

29-31

数字电影

计算处理系统

工作站机群

计算机科学2000Vol. 27No. 2

电影制作

## 基于工作站机群的数字电影计算处理系统\*

A Digital Film Processing System Based on Workstation Cluster

毛希平 温冬婵 董春雷

J969

(清华大学计算机科学与技术系 北京100084)

**Abstract** Computerized processing is the most important part of the Digital Film System. In this paper, we have introduced the organization and implementation of a computerized processing system. The system has used workstation cluster for rendering instead of expensive mainframe because it owns higher cost performance, utilization ratio and scalability.

**Keywords** Digital film system, Workstation cluster, Rendering, Cost performance, Load-balancing, PVM

八十年代末,高精度连续图像数字处理技术日渐成熟,在电影制作中被大量采用并形成了完整的体系,即数字电影系统。尽管目前世界上的数字电影系统多种多样,但从功能上讲它们都由三部分组成:透射式胶片扫描输入系统、数据存储及处理系统(计算处理系统)和胶片记录输出系统。

计算处理系统作为数字电影系统最重要的组成部分,承担着数字电影全部的数据处理和数据传输工作,其性能的好坏直接关系到整个系统的成败。本文主要介绍了一个基于工作站机群的计算处理系统,它以工作站机群替代昂贵的大型服务器进行连续帧的生成(渲染)工作,使得整个数字电影系统性价比更高,更适合我国国情。

## 1 计算处理系统机群解决方案的提出

### 1.1 数字电影系统的主要技术难点

在数字电影中,称美工工作台上处理的单帧数据图像文件为关键帧(Key-Frame),主要的特技效果、三维模型都要首先制作成关键帧,当关键帧的效果通过导演工作台的审查后,利用计算机运算自动获得同一镜头的其他帧图像,这一工作通常称为渲染(Rendering),在标准配置的O2工作站上生成单帧图像(4K×4K)约需要1~1.5个小时,生成一个5秒钟的特技镜头要连续运算220多个小时,渲染的计算量极其惊人。一段20分钟的普通35毫米电影胶片的信息量为1200GB,远超过一般多媒体图像处理水平<sup>[1]</sup>。由此可见,数据

处理能力是数字电影系统性能的重要标志,也是主要技术难点之一,国外的系统通常采用多CPU的大型服务器(如 Origin2000, Challenge4000等)专门从事渲染工作。

### 1.2 计算处理系统机群解决方案的可行性

随着工作站性能的迅速提高和价格的日益下降,以及高速网络产品的陆续问世,工作站机群系统应运而生,它以某种结构将一群工作站用网络连接起来,充分利用各工作站资源,统一调度,协调管理,以实现高效、并行的计算。

目前我国的数字电影制作才刚刚起步,仍处于技术跟踪阶段,与国外相比,特别是资金方面缺口很大,因此不能照搬国外的模式,为此我们提出了以工作站机群替代价格昂贵的大型服务器进行渲染工作的设想,这是有充分依据的:

(1)机群与服务器的渲染结果一致。一般来说,只要采用相同的应用软件及版本,二者渲染的结果是一致的。

(2)关键帧处理与渲染工作在时间上不冲突。在数字电影处理过程中,关键帧的制作与渲染工作在时间上相隔比较长,并且渲染过程不需要有人看守,因此,如果利用同一台或同一组机器从事这两项工作的话,在时间上是不会冲突的。

(3)采用数据相关性少的连续帧生成算法。在生成连续帧时,数字电影制作软件会有许多种不同的算法。由于电影图像相邻的帧在画面上比较相似,因此有可

\* 本课题得到“九五”国家重点科技攻关项目资助,毛希平 博士后,研究方向为图像的处理和并行计算,温冬婵 教授,研究方向为计算机体系结构和并行处理,董春雷 博士研究生。

能采用下一帧图像使用上一帧数据的相关性算法,这会使机群系统由于机器间频繁的数据交换而降低效率。

我们使用的相关制作软件在连续帧生成方面均采用“脚本+关键帧”的算法,即任意一帧图像的生成只会用到关键帧及记录制作过程中所用操作的脚本中的数据,而与其它帧无关,这就为机群替代服务器创造了条件。

(4)机群系统的工作效率与服务器接近。以O2工作站为例,它的R10000 CPU与Origin2000的CPU在性能上并无差别,不同的只是后者有8个CPU。在进行多机分布计算时,只要网络的数据传输速度足够快,几十台O2的计算能力远远高于8CPU的服务器。

(5)有相应的软件支持。多机分布计算与单机操作有很大差别,为此我们开发了专门的软件来进行系统管理、任务分配、负载均衡、进程监视等系统维护工作。

## 2 基于工作站机群的计算处理系统的特点

采用工作站机群技术的计算处理系统有许多显著的特点:

(1)性能价格比高 以20台O2组成的机群系统为例,每台O2均采用R10000芯片等较高的系统配置,使其在单CPU性能上不落后于Origin2000;加上ATM快速网络及完善的并行技术,其整体性能也不落后。单台O2工作站的报价在1万美元左右,而仅单台Origin-2000的报价就不低于160万美元,由此可见,该系统的性价比是很高的。

(2)设备利用率高 在计算处理系统中,工作站既要作为美工台单独进行关键帧的处理等工作,还要在最后的渲染阶段进行大规模并行计算,这样就减少了设备的闲置率,使其充分运转,产生最大的价值。如果采用大型服务器,不仅购置、保养需要大笔经费,而且在时间较长的中前期制作阶段,它基本处于闲置状态。

(3)可扩展性好 机群系统中工作站的数目是不受限制的,可以是20台,也可以扩展到40、80、100台。而Origin2000的CPU数目最多为64个,并且随着CPU数目的增多,其价格急剧上升。今后若需要Origin2000这样的服务器,可随时接入系统,无需对系统进行任何改造。

(4)系统具有容错性 机群系统的一个重要特征就是容错性。当某台工作站出现故障不能继续工作时,并不影响其它工作站的正常运行,系统监视程序只需将它的任务转移到其它机器上即可。若服务器的一个CPU发生故障,将很容易导致整个系统的瘫痪,而多CPU服务器的故障排除是极费时间和金钱的。

(5)可动态监控工作进程,进行负载均衡 机群系

统的系统管理程序运行在应用软件及操作系统之上,可进行各级工作站之间的任务分配,并随时监视其执行情况,按任务的特点改变分配策略。而多CPU的服务器办不到这一点,它只能根据固定的策略分配负载,操作人员无法干涉及控制,不能灵活地改变已有的设置。

## 3 基于工作站机群的计算处理系统的组成

计算处理系统的硬件设备主要包括图像处理工作站、高速宽带网络、大容量磁盘阵列等。其中,作为美工台和导演台的工作站数目最多,功能也最复杂,主要任务有前期制作、三维模型及特技效果的制作、后期制作及合成、渲染等;网络用来连接各个处理单元,实现数据传输及通信;磁盘阵列用于存放处理过程中大量的数据文件和中间结果。

计算处理系统的软件主要由并行程序运行环境PVM、并行任务提交系统、负载均衡系统、可视化人机交互集成开发环境、容错处理系统等部分组成,能够提供给用户完善的功能和友好的使用环境<sup>[2]</sup>。

### 3.1 并行程序运行环境 PVM

PVM(Parallel Virtual Machine)是一个能适应大量硬件结构平台和不同数据表示的分布操作环境,主要目的在于开发异构型并行计算。PVM支持应用程序、机器及网络级的异构,与原有操作系统的关系如图1。PVM软件由两部分组成:一部分是称为PVMD的daemon,在机群系统的每台主机上都有一个daemon,由它负责进程之间的控制与通信;另一部分是一组PVM接口库PVMLIB,由用户在应用程序源码中调用以传递消息。



图1 PVM并行运行环境示意图

### 3.2 并行任务提交系统

在并行程序运行环境的统一管理下,机器对任何用户来讲都是一个具有强大计算能力的并行系统。为使并行系统象单机系统一样便于操作,还需要一个能自动进行并行任务分配的系统,它按照一定的负载均衡策略,将并行任务分配到机群中执行,并管理所有的处理数据(帧文件)。

如图2所示,渲染主控可以是导演台,也可以是众多美工台中的一台。渲染主控平时从事自己单机的工作,在需要进行渲染时,首先由它启动PVM环境,使机器进入并行状态;之后将脚本和关键帧及相关参数通过网络传递给各美工台,在各美工台进行计算时,它还要动态地监视其工作情况,随时进行负载均衡及任务调度。

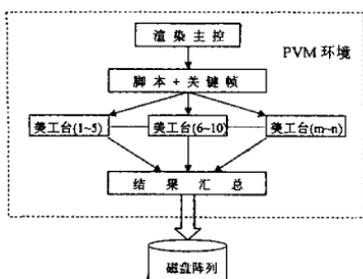


图2 并行任务提交系统示意图

### 3.3 并行任务负载均衡系统

负载均衡是并行计算中的一个重要问题,由于计算任务被分配到各个结点并行执行,若各结点上的任务负载不均衡,就会导致系统性能的显著下降<sup>[1]</sup>。为了充分利用高度并行的系统资源,提高整个系统的吞吐量,缩短总任务的响应时间,需要负载均衡技术的支持。所谓负载均衡,就是各个任务比较均衡地分布在不同的处理结点上,从而使各处理结点的利用率达到最大。

### 3.4 可视化人机交互集成开发环境

要充分发挥工作站机群的优点,建立方便高效的并行程序开发环境也是一项重要工作。所谓可视化的人机交互集成开发环境,实际上是一组图形界面工具,它使程序员在多机环境下能够实时地监控不同机器上同步运行的程序,具体地说,它为应用程序员和最终用户提供编辑、各种语言的编译链接装配、并行环境配置、各结点资源使用情况显示、并行任务的加载运行、状态监视和控制、性能评测和并行调试、负载均衡、容错处理等工具以及联机帮助等功能,目的是使用户不出此环境即能完成所要做的工作。它可用来监视并行程序运行过程、分析比较并行算法、找出性能瓶颈、优化并行算法设计、辅助程序调试、辅助对并行体系结构

的性能评测等,它与实际应用紧密结合,使并行系统实用化。

### 3.5 容错处理系统

随着机群规模的不断扩大,其在计算过程中发生故障的几率会以指数递增,一个结点上的人为操作失误或软硬件故障就会导致本次并行计算的彻底失败,故障前的大量计算结果丢失。

为了避免在系统发生故障时由于从头开始执行而引起的计算上的大量浪费,充分提高机群系统的可用性,容错处理是必不可少的,在系统正常运行的适当时刻设置检查点(checkpointing),保存系统当时的规范状态,并对各进程进行相关性跟踪和记录。当系统发生故障并经修复后,将相关进程回卷(rollback)到故障前的一致状态(检查点),状态恢复后从该检查点处重新执行(replay),而不是从程序开始处执行,因此节省了大量重复计算时间,充分体现了机群系统的并行性能。与此同时,基于checkpoint的后向恢复技术可对系统瞬时故障进行自动恢复,这也是恢复未知故障的唯一手段。

小结 基于上述的应用背景,我们在PVM并行环境下开发了诸如负载均衡、容错处理、多机进程管理、多机并行调试器等配套软件,构造了4台SGI O2+10Mbps Ethernet的数字电影计算处理实验系统,对常用的数字电影处理软件(PowerAnimator, Composer, Studio Paint, Soft Image, Interactive Photorealistic Rendering等)进行了测试。它们都可以在机群系统上运行,并行效率很高。

工作站机群作为当前并行计算的一种主流技术,正日益受到广泛重视,并必将获得进一步发展。它在数字电影计算处理系统中的成功应用说明了,机群系统以其可接受的价格提供显著的计算能力,在图形图像处理、多媒体、科学计算等许多领域都有广阔的应用前景。

### 参考文献

- 1 毛希平,温冬梅,等.高精度连续图像数字处理(数字电影)系统.计算机科学,1999(5)
- 2 Hipper G, Tavangarian D. Advanced workstation cluster architectures for parallel computing. Journal of Systems Architecture, 1997, 44(3/4): 207~226
- 3 Hamdi M, Lee C K. Dynamic load-balancing of image processing applications on clusters of workstations. Parallel Computing, 1997, 22: 1477~1492