Research direction: Multimedia technologies.

(100084 清华大学电子工程系)王熹微 唐昆 崔慧娟
(Department of Electronic Engineering, Tsinghua University, Beijing 100084, China) Wang, Xiwei Tang, Kun Cui, Huijuan

通信地址:

(100084 清华大学东主楼 11#418) 王熹微

(投稿日期:2005.3.5) (修稿日期:2005.3.11)