

引用格式:杨莫洋,蒋剑飞,王琴. 基于传输触发架构的 IIR 型高斯滤波实现[J]. 微电子学与计算机,2020,37(7):21-24.
[YANG M Y,JIANG J F, WANG Q. Implementation of IIR gaussian filter based on transport triggered architecture[J]. Microelectronics & Computer,2020,37(7):21-24.]

基于传输触发架构的 IIR 型高斯滤波实现

杨莫洋,蒋剑飞,王 琴

(上海交通大学 电子信息与电气工程学院,上海 200240)

摘 要: 为了满足嵌入式领域需求,针对无限冲激响应(IIR)高斯滤波器,提出了一种定点化策略,能够有效地替代浮点数实现版本. 为了进一步提高滤波速度,设计了一款传输触发架构(TTA)处理器,并针对该算法特点,定制特殊功能单元,优化存储结构. 实验结果表明,对于大小为 640×480 的 24 位真彩图像,IIR 型高斯滤波的 TTA 实现速度是通用 DSP 的 4.12 倍.

关键词: 无限冲激响应;高斯滤波;定点化;传输触发架构

中图分类号: TN402

文献标识码: A

文章编号: 1000-7180(2020)07-0021-04

Implementation of IIR gaussian filter based on transport triggered architecture

YANG Mo-yang,JIANG Jian-fei, WANG Qin

(School of Electronic Information and Electrical Engineering,
Shanghai Jiao Tong University,Shanghai 200240,China)

Abstract: In order to meet the needs of the embedded field, a fixed-point strategy is proposed for the infinite impulse response Gaussian filter, which can effectively replace the floating-point implementation version. In order to further improve the speed of filtering, a Transport Triggered Architecture processor is designed, and special function units are customized for the characteristics of the algorithm, and the storage structure is optimized. The experimental results show that for a 24-bit true color image with a size of 640×480 , the Transport Triggered Architecture implementation speed of the IIR Gaussian filter is 4.12 times that of the general-purpose DSP.

Key words: infinite impulse response;Gaussian filter; fixed-point;transport triggered architecture

1 引言

在数字图像处理领域,高斯滤波是一种常见的预处理手段^[1]. 通过高斯滤波,可以很好的抑制图像输入过程中引入的高斯噪声,从而得到信噪比较高的图像. 对于二维图像的高斯滤波,一般直接用离散化高斯核进行卷积,或者将二维高斯滤波拆分为两次一维的高斯滤波. 但是这种方法随着高斯半径的增大,计算复杂度会越来越大. 为了提高高斯滤波算

法的效率,许多近似算法被提出,本文选取了一种 IIR 型高斯滤波^[2]方法作为研究对象,这种算法的计算复杂度独立于高斯半径,文献[3]表明该算法可以很好地替代传统的线性高斯滤波. IIR 型高斯滤波采用递归结构,其中间结果需要用浮点数来表示,这限制了其在嵌入式领域的运用.

本文首先分析不同高斯滤波方法的原理和计算复杂度,然后对 IIR 型高斯滤波的定点化策略进行讨论,最后提出了一种基于 TTA 架构的 IIR 型高

斯滤波实现方式,并与现有的通用 DSP 进行比较。

2 高斯滤波原理

2.1 直接高斯滤波

数字图像通常用矩阵的形式进行存储,图像中各像素点的值与矩阵中的元素对应。二维高斯滤波通过离散滑动窗口与原图像进行卷积来实现,这里的离散窗口由高斯核函数决定,二维高斯函数的定义如公式(1)所示。

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (1)$$

式中, σ 是正态分布的标准差, σ 越大,代表边缘像素对中心像素的影响越大,结果也就越平滑。高斯窗口大小也与 σ 有关,在计算高斯函数的离散近似值时,距中心点 3σ 以外的值接近于 0,可以忽略这些值对中心像素点的影响,所以高斯窗口大小一般取 $6\sigma \times 6\sigma$,另外为了使窗口有明确的中心点,大小通常取奇数。

对于 3×3 的卷积核,每完成一个点的滤波,需要 9 次乘法,8 次加法。二维高斯滤波的计算复杂度为 $O(M \times N \times K \times K)$,其中 M 和 N 表示原始图像的尺寸, K 表示高斯窗口的大小。可以看到随着高斯窗口的增大,计算复杂度呈指数增长。为了降低复杂度,一般采用可分离的高斯滤波。公式(2)展现了高斯函数的可分离性, $K \times K$ 的二维高斯滤波可以分成两步进行,首先用长度为 K 的一维高斯模板对原始图像进行水平方向的卷积,再用同样的模板对中间结果进行竖直方向的卷积,这样得到的滤波结果在数学上与直接进行二维高斯滤波等价,但是计算复杂度降低到了 $O(2 \times M \times N \times K)$ 。

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \exp\left(-\frac{y^2}{2\sigma^2}\right) \quad (2)$$

2.2 IIR 型高斯滤波

由于直接高斯滤波的计算复杂度与标准差 σ 有关,当 σ 增大时,计算会越来越复杂。在实时性要求高的场合,直接高斯滤波很难满足需要。于是 Vliet 等人^[2]提出了一种 IIR 型高斯滤波方法,这种方法首先在傅里叶域中用六次多项式逼近高斯函数,然后求拉普拉斯变换,最后转换到 z 域中。由于表达式的分子中不含有多项式,所以滤波器是纯递归结构。

具体到实际计算, IIR 型高斯滤波可以分为 2 个步骤,公式(3)表示前向迭代过程,公式(4)表示反向迭代。其中, w 存储计算的中间结果,滤波参数 B , b_1, b_2, b_3 取决于 σ 的值。

$$w[n] = B_{in}[n] + b_1 w[n-1] + b_2 w[n-2] + b_3 w[n-3] \quad (3)$$

$$\text{out}[n] = B_w[n] + b_1 \text{out}[n+1] + b_2 \text{out}[n+2] + b_3 \text{out}[n+3] \quad (4)$$

IIR 型高斯滤波每维度只有 6 次乘加运算,对于大小为 $M \times N$ 的图像,其计算复杂度为 $O(M \times N)$,与高斯窗口大小无关。相关研究表明^[3],这种递归高斯滤波方法,能很好的代替直接高斯滤波,且精度随着 σ 的增大而增大。

2.3 IIR 型高斯滤波定点化

在嵌入式领域,对算法进行定点化处理是一种常见的近似手段。出于功耗、计算速度还有资源消耗等方面考虑,定点化算法在 DSP、FPGA 以及嵌入式 CPU 中得到了广泛应用。

对于直接高斯滤波而言,由于高斯核严格为正,并且有界,所以很容易进行定点化^[4]。而对于 IIR 型高斯滤波,首先滤波系数 b_2 恒定为负数,所以需要 1 位符号位;其次,滤波基于递归结构,中间结果的舍入误差可能会产生累积效应,所以对精度有较高的要求。

本文对 IIR 型高斯滤波分别进行了不同位宽的定点化:

(1)采用 32 位定点数表示,其中 1 位表示符号位,10 位表示整数位,21 位表示小数位;

(2)采用 32 位定点数表示,其中 1 位表示符号位,15 位表示整数位,16 位表示小数位;

(3)采用 32 位定点数表示,其中 1 位表示符号位,23 位表示整数位,8 位表示小数位;

(4)采用 16 位定点数表示,其中 1 位表示符号位,8 位表示整数位,7 位表示小数位。

考虑将定点化实现的 IIR 型滤波与单精度浮点数实现的滤波相比较,方法是分别将两种滤波作用于测试图像,然后进行定量分析。本文选取的测试图像如图 1 所示,这是一幅随机生成的高斯噪声图像,大小为 320×240 ,位深度为 24。

采用公式(5)表示归一化的均方根对定点化精度损失进行定量表示。其中 I_h 表示定点化滤波后的测试图像, I_g 表示单精度浮点滤波后的测试图像, N 表示图像中的总像素点数量。

$$\Delta = \frac{1}{N} \sum_{(i,j)} \frac{\sqrt{(I_h(i,j) - I_g(i,j))^2}}{I_g(i,j)} \quad (5)$$

针对每种定点化方案,分别测试了 σ 为 1、2、5 和 10 的误差百分比,结果在表 1 中给出。由结果可知,32 位定点化时,用 21 位表示小数,误差在 0.1%

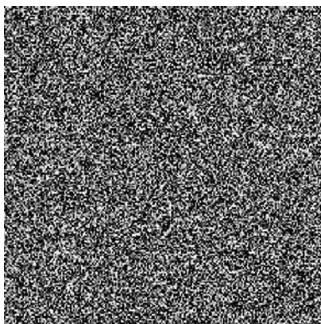


图 1 随机高斯噪声

以内,用 16 位来表示小数,误差不超过 2%,这表明用 32 位定点化来代替单精度浮点数是一种有效的方案.但对于 16 位定点化来说,在 σ 大于 1 的时候,误差变得不可接受,表明 16 位定点化的精度不足以近似 IIR 高斯滤波.

表 1 定点化误差百分比

定点化方案	$\sigma=1$	$\sigma=2$	$\sigma=5$	$\sigma=10$
1:10:21	0.000 4	0.001 6	0.008 5	0.043 8
1:15:16	0.011 4	0.027 5	0.232 4	1.634 9
1:23:8	3.043 6	7.253 2	83.566	52.880
1:8:7	0.291 3	54.946	56.383	100.00

值得注意的是,由于图像像素值在 0~255 之间,整数部分至少要用 8 位来表示.在滤波过程中做乘加运算时,如果目标硬件仅支持 32 位加法,需要考虑在 σ 过大时,结果可能会溢出,如果目标硬件支持 64 位加法,则不用考虑溢出问题.对于 1:10:21 的定点化方案,在仅支持 32 位加法的硬件上, σ^2 应小于 170.可以根据 σ 大小灵活选择整数位宽来避免中间结果溢出.

3 TTA 架构下的算法实现

3.1 TTA 架构处理器

TTA 架构可以看作是超长指令字(VLIW)的一种改进结构^[5],超长指令字通过多个计算单元同时工作,达到指令级并行的效果.但是超长指令字结构的每个计算单元都需要与寄存器堆相连,当计算单元变多时,这些单元和寄存器堆的接口需要增加,控制逻辑也会越来越复杂.而 TTA 架构中的计算单元本身就包含输入和输出寄存器,功能单元之间的数据传输不再通过寄存器堆,而是通过互联网络来完成.这样不仅减少了数据通路的复杂度,还支持软件旁路方式,即总线上的数据对每个功能单元来说都是可见的,一个功能单元的输出可以通过互联

网络直接移动到另一个功能单元的输入,从而减少寄存器堆的访问.

TTA 处理器的基本结构如图 2 所示,首先它是基于哈佛结构,取指单元从指令存储器中取出指令,然后经过译码,将控制信号或者立即数传递给互联网络,功能单元和访存单元(LSU)通过互联网络来交换数据. TTA 是一种模块化的结构,可以根据特定应用自由改变总线和功能单元数量,并且可以定制特殊计算单元来加速运算.

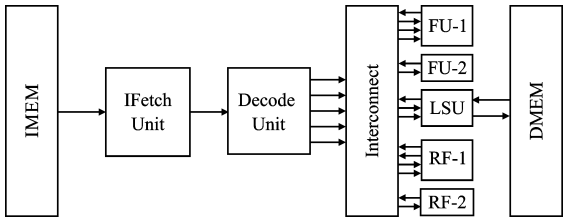


图 2 TTA 处理器基本结构

TTA 协同设计环境(TCE)是一个开源工具集^[6],该工具集由芬兰坦佩雷大学的定制并行计算(CPC)组维护,工具集包括编译器、仿真器、代码生成器等一系列软硬件协同工作环境.本文在 TCE 工具集下,设计了一款 TTA 处理器,用于对 IIR 高斯滤波算法进行加速.

整体设计如图 3 所示,该处理器包含 1 个全局控制单元(负责取指、译码),4 个 Load/Store 单元,若干个基本计算单元,2 个 16~32 的寄存器堆,1 个布尔寄存器,1 个自定义计算单元.其中互联网络由总线、Socket 和连接点组成,一共有 11 条总线.存储器包括一个指令存储器和两个数据存储器.

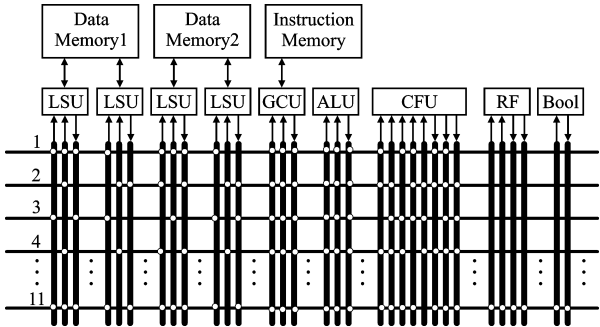


图 3 本文设计的 TTA 处理器

3.2 自定义计算单元

使用 TTA 处理器对特定算法加速,首先要确定基本计算单元和总线的数目,当基本单元增加到一定数目时,算法并行度潜力耗尽,继续增加基本单元不会对运行周期产生任何影响.想要进一步降低

运算周期,则需要使用自定义计算单元来进行加速,自定义计算单元也是 TTA 处理器的核心.

分析 IIR 高斯滤波计算过程,对于一维滤波,分为前向迭代和反向迭代,单次迭代需要 4 次乘法,3 次加法. 且反向迭代过程依赖于前向迭代的结果,所以两次迭代不能并行. 对于二维滤波,即对原图像分别进行两次一维滤波即可. 滤波计算的中间结果用 32 位定点数表示,原始图像和目标图像像素点均为 8 位无符号整形. 针对算法特点,为了尽可能的复用加法器和乘法器,定制了图 4 所示的计算单元.

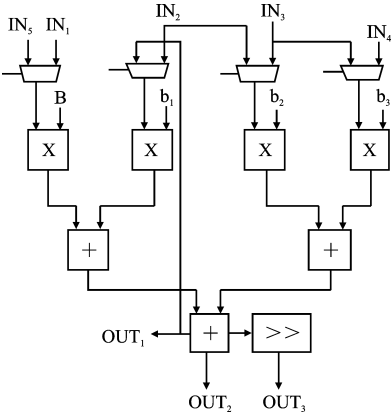


图 4 定制计算单元

此计算单元一共有 5 个输入端口,3 个输出端口,其中第 4 个输入端口为触发端口. 由于内部采用两级流水结构,所以实际共使用 8 个 32 位乘法器,6 个 32 位加法器. 计算过程如下:

(1)对于一维水平方向滤波,由于访存开销是主要计算瓶颈,所以考虑 load 进 5 个数之后开始执行运算,第一个周期计算公式(6),结果由输出端口 1 输出到总线上,第二个周期计算公式(7),结果由输出端口 2 输出,这样一共复用了 3 个操作数.

$$OUT_1 = B * IN_1 + b_1 * IN_2 + b_2 * IN_3 + b_3 * IN_4 \tag{6}$$

$$OUT_2 = B * IN_5 + b_1 * OUT_1 + b_2 * IN_2 + b_3 * IN_3 \tag{7}$$

(2)对于一维纵向滤波,从上到下直接进行 4 输入乘加运算,结果由输出端口 1 输出;从下到上时,把 32 位定点转为 8 位无符号整形的过程与滤波过程合并,滤波结果通过输出端口 1 输出并参与下一次迭代过程,同时 8 位无符号整形即为目标图像像素点,通过输出端口 3 输出到总线上.

为了减小访存开销,本设计使用了 2 个数据存储寄存器,每个寄存器可由 2 个 LSU 访问. 原始图像存在存储器 1 中,前向滤波过程中,从 1 中取数到自定

义单元进行运算,结果存到存储器 2 中,反向滤波过程中,从存储器 2 中取数,计算结果保存到存储器 1 中.

4 实验结果

为了验证本文所设计的 TTA 处理器的性能,选取了 TI 公司的 C6678 作为比较对象,C6678 是一款高性能 DSP^[7],每个核心每个周期能够执行 4 个 32×32 bit 乘法或者 4 个单精度浮点乘法,主频最高能达到 1.25 GHz.

本文所设计 TTA 处理器在 SMIC40LL 工艺下综合频率达到 200 MHz. 使用的 DSP 仿真器版本为 CCS5.5,测试图像大小为 640×480,位深度为 24. 选择 32 位定点化 IIR 高斯滤波算法,其中 21 位表示小数. 实验结果如表 2 所示,可见,与单核 C6678DSP 相比,单核 TTA 处理器的加速比为 4.12.

表 2 本文设计 TTA 处理器与 C6678 单核性能对比			
方案	周期数	频率/Mhz	时间/ms
TTA	11 553 650	2 001 250	57.77
DSP	297 540 027		238.03

5 结束语

本文针对 IIR 型高斯滤波中间结果采用浮点数这一特点,提出了有效的定点化策略,能很好地替代原浮点数实现版本. 并基于 TTA 架构,提出了一种定点化的 IIR 型高斯滤波实现方式,与传统的 DSP 相比,显著调高了计算效率,且该加速器结构具有易复用,易扩展的特点.

参考文献:

[1] CANNY J. A Computational Approach To Edge Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, PAMI-8 (6): 679-698.

[2] YOUNG I T, VLIET L J V. Recursive implementation of the Gaussian filter[J]. Signal Processing, 1995, 44(2):139-151.

[3] TAN S, DALE J L, JOHNSTON A. Performance of three recursive algorithms for fast space-variant Gaussian filtering[J]. Real-Time Imaging, 2003, 9 (3):215-228.

[6] ABIDI W, EZZEDINE T. Fuzzy Cluster Head Election Algorithm based on LEACH protocol for Wireless Sensor Networks[M]. International Wireless Communications and Mobile Computing Conference, NEW YORK:IEEE, 2017: 993-997.

[7] MOSTAFAEI H. Energy-Efficient Algorithm for Reliable Routing of Wireless Sensor Networks[J]. IEEE Transactions on Industrial Electronics, 2019, 66(7): 5567-5575.

[8] 胡中栋, 易涛, 王振东. 基于最优分簇的能量异构无线传感器网络路由协议[J]. 传感技术学报, 2018(8): 1240-1245.

[9] JOTHIKUMAR C, VENKATARAMAN R, RAJ T S, et al. An energy efficient cluster based routing approach to minimize energy consumption using CORP in wireless sensor networks[J]. Journal of Intelligent &

Fuzzy Sysytems, 2019, 36(6): 5835-5844.

[10] MEZGHANI M. An efficient multi-hops clustering and data routing for WSNs based on Khalimsky shortest paths [J]. Journal of Ambient Intelligence and Humanized Computing, 2019, 10(4SI): 1275-1288.

作者简介:

蒋 华 男,(1963-),博士,教授. 研究方向为数据库系统及信息安全.

王 瑶(通讯作者) 女,(1995-),硕士研究生. 研究方向为无线传感器网络. E-mail: ts5748@126. com.

王慧娇 女,(1976-),硕士,副教授. 研究方向为无线传感器网络及信息安全.

王 鑫 男,(1976-),硕士,教授. 研究方向为无线传感器网络及云计算技术.

(上接第 24 页)

[4] LAMPERT C H, WIRJADI O. Anisotropic Gaussian Filtering using Fixed Point Arithmetic[C]// 2006 International Conference on Image Processing. Atlanta, Georgia, USA: IEEE, 2007:1565-1568.

[5] CORPORAAL H. Microprocessor Architectures - from VLIW to TTA[M]. Chichester :John Wiley & Sons, 1997:101-125.

[6] ESKO O, PEKKA J, HUERTA P, et al. Customized Exposed Datapath Soft-Core Design Flow with Compiler Support [C]// International Conference on Field Programmable Logic and Applications. Milano, Italy: IEEE Computer Society, 2010:217-222.

[7] RAMESH B, BHARDWAJ A, RICHARDSON J, et al. Optimization and evaluation of image- and signal-processing kernels on the TI C6678 multi-core DSP [C]// High Performance Extreme Computing Conference. Waltham, MA USA: IEEE, 2015:1-6.

作者简介:

杨莫洋 男,(1995-),硕士研究生. 研究方向为 SoC 设计与系统集成. E-mail: vanceyang@sjtu. edu. cn.

蒋剑飞 男,(1979-),助理研究员. 研究方向为高速数字集成电路设计、SoC 设计与系统集成.

王 琴 女,(1975-),副教授. 研究方向为大规模集成电路设计、先进集成电路设计方法.