

高性能图像匹配电路乘累加性能分析

周 泉, 曹 辉, 闫 博, 杨 靓

(西安微电子技术研究所, 陕西 西安 710054)

摘 要: 基于归一化互相关的图像匹配在导航制导和模式识别中应用广泛, 由于计算量大, 应用中通常采用硬件实现. 对级联乘累加和并发自累加两种典型结构电路的计算时间性能和资源利用率作了对比分析, 从而明确了如何根据具体的匹配尺寸择优选取一种电路完成计算, 以获得更好的实时性能和更好的资源利用率, 最后从数据复用的角度对存储模型作了说明, 并给出了计算核的带宽计算方法.

关键词: 级联乘累加; 并发自累加; 加速比; 数据复用; 冗余访问因子

中图分类号: TN402

文献标识码: A

文章编号: 1000-7180(2014)08-0056-05

Performance Analysis of Multiplication and Addition of High-performance Block Matching Circuit

ZHOU Quan, CAO Hui, YAN Bo, YANG Liang

(Xi'an Microelectronics Technology Institute, Xi'an 710054, China)

Abstract: The image matching based on normalized cross-correlation has been widely applied in navigation guidance and pattern recognition, because of the large amount of calculation, the application of the hardware implementation is normally used. There's a comparative analysis between the two typical structural circuits Cascade accumulation and Concurrent accumulation on computing time performance and resource utilization. So as to make clear how to pick over a circuit to complete the calculation according to the concrete matched sizes, which help to obtain a better real-time performance and better resource utilization. At the end, there's an explanation on storage model from the perspective of data reuse. And it gives out the computing method of the bandwidth of computing kernel in the end.

Key words: cascade accumulation; concurrent accumulation; speed-up ratio; data reuse; redundant access factor

1 引言

图像匹配在导航、制导、侦查等领域应用广泛. 实现图像匹配有很多种衡量配准的方式, 其中基于归一化互相关系数的图像匹配具有匹配精度高, 但是计算量有可能大到超过千亿次, 而其中乘加运算占据绝大多数. 随着图像匹配应用领域的不断拓宽, 对匹配的实时性提出了更多的挑战, 而在归一化互相关匹配中, 如何缩短乘累加的时间是提高归一化互相关匹配的实时性能的关键所在.

目前的研究多数都把注意力放在如何从算法上进行改进, 减少归一化互相关的计算量, 提出了多种

快速互相关匹配算法^[1-3], 从硬件实现的角度做定量的性能分析则较少见. 实际上, 对实现归一化互相关计算的硬件电路进行性能分析并给出不同硬件电路的性能对比, 有助于设计出更高效的硬件实现电路, 提出更好的数据复用方式, 从而从硬件实现上去提高计算的实时性能.

随着超大规模集成电路技术的发展, 研究者们提出并实现了很多集成电路实现图像匹配的计算模式. 本文针对归一化互相关系数匹配的级联乘累加和并发自累加两种广泛应用的计算结构做了时间和资源性能对比分析, 并给出了数据复用下存储器带宽的计算方式.

2 块匹配的两典型互联结构及性能参数建模

2.1 两种典型的互联结构

将乘法运算和加法运算单元放在一起可以形成一个独立的处理单元(Process Element, PE),一般这样的 PE 中,加法单元的一个加数来自乘法单元的输出,根据加法单元另一个加数来源的不同可以将 PE 分为两种类型:进位累加型(加法单元的另一个加数来自 PE 的进位输入端)^[4]和自累加型(加法单元的另一个加数来自加法单元自身的上一节拍输出)^[4-5],这两种类型的 PE 都具有内部结构简单,外部端口较少的特点,因此适合用来从空间上扩展形成大规模的互联阵列,将进位累加型的 PE 通过一定方式的组织能形成级联乘累加阵列^[4-8],而将自累加型的 PE 通过一定的方式组织能够形成并发自累加阵列^[4-5],这两种阵列能够让大量的乘累加运算采用流水并行执行,因此很适合用于互相关匹配算法中,从而极大程度上提高计算的实时性能.另外,级联乘累加还在有限脉冲响应滤波器中也有广泛的应用^[9].

2.2 时间和资源利用性能参数化模型

2.2.1 级联乘累加时间模型

级联乘累加计算模式的一维处理链^[8]模型如图1所示.计算时先由 Bin 端串行加载模板图数据,每个 PE 中对应保存一个模板图像素,且计算过程中保持不变.模板图数据加载完成后启动并发数据端 Ain 的参考图加载,每个时钟节拍加载一个参考图数据,并发完成与模板图相乘,乘积结果送入加法器完成与上一级的进位输入端的值相加,整个处理链以流水的方式计算,流水线建立后会在 Cout 端每个时钟节拍得到一个结果.在级联乘累加计算模式下可以看作是模板图(或它的一部分)在计算阵列中完成与整个参考图的匹配运算.如果模板图不能一次全部加载到阵列中,可以采用多次循环的方式,直到模板图加载完成为止,只是在下一次循环中需要将上次循环所得的乘累加结果以流水方式加载到整个

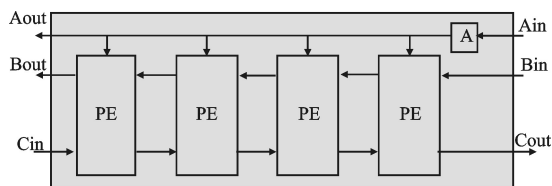


图1 级联乘累加处理链

阵列第一个 PE 的进位输入端.当所有循环都计算完成时,阵列输出的就是整个模板图与整个参考图的匹配乘累加结果.

对于图1采用的级联乘累加结构,假设由 $C(C \leq N_x$, 且能整除 N_x) 个 PE 级联,完成模板图 $N_x \times N_y$ 和参考图 $M_x \times M_y$ 的匹配需要 $N_x \times N_y / C$ 个循环,每个循环都得到 $(M_x - N_x + 1) \times (M_y - N_y + 1)$ 的乘累加结果矩阵,将结果矩阵表示成 $R_x \times R_y$ (这里 $R_x = M_x - N_x + 1, R_y = M_y - N_y + 1$),每次循环过程中加载模板数据需要 C 个时钟节拍,计算结果矩阵每行的第一个结果,需要 C 个时钟节拍,同时完成流水线建立,以后每拍计算一个结果,计算一行中剩下的 $M_x - N_x$ 个结果需要 $M_x - N_x$ 个时钟节拍,所以完成参考图的每一行的匹配需要 $M_x - N_x + C$ 个时钟节拍,而每个循环要完成 R_y 行的匹配,因此每个循环需要 $(M_x - N_x + C) \times R_y + C$ 个时钟节拍来完成,上式中第二项 C 是加载 C 个模板数据的节拍数,由此可以计算出完成整幅图像匹配的节拍时间为

$$T_1 = [(M_x - N_x + C)R_y + C] \frac{N_x N_y}{C}.$$

2.2.2 并发自累加时间模型

并发自累加计算模式的一维处理链^[4]模型如图2所示.计算时先由 Bin 端串行加载参考图数据,每个 PE 中对应保存一个参考图像素,且计算过程中参考图像素需要继续串行加载.参考图数据预加载完成后,启动并发数据端 Ain 的模板图加载,每个时钟节拍加载一个模板图数据,同时继续在每个时钟节拍加载一个参考图数据,模板图数据并发完成与参考图数据相乘,乘积结果送入加法器完成与上一个时钟节拍本地加法器的输出结果相加,整个处理链以并发方式计算,且每次循环都能得到一部分最终结果,这样将整个结果矩阵分成若干部分来计算,每次循环计算一部分,直到所有结果都计算完成.

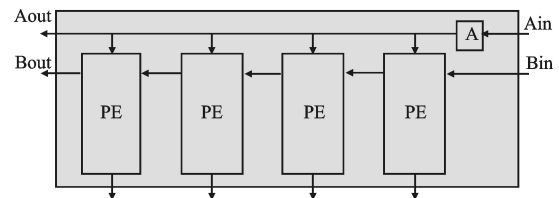


图2 并发自累加处理链

对于图2所示的并发自累加结构,假设由 $C(C \leq R_x$ 且能整除 R_x) 个 PE 级联,完成模板图 $N_x \times N_y$

和参考图 $M_x \times M_y$ 的匹配需要 $R_x R_y / C$ 个循环. 匹配过程中每次加载参考图数据要用 $C-1$ 个时钟节拍, 第 C 拍加载第 C 个参考图数据同时启动计算, 之后级联方向每拍加载一个新的参考图数据, 经过 N_x 拍的计算后, 级联运算单元中的参考图数据全部重新加载. 每个循环有 N_y 次加载和计算的过程, 计算的模板图数据由并发方向加载, 因此每个循环需要 $(N_x + C - 1) \times N_y$ 个时钟节拍, 由此可以计算出完成整幅图像匹配的节拍时间为

$$T_2 = (N_x + C - 1) N_y \frac{R_x R_y}{C}.$$

2.2.3 两种结构的时间模型对比分析

我们引入一个加速比 $S^{[10]}$, 它的值等于用级联乘累加结构和并发自累加结构两种电路完成模板图 $N_x \times N_y$ 配参考图 $M_x \times M_y$ 各自所用乘累加总时间之比, 即

$$S = \frac{T_1}{T_2} = \frac{1 + \frac{C-1}{R_x} + \frac{C}{R_x R_y}}{1 + \frac{C-1}{N_x}}, \text{ 实际应用中 } \frac{C}{R_x R_y} \approx 0, \text{ 因此 } S \approx \frac{1 + \frac{C-1}{R_x}}{1 + \frac{C-1}{N_x}};$$

可对 S 作出如下分类:

(1) 若 $N_x \approx R_x$, 此时 $M_x \approx 2N_x$, $S \approx 1$, 因此用两种结构电路完成这样的一次匹配速度相当.

(2) 若 $N_x \gg R_x$, 此时 $M_x \ll 2N_x$, $S \gg 1$, 因此用并发自累加结构电路完成这样的一次匹配比用级联乘累加结构速度快.

(3) 若 $N_x \ll R_x$, 此时 $M_x \gg 2N_x$, $S \ll 1$, 因此用级联乘累加结构电路完成这样的一次匹配比用并发自累加结构速度快.

2.2.4 两种结构的资源模型对比分析

S 不仅能反映出两种电路的时间性能, 还能反映出两种电路资源利用率, 将 S 的表达式改写成如下形式:

$$S = \frac{T_1}{T_2} = \frac{(2N_x N_y - 1) R_x R_y}{T_2 C} \div \frac{(2N_x N_y - 1) R_x R_y}{T_1 C} = \frac{D_2}{D_1};$$

其中 $D_1 = \frac{(2N_x N_y - 1) R_x R_y}{T_1 C}$,

$$D_2 = \frac{(2N_x N_y - 1) R_x R_y}{T_2 C}, \text{ 而 } (2N_x N_y - 1) R_x R_y \text{ 表示乘累加的计算量, 即完成乘法和加法的}$$

计算次数总和.

因此 D_1 、 D_2 分别表示级联乘累加电路和并发自累加电路在单位节拍内每个 PE 的计算量, 称之为计算密集度, 计算密集度可以反应出电路的资源利用效率, 计算密度越大, 资源利用效率越高, 反之就越小. 由 $S = \frac{T_1}{T_2} = \frac{D_2}{D_1}$ 可知两种电路匹配所用时间之比与资源利用率成反比, 直观的看就是用相同的级联规模对同样尺寸大小的图像做匹配运算所用时间短的资源利用率越高, 反之资源利用率越小. 按照上面对 S 的讨论, 同样可以用 $M_x = 2N_x$ 作为分界点, 来比较两种电路的资源利用效率:

(1) 若 $N_x \approx R_x$, 此时 $M_x \approx 2N_x$, $S \approx 1$, $D_1 \approx D_2$, 两种电路的资源利用率相当.

(2) 若 $N_x \gg R_x$, 此时 $M_x \ll 2N_x$, $S \gg 1$, $D_1 \ll D_2$, 并发自累加电路资源利用率比级联乘累加电路高.

(3) 若 $N_x \ll R_x$, 此时 $M_x \gg 2N_x$, $S \ll 1$, $D_1 \gg D_2$, 级联乘累加电路资源利用率比并发自累加电路高.

3 计算核带宽分析

3.1 冗余访问因子及存储模型

我们把计算核在计算时为了获取像素数据而访问存储器的总次数与取得的图像像素总数的比值称为冗余访问因子^[11], 用 R 表示, 即:

$$R = \frac{\text{访问存储器的总次数}}{\text{图像像素总数}} = \frac{A}{I},$$

式中, A 代表访问存储器的总次数, I 代表图像像素总数.

比值 R 总是大于等于 1 的, 它实质上代表了该图像中的每个像素被访问的平均次数, 用它能恰当的反映访问存储器的效率, R 越大, 该图像中的像素被访问的平均次数越多, 即访问的冗余度越大, 反之则图像中每个像素被访问的次数越少, 即访问的冗余度越小. 引入存储器冗余访问因子后, 单个存储器的带宽就可以表示为

$$BW = fXYR,$$

式中, f 是单位时间内完成的图像匹配次数, X 、 Y 分别代表图像的行数和列数. 通过上式可以看出, 在给定每秒钟内需要完成的图像匹配次数和图像维数之后, 存储器带宽取决于冗余访问因子, 降低计算和对存储器的带宽需求, 只有靠降低冗余访问因子, 即减少对图像的重复访问, 由此我们在匹配处理器设计

时可以采用“核内本地寄存器 + 外围模板图存储器 + 外围参考图存储器 + 中间结果存储器”的存储模式,存储器模型如图3所示。

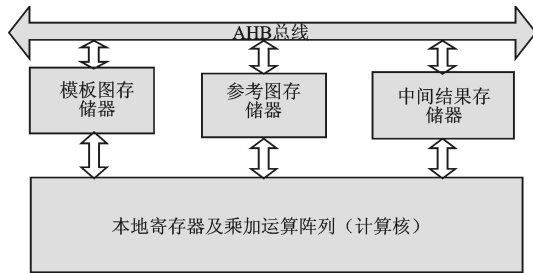


图3 块匹配电路存储模型

该种设计模型是基于数据复用而提出的。在图像匹配过程中,由于大量数据的重复使用,可增设核内本地寄存器用于存储能够多次重复使用的数据,这样能减少对参考图和模板图存储器的访问次数,设计时在构成内核阵列的每个PE中都包含了两个8 bit的本地寄存器,分别用于存放8 bit的模板图和参考图像素数据。

3.2 级联乘累加带宽计算方式

对于级联乘累加结构,在每次循环计算之前,先串行加载模板图数据,在一次循环计算过程中,这些模板图数据作为复用数据被核内本地寄存器一直保存,并与并行端逐个加载的参考图数据相乘,直到核内本地寄存器中的模板图数据遍历完所有可能与之产生相乘关系的参考图数据,每次所得乘积与前一个级联单元的结果寄存器的输出相加,乘累加的和最终以流水的方式输出到中间结果存储器保存,然后又重新加载本地寄存器中的模板图数据,进入下一次循环,下一次循环中要用到的中间结果,又从中间结果存储器中以流水方式读出。

计算核所需的存储器带宽为模板图存储器、参考图存储器、中间结果存储器的带宽之和。设计时,中间结果存储器采用32 bit存储一个结果数据,可以根据级联乘累加计算过程分别计算出该方式下模板图存储器、参考图存储器、中间结果存储器的冗余访问因子 R_{m1} 、 R_{r1} 、 R_{z1} ,然后可得总带宽 BW_1 (单位:字节)的计算方式是

$$BW_1 = fN_xN_yR_{m1} + fM_xM_yR_{r1} + 4fR_xR_yR_{z1} = fN_xN_y + \frac{1}{C}f(M_x - N_x + C)R_yN_xN_y + 4fR_xR_y(2\frac{N_xN_y}{C} - 1).$$

3.3 并发自累加带宽计算方式

对于自累加结构,同样在每个PE中也引入本

地寄存器,在每次循环计算之前先串行加载参考图数据,在计算过程中,这些参考图数据随着时钟节拍顺次往前移动,进入下一个PE的本地寄存器被再次使用,每个节拍移除一个旧的参考图数据,同时移入一个新的参考图数据,这些新的参考图数据就是在这样的一个动态移位过程中得到复用的。从一个新的参考图数据被移入第一个PE的本地寄存器开始,到从最后一个PE的本地寄存器中移除,每一拍都被用来与并发加载的模板图数据相乘,产生一个有用的乘积作为加法器的一个加数,加法器的另一个加数则来自上一拍本地加法器自身的输出。根据自累加结构的计算过程可知,每次循环所得的中间结果保存在核内每个PE的本地结果寄存器中作为复用数据,只在每次循环结束之后将所得最终结果输出,且每个循环输出C个结果,结果存储器采用32 bit存储一个结果数据,我们同样可以根据并发自累加计算过程分别计算出该方式下模板图存储器、参考图存储器、中间结果存储器的冗余访问因子 R_{m2} 、 R_{r2} 、 R_{z2} ,然后可得总带宽 BW_2 (单位:字节)的计算方式是

$$BW_2 = fN_xN_yR_{m2} + fM_xM_yR_{r2} + 4fR_xR_yR_{z2} = f\left[\frac{N_y}{C}(2N_x + C - 1) + 4\right]R_xR_y.$$

3.4 两种结构的带宽对比计算

表1中给出了两种电路的三种不同参数的带宽对比计算,其中 f 的单位为:次/秒, R_m 、 R_r 、 R_z 分别代表模板图存储器、参考图存储器、中间结果存储器的冗余访问因子,冗余访问因子为1的存储器就算是从最优层面上做了数据复用考虑。

表1 两种电路的带宽对比计算

计算参数	计算模式	R_m	R_r	R_z	BW/GB
$M_x=127, M_y=128,$ $N_x=32, N_y=32,$ $C=32, f=25$	级联乘累加	1	24.25	63	0.064
	并发自累加	291	36.09	1	0.021
$M_x=255, M_y=256,$ $N_x=64, N_y=64,$ $C=64, f=25$	级联乘累加	1	48.25	127	0.512
	并发自累加	579	72.09	1	0.168
$M_x=511, M_y=512,$ $N_x=128, N_y=128,$ $C=128, f=25$	级联乘累加	1	96.25	255	4.098
	并发自累加	1155	144.09	1	1.332

从表1中的数据对比可以看出两种电路都在最高层面上对数据作了复用,其中级联乘累加从最高层面上复用了模板图数据,并发自累加从最高层面上复用了结果数据。总体上并发自累加计算时的带

宽需求会小于级联乘累加,虽然级联乘累加模式的冗余访问因子 R_m 、 R_r 都会小于自累加模式,但是自累加从最高层面上复用了结果数据, R_s 远小于级联乘累加模式,并且结果数据是 32 bit 位宽,这样对结果数据复用降低带宽的优势会迅速体现出来,因此在计算核设计中,计算结果的复用对降低数据带宽要求占据主导性的作用。

4 结束语

归一化互相关匹配中涉及到大量的乘累加运算,完成乘累加运算可以采用级联乘累加阵列和并发自累加阵列两种电路来完成,但是这两种电路在不同匹配任务下,计算性能会有较大差别。

在 PE 固化之后,运算的时间性能和资源利用效率都会随着匹配中参考图和模板图的横向尺寸 M_x 与 N_x 的大小关系的变化而变化,文中为了探究级联乘累加和并发自累加两种电路的运算时间性能和资源利用效率随着 M_x 和 N_x 的大小关系变化的规律,引入了加速比的概念;为了分析计算核对带宽的要求,引入了冗余访问因子的概念,同时说明了为了降低冗余访问因子而采用数据复用以及核内本地寄存器的优势,并给出了两种阵列模式的带宽计算方法,以期在实际应用中对计算核的带宽有个较明确的把握。

参考文献:

- [1] 邓汉华,杨靓,黄士坦. 基于归一化互相关系数的图像匹配的连续消算法[J]. 微电子学与计算机, 2006, 23(12): 227-230.
- [2] Mori M, Kashino K. Fast template matching based on normalized cross correlation using adaptive block partitioning and initial threshold estimation[C]// Multimedia (ISM), 2010 IEEE International Symposium on, China: Taichung, IEEE, 2010: 196-203.
- [3] Pei L, Xie Z, Dai J. Fast normalized cross-correlation image matching based on multiscale edge information [C]// Computer Application and System Modeling (ICCASM), 2010 International Conference on, China: Taiyuan, IEEE, 2010(10): 507-511.
- [4] Cheng Hengda, Lin Weichung, Fu Kingsun. Space-time domain expansion approach to VLSI and its appli-

cation to hierarchical scene matching[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1985, PAMI-7(3): 306-311.

- [5] Donald G Bailey. 基于 FPGA 的嵌入式图像处理系统设计[M]. 原奎,何文浩,肖晗,译. 北京:电子工业出版社,2013:262-263.
- [6] Almudena Lindoso. Luis Entrena. High performance FPGA-based image correlation[J]. Real-Time Image, 2007(2):224-226.
- [7] Lindoso A, Entrena L, López-Ongil C, et al. Correlation-based fingerprint matching using FPGAs [C] // Field-Programmable Technology, Proceedings. 2005 IEEE International Conference on, Singapore, IEEE, 2005: 87-94.
- [8] 张遂南,黄世坦,催建生. 一种面向图像匹配算法的高性能专用集成电路设计[J]. 北京师范大学学报:自然科学版,2002, 38(1):48-52.
- [9] Joanblanc, Christophe, Patrice Senn, et al. A 54-MHz CMOS programmable video signal processor for HDTV applications[J]. Solid-State Circuits, 1990, 25(3): 730-734.
- [10] 李学干. 计算机系统结构[M]. 4 版. 西安:西安电子科技大学出版社,2006:160-162.
- [11] Tuan Jenchieh, Chang Tiansheuan, Jen Cheinwei. On the data reuse and memory bandwidth analysis for full-search block-matching VLSI architecture [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(1):61-65.

作者简介:



周 泉 男,(1987-),硕士研究生.研究方向为集成电路设计.

闫 博 男,(1987-),硕士研究生.研究方向为数字集成电路设计.

曹 辉 男,(1984-),博士研究生.研究方向为数字集成电路设计.

杨 靓 男,(1975-),博士生导师,研究员.研究方向为数字图像匹配、数字集成电路设计.