



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2013
09
VOL.18

ISSN1006-8961
CN11-3758/TB



城市交通规划 P1197



第18卷第9期（总第209期）

2013年9月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

2013 all rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院

主办单位 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 李小文

编辑出版《中国图象图形学报》编辑出版委员会

邮政信箱 北京9718信箱

邮 编 100101

电子信箱 jig@irsa.ac.cn

电 话 010-64807995 010-82614429

网 址 www.cjig.cn

广告经营许可证 京朝工商广字第0361号

总 发 行 北京报刊发行局

订 购 全国各地邮局

海外发行 中国国际图书贸易集团有限公司

（邮政信箱：北京399信箱 邮编：100048）

印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Institute of Remote Sensing and Digital Earth, CAS
China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief LI Xiaowen

Editor, Publisher Editorial and Publishing Board of Journal of
Image and Graphics

P.O.Box 9718, Beijing, P.R.China

Zip code 100101

E-mail jig@irsa.ac.cn

Telephone 010-64807995 010-82614429

Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China))

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

ISSN 1006-8961

CODEN ZTTXFZ

国外发行代号 M1406

国内邮发代号 82-831

国内定价 50.00元

综述

高分辨率遥感影像空间结构特征建模方法综述

秦昆, 陈一祥, 甘顺子, 冯霞, 任文力

1055

简缩极化SAR数据信息提取与应用

张红, 谢镭, 王超, 张波, 吴樊, 汤益先

1065

图像处理和编码

融合CCCTI码和谱系聚类的仿射配准

张桂梅, 江少波, 刘丕玉, 张松

1074

面向虚拟视点绘制的深度图滤波及上采样方法

高凯, 陈贺新, 赵岩, 耿英楠, 王钢

1085

图像分析和识别

包空间多示例图像自动分类

王科平, 杨艺, 王新良

1093

表情和姿态的双模态情感识别

闫静杰, 郑文明, 辛明海, 邱伟

1101

等级关联的粒子群迭代多目标检测跟踪

程旭, 李拟珺, 张索非, 吴镇扬

1107

改进后的TLD视频目标跟踪方法

周鑫, 钱秋朦, 叶永强, 王从庆

1115

多特征稳健主成分分析的视频运动目标分割

甘超, 王莹, 王向阳

1124

融合全局最好和声搜索算法的模糊C均值聚类图像分割

崔兆华, 高立群, 欧阳海滨, 李文娜

1133

从运动中恢复目标结构的改进因子分解法

邱少华, 文贡坚, 回丙伟, 张鹏

1142

计算机图形学

并行扫描转换结构中的状态管理

殷诚信, 韩俊刚, 黄虎才

1149

医学图像处理

3维颅骨表面模型的复杂孔洞修补

包佳蕊, 梁荣华, 吴福理, 黄鲜萍, 蒋莉, 陈矛

1156

归一化Radon变换和统计测度的超声图像肌束方向自动跟踪

李乔亮, 李振宇, 张会生, 张小飞, 陈昕, 陈思平, 汪天富

1164

遥感图像处理

改进的多光谱双边滤波图像融合

吴川, 杨冬

1170

长度归一化扫描的合成孔径雷达图像船舶尾迹恒虚警率检测

南杰, 张红, 王超, 吴樊, 张波

1176

地理信息技术

3D可视化输电线路中杆塔空间挂点计算

何国辉, 周聪

1184

地图符号服务与地图服务的耦合

焦东来, 张海涛, 闫国年, 姜杰

1190

不同表达粒度对城市路网结构健壮性评价的影响

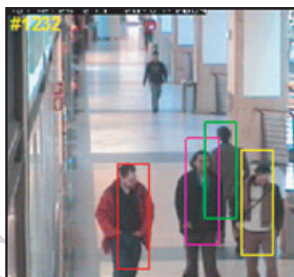
段滢滢, 陆锋

1197

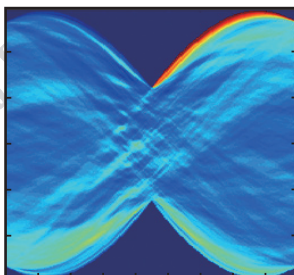
DEM保真精度量化描述模型

张锦明, 游雄

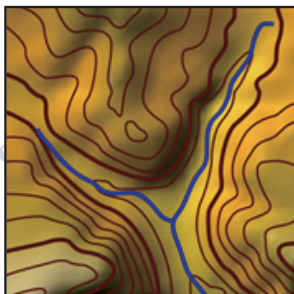
1206



等级关联的粒子群迭代多目标检测跟踪(第1107页)



归一化Radon变换和统计测度的超声图像肌束方向自动跟踪(第1164页)



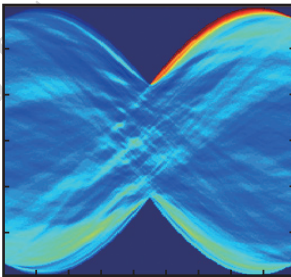
DEM保真精度量化描述模型(第1206页)

CONTENTS

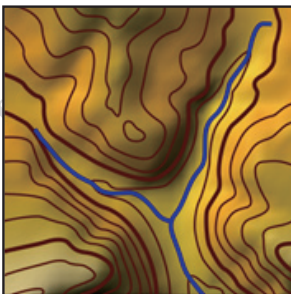
JOURNAL OF IMAGE AND GRAPHICS



Hierarchical associative multi-target detection and tracking with particle swarm optimization(P1107)



Automatic tracking of muscle fascicle orientation in ultrasound images: a approach based on a normalized Radon transform and statistical measure (P1164)



Quantitative descriptive model of DEM fidelity accuracy(P1206)

Review

Review on methods of spatial structural feature modeling of high resolution remote sensing images

Qin Kun, Chen Yixiang, Gan Shunzi, Feng Xia, Ren Wenli 1055

Information extraction and application of compact polarimetric SAR data

Zhang Hong, Xie Lei, Wang Chao, Zhang Bo, Wu Fan, Tang Yixian 1065

Image Processing and Coding

Affine registration based on CCCTI and hierarchical clustering

Zhang Guimei, Jiang Shaobo, Liu Piyu, Zhang Song 1074

Depth map filtering and upsampling method for virtual view rendering

Gao Kai, Chen Hexin, Zhao Yan, Geng Yingnan, Wang Gang 1085

Image Analysis and Recognition

Automatic classification of multiple-instance image based on the bag space

Wang Keping, Yang Yi, Wang Xinliang 1093

Bimodal emotion recognition based on body gesture and facial expression

Yan Jingjie, Zheng Wenming, Xin Minghai, Qiu Wei 1101

Hierarchical associative multi-target detection and tracking with particle swarm optimization

Cheng Xu, Li Nijun, Zhang Suofei, Wu Zhenyang 1107

Improved TLD visual target tracking algorithm

Zhou Xin, Qian Qiumeng, Ye Yongqiang, Wang Congqing 1115

Multi-feature robust principal component analysis for video moving object segmentation

Gan Chao, Wang Ying, Wang Xiangyang 1124

Improved fuzzy C-means clustering combined with the global best harmony search algorithm for image segmentation

Cui Zhaozhuo, Gao Liqun, Ouyang Haibin, Li Wenna 1133

Improved factorization method for structure from motion

Qiu Shaohua, Wen Gongjian, Hui Bingwei, Zhang Peng 1142

Computer Graphics

Efficient state management for parallel rasterization

Yin Chengxin, Han Jungang, Huang Hucai 1149

Medical Image Processing

Various hole repairing algorithm for 3D mesh surface models of human skull

Bao Jiarui, Liang Ronghua, Wu Fuli, Huang Xianping, Jiang Li, Chen Mao 1156

Automatic tracking of muscle fascicle orientation in ultrasound images: a approach based on a normalized Radon transform and statistical measure

Li Qiaoliang, Li Zhenyu, Zhang Huisheng, Zhang Xiaofei, Chen Xin, Chen Siping, Wang Tianfu 1164

Remote Sensing Image Processing

Improved multispectral bilateral filter video fusion algorithm

Wu Chuan, Yang Dong 1170

Ship wake constant false alarm rate detection algorithm in synthetic aperture radar images based on length normalized scanning

Nan Jie, Zhang Hong, Wang Chao, Wu Fan, Zhang Bo 1176

Geoinformatics

Tower hanging point calculation for transmission line on three-dimensional visualization

He Guohui, Zhou Cong 1184

Coupling of map symbol web service and web map service

Jiao Donglai, Zhang Haitao, Lv Guonian, Jiang Jie 1190

The impact of different granularity representations on robustness evaluation of city road network

Duan Yingying, Lu Feng 1197

Quantitative descriptive model of DEM fidelity accuracy

Zhang Jinming, You Xiong 1206

中图法分类号: TP391.9 文献标识码: A 文章编号: 1006-8961(2013)09-1149-07

论文引用格式: 殷诚信, 韩俊刚, 黄虎才. 并行扫描转换结构中的状态管理[J]. 中国图象图形学报, 2013, 18(9): 1149-1155. [DOI: 10.11834/jig.20130912]

并行扫描转换结构中的状态管理

殷诚信, 韩俊刚, 黄虎才

西安邮电大学计算机学院, 西安 710061

摘要: 扫描转换为图形流水线的核心部分, 许多高性能的图形硬件都是采用并行的扫描转换结构。现代 GPU (graphic processing unit) 用于高性能计算时, 主要依靠其极大的系统吞吐率, 但是用基于传统的状态管理方法维持命令顺序性时会造成处理单元的停顿, 降低系统的吞吐能力。提出一种新的状态管理方法, 将命令分为计算与控制两类, 对于控制类的命令由状态处理单元生成状态掩模 (state-mask), 减少了状态管理中的广播通信, 降低了保持命令顺序性和状态管理的复杂性。仿真结果表明所提出状态管理方法同传统的状态管理方法相比, 在游戏等后端状态变化明显的图形应用中, 扫描引擎利用率提高了 5%~9%。

关键词: 状态管理及命令顺序性; 扫描转换; 图形处理器; 并行处理

Efficient state management for parallel rasterization

Yin Chengxin, Han Jungang, Huang Hucai

School of Computer Science and Technology, Xi'an University of Posts & Telecommunications, Xi'an 710061, China

Abstract: Scan-conversion is the essential part of the graphic pipeline. It takes full advantage of data parallel processing. High performance computing mainly relies on large throughput, but the conventional state management method in the graphics pipeline reduces the throughput because of processing unit stalling. An efficient state management method is proposed in this paper. In order to maintain command ordering, the proposed method divides the commands into two classes: compute command and control command. The method reduces the complexity of maintaining ordering and state management compared with conventional method, and the state-mask generated by state processing unit decreases the count of the broadcast communication and reduces the bandwidth needed for state management. Experimental results show that for applications such as computer games, in which the backend states change frequently, the rate of raster engine usage is improved by 5%~9%.

Key words: state management and ordering; scan-convention; GPU (graphic processing unit); parallel processing

0 引言

并行处理是图形处理的基础, 这是因为图元、顶点、片元和像素存在极大的独立性, 这些实体的集合可以并行处理^[1]。三角形的扫描转换是图形流水

线的核心部分^[2], 也是现代 GPU (graphic processing unit) 中固定功能单元之一^[3], 许多高性能的图形硬件都是采用并行的扫描转换结构^[4-6]。然而某些图形 application programming interface (API) (例如 OpenGL^[7]) 要求图形命令语义上的顺序性, 支持此类图形 API 的图形硬件也必须实现相应的状态管理

收稿日期: 2012-08-19; 修回日期: 2013-01-14

基金项目: 国家自然科学基金重点项目 (61136002)

第一作者简介: 殷诚信 (1985—), 男, 2012 年于西安邮电大学获计算机应用技术专业硕士学位, 主要研究方向为计算机图形学, 图形硬件结构设计。E-mail: yincx_xupt@126.com

机制和维持命令的语义顺序性。

现代 GPU 发展高性能计算时,主要依靠其极大的系统吞吐率^[8]。但是传统状态管理方法在维持命令顺序性时会造成处理单元的停顿,从而降低了系统的吞吐能力。另外,传统状态管理方法需要广播通信,这会限制图形处理结构的可裁剪性^[9]。

为了实现并行扫描转换结构下的状态管理,将命令按照计算与控制分成两类,从而降低了保持命令顺序性和状态管理的复杂性。对于直接传递的控制类命令则直接存入控制命令 first in first out (FIFO) 中,并更新状态掩膜 (state-mask) 中 upstate-mask,能够避免并行扫描转换结构由于状态管理引起的停顿,提高并行扫描转换处理单元的利用率,从而提高并行扫描转换结构的吞吐率;对于自身相关的控制类命令由状态处理单元生成 state-mask,减少状态管理中的广播通信和降低状态管理所需的带宽需求。

1 相关工作

状态管理是并行图形结构的一个关键特性。早在 1987 年, Torborg^[10] 采用并行硬件加速几何处理时就实现了该结构中的状态管理,他将命令分成 6 类分别做不同的处理,依靠广播通信来实现状态更新的命令同步。Cox^[11] 在其博士论文中为并行实现 Pixar's RenderMan 提出了一种优化的状态管理算法,由于在 RenderMan 中计算状态元素的不确定性是等同的,因此该方法需要发送较多的数据。Ellsworth 等人^[12] 描述了一个系统中仅仅需要发送相关状态元素给渲染节点,但 Cox 指出 Ellsworth 的系统仅限于支持保留模式下的图形应用。Reality-Engine^[13] 将命令分为不经常改变的状态命令和经常改变的命令,对于不经常改变的命令通过广播的方式来更新,而对于经常改变的命令则附属在几何图元上,避免不必要广播通信。Eldridge 等人^[14] 提出 Pomegranate 结构时利用硬件 barrier 命令和 nextR 命令来实现命令顺序性,通过广播通信来管理状态,但是硬件同步命令会导致流水线的停顿降低图形处理器的性能。Antochi 等人^[15] 针对 tile 结构的扫描转换,提出了按照静态状态信息和动态状态信息分类管理,为了提高并行度还提出延迟执行。

除了并行图形硬件结构需要处理状态管理,一

些并行图形系统也提出了一些状态管理方法。例如, Buck 等人^[16] 在网络并行图形系统 WireGL 中提出了 state tracking 和 lazy state update 方法,这两种方法在后来针对集群的并行图形系统 Chromium^[17] 中和分布图形系统 AnyGL^[18] 中都有沿用。为了得到更高的可扩展性能, AnyGL 采用基于逻辑时间戳的 state tracking。

2 并行扫描转换结构及 state-mask

本文对传统的并行扫描转换结构进行适当改进,增加了状态命令处理模块并采用集中式控制命令 FIFO。并行扫描转换整体结构主要由命令接口、图元调度、状态命令处理和扫描引擎组成(图 1)。命令接口模块处理输入命令和输出命令;图元调度模块完成扫描转换的初始化 (polygon setup)、图元覆盖区域的任务划分 (polygon overlap), 以及任务的调度与图元属性的更新;状态命令处理模块则是处理状态命令;扫描引擎完成扫描转换中内点测试 (point-in-polygon tests) 和属性插值。

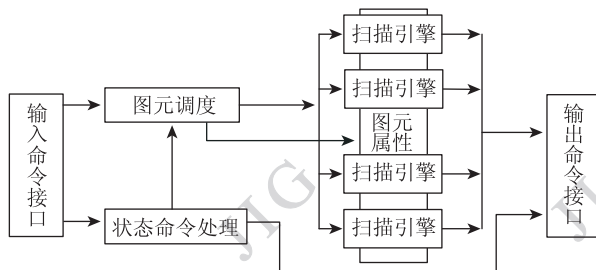


图 1 并行扫描转换结构

Fig. 1 Parallel rasterization architecture

控制命令 FIFO 用于保存后续单元所需的命令,文献[10]中控制命令 FIFO 分布于所有的扫描引擎中,这样一来就存在冗余的状态命令信息。本文中控制命令 FIFO 仅存在于状态命令处理单元中,可以减少冗余状态命令带来的额外存储。同时,仅需要维护一个控制命令 FIFO,使状态管理和维持命令顺序性更加简洁。

扫描转换的状态涉及当前模块状态更新的使能与禁止 (enable/disable)、插值模式 (interpolation model)、激活的纹理单元数目 (active texture count)、光滑多边形 (polygon-smooth)、多边形点画模式 (polygon-stipple), 以及后续模块是否更新 (update state)、更新状态命令列表的索引号 (updatestate id)

和状态命令列表中命令数目 (state list length)。将这些状态组织成 state-mask (图 2)。

selfstate-mask	enable/disable (3 bit)
	interpolation model (1 bit)
	active text count (3 bit)
	stipple address (8 bit)
upstate-mask	update state (1 bit)
	update state id (8 bit)
	state list length (8 bit)

图 2 state-mask 格式

Fig. 2 The format of state-mask

3 命令顺序性和状态管理

3.1 命令的分类和处理

现代图形处理器是基于图形流水线实现的, 图形命令对某一级单元来说或者引入新的计算, 或者改变其状态。为此, 将图形命令分为计算和控制两大类。

在扫描转换中, 计算命令是指图元命令 $Draw$, 控制命令是指除计算类外所有流经扫描转换的命令。从对扫描转换的状态影响作用角度, 可将控制命令再分为消费命令 $State_{consuming}$ 和传递命令 $State_{passdown}$ 。消费命令是指改变扫描转换状态的命令, 传递命令是指消费命令外所有的控制命令。本文的状态管理与命令顺序性的维护集中在 $State_{consuming}$ 和 $State_{passdown}$, 称之为消费-传递 (CP) 方法。

计算命令和控制命令的解析是由输入命令接口完成。若输入命令是 $Draw$, 则交由图元调度单元处理; 否则交由状态命令处理单元处理。状态命令处理单元将 $State_{consuming}$ 生成 state-mask 中 selfstate-mask, 而 $State_{passdown}$ 则直接存入控制命令 FIFO 中, 并更新 state-mask 中 upstate-mask 部分。对于全局命令 ($glFlush$, $glFinish$, $SwapBuffer$) 将通过内部总线广播给所有扫描引擎。

3.2 CP 状态更新策略

对于图形应用程序 P , 采用命令流的形式化描述为

$$P = Draw_0 State_1 \cdots Draw_i State_j \cdots State_{m-1} Draw_{n-1}$$

对于每一个命令 $Draw_i$, 由调度单元以 Tile 的形式分配到并行扫描转换引擎并由扫描引擎完成后续扫描转换的各项工作。

状态更新涉及 $State_{passdown}$ 、 $State_{consuming}$ 的处理和维持命令流的顺序性。状态更新主要由状态处理单元和图元调度单元完成。根据 $State_{consuming}$ 命令由状态命令处理单元生成 state-mask 中 selfstate-mask。state-mask 作为图元测试命令的组成部分由调度单元分配给扫描引擎处理。同时通过 state-mask, 将 $State_{consuming}$ 转换为计算类命令处理, 这样避免了状态更新时处理自身状态更新的过程。

状态更新另外一项工作就是维持命令的顺序性。命令的顺序性主要是保证处理的结果同用户输入命令流的结果一致。传统的处理方式是等待所有的 $Draw$ 全部处理完毕, 再将 $State_{passdown}$ 传送给后续模块 (图 3)。

通过上述分析, 可以发现扫描引擎的闲置完全是由于作用于后续模块的 $State_{passdown}$ 造成的。为此, 针对 $State_{passdown}$ 的处理做了一些改进 (图 4)。

改进后, 当 $State_{passdown}$ 进入扫描转换模块中, 输入命令接口将其交由状态命令处理单元处理。此时状态命令处理单元不是等待扫描引擎全部完成当前任务后将 $State_{passdown}$ 发送给后续模块, 而是将其存入控制命令 FIFO 中并更新 upstate-mask。后续的 $Draw$ 可以继续进入扫描转换模块由图元调度单元处理, 生成图元测试任务命令。扫描引擎输出生成片元时, 会将 upstate-mask 复制到片元结构中。这些片元在执行片元染色过程前, 线程管理处理器会根据 upstate-mask 读取状态命令 FIFO 中命令形成新的状态变量, 并根据片元染色程序 (fragment shader) 编译的结果生成片元染色线程。

新的状态变量可以存储在局部共享存储内, 线程处理器可以利用重排序机制^[4]将 upstate-id 相同的片元组织成线程数组 (thread array)^[19]或者线程组 (thread group)^[20], 以此共享新的状态变量; 也可以将新的状态变量保存在统一变量缓存中 (uniform cache)^[21], 片元染色线程通过 upstate-id 来读取。同样, raster operations (ROP) 阶段同样可以利用这种分离的方式来管理状态命令, 这样可以将 $State_{passdown}$ 推迟到需要处理 $State_{passdown}$ 的模块中再进行处理。



图 3 传统的状态处理过程
Fig. 3 The conventional method scheme

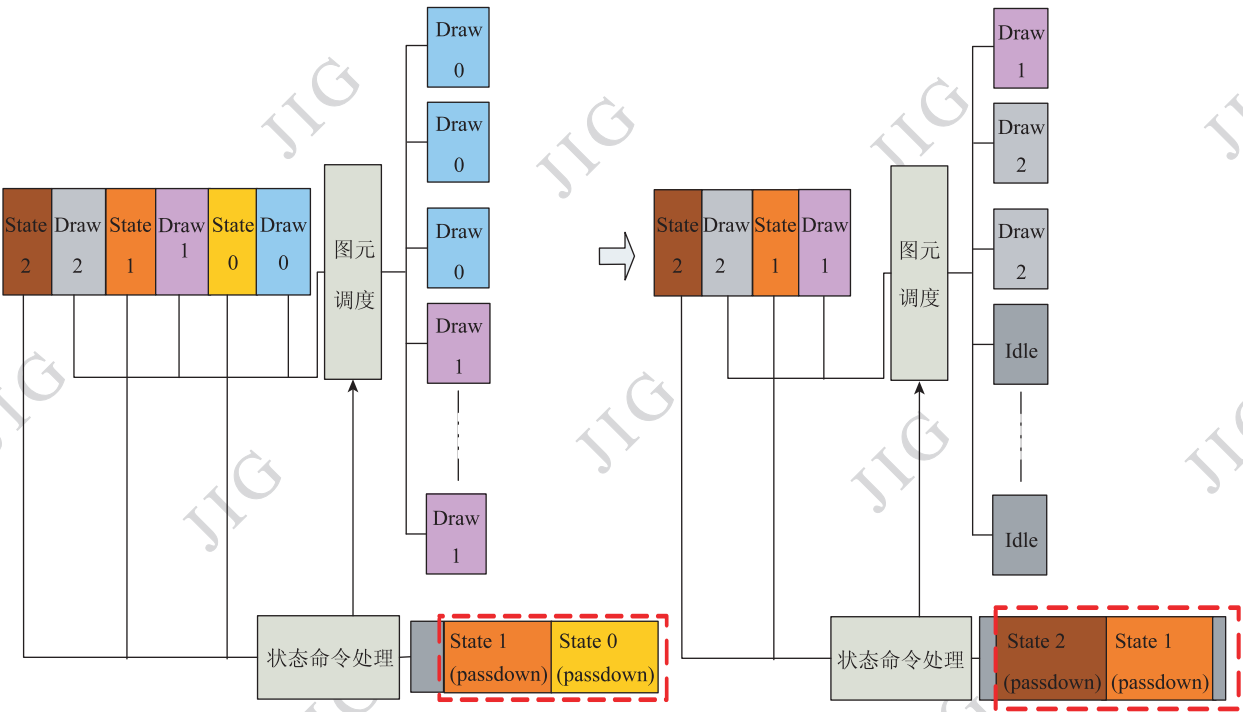


图 4 CP 状态处理过程
Fig. 4 The CP method scheme

4 理论分析与实验

4.1 性能提升的理论分析

为了便于理论上的分析, 首先标识一些变量。记 N_{SIMD} 为扫描转换引擎数目, T_{fragment} 为单个片元的扫描转换时间, T_{tile} 为单个 Tile 的扫描转换时间, T_b 为所有扫描引擎单次广播处理时间, U_{SIMD} 为扫描转换的利用率, TP_{SIMD} 为扫描转换的系统吞吐率。 $F(n)$ 为 P 的总片元数目, $f(n)$ 为 P 实际占用的扫描引擎数。

论点 1 提高 U_{SIMD} 可以提高 TP_{SIMD}

假设完成 P 的扫描转任务的时间为 T_{total} , 则有

$$TP_{\text{SIMD}} = \frac{n}{T_{\text{total}}}, U_{\text{SIMD}} = \frac{f(n)}{T_{\text{total}} N_{\text{SIMD}}}. \text{ 故 } TP_{\text{SIMD}} = \frac{n U_{\text{SIMD}} N_{\text{SIMD}}}{f(n)}, \text{ 易知当 } U_{\text{SIMD}} \uparrow, TP_{\text{SIMD}} \uparrow.$$

论点 2 CP 状态更新策略提升利用率与状态命令数成正比, 即 $U_{\text{SIMD}} \propto m$ 。

假设 $Draw_i$ 的 Tile 数目为 k_i 。 $N_{\text{SIMD}} + r_i$, 分别就 k_i 和 r_i 的 3 种情况下的传统状态管理方法进行分析。

Case1: $k_i \neq 0, r_i = 0$ 。此时, 可以认为 $Draw_i$ 恰好能使并行扫描转换引擎同时完成, 同时结束。处理 $State_j$ 需要一次广播通信, 所有扫描引擎需要停顿 T_b 。

Case2: $k_i = 0, r_i \neq 0$ 。此时, $State_j$ 需要等待 r_i 个引擎全部完成扫描转换才能将 $State_j$ 广播给所有扫描转换引擎。故至少有 $N_{\text{SIMD}} - r_i$ 个并行扫描引擎闲置, 至少等待 T_{fragment} , 至多等待 T_{tile} 。

Case3: $k_i \neq 0, r_i \neq 0$ 。此时, 同 Case2, 至少有 $N_{\text{SIMD}} - r_i$ 个并行扫描引擎闲置, 至少等待 T_{fragment} , 至多等待 T_{tile} 。

同传统方法相比, CP 状态管理方法在性能上的提高主要来源两方面: 一是将处理 $State_j$ 前 $N_{\text{SIMD}} - r_i$ 个并行扫描引擎利用起来, 二是可以减少 $State_j$ 中的广播通信时间。

假设 N_{Δ} 为 CP 方法增加的扫描引擎数目, α 为 N_{Δ} 利用率, β 为全局命令的比例, T'_{total} 为 CP 方法时完成 P 扫描转任务的时间。则有

$$T'_{\text{total}} = \frac{F(n)}{T_{\text{total}} N_{\text{SIMD}} U_{\text{SIMD}} + \alpha N_{\Delta}} - (1 - \beta) m T_b = \frac{F(n)}{T_{\text{total}} N_{\text{SIMD}} U_{\text{SIMD}} + \alpha \sum_m (N_{\text{SIMD}} - r_i)} - (1 - \beta) m T_b,$$

易知 $m \uparrow, T'_{\text{total}} \downarrow$, 而 $U_{\text{SIMD}} \uparrow$, 故 $U_{\text{SIMD}} \propto m$ 。其中 α

由图元调度算法决定^[22], 为了严谨地验证 CP 方法带来的 U_{SIMD} 的提升, 实验时 CP 方法与传统方法采用了同样的图元调度算法。

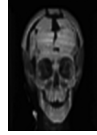
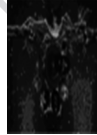
4.2 实验数据分析

为了分析图形应用对流水线中模块的影响, 定义了两个变量: 顺序状态命令比例 PPD (percentage of ordering pass-down command) 和有效的顺序状态命令比例 EPPD (efficient percentage of ordering pass-down command)。PPD 为 $State_{\text{passdown}}$ 总数与图形应用命令流总数比值, EPPD 为 $State_{\text{passdown}}$ 减去初始化后端的命令后与图形应用命令流总数的比值, EPPD 反映了 P 中 m 的比重。

以表 1 中的图形应用 (graphics application) 作为测试数据, 并且给出这些图形应用的 PDD 值和 EPPD 值。

表 1 实验数据

Table 1 The experimental data of the CP method

	Volume	Maiden	Fire	Quake 3
				
PPD/%	1.1	0.05	2.37	31.5
EPPD/%	0.000 1	0.002 2	2.23	23.6

4.3 实验流程及实验结果

为了分析并行扫描引擎的性能, 利用时钟级图形处理器性能仿真平台 XSP-3D^[23], 实现传统的状态管理方法和本文方法。通过 gltrace 记录图形应用命令流, 并生成命令文件, 然后利用这些命令文件作为仿真平台的输入, 最后在仿真时记录了各状态管理模型下的利用率 (图 5)。

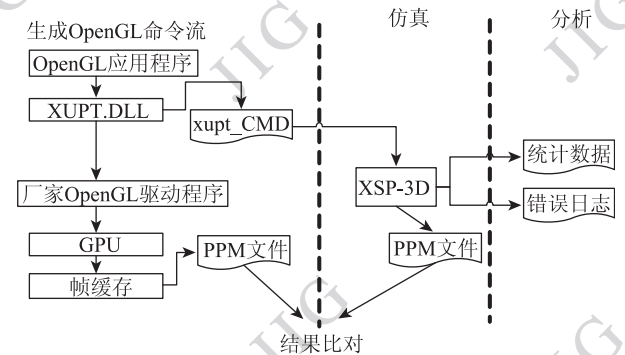


图 5 实验流程

Fig. 5 The framework of function and performance experiment

图6给出了4个图形应用分别在扫描引擎为2、4、8、16、32、64、128情况下,传统状态管理模型和提出的状态管理模型下的利用率。可以发现虽然Volume的PDD较Maiden大,但是由于Volume的EPPD很低造成本文模型对Volume中扫描引擎的利用率提升反而不如Maiden。对于EPPD较大的Quake 3而言,本文模型对扫描引擎的利用率提高了9%。

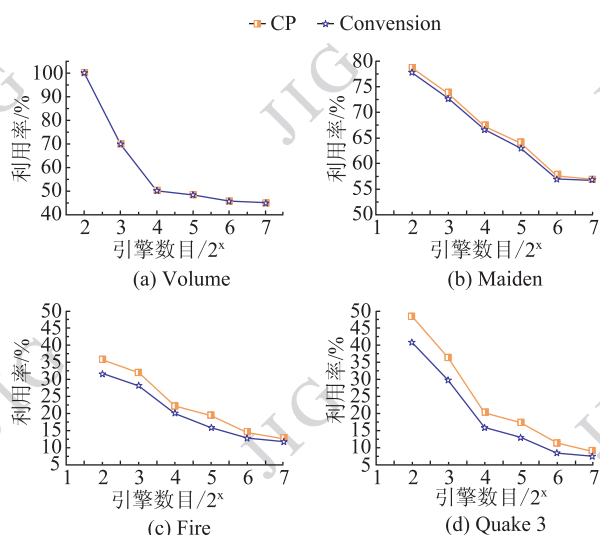


图6 扫描引擎利用率比较

Fig. 6 The rates of raster engine usage with different methods

仿真平台在记录扫描引擎的利用率的同时还记录了各图形应用的加速比(图7显示了本文模型下各图形应用的加速比),分析各图形应用的加速比曲线,可以发现在扫描引擎数目为64时达到最大的加速比(9~11倍)。

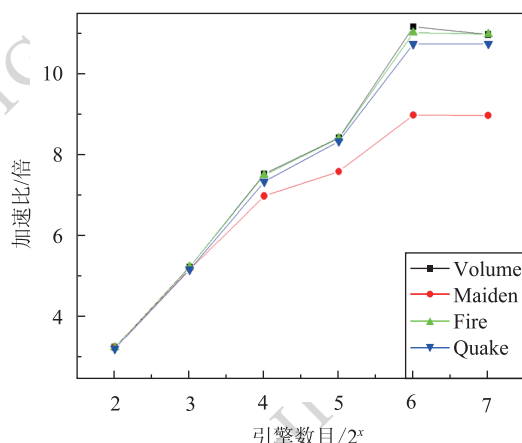


图7 并行扫描转换结构加速比

Fig. 7 The speedups of raster engine with the CP method

5 结论

并行扫描转换是现代图形处理器的一个主要组成部分。传统的状态管理方法会造成并行扫描引擎的停顿,降低图形处理器的性能。本文针对并行扫描转换结构提出的状态管理方法能有效避免由于状态管理造成扫描引擎停顿,提高并行扫描结构的性能。另外,本文方法采用 state-mask 能减少状态更新所需的广播通信,降低通信带宽,能够明显提高EPPD值较大的图形应用中扫描引擎的效率,但对于EPPD值较小的图形应用性能的提高不明显,并且目前还不能完全脱离广播通信。这也是将来研究工作的内容。

参考文献 (References)

- [1] Blythe D. Rise of the graphics processor [C]//Proceedings of the IEEE. Raleigh, NC: IEEE Xplore, 2008, 96: 761-778.
- [2] Beaudoin P, Poulin P. Compressed multisampling for efficient hardware edge anti-aliasing [C]//Proceedings of Graphics Interface 2004. Waterloo, Canadian: Canadian Human-Computer Communications Society, 2004: 169-176.
- [3] Fatahalian K, Houston M. A closer look at GPUs [J]. Communications of the ACM, 2008, 51(10): 50-57.
- [4] Fuchs H, Poulton J, Eyles J, et al. Pixel-Planes 5: a heterogeneous multiprocessor graphics system using processor-enhanced memories [C]//Proceedings of SIGGRAPH '89. New York, USA: ACM Press, 1989: 79-88.
- [5] Houston M. Anatomy of AMD's TeraScale graphics engine [EB/OL]. (2008-08-14) [2012-12-26]. <http://s08.idav.ucdavis.edu>.
- [6] NVIDIA Corporation. NVIDIA GF100 Whitepaper [EB/OL]. [2012-12-26]. http://www.nvidia.com/object/IO_89569.html.
- [7] Segal M, Akeley K. Version 4.2 (Core Profile) The OpenGL graphics system: a specification [S]. Oregon, USA: The Khronos Group Inc, 2012.
- [8] Keckler S W. GPU computing and the road to extreme-scale parallel systems [C]//The 2011 IEEE International Symposium on Workload Characterization. Washington, USA: IEEE Computer Society, 2011: 1-1.
- [9] Eldridge M. Designing graphics architectures around scalability and communication [D]. Stanford, USA: Stanford University, 2001.
- [10] Torborg J. A parallel processor architecture for graphics arithmetic operations [C]//Proceedings of SIGGRAPH '87. New York, USA: ACM Press, 1987: 197-204.
- [11] Cox M. Algorithms for parallel rendering [D]. Princeton, USA: Princeton University, 1995.

- [12] Ellsworth D, Good H, Tebbs B. Distributing display lists on a multicomputer [C]//The ACM Symposium on Interactive 3D Graphics. New York, USA: ACM Press, 1990;147-154.
- [13] Akeley, K. Realityengine graphics [C]//Proceedings of SIGGRAPH'93. New York, USA: ACM Press, 1993;109-116.
- [14] Eldridge M, Igehy H, Hanrahan P. Pomegranate: a fully scalable graphics architecture [C]//Proceedings of SIGGRAPH'2000. New York, USA: ACM Press/Addison-Wesley Publishing Co, 2000;443-454.
- [15] Antochi I, Juurlink B, Stamatis, et al. Efficient state management for tile-based 3D graphics architectures [C]//Proceedings of the 15th Annual Workshop on Circuits, Systems and Signal Processing. Veldhoven, Netherlands: Technology Foundation, STW, 2004;336-340.
- [16] Buck I, Humphreys G, Hanrahan P. Tracking graphics state for networked rendering [C]//Proceedings of the ACM SIGGRAPH/Eurographics Workshop on Graphics Hardware. New York, USA: ACM Press, 2000;87-95.
- [17] Humphreys G, Houston M, Ng R, et al. Chromium: a stream-processing framework for interactive rendering on clusters [J]. ACM Transactions on Graphics, 2002, 21(3): 693-702.
- [18] Yang J, Shi J Y, Jin Z F, et al. Design and implementation of a large-scale hybrid distributed graphics system [C]//Proceedings of the 4th Eurographics Workshop on Parallel Graphics and Visualization. Aire-la-Ville, Switzerland: Eurographics Association, 2002;39-49.
- [19] Tumeo A, Milano P. Massively parallel computing with CUDA [EB/OL]. (2009-03-04) [2012-12-26]. http://www.ogf.org/OGF/materials/1605/CUDA_Programming.pdf
- [20] Thibieroz N, Cebenoyan C. DirectCompute performance on DX11 hardware [EB/OL]. (2010-03-9) [2012-12-26]. http://developer.download.nvidia.com/presentations/2010/gdc/DirectCompute_Performance.pdf
- [21] NVIDIA Corporation. Next generation GPU architecture——Fermi [EB/OL] [2012-12-26] http://www.nvidia.com/object/fermi_architecture.html
- [22] Han J G, Jiang L, Du H M, et al. Hardware accelerator and 3D pixel shader architecture for computer graphics [J]. Journal of Computer-Aided Design & Computer Graphics, 2010, 22(03): 363-372. [韩俊刚, 蒋林, 杜慧敏, 等. 一种图形加速器和着色器的体系结构 [J]. 计算机辅助设计与图形学学报, 2010, 22(03): 363-372.]
- [23] Yin C X, Han J G. Design and implementation of a fragment generator in cycle level GPU simulator [C] //Processings of CAD/CG' 2010. Taiyuan: Tsinghua University Press, 2010; 529-536. [殷诚信, 韩俊刚. 时钟级图形处理器仿真平台中片元生成器设计与实现 [C]//全国第十六届计算机辅助设计与图形学学术会议论文集. 太原: 清华大学出版社, 2010; 529-536.]