

利用 CUDA 的剪切波数据三维可视化

刘金硕¹ 程力¹ 王丽娜¹ 郑勇²

(1 武汉大学计算机学院, 武汉市东湖南路 8 号, 430079)

(2 中国科学院测量与地球物理研究所大地测量与地球动力学国家重点实验室, 武汉市徐东大街 340 号, 430077)

摘要:提出了一种基于 CUDA 的并行体绘制技术,将 CPU 与 GPU 进行协同工作,将光线投射这一高度并行的过程交由 GPU 来处理,同时设计了相应的体绘制数据访存优化策略来提高并行加速比。实验证明本方法可以将串行绘制速度提升三十倍。通过中国大陆南北带地区的剪切波数据三维可视化实验结果,可以直观看出低温物体的下沉俯冲带的分布。从三维图像可以看出,由于俯冲带的拖曳作用,使得软流圈的物质存在着向上流动的趋势,该高温物质在上升的过程中,使研究区两侧受到热物质的侵蚀而产生拉张作用,在一定程度上导致了该两侧的构造较为活跃,形成了类似断层、裂谷等拉张构造。相对于传统的显示方法,剪切波速度三维可视化更加高效、直观地展示了太平洋俯冲板块对南北带地区和华北克拉通动力学构造的影响,在地球物理学研究中得到良好的应用。

关键词:体绘制技术;CUDA;剪切波;三维可视化;固体地球物理学

中图分类号:P31

剪切波是地震波中与传播路径垂直的波,简称为 S 波^[1];地震波中伴随着断层两侧岩体快速移动产生的推力而形成的压缩波,被称为 P 波。分析地表接收到的地震波到达的时间及形态,可以对地下岩层的结构、深度、形态等做出精确推断,从而实现地震波速度和地球密度、温度等信息的探测。

三维可视化技术是对三维地震数据进行展示和分析的有效方法,为科研工作者提供直观、生动的感受同时能缩短解释时间^[2]。目前三维可视化技术在地震勘探领域应用广泛。文献[3]针对石油和天然气勘探设计了一种体数据漫游系统。文献[4-5]通过加入光照模型对地震数据的显示质量进行了深入研究。

当前国内外对于剪切波的研究主要集中在数据的处理和反演上^[6],对于剪切波的解释和阐述大部分还是定性的结果^[7-8]。二维静态的图像显示方法不能给出整体的三维构架状态,不够直观,使得了解和分析地下构造和动力学机制存在一定的困难。如果能够定量地给出三维的可视化图像,特别是动态的三维剪切波演化图像,将在很大

程度上增加对大陆形变、动力学演化以及地震孕育机理的认识和理解。

1 相关理论背景

1.1 光线投射算法

光线投射(ray casting)算法是最重要也是应用最为广泛的一类直接体绘制算法^[9-11]。如图 1 所示,光线投射算法从图像的每一个像素出发,沿固定方向(通常是视线方向)发射一条光线,光线在穿过整个三维数据场的过程中,选择 N 个等距的采样点,对每个采样点最近的 8 个体数据点的颜色值和不透明值进行三维线性插值,从而得到该采样点的颜色值和不透明值,直到光线穿过整个三维数据场。最后,将这个 N 个采样点的颜色值和不透明值进行由前向后或者由后向前的叠加,得到发射该光线的像素的颜色值,从而得到最终显示在屏幕上的图像。由于该算法需要对整个三维数据场进行遍历,而且当转动观察角度时,数据场中的每个体素之间的位置关系也发生了变化,需要重新采样计算,因此其计算量十分庞大。

收稿日期:2013-06-20。

项目来源:中国地震局国家地震行业重大专项资助项目(201008007);国家自然科学基金资助项目(40974034, 41174086)。

1.2 CUDA 计算模型

统一计算架构^[12](compute unified device architecture, CUDA)是由 NVIDIA 公司在 2007 年推出的一种整合架构,通过 CUDA 架构,使用者可以利用 NVIDIA 公司的较新 GeForce 8 和 Quadro GPU 进行高性能的并行计算。CUDA 将软件与硬件结合起来,使 GPU(graphic processing unit, GPU)形成一个统一的数据并行计算设备,跳过数据映射纹理步骤,使 GPU 能直接进行计算;它还对传统的 GPU 显卡架构进行调整,在芯片内加入了共享存储器,使得计算单元之间、并行线程之间能相互通信;它采用了统一的处理架构,使得定点渲染器和像素渲染器的利用大大提升;最为重要的是,它提高了并行编程的灵活性。

2 基于 CUDA 的并行体绘制技术

CPU 的优势在于处理串行计算以及逻辑事件,而 GPU 相比 CPU 能更好地处理高度并行计算任务。GPU 端并行处理光线投射算法中,每条光线均参与体数据计算的过程。算法中 CPU 负责初始化环境与设备,读取体数据,通过用户交互计算视角矩阵等逻辑性较强且内部耦合性较高的部分。

2.1 算法流程

2.1.1 CPU 端设计

1) 体数据读取与传递。

CPU 端以 void 类型将体数据读入主机内存,同时建立一个 cudaChannelFormatDesc CUDA 通道类型来对体数据类型进行描述,确定体数据的存储类型,再使用 cudaMalloc3DArray() 调用该描述,在设备内存中分配一个 volumeSize 大小、类型为 cudaChannelFormatDesc 的内存空间将体数据存储到设备内存。接着调用 cudaMemcpy3D() 将体数据从主机内存以 CUDA 内建的 3D Array 类型复制到设备内存。最后将 3D Array 类型的体数据绑定到 3D Texture。

2) 视角矩阵的传递。

在 CUDA 核函数中,视点与射线方向都需要与视角矩阵进行计算才能得到最终符合用户交互的绘制结果。本文采用常量存储器进行视角矩阵的传递。同时,对常量内存的单次读操作可以广播到其他附近线程以节省大量的内存读写过程,从而提高性能。

3) 建立 PBO。

像素缓冲对象(pixel buffer object, PBO)是 OpenGL 提供的缓冲对象(buffer object)中的一

种。PBO 提供了一种高效的数据操作及吞吐方式,这种操作涉及整个渲染管线的多个环节,可以提高 OpenGL 的实现效率。在 CUDA 核函数运行结束后,直接调用 PBO 绘制出结果图像,省去了数据由设备内存到主机内存的传递。

4) 设定内核函数调用参数。

由于体数据的数据量较大,此处直接使用 CUDA 设备支持的最大 blockSize 的倍数作为 grid 的数量,blockSize 为 block 的数量,来发挥设备的最大并行处理能力。

5) 图像显示。

图像通过视角矩阵调整 OpenGL 的摄像机位置,使最终图像能够按照用户需要的显示方式进行绘制。设定好启动 CUDA 运行时的参数,将 PBO 指针、图像高宽、图像亮度等参数传递给内核函数,设备运行时启动核函数进行光线投射的并行计算。计算完毕后,使用建立并绑定好的 PBO 进行图像绘制。用 glBindBufferARB() 将 PBO 与二维纹理绑定,打开 GL_TEXTURE_2D 开关,通过绘制四边形的方式,将纹理绘制出来。

6) 内存回收。

程序运行完毕后需要对已分配的内存进行回收。调用 CUDA 提供的 freeCudaBuffers() 函数回收在设备内存上分配的空间,cudaGraphicsUnregisterResource() 解除 PBO 与 cudaGraphicsResource 的绑定,glDeleteBuffersARB() 删除 PBO,glDeleteTextures() 删除纹理。

2.1.2 GPU 端设计

GPU 端的主要功能是实现并行处理光线与体数据的计算,是算法的核心部分。

1) 算索引。

进入并行部分后,每一个 thread 通过方法 $index = (block\ index) \cdot (block\ size) + (thread\ index)$ 计算出分配给该 thread 的屏幕像素点的坐标 (x, y) ,再通过 $2 \cdot (x/imageWidth) - 1, 2 \cdot (y/imageHeight) - 1$ 的方法将该坐标映射到 $(-1, -1, -1) \sim (1, 1, 1)$ 的三维空间中。

2) 计算相机位置及射线法向量。

GPU 端需要调用在 CPU 端实现中的视角矩阵,与 Z 轴正方向的向量相乘得到 OpenGL 的摄像机即算法模型中人眼的位置。再用该矩阵与摄像机到屏幕像素点的法向量相乘,得到射线在该视角矩阵下的方向。

3) Ray-box Intersection 算法计算管线与体数据的交点。

将体数据视为一个长方体盒子,有一束光线

射入该三维空间,该射线穿过边界时,会得到最近交点和最远交点。如果盒子最小点 B_l 为 (X_l, Y_l, Z_l) ,最大点 B_h 为 (X_h, Y_h, Z_h) ,设从眼睛发出的射线起始点 R_o 为 (x_o, y_o, z_o) ,终点 R_d 为 (x_d, y_d, z_d) ,射线与成对的边界面相交的最近交点和最远交点为 $tNear, tFar$ 。判断最大的最近交点与最小的最远交点的关系。若前者大于后者,则从眼睛发出的射线没有穿过该体数据,反之则穿过该体数据。

2.1.3 体数据采样

沿着射线方向,以 $tStep$ 的距离从最近交点 $tNear$ 向最远交点 $tFar$ 步进,使用 $tex3D$ 和 $tex1D$ 纹理拾取函数对每一个采样点通过硬件插值滤波的方式取值,再从前往后进行 $RGBA$ 混合。在遍历的时候,判断采样点的灰度值是否超过阈值,来确定该点是否是非空体素,是否对该点进行颜色混合,若非,则跳过该点。只对非空体素进行插值、采样、光照等运算,来提高运算速度。

2.2 利用 GPU 访存优化策略进一步提升加速比

2.2.1 利用全局+共享存储器提升体数据采样

提出一种共享存储器与全局存储器相结合的策略,以解决对体数据采样时访问时段不对齐和间隔访问带来的访问效率下降和全局存储器带宽浪费的问题:当对全局存储器发生非合并访问时,通过共享存储器充当缓冲区域,调整 $Half-warp$ 中线程的访问顺序,从而优化为合并访问。利用共享存储器及 $CUDA$ 提供的线程同步机制,可以将对 $float3$ 型数据的一次性读取的非合并访问转化为合并访问。

2.2.2 利用常量存储器提升视角矩阵转换

利用常量存储器策略提升用于控制图像视角变换的视角矩阵的运算速度。在光线投射算法中,每一条射线在计算采样点时都需要根据当前的视角矩阵计算出眼睛所在的位置以及射线的归一化法向量。利用 $CUDA$ 架构提供的 64 KB 大小的常量存储器存储频繁访问的只读数据。一个 $Half-warp$ 的线程只需要一个周期即可访问常量存储器并得到所需数据,效率大大高于从显存中读取数据。

3 实验及分析

实验开发环境为 32 位 Windows 7,开发工具为 Visual Studio 2008,开发语言为 $C++$, $CUDA$ 3.2 版。本文采用了中国科学院测量与地球物理所提供的中国大陆南北带地区($28^\circ \sim 42^\circ N, 96^\circ \sim$

$113^\circ E$)的剪切波数据,实现了 S 波三维可视化方法。由图 1 可知,中国大陆南北带区域构造非常复杂,存在着低温物体的下沉俯冲带的分布,在鄂尔多斯地区,能够看到大面积的高速块体存在于地壳和上地幔岩石圈中,清楚地展示了该区域的高速特征,也回答了鄂尔多斯地块稳定和地震少的原因。在南北带东部地区,有来自于太平洋俯冲板块的存在,该俯冲板块的前端地区能够看到地浅部地区存在着低速异常体,展示了太平洋俯冲板块是如何影响鄂尔多斯东部构造,进而造成华北克拉通的拉张的图像。由于俯冲带的拖曳作用,使得软流圈物质存在着向上流动的趋势,该高温物质在上升的过程中,使得研究区两侧受到热物质的侵蚀而产生拉张作用,在一定程度上导致了该两侧的构造较为活跃,形成了类似断层、裂谷等拉张构造;特别是山西裂谷带的形成,与该作用密切相关。这些构造特征都能从三维图像及其对应的切片上清楚地展示出来。

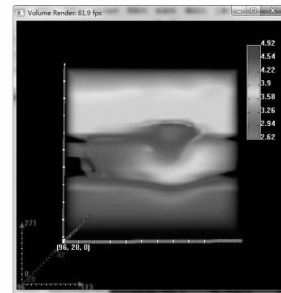


图 1 南北带地区剪切波数据三维可视化

Fig. 1 Visualization of 3D Shear Wave of North-South Earthquake Belt

为验证基于 $CUDA$ 的光线投射算法的效率,本文以 CPU 与 GPU 实现的光线投射算法为比较,以 4 种不同的标准体数据为例,测试两者的效率,结果见图 2。图 2 中左边阴影的柱状体代表 CPU 实现的光线投射算法的绘制效率,右边黑色的柱状体表示基于 $CUDA$ 的光线投射算法的效率。对于 $512 \times 512 \times 512$ 的三维图形, CPU 串行绘制时间为 27.4 s , GPU 并行绘制时间为 110.2 s 。光线投射算法在 $CUDA$ 上的绘制效率要明显高于在 CPU 上的。在低密度的体数据上, $CUDA$ 的效率是 CPU 的一倍,而随着体数据密度的增加, CPU 的绘制已不适合实时交互,而此时 $CUDA$ 绘制效率依然很高,几乎为 CPU 的三倍左右。此外,本文纵向比较了针对体数据的三种策略在相同条件下的绘制效率。以 $128 \times 128 \times 128$ 的体数据为例,实验结果如表 1 所示。由表 1 可知,全局存储器策略绘制效率太低,导致算法性能

不理想,因此可得出结论:全局存储器策略不适合基于 CUDA 的光线投射算法;纹理存储器策略和全局存储器策略绘制效率不错,在对体数据读取不频繁的情况下,纹理存储器策略绘制效率比全局+共享策略略高。

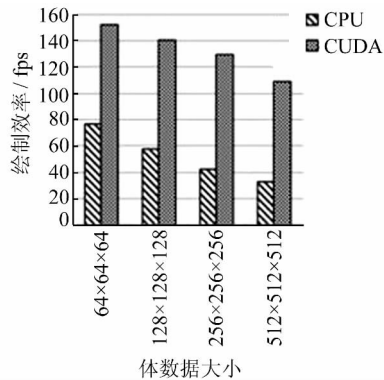


图2 CPU与GPU加速效率比较

Fig. 2 Comparison of Effects of CPU and GPU

表1 不同访存策略的绘制效率比较

Tab. 1 Effects of Different Strategies of Memory Access

访存策略	绘制效率/fps
纹理存储器	137.2
全局存储器	90
全局+共享	130

针对纹理策略在对纹理存储器访问频繁的情况下效率降低的现象,设计实验如下。逐步减小对体数据的采样间距,增加对体数据读取的次数,观察纹理策略和全局+共享策略的绘制效率。实验结果如图3所示。随着对体数据采样间距的缩小,即对存储器的读取次数逐渐增加的情况下,纹理策略的绘制效率下降非常快,而全局+共享策略对实验条件的变化没有产生较大变化。从此实验可知,当算法需要对体数据频繁读取时,全局+共享策略要优于纹理策略。

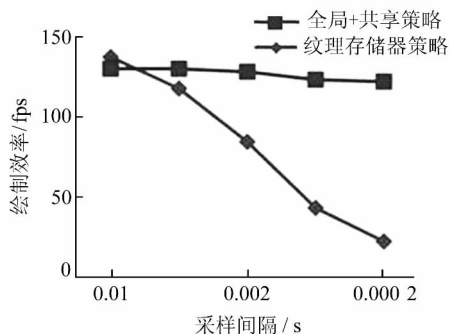


图3 纹理策略与全局+共享策略的效率比较

Fig. 3 Combination of Texture Memory and Global Plus Shared Memory

4 结 语

三维可视化技术是地震图像领域的一个研究热点。剪切波作为地震数据的一种,可以用于地下结构和地球动力学信息的勘测。提出了一种基于 CUDA 的并行体绘制技术,利用 CUDA 实现剪切波数据的三维可视化,从而更好地实现地下结构的展示。为了提高并行体数据的执行效率,设计了不同的访问策略:纹理存储器策略、全局存储器策略、全局与共享存储器策略以及常量存储器策略,并设计实验验证本文方法的绘制效率较之前提升了三倍。

参 考 文 献

- [1] Wikipedia. S 波 [OL]. <http://zh.wikipedia.org/wiki/S%E6%B3%A2>, 2003
- [2] Bnoit M H, Nyblade A A, Decar J C V. Upper Mantle P Wave Speed Variations Beneath Ethiopia and the Origin of the Afar Hotspot[J]. *Geology*, 2006, 34(2): 329-332
- [3] James H, Kostrova T. Use of Lighting in the 3D Display of Seismic Data[J]. *First Break*, 2005, 23(3): 53-57
- [4] Castanie L, Levy B, Bosquet F. Volume Explorer: Roaming Large Volumes to Couple Visualization and Data Processing for Oil and Gas Exploration [C]. *The 16th IEEE Visualization*, Minneapolis, USA, 2005
- [5] Kidd G D. Fundamentals of 3D Seismic Volume Visualization[J]. *The Leading Edge*, 1999, 18(6): 702-709
- [6] 张学民,刁桂苓,赵英萍,等. 华北地区深部地幔剪切波速度结构研究[J]. *地球物理学报*, 2006, 49(6): 1709-1719
- [7] 王椿镛,楼海,吕智勇,等. 青藏高原东部地壳上地幔 S 波速度结构—下地壳流的深部环境[J]. *中国科学(D 辑:地球科学)*, 2008, 38(1): 22-32
- [8] 郑勇,葛黎,谢祖军,等. 芦山与汶川地震震区地壳上地幔结构及深部孕震环境[J]. *中国科学(D 辑:地球科学)*, 2013, 43(6): 1027-1037
- [9] Strengert M, Magallón M, Weiskopf D, et al. Large Volume Visualization of Compressed Time-Dependent Datasets on GPU Clusters[J]. *Parallel Computing*, 2005, 31(2): 205-219
- [10] Strengert M, Magallón M, Weiskopf D, et al. Hierarchical Visualization and Compression of Large Volume Datasets Using GPU Clusters[C]. *Eurographics Symposium on Parallel Graphics and Visu-*

- alization(EGPGV), Switzerland, 2004
- [11] Levoy M. Display of Surfaces from Volume Data [J]. IEEE Computer Graphics and Applications, 1998,8(3): 29-37
- [12] NVIDIA Corporation. NVIDIA CUDA Compute Unified Device Architecture Programming Guide [OL]. <http://moss.csc.ncsu.edu/~mueller/clus->

[ter/nvidia/0.8/NVIDIA_CUDA_Programming_Guide_0.8.2.pdf](http://moss.csc.ncsu.edu/~mueller/clus-ter/nvidia/0.8/NVIDIA_CUDA_Programming_Guide_0.8.2.pdf),2008

第一作者简介:刘金硕,博士,副教授,主要从事高性能计算与科学可视化研究。

E-mail:liujinshuo@whu.edu.cn

3D Visualization of Shear Wave Data Based on CUDA

LIU Jinshuo¹ CHENG Li¹ WANG Lina¹ ZHENG Yong²

(1 Computer School, Wuhan University, 8 South Donghu Road, Wuhan 430072, China)

(2 State Key Laboratory of Geodesy and Earth's Dynamics, Geodesy and Geophysics Institute, Chinese Academy of Science, 340 Xudong Street, Wuhan 430077, China)

Abstract: By analyzing the arrival time and the morphology of a shear wave, we can make accurate predictions of underground structures, depth, and morphology. Explanation and elaboration of a shear wave are currently based on qualitative results and two-dimensional display of the images, and not beneficial for analysis of underground structures directly. This paper proposes a parallel volume rendering technology based on CUDA, which makes a CPU and GPU work together. Highly parallel processes such as light sampling are handled by the GPU. Further, we also designed a corresponding visit and access optimizing strategy for rendering volume data; texture register strategies, global register policies, global and shared register strategies, and constant registers strategies. An experiment demonstrates that this model is three times faster than the serial programming with improved visualization results. In this experiment, at the center of the corresponding 3D image, the trends for sinking low-temperature objects and rising sthenospheric material can be observed. During a trend, two sides generate a tensile function when experiencing erosion from hot material, causing the structure of the two sides to become more active, forming tensile tectonics such as faults and rifts. Using our proposed technology, underground structures and geodynamic information can be explored more effectively.

Key words: volume rendering; CUDA; shear wave; 3D visualization; solid earth geophysics

About the first author: LIU Jinshuo, Ph D, associate professor. Her major interests are in high performance computing and scientific visualization.

E-mail: liujinshuo@whu.edu.cn