

基于图像处理器的传声器阵列实时声成像方法

余立志^{1,2}, 杨亦春², 滕鹏晓², 郑金华¹, 易灵芝¹

(1. 湘潭大学信息工程学院电气工程系, 湖南湘潭 411105; 2. 中国科学院声学研究所, 北京 100190)

摘要: 为了解决传声器阵列用于声场分析时的精确且快速声成像的技术难题, 提出了一种利用图像处理器(Graphics Processing Unit, GPU)的通用计算技术实现快速声成像算法。可控响应功率算法是广泛应用的一种声源定位算法, 但计算量巨大, 阻碍了它在实际场合中的应用。通过将可控响应功率算法进行任务分解及线程映射, 实现了利用计算统一设备架构实现的基于 GPU 的可控响应功率声源成像定位算法, 在特定阵元通道数和信号长度情况下, 与基于 CPU 的声源定位计算方法相比, 综合计算效率提高了约 20 倍, 并将其成功地应用于平面螺旋阵的声成像应用中, 实现了实时声成像定位。

关键词: 波束形成; 可控响应功率; 图像处理器; 计算统一设备架构; 并行计算

中图分类号: TP311

文献标识码: A

文章编号: 1000-3630(2014)-04-0367-05

DOI 编码: 10.3969/j.issn1000-3630.2014.04.017

A GPU based realtime acoustic imaging method for microphone array

YU Li-zhi^{1,2}, YANG Yi-chun², TENG Peng-xiao², ZHENG Jin-hua¹, YI Ling-zhi¹

(1. Department of Electrical Engineering, College of Information Engineering, Xiangtan University, Xiangtan 411105, Hu'nan, China;

2. Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: To implement an accurate and rapid acoustic imaging method, an algorithm based on general-purpose computing technology of GPU is used to make fast acoustic imaging calculation, which is a key technical issue for sound field analysis of microphone array. One well-known steered response power (SRP) method is widely used in sound source localization, but due to high computational expense its practical application to real-time processing is limited. The SRP algorithm is implemented in real time based on CUDA (Compute Unified Device Architecture) frame using GPU task decomposition and thread mapping, which gains almost 20 times total efficiency improvement compared to CPU based SRP sound source localization algorithm under condition of certain number of channel and length of signal. And, it has been successfully applied to the acoustic imaging of a plane spiral array to achieve real-time image positioning.

Key words: beamforming; steered response power; graphics processing unit; compute unified device architecture; parallel computing

0 引言

声源定位是传声器阵列的一个重要应用领域。声成像分析是目前国内外研究的热点方法, 并且在各行业得到了广泛的应用。针对声源特性的不同可以有不同的分析方法, 如有学者提出了抗混响声源定位算法, DiBiase^[1,2]提出的可控响应功率和相位变换的声源定位方法, 将可控响应功率方法固有的鲁

棒性、短时分析特性和时延估计中相位变换方法对周围环境的不敏感性相结合。使用可控响应功率法的声源定位具有一定的抗噪性、抗混响性和鲁棒性。但是因为可控响应功率空间有很多局部极值, 需要使用网格搜索法找出全局最大值, 所以可控响应功率算法计算量较大, 与传声器的数量及声图像的分辨率成正比。一些文章对此提出了解决办法, 分层搜索^[3]可控响应功率算法减小了计算量, 但是当混响时间大于 300 ms 时对混响极为敏感; Peterson^[4]的适用于混响环境中的混合定位算法虽然克服了分层搜索可控响应功率算法的缺点, 但只适用于近场范围。GPU 本是计算机显卡的核心硬件, 其运行性能有别于 CPU。近几年, 随着大型计算的出现, GPU 在并行信号处理方面的潜能逐渐被挖掘, 成为信号

收稿日期: 2013-03-06; 修回日期: 2013-06-17

基金项目: 国家自然科学基金面上项目(11074278)、国家自然科学基金面上项目(11174320)

作者简介: 余立志(1988—), 男, 湖南岳阳人, 硕士研究生, 研究方向为计算机测控技术与应用。

通讯作者: 余立志, E-mail: yulizhi81@126.com

处理领域的一个热点。

本文将 GPU 引入传声器阵列实时成像算法的实现中,是 GPU 在声学计算方面的一个应用,具有较大的应用价值。

快速计算声像可以精确反映出声音变化的细节,有利于解释许多复杂多变的声学现象,因而是传声器阵列声学测量分析需要解决的重要技术难点。

与 CPU 相比,利用 GPU 进行通用计算具有成本低、性能高的特点,与 DSP 和 FPGA 平台相比具有便捷高效的特点。为此,本文提出一种基于 GPU 的可控响应功率算法实现,基于 NVIDIA 公司的计算统一设备架构(Compute Unified Device Architecture, CUDA)^[5,6]通用并行编程模型,利用 C 语言编程,使算法的密集计算部分并行化来实现加速。并将其应用于平面螺旋阵的声成像应用中,充分利用了可控响应功率算法的抗混响性和鲁棒性。从实验结果可见,整个系统达到了实时声成像的要求,与基于 CPU 的可控响应功率声源定位算法相比综合效率提升 20 倍。

1 可控响应功率算法

在一个 M 元传声器阵列上,各个阵元的接收信号为 $x_m(t)$, $m=0, 1, \dots, M-1$, 在直角坐标系下声源传播到各个阵元的信号表示为

$$x_m(t) = \alpha_m(x_s, y_s, z_s) s [t - \tau_{0,m}(x_s, y_s, z_s)] + v_m(t) \quad (1)$$

式(1)中: s 是信号源; $v_m(t)$ 是传声器 m 接收的额外噪声,包括背景噪声,传感器噪声等等; α_m 由传播衰减决定; τ_{ij} 是与声源位置相关的量,即传声器 i 、 j 之间的相对时延:

$$\tau_{ij} = \frac{1}{c} [d_{s,j}(x_s, y_s, z_s) - d_{s,i}(x_s, y_s, z_s)] \quad (2)$$

其中: c 是声速; $d_{s,j}$ 是声源到传声器之间的距离。

SRP^[7,8]算法包括两个步骤:互相关函数的计算和声源定位时的搜索过程。

第一阶段,计算所有的传声器对的互相关函数 $R_{xi,xj}(\tau)$,这个过程通常通过傅里叶变换(FFT)在频域内进行,

$$R_{xi,xj}(\tau) = \sum_{k=0}^{N_j-1} X_i(k) X_j^*(k) e^{i2\pi \frac{k}{N} F_s \tau} \quad (3)$$

式中: R 是 i 、 j 通道间的互功率谱; k 是离散频率点索引; F_s 是采样频率; N 是 FFT 计算中信号的样本长度; $X_i(k)$ 是信号 $x_i(n)$ 的 FFT 结果。

第二阶段为可控响应功率查找。逐个遍历每个扫描点,即候选的声源位置。对于每一个扫描点有 M 个相对时延。则 (x, y, z) 处的可控响应功率为

$$S^{SRP}(x, y, z) = \sum_{i=0}^{M-1} \sum_{j=i+1}^{M-1} R_{xi,xj}(\tau_{x,y,z}^{i,j}) \quad (4)$$

2 声成像计算的多线程并行算法

利用 GPU 的多核结构实现可控响应功率算法计算是一种解决传声器阵列声成像实时运算的有潜力的方法。GPU 具有多线程的并行性, NVIDIA 等公司对硬件和软件的改进使 GPU 的可编程能力不断提高, GPU 通用计算应运而生^[6]。GPU 的浮点计算能力和内存带宽都远超过 CPU,原因在于 GPU 专为计算密集型、高度并行化的计算而设计,能使更多晶体管用于数据处理,而非数据缓存和流控制。

CUDA 是一个 SIMT 架构,即使用单指令来控制多个处理元素,指定单一线程的执行和分支行为^[9]。在 GPU 上以并行线程执行的程序称为核(kernel), GPU 线程以网格(grid)的方式组织,而每个网格中又包含若干个线程块。同一线程块中的众多线程不仅能够并行执行,而且能够通过共享存储器(shared memory)和栅栏(barrier)通信。线程按照粗粒度的线程块和细粒度的线程两个层次进行组织,在细粒度并行的层次通过共享存储器和栅栏同步实现通信,这就是 CUDA 的双层线程模型。对于使用可控响应功率算法的声成像计算来说,很自然地按扫描点划分计算任务,因为对每个扫描点的处理是独立并且相同的。所以声成像算法的初始 kernel 任务划分为对 k 个扫描点的处理,如图 1 所示。

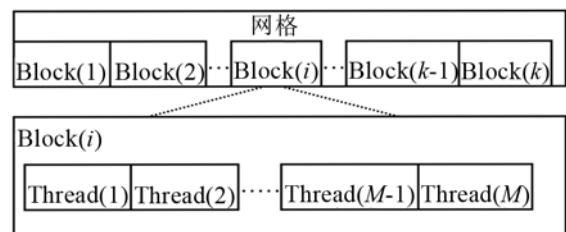


图 1 传声器阵列声像计算的 CUDA 核分配方法
Fig.1 Distribution of CUDA kernels for sound imaging calculation of a microphone array

初步划分任务后,应用 CUDA 优化技巧作为编写核函数的原则:对核中相加的问题采用了类似二叉树的并行计算方法;利用共享存储器完成中间复杂的计算过程;对常量的访问采取常量存储器(constant memory)的访问方法可以隐藏延时、提高速度;图像处理部分应用纹理存储器;对核中每个线程(thread)执行的程序规模要降到最低,因为线程上的程序是串行运行的,线程块(block)之间是并行执行的,所以要增多线程块的数目;同时要减少数

据在设备(device)和主机(host)之间的拷贝次数。

3 可控响应功率算法的 CUDA 实现

可控响应功率算法中每个候选扫描点的累加响应功率的计算是相互独立的，与其他扫描点没有依赖关系，具有很好的数据并行性。但简单的按扫描点划分计算线程没有充分利用 GPU 强大的计算资源，故本文针对算法的每个阶段分配合适的线程块和线程，以达到最优的效率。图 2 给出的是利用计算统一设备架构实现的基于 GPU 的 SRP 成像定位算法实现的流程图。

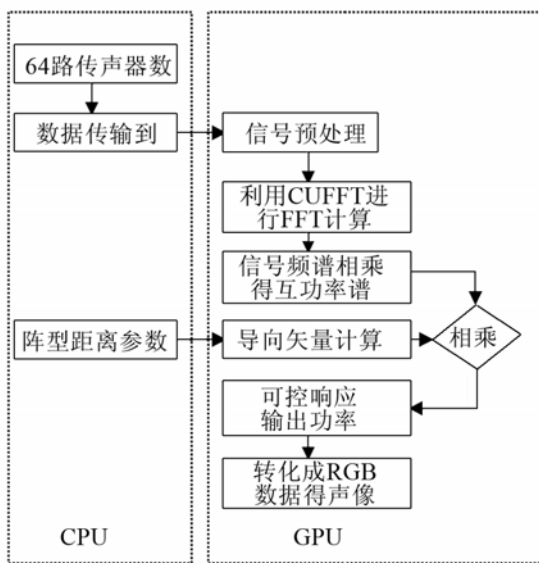


图 2 利用 CUDA 实现的基于 GPU 的可控响应功率成像定位算法实现流程
Fig.2 Implementing steps of GPU based steered response power sound imaging localization algorithm using CUDA

由传声器阵列采集的 64 路数据经过以下几个步骤的处理：

第一步，将输入数据进行预处理和 FFT 计算得到信号频谱。数据预处理包括信号的去均值、信号加窗和补零，每一路信号可以独立加窗，处理线程的数量为通道数乘以信号长度。同时利用 GPU 计算超越函数的优势，信号处理中的 FFT 计算利用 CUDA 自带的 CUFFT 库来实现，采用能同时计算多路相同点数 FFT 的库函数计算 64 路信号的频谱。以下是在 GPU 上对 64 路数据进行 FFT 计算的过程：

- (1) 为 kernel 函数申请显存空间，包括采集的声音数据，存放 FFT 结果的实部和虚部。
- (2) 将声音数据从内存传递到显存。
- (3) 为 FFT 计算初始化一个 plan 句柄。
- (4) 计算执行配置。CUFFT 库支持复数 FFT 和

实数 FFT，cufftType_t 是 CUFFT 库支持的转换数据类型枚举体。cufftType_t 定义如下：

```
typedef enum cufftType_t {
    CUFFT_R2C = 0x2a, // Real to complex (interleaved)
    CUFFT_C2R = 0x2c, // Complex (interleaved) to real
    CUFFT_C2C = 0x29, // Complex to complex (interleaved)
    CUFFT_D2Z = 0x6a, // Double to double-complex
    CUFFT_Z2D = 0x6c, // Double-complex to double
    CUFFT_Z2Z = 0x69 // Double-complex to double-complex
} cufftType;
```

由于输入的数据为实数，将数据对应的虚部设为零，配置为 CUFFT_C2C，正向 FFT 变换 CUFFT_FORWARD。

- (5) 将实数输入转变为复数输入。
- (6) 执行 FFT。
- (7) 释放 plan 和相关占用的显存。

第二步，将一路信号频谱乘以另一路信号频谱的共轭得到互功率谱，并去除互功率谱矩阵的主对角线元素，得到新的互功率谱矩阵。这一步包括大量的矩阵运算，需充分利用共享存储器，减少对延迟较大的显存的访问。对于两个 $n \times n$ 的矩阵相乘，将矩阵放在显存中直接计算时，矩阵 A 的一行数据在每次进行计算时都被重复读入，需要读取 $2n^3$ 次显存，把整个矩阵乘法的运算，划分成很多小矩阵的乘法，可以避免耗费大量的存取时间。因此可以把两个小矩阵加载到共享存储器，小矩阵本身的乘法就不需要再存取任何外部的内存。假设小矩阵的大小是 k ，则实际上需要的内存存取次数约为 $2k^2(n/k)^3 = 2n^3/k$ ，大大降低了显存的访问次数，提高了矩阵乘法的速度。

第三步，当阵型确定后，声源到阵面的距离决定了成像面上各扫描点到阵元的相位延迟量，也就是导向矢量矩阵。优化线程分配，在计算量大时增加线程的分配量，确保充分利用 GPU 的计算能力。

第四步，导向矢量与互功率谱矩阵相乘后累加得到按扫描点分布的可控响应输出功率。可以得到与扫描点数对应的能量数据，并将数据放入共享存储器，作后续相乘累加运算，最后做归约运算得到每点的能量值。CUDA 的归约代码如下所示：

```
for(unsigned int s=blockDim.x/2;s>0;s>>=1)
{
    if(tid<s)
    {
        sdata[tid]+=sdata[tid+s];
    }
    _syncthreads();
}
if(tid==0) result[tid]=sdata[0];
```

第五步，通过插值得到一定密度的可控响应输

出功率。CUDA 架构的 GPU 具有快速图像渲染功能,可同时使用 GPU 的可编程单元和非可编程单元,使用非可编程单元不仅能够减少可编程单元的负担,而且能够提升运算速度。利用 GPU 的纹理存储器进行图像插值的步骤如下:

(1) 在所有函数之前添加一个纹理参照系的声明。

(2) 将输入数据拷贝到二维 cuda 数组,并进行纹理绑定。在 kernel 能用纹理参照系从纹理内存中读数据前,纹理参照系必须通过 `cudaBind TextureToArray()` 绑定到纹理上。

(3) 在核函数中利用纹理拾取进行插值。

第六步,将可控响应输出功率转换为 RGB 数据,得到声像图。

4 实验测试与结果

为了验证该算法的可行性和高效性,在普通台式机上使用 CUDA 实现了基于 GPU 的可控响应功率成像定位算法。实验中 CPU 为双核的 AMD 速龙 250, GPU 为 NVIDIA 公司的 GTX260,拥有 24 个流多处理器核心。基于该 GPU 的传声器阵列系统^[10]声传播测量实验如图 3 所示。



图 3 螺旋阵测量声传播实验

Fig.3 Experiment of sound propagation measured by a spiral array

将该算法应用到一个 64 阵元的声成像计算,假设信号源是一个频带为 1500~1800 Hz、信噪比为 10 dB 的窄带信号,采样率为 12 kHz,FFT 计算的信号样本长度为 4096 点;波束扫描范围设为横向 66°,纵向 48°。分别在 CPU 和 GPU 上运行可控响应功率成像定位计算,比较两者关键步骤的运算速度,得出如表 1 所示的运行时间及加速比。

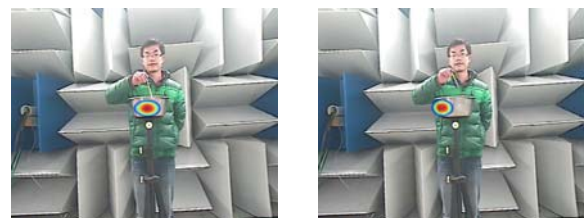
表 1 两种 SRP 算法的运行时间及加速比

Table 1 Consumption time of two kinds of SRP algorithms and speeding ratios

	CPU/ms	GPU/ms	GPU/CPU 加速比
信号预处理及 FFT	18.10	0.56	32.3
导向矢量计算	2468.40	29.12	84.7
可控响应功率计算	438.80	13.14	33.4
总时间	2925.3	42.82	68.3

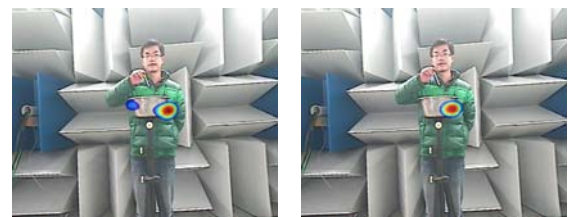
从表 1 可以看出,基于 GPU 的 SRP 算法的各个部分都比 CPU 版本的 SRP 算法有 30~85 倍的速度提升。其中一个主要的原因是利用了 GPU 计算超越函数的功能, GPU 在计算量较充分的情况下, GTX260 计算超越函数的速度大致是高阶(2.8 GHz 以上)CPU 的 9 倍以上,产生一个扫描点的声像需要近万次计算超越函数,因而使用 GPU 能够显著提高计算效率。声场中各个扫描点的声像是相互独立的,可以利用 GPU 的并行性同步完成各个扫描点的声像合成,使每一帧声像图的计算速度得到了极大的提高。

图 4 是基于 GPU 的 SRP 成像算法应用到平面螺旋声相仪测量声传播中的实验结果。从图 4 可以看出,当从盆中央敲击之后,声音首先从盆中央辐射,然后转移到盆的一侧;到第三阶段,盆一侧的声辐射开始向另一侧转移;到第四阶段声音转移到另一侧。这个可以看成由盆的介质分布不均而形成的偶极子模型。



(a) 第一阶段

(b) 第二阶段



(c) 第三阶段

(d) 第四阶段

图 4 螺旋阵测量声传播实验结果

Fig.4 Experimental results of sound propagation measured by a spiral array

图 5 是基于 GPU 的 SRP 成像算法应用到平面螺旋声相仪测量瞬态声源中的实验结果。当右手打



(a) 右手打响指

(b) 左手打响指

图 5 螺旋阵测量瞬态声源实验结果

Fig.5 Experimental results of transient sound source measured by a spiral array

响指时, 基于 GPU 的声相仪能快速定位到右手拇指根处的冲击声, 左手打响指也是一样。而应用基于 CPU 的成像算法的声相仪由于实时性不够, 无法捕捉到打响指这一类的冲击声源。

综上所述, 基于 GPU 成像算法的平面螺旋声相仪不仅可以跟踪运动声源, 还可以实时反映变化的声场。

5 结束语

利用 CUDA 架构开发出基于 GPU 的声成像定位软件后, 程序初始化只需 30 ms, 每秒可计算的声像图由原来的 1 帧变为现在的 20 帧。从而基本达到连续的声像图效果, 及时地反映出声场的变化, 实现了对高速运动声源的定位能力, 扩展了平面螺旋阵声相仪的应用范围。实验测量时只需配置有支持 CUDA 的显卡的计算机即可, 硬件配置简单方便。目前的算法是基于单显卡的算法, 当传声器数量增大 10 倍以后, 可以设计多 GPU 联合运算以保证声像计算的连续性。

参考文献

- [1] D J A. High-accuracy, low-latency technique for talker localization in reverberant environment[D]. Ph.D., Brown University, Providence RI, USA, 2000.
- [2] Christensen J J, Hald J. Beamforming technical review[J]. Brüel & Kjær Technical Review, 2007(1): 3-6.
- [3] Zotkin D N, Duraiswami R. Accelerated speech source localization via a hierarchical search of steered response power[J]. IEEE Trans. Speech Audio Process, 2004, 12(5): 499-508.
- [4] Peterson J, Kyriakakis C. Hybrid algorithm for robust, real-time source localization in reverberant environments[C]// IEEE Int. Conf. Acoust. Speech, Signal Processing, 2005: 1053-1056.
- [5] NVIDIA Corporation. CUDA Programming Guide 4.0[EB/OL]. <http://www.nvidia.com>, Jun, 2011.
- [6] 张舒, 褚艳利. GPU 高性能运算之 CUDA[M]. 北京: 中国水利水电出版社, 2009.
- [7] ZHANG Shu, CHU Yanli. High performance computing of GPU by CUDA[M]. Beijing: China WaterPower Press, 2009.
- [8] Mohammad J Taghizadeh, Philip N Garner, Hervé Bourlard, et al. An integrated framework for multi-channel multi-source localization and voice activity detection[C]// 2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays. 2011, 5: 91-97.
- [9] Jacek P Dmochowski, Jacob Benesty, Sofiène Affes. A generalized steered response power method for computationally viable source localization[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(8): 2510-2526.
- [10] 迟学斌. 高性能并行计算[R]. 中国科学院计算所, 2005.
- [11] CHU Xuebin. High performance parallel computing[R]. Institute of Computing Technology Chinese Academy of Sciences, 2005.
- [12] 刘哲, 陈日林, 滕鹏晓, 等. 基于平面传声器阵列的声源定位系统[J]. 声学技术, 2011, 30(2): 123-128.
- [13] LIU Zhe, CHEN Rilun, TENG Pengxiao, et al. Sound source localization system based on planar microphone array[J]. Technical Acoustics. 2011, 30(2): 123-128.