

科学计算可视化

网络环境

网络环境

5

24-26,SS

高速网络环境中的科学计算可视化

蔡勋 黄朝晖 李晓梅

TP391.41

(国防科技大学计算机系 长沙 410073)

摘要 The visualization techniques have vastly increased the researcher's ability to analyze or understand data. A close link between scientific visualization and high speed networks provides an efficient supercomputing environment. This paper introduces some problems met in scientific visualization and provides a solution to workload distribution in high-speed network environments. Finally, the research for supercomputing visualization environments to support future computational requirements is summarized.

关键词 Visualization, High-Speed Network, Workload distribution, Distributed visualization

一、引言

近年来,科学技术飞速发展,人类每天都产生了大量数据,由于缺乏大体积数据的有效分析手段,百分之九十九以上的数据都被白白浪费掉。斯坦福大学和哈佛大学的研究人员都已发现复杂的数据以视觉形式来表现时最容易理解,这是因为图像是沟通思维的最自然手段。事实上,大脑具有的大约一百亿条神经元中有一半专为视觉服务;但是人类产生数据的能力远远大于自身把数据信息转换成图像的能力。在这种形式下,美国计算机成像专业应美国国家科学基金的要求,研究并提出了解决的方法——“可视化”方法,并于1987年11月提交了“科学计算可视化”报告。报告宣称,“科学计算可视化对科学生产力和重大科学突破产生重大影响,这种影响可与巨型机的影响相比拟”。因此,有人称科学计算可视化第二次计算机革命。科学可视化就是用图画来描述物理现象,将看不见的变成可见的,帮助科研人员对他们研究的问题形成一个清晰的图像。它丰富了科学发现的过程,给予人们意想不到的深入事物内部的洞察力。自Visc报告发表至今,科学计算可视化已取得了很大进展。与此同时,提供高效超级计算环境的网络技术也有了很大进展,科学计算可视化和高速网络的结合为超级计算提供了一种有效的分析计算环境。随着计算模型日趋复杂,以及超级计算所产生的数据量的不断增大,上述两种技术的结合将更为紧密。

二、网络的引入

在过去的科学计算可视化研究中通常是采用图1模型进行工作。但是这种模型方式的最大缺点就是共享信息必须局限在某一范围,如果合作人员分布在不同区域里,则这种交互式的合作参与开发是不可能的。在网络环境中,我们采用图2的模型就可以解决这个问题。但是这里有一个计算环境可能不相同的问题,有些人可能在一簇工作站环境下工作,而有些人可能在PC机构成的局域网下工作,还有人也可以在联着超级计算机的图形工作站上运行他们的作业。如何有效地在各种网络资源中分配工作负荷是可视化技术在网络中运行所需解决的重大问题之一。下面我们将进行详细讨论,虽然我们不知道网络的带宽也是可视化在网络中运行的瓶颈,但是随着光纤网络的发展,这个问题将会逐渐得到解决。

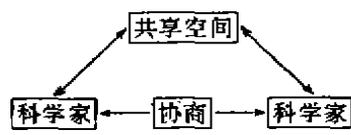


图1 直接合作环境

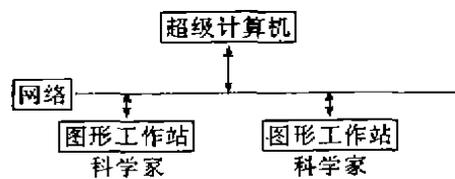


图2 网络合作环境

三、负载分配方法

首先,我们列举一些可能的工作负荷分布环境,见图 3(a),(b),(c)。下面将针对图 3(a)为模型讨论网络中工作任务分配问题。我们可以把可视化过程理解成数据的转换过程,见图 4。可视化技术应用在标量和矢量域,生成大批的原始数据,并且把它们形式化为几何数据,几何数据再描绘成图象数据。通过在超级计算机和图形工作站之间划分计算任务,使分布式处理技术应用于可视化过程。如何分配任务量,原则上取决于可视化系统是否提供了一个最优发挥超级计算机、图形工作站和网络能力的环境。我们给出几种可能的工作负荷划分方式(图 5): (1)分布式文件系统划分;(2)几何划分;(3)分布式的图形库划分;(4)帧缓存划分。总体看一下可视化过程,分布式文件系统划分接受一个远程机器中的计算输入,而帧缓冲划分输出一些数据到远程机器上,分布式的图形库和几何划分是使图形处理分离成局部和全程任务。我们将对这四种划分方案进行讨论。

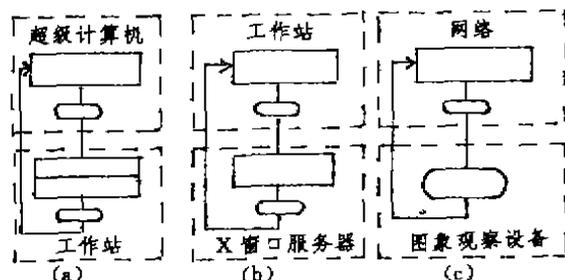


图 3 工作负荷分布环境
(a)与超级计算机相连接的工作站网
(b)带有一个 X 窗口服务器的一簇工作站
(c)带有观察图象文件设备的通用网络

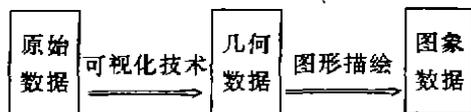


图 4 数据转换过程

NFS(Network File System)是分布处理中分布式文件系统的典型例子。对一个可视化应用程序,原

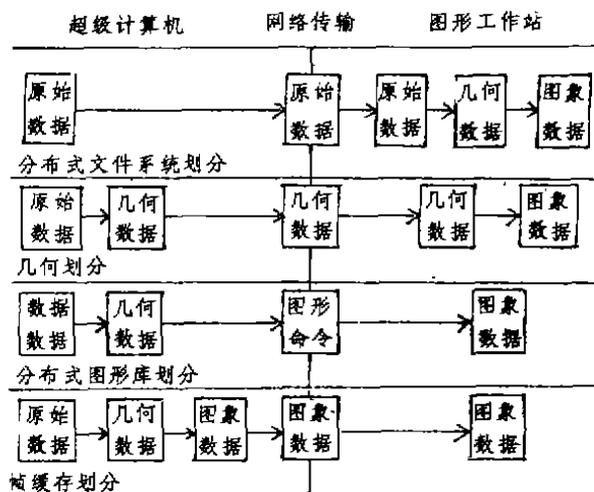


图 5 可视化负载分配方法

始数据在超级计算机中产生,而图形的处理过程则在用户的工作站上完成,用这种方式进行分布式处理允许用户使用超级计算机的大容量存储资源,在这里还有一个小问题就是原始数据的传送有两种方式,一种就是上面所述,先存入巨型机中的磁盘设备中,然后再传送给用户;另一种就是原始数据一旦产生就立刻直接传到用户中去,也就是说前一种保留数据副本,后一种不保留。分布式文件系统划分的一个优点就是节约了磁盘的存取时间和空间,然而可视化应用的这种划分过程仅仅利用了超级计算机的高速磁盘到内存的速度,而把大量的数据处理过程留给工作站去做,这显然没有合理利用超级计算机、图形工作站和高速网络的功能。在分布式处理过程中,原始数据到几何数据的转换必须由超级计算机执行是一条重要的原则,这样做,可以获得显而易见的好处。

在帧缓存划分中,可视化处理的所有过程都是在超级计算机中完成,图象以象素数据的方式只是传送到图形工作站上去显示。这种在不同机器环境和位置的图象显示显然比图象的计算和描绘因物理位置的限制而必须局限在与用户比较接近的环境中计算要好,但是,我们发现图象的传输将引起网络带宽的瓶颈问题,假设每个象素点有 24 位,每一帧图象有 1024×1280 个象素点,在网络中的数据传输率是每秒 24 帧,这样每秒就要传输大约 90MB。在两个或更多个工作站之间维持这样的传输速率显然是很困难的,当然用视频信号传输图象能够减少数据的传送量,但这将引起图象质量的损失即图象失真,虽

然以 Gbit/s 传输的高速网可以部分地解决这种分布处理的瓶颈问题,但在这里高速网仍然不是一个很好的解决方法,即使网络以每秒一个 Gbit 速率传输数据,并且高性能的图形工作站也具有以这个速度传输或接受数据的能力,但对于图象数据以动画速度连续传输同时具备交互能力那就显得无能为力了。

分布式文件系统和帧缓存划分都允许整个可视化过程在一台机器上完成,所不同的只是前一种需要从远程计算机上接受原始数据,后一种需要把输出结果传输到一台远程机器上,这两种划分实现方法都对灵活使用超级计算机和图形工作站这一整体处理有所限制。

第三种分布式处理任务划分是用分布式图形库任务划分来进行几何数据到图象数据的转换。原始数据被产生并且形式化为几何数据,网络传输图形库调用命令序列,而图象在另一台机器上生成,这种任务划分的优点就是程序设计的所有工作都在超级计算机上执行,而图形调用语句被传送过去,这样就象图象的生成是在当地,而我们不必管它是在哪一台计算机上执行的,当然,这种任务划分需要一个能实现网络事务处理所需要的特定分布式图形库, NWS(Network Window System)就是利用这种划分的一个例子。分布式图形库任务划分对一个需要在分布式处理中修改的应用程序的底层设计有很小的影响。因此,这种任务划分也具有一定的局限性,但是利用分布式图形库任务划分却成功地发挥了超级计算机和图形工作站的能力。

交互式控制从几何数据生成图象的过程需要网络事务处理,并且还要具有一个分布式图形库任务划分。网络的事务处理和相关的额外工作可以通过把几何数据传输到图形工作站的方式得以减少,这就是我们所关注的几何任务划分。这种任务划分使可视化过程中的整个描绘阶段全部由图形工作站完成,而不必涉及命令信息和图形数据的网络传输,几何任务划分对分布式可视化应用是一个非常灵活的框架模型,它不同于分布式图形库任务划分,几何任务划分把网络处理同图形处理分隔开。几何任务划分允许可视化过程中的描绘控制完全局限于图形工作站上,假设构成一幅完整动画的几何数据都存贮在图形工作站上,则动画的整个可视过程包括颜色控制、几何变换等其它描绘控制完全局限在工作站上,而不必引起网络处理中的额外开销。为了便于更好地交互控制可视化应用结果,我们还采用一个会

谈管理系统(Conference Manager System)。通过上面讨论我们可以把图 2 改成图 6 的形式。美国的 NASA Ames Research Center 已经成功地在网络环境下开发出 NAS(Numerical Aerodynamic Simulation)系统,并且已运用在 CFD(Computational Fluid Dynamics)领域。

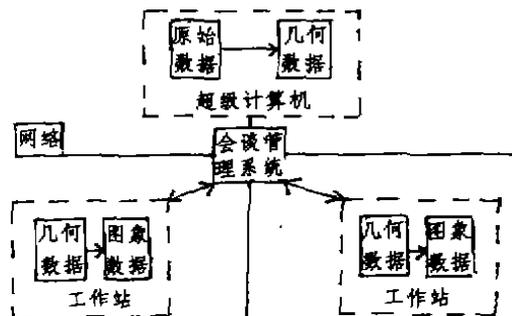


图 6 可视化几何任务划分模型

四、未来发展方向

我们现在讨论一下可视化技术在网络环境中的未来发展方向。由于可视化技术飞速发展,针对 HPC 计算能力的可视化需求已经提出,每秒兆兆浮点(tera-FLOPS)处理能力的超级计算机可望在 2000 年实现,如果工作站的处理能力和超级计算机一样线性增长,那么目前相对于超级计算机性能 1:100 的计算速度的工作站到 2000 年将达到 10GFLOPS。在 TFLOPS 时代,超级计算的环境应该提供一个完整理解模拟计算过程的能力,在不同的模拟阶段,不同的交互式选择层次都应被支持,我们称这种模型为 PSM(Perceptual Supercomputing Model),它将支持模拟控制、交互可视化、可视分析及便于人类理解的多媒体计算。

四个技术领域将对未来的科学计算可视化平台有着重要影响,它们是:(1)可视化环境;(2)虚拟真实界面;(3)体可视化;(4)分布/并行系统。基于这些技术的 PSM 平台将生成,这个系统包括一个具有虚拟真实界面的可视化环境,对于 3D 数据集的可视描述和分析的主要手段是体可视化技术,而运行的环境将是一个由分布和并行系统组成的拥有不同资源的网络。虚拟真实界面将解决目前在深度感(Depth Perception)上的局限性,这一问题是由 3D 数据集映射到 2D 显示屏所引起的。可视化环境提供了目前最有效的对交互控制科学计算和可视化应用原型工

(下转第 55 页)