

虚拟现实 虚拟环境系统 系统设计 (25)

计算机科学 1998 Vol. 25 No. 6

# 分布式虚拟环境系统的设计与评价

Design and Evaluation of Distributed Virtual Environment System

王勇军 李思昆 胡守仁

(国防科技大学计算机系 长沙 410073)

**摘要** This paper proposes some design issues of distributed virtual environment system. Based on these issues, we evaluate and compare current famous distributed virtual environment systems.

**关键词** Distributed Virtual Environment System, Virtual Reality, Evaluation.

虚拟现实技术作为计算机技术、自动控制技术、仿真技术、传感器技术等完美结合,以及作为人和技术的完美结合,已成为人们所追求的最高技术境界。

典型的虚拟现实应用可分成三个单元:计算、交互、和表现。所谓计算单元,通常指针对不同应用建立的仿真模型的计算。交互单元协调管理所有来自用户的(可包括各种感觉通道)输入和反馈给用户的输出。表现单元则是基于应用数据产生所有反馈给用户(可包括各种感觉通道)的输出,如所能看到的图象、所能听到的声音以及能感受到的力量的反馈等。虚拟现实系统既要以比较高的真实感来模拟现实世界,又要给用户实时的沉浸感,那么各个应用单元都将是密集型的计算,这种强大的计算压力只能依靠并行与分布处理技术来解决。

目前,随着网络技术的飞速发展,分布式(网络)虚拟环境系统的研究正在成为引人注目的热点。除了解决并行计算的问题外,分布式虚拟环境还使得不同地区的多个用户身临其境地共享同一个虚拟环境并且协同工作,它已被很好地应用在远程会议、合作设计以及各种各样的训练仿真、游戏和模拟旅游等领域,具有广泛的应用前景。分布式虚拟环境必将渗入到社会的各个方面,使人们的生活方式发生巨大改观。

## 1 系统的设计

### 1.1 系统目标

构造分布式虚拟环境最终要达到的目标是:实时性和真实感。

一个虚拟现实系统的最大约束是它必须实时地操作,即要求各种感觉通道的反馈(如视觉显示)能够足够地快,以允许参与者和虚拟环境及其它参与者有效的交互。这里的实时并不要求与传统的实时系统严格地等同,它的主要目标是尽量降低人的行为和仿真反应之间的延迟,使仿真达到足够的真实感。

文献[1]将分布虚拟环境中的真实感分为:感觉真实感、物理真实感、模型真实感、行为真实感、时间真实感、信息真实感、输入设备真实感、系统真实感。所谓感觉真实感是指用户在与虚拟环境交互时,在视觉、听觉、触觉等感觉通道的反馈与现实世界相同;物理真实感是指虚拟环境所遵循的物理规则表现得与现实世界相同;模型真实感是指虚拟环境中的虚拟物体在几何模型的尺寸、相对距离等方面表现得与现实世界相同;所谓行为真实感是指虚拟环境中的智能实体表现出来的智能行为与现实世界相同;时间真实感是指虚拟环境中的时间特性表现得与现实世界相同;信息真实感是指在虚拟环境中用于决策的信息的复杂度和细微程度表现得与现实世界相同;而所谓输入设备真实感是指用于用户与虚拟环境交互的输入设备与现实情况相同;系统真实感是指人机一体的整个系统所表现出来的整体真实感。当然,真正严格意义上的真实是不可能的,甚至无法精确衡量,但是,只要在一定程度上差别不会

王勇军 博士生,主要研究方向为虚拟现实和并行分布处理技术。胡守仁 教授,博士生导师,主要研究方向为高性能机体系结构及并行分布处理技术。李思昆 教授,博士生导师,研究领域为 CAD 技术和虚拟现实技术。

被人察觉出来,这就达到足够的真实感了。

## 1.2 设计要素

在进行一个分布式虚拟环境系统的设计时,要考虑以下要素:

1.2.1 实时性和真实感的折衷 不同的应用对不同层次的真实感的要求是不同的,常常仅需侧重于某些真实感的表现。但是,从目前技术实现的角度看,实时性和真实感二者是矛盾的,即如要达到较高的实时性,则要以损失一定的真实感为代价,反之亦然。因此,对于不同的应用,首先要在二者之间进行折衷,以选择合适的实现技术和硬件设备。这将决定设计的其它要素。

1.2.2 通信结构 同样,通讯结构的选择与虚拟环境系统所面向的应用紧密相关,不同的应用在网络上所要求传输的数据量的大小,以及所要求的通信延迟是不同的,这对系统设计有着至关重要的影响。因此要选择合适的硬件网络环境、通信协议以及通信模型,使得应用所需要的网络带宽和传输速度能够得到满足。但是由于受技术和经费的限制,大多数应用只能基于现有的硬件网络设施,这时就要选择或研究新的通信协议和通信模型来改善通信的质量。

目前,可选择的通信协议有点到点传输、广播、多信道广播等。点到点传输通常建立发送者和接收者之间的连接,能够提供可靠的服务。但是,要维护可靠连接需要大量的主机开销,降低了网络速度。广播是一个相反的机制,它的好处是不需要保持大量的连接,但它是不可靠的,而且也会给网络带来信息泛滥。对广播的发展是多信道广播协议,但目前只有美国的 Mbone 高速网上的网络层提供该服务,而且还不可靠。

用于通信的模型有三种:客户/服务器、端对端、层次模型。其中,客户/服务器模型能很好地共享资源,但如果客户的数量太多了,服务器就会变成瓶颈,而层次模型则较为灵活。

1.2.3 数据和计算管理 我们所讲的数据是指虚拟环境的各种属性,如虚拟场景的地貌数据、虚拟实体的几何模型、物理属性以及其它的属性等。这些属性状态可以是静态的,也可以是动态的。所谓计算,指可以对虚拟环境数据进行处理的操作,如对虚拟实体状态产生修改的行为计算、对虚拟场景进行的绘制计算等。分布式虚拟环境的运行效果如何,本质上很大程度依赖于它对数据和计算的分布的管理上。

数据的分布有三种:完全分布、部分复本、全复本。完全分布是将虚拟环境的属性数据完全地分布在网络的各个结点机上,如果某结点需要某些数据,就要同拥有它的结点通讯。由于某数据只能在一个地方得到,这种分布所需的网络开销很大,而且缺乏容错性,因此这种技术只能应用于近距离/紧耦合或近距离/松耦合层次的分布系统。当使用部分复本时,只有部分虚拟环境的数据在仅当需要时被局部保存,它又分为主动式和被动式两种。主动式是指当某结点机要对某些数据进行修改时才要求对其局部拷贝,修改信息被传给该数据的拥有者;而被动式是指在虚拟环境运行前,某结点机就说明对某些数据有兴趣,此后,一旦其状态被某些结点改变,都会有局部拷贝发送到该结点上,而本节点对其状态的修改可以在本地进行,并将结果发送给拥有者。当然,复本问题必然会导致一致性问题,当采取读写锁的机制,会产生死锁的问题。但是,如果所有的修改都由拥有者来作,就不存在这个问题,这种技术通常被应用于近距离/松耦合层次的分布系统,如果频带足够高时也可应用于远距离/松耦合层次分布系统。相对于数据的全分布,数据的全复本是在每一个结点机上都保留一份完全的虚拟环境的数据。这样作的主要原因有两个,第一,某结点的计算基于所有或大多数的信息;第二,结点间的距离太大,延迟成为很大的问题,如远距离/松耦合层次的分布系统。因为每个结点都要保持其它结点的修改信息,网络带宽迅速耗竭,锁的分配也不太可行,所以这种方法的伸缩性很差。

当数据量小时,每个进程都能直接访问,而当数据量增大时,对资源的压力也增大了;内存和外存的容量急速下降,处理数据的计算量急速提高。当数据从一个结点向另一个结点移动时,网络的拥塞程度也提高了。因此,为了快速访问数据,采用数据局部化原则。有一种比较好的方法是用某种逻辑的方法对数据进行分段,使得进程在某个时间或大部分的时间内只对一个段感兴趣。最通常的分段方法是空间,当一个虚拟环境跨越了一个较大的虚拟距离时,它被划分为许多领域,这些领域由一个领域管理器分配和控制。领域的划分依赖于许多因素,比如可见性、内存、虚拟环境中移动的速度等,还依赖于输出的感觉通道,如视觉通道或听觉通道等。每个领域基于简洁性的考虑,经常保持一致的形状,如长方形、六边形。

类似的,计算分布有四种:完全分布、部分分布、

部分复本、全复本。

完全分布是指在数据完全分布的情况下对某些数据的计算都在拥有数据的结点机上进行,如果需要对别的数据进行计算,就必须向其它结点发送消息。这种方法在近距离/紧耦合系统上工作得很好,但在远距离/松耦合系统上的延迟和低带宽将会降低它的效率。

部分分布类似于完全分布,唯一的不同是通常使用被动或主动的数据复本技术来获取数据、修改在本地进行。

部分复本是为了弥补更慢的通信链路的不足,如近距离/松耦合的分布系统,将对一些数据的计算复制到一些或全部的结点上。如航位推算技术<sup>[2]</sup>就是一种部分复本,它利用虚拟实体的简化模型来进行计算(主要的关键状态变量是位置和速度)得到实体的近似行为,从而非常有效地减小了带宽的消耗,但是实体的行为和正常的相比有些差别。当然可以通过提高近似的复杂度来改进这一点,但这会引起计算量的增加。

全复本是指,如果仿真在非常远的距离(远距离/松耦合的分布系统)的情况下运行,就有可能要求在每个结点上仿真每个虚拟实体。因为当带宽很珍贵的时候,使用被动复本从别的结点接收周期性修改信息是不现实的。

计算的分布可以预先设定,也可以在系统运行时动态调节。计算的分布要采用负载均衡算法来进行优化,通过管理资源消耗和通信模式可以产生优化分配,这样可以使得大多数互相通讯的对象驻留在同一结点上,计算的移动通常被称为迁移,负载均衡和计算迁移在紧耦合系统中得到最好的应用。当面向的应用计算较为复杂时,提供对计算迁移的支持将大大增强系统的灵活性和可伸缩性。

1.2.4 对多个环境的支持 如果分布式虚拟环境系统允许多个使用同样或者不同模型的虚拟环境的并行运行,我们称之为支持并行环境。通过使用同样的虚拟环境,可以最大限度地利用特定的对象或者虚拟环境中的某个区域。比如,一个虚拟的城市大厅能供不同的人群同时集会。如果能够支持多个不同虚拟环境,那就最大限度地利用系统的资源。这种能力提出了关于更高层次的调度、负载均衡等问题。如果系统支持一个虚拟实体能从一个环境移动到另一个环境的功能,我们称这种并行环境为多环境。

1.2.5 对多用户的支持 在虚拟环境运行过

程中,支持多个用户是分布式虚拟环境系统的最大特色,这意味着要在各仿真结点间传送大量的用户行为信息,在大规模的虚拟环境中,额外的通信开销将是分布式虚拟环境系统的主要问题。因此,能有效降低带宽的航位推算技术和数据局部化技术也成为解决这一问题最为实用有效的技术。

1.2.6 时间管理 虚拟环境的时间真实感是一个非常重要的问题,仿真时钟作为虚拟环境的同步的基础,它可以采用隐式或显式的推进方法。隐式的推进是指,当系统时钟改变时,仿真时间也发生改变。而显式的推进是通过远程资源的通知来进行修改,如一个消息时标或一个特殊的消息,它可能在每个时间步的开始发生。在分布式系统中,当使用隐式的推进时,必须确保每个结点的实时时钟是同步的。一个可能的解决方法是GPS<sup>[3]</sup>接收器,可将它内建进每个结点。另外,也可以使用软件算法来进行时间同步<sup>[4]</sup>。

## 2 现有系统的评价

当前存在的一些较为典型的分布式虚拟环境系统,按应用目标大致可分为三类:面向分布式交互仿真(DIS)的系统、面向人机交互、计算机协同(CSCW)的系统和通用系统。在这里我们列出的系统有:SIMNET<sup>[5]</sup>、NPSNET-IV<sup>[6]</sup>、MR<sup>[7]</sup>、DIVE<sup>[8]</sup>、dVS<sup>[9]</sup>、WAVES<sup>[10]</sup>、AVIARY<sup>[11]</sup>。其中,SIMNET、NPSNET-IV是面向分布式交互仿真的系统,MR、DIVE是面向人机交互、CSCW的系统,其它的可看作是通用系统。所谓通用系统是指,该系统并不面向任何特定的应用领域,而是为开发分布式虚拟环境提供基本工具(如MR),或侧重于先进的分布式虚拟环境的体系结构和软件平台的研究(如dVS、WAVES、AVIARY)。

我们根据第1节中归纳的设计要素,将各系统列表如表1。

一般说来,前两类系统由于基于特定的应用设计实现,因而性能较好,尤其NPSNET-IV是最成功的分布式虚拟环境系统,它利用多信道广播协议和数据局部化的技术大大提高了系统的可伸缩性,这些技术将成为分布式虚拟环境系统的标准技术。但面向分布交互仿真的系统受限于军事应用,缺乏通用性。DIVE则在面向人机交互、计算机协同(CSCW)的系统开发方面提供了灵活而有效的系统支持,但DIVE仅提供基本的计算迁移机制,控制工作须由应用开发者完成。