

· 计算机工程与应用 ·
特约论文

基于HDR的延迟光照技术

陈雷霆, 蔡洪斌, 王国锦, 何明耘

(电子科技大学计算机科学与工程学院 成都 610054)

【摘要】在研究光照模型的基础上,重点分析了延迟光照技术和高动态范围(HDR)技术,并研究了基于高动态范围技术的延迟光照技术,在延迟光照的后处理阶段引入高动态范围技术,发挥两者的优点,基于GPU实现了加速绘制。实验证明,使用基于高动态范围技术的延迟光照技术,表现的场景更加真实,细节表现更加丰富。在延迟光照中,使用高动态范围技术,虽然降低了绘制速度,但仍然达到了25FP/s。

关键词 延迟光照; 高动态范围; 色调映射;
中图分类号 TP391 文献标识码 A

HDR-Based Deferred Lighting Technology

CHEN Lei-ting, CAI Hong-bin, WANG Guo-jin, HE Ming-yun

(School of Computer Science and Engineering, University of Electronic Science and Technology of China Chengdu 610054)

Abstract With the developing of the compute game industry and virtual reality technology, people need realistic effect in 3D scene in the runtime interactive system, such as game software, simulation of fly and simulation of battle. To improve the realistic effect in complex 3D scene, we study geometry scene, material and illumination. This paper focuses on illumination model. Deferred lighting and HDR technologies are analyzed, and a combination approach is proposed. Experimental results show that HDR-based deferred lighting technology can help to represent more detailed scenes. Even though this approach slows down the rendering speed, it is still above 25FP/s.

Key words deferred lighting; high dynamic range; tone mapping

1 光照模型及其绘制方程

在三维复杂场景中,表现图形真实感的绘制技术主要涉及几何场景、物体材质和光照模型三个方面。光照模型描述的是光源与被绘制模型表面的颜色的关系,可以分为局部光照模型和全局光照模型。

文献[1]提出了绘制方程(Rendering Equation),该绘制方程根据光从一点照射到另一点的亮度,以及从其他所有照射到第一点并被反射到第二点的光亮度,表示光从一点到另一点的转换。文献[1]将此绘制方程表示为:

$$I(x, x') = g(x, x') [\epsilon(x, x') + \int_s \rho(x, x', x'') I(x', x'') dx''] \quad (1)$$

式中 x 、 x' 和 x'' 是环境中的点; $I(x, x')$ 是从 x' 到 x 的亮度; $g(x, x')$ 是几何形状项指由 x' 发射到 x 的光的亮度,当 x 与 x' 相互不可见时为0,可见时为 $1/r^2$ (r 是 x 与 x' 之

间的距离)。方程对所有面 S 上的所有点进行积分, $\rho(x, x', x'')$ 是从 x'' 反射并经由 x' 发射到 x 的光的亮度(包括镜面反射和漫反射)。绘制方程表明从 x' 到达 x 的光,包括了它自身发射的和它所反射的从其他表面来的光,这些自身发光递归地反射从其他表面来的光。光照模型的基本思想就是将所有可能路径加起来得到给定方向的辐射率,如果希望得到更高级的真实感,那么可以增加更多的路径来计算。全局光照的研究重点在于,如何有效地计算出各种不同光线路径的效果。有辐射度算法和光线跟踪算法两种基本的算法:(1)辐射度算法^[2-3]是基于热能工程模型对热辐射的发散和反射进行描绘,为物体间的多重反射提供更加精确的处理方法,避免了直接处理漫反射光线;(2)光线跟踪算法^[4-5]使用光线判断各种元素的可见性。

收稿日期:2007-05-15

基金项目:国家863计划目标导向资助项目(2006AA01Z335)

作者简介:陈雷霆(1966-),男,博士,教授,博士生导师,主要从事计算机图形学、数字媒体技术及虚拟现实技术方面的研究。

2 延迟光照技术

为了得到更真实的场景,在计算机游戏等应用中常运用在多个物体上使用多个光源的技术,并要求能实时执行光照计算。实时光照计算有三种方法:(1) 在一个绘制管道中使用多个光源;(2) 在多个绘制管道中使用多个光源;(3) 延迟光照。

1) 在一个绘制管道中使用多个光源可以用一个物体一个绘制管道来实现,即在一个绘制管道中计算所有光源对一个物体的影响。该方法简单,但是存在如下问题:(1) 隐藏面将引起多余的绘制;(2) 使用可编程管线时,需要在在一个绘制程序中管理多个光源;(3) 很难整合阴影算法。

2) 在多个绘制管道中使用多个光源,对所有物体都计算每个光源的影响。由于是在每个物体上对所有的光源进行光照计算,很可能引起隐藏面的着色计算,造成多余的计算,如顶点转换、顶点组合成三角形和各异向性过滤等可能被重复计算。因为对每个光源和对每个物体的排序是相互对立的,所以很难用于成组操作。

一个管道多个光源和多个管道多个光源是两种最常用的实时照亮虚拟场景的技术,但存在一个严重问题,即最坏计算复杂度 O 是物体的数量乘以光源的数量(物体数 \times 光源数)。由于方法本身的限制,必须重复进行一些主要的绘制计算。另外,用这两种技术进行复杂的着色需要更多的绘制管道产生最后的像素颜色。

延迟光照技术在传统光照计算中融入图形光栅化后的图像处理技术。传统的光照计算主要集中在对物体顶点集合的光照处理上,在获取了各顶点的光照计算结果之后,通过线性插值计算出面上各个像素点的光照结果。顶点光照计算比较直接,但存在两个明显的问题:(1) 必须针对场景中所有顶点实施光照计算,或者采用BSP等预处理可见性判别方式获取空间中的可见顶点集合;(2) 中间像素颜色需要通过插值的方法获取,影响了绘制效果细节的表达。延迟光照技术将图形的光照计算转移到图形光栅化处理之后,与传统的着色技术不同的是,传统的着色技术提交几何图元,立即运用着色等技术转移到光栅化图元上。而延迟光照技术只需提交场景的几何图元一次,在几何缓存G-Buffer(Geometry Buffer)中按照1:1的比率存取每个像素的属性,所以每个像素是各自独立着色的。因此,比起传统的着色技术,延迟光照技术的最大优点是最坏计算复杂度为 O 。

3) 延迟光照计算有几何阶段、光照阶段、后期

处理阶段和最后阶段四个阶段。

(1) 几何阶段。几何阶段是唯一实际使用3D物体几何数据的阶段,这一阶段的输入是场景需要被绘制的物体,将绘制所有像素所需要的信息写到G-Buffer中。

为了把所需的信息填充到G-Buffer中,需要将G-Buffer设置为当前的绘制目标。一旦G-Buffer中装有数据,场景将被绘制。首先,使用固定功能管线把原始的几何数据送到显存中(只需要更新深度缓存)。其次,场景的材质和几何信息被送到显存和片元shader中,就在G-Buffer中填充了其余的数据。最后,解除G-Buffer的绑定。

(2) 光照阶段。经过几何阶段的材质管道处理后,会有4个纹理图(G-Buffer内容):法向量纹理图、漫反射纹理图、镜面反射纹理图和深度纹理图,它们及其他与光源有关的信息(如位置、颜色和半径)作为光照阶段的输入。

光照阶段将把所有的管道绘制到P-Buffer0(Pixel Buffer),并且绑定P-Buffer0用于绘制。以环境光管道开始,用漫反射纹理调节环境光部分。对于每个光源,计算屏幕空间的包围裁剪矩形,并添加光的效果。最重要的是,光照阶段是在图像空间与2D正交投影相机一起工作。在这一阶段,已经体现出了延迟光照技术的主要优点,完全把光照和几何体分开了。接下来,重要的是怎样计算与每个光照有关的几何图元。对于动态光源和动态几何体,这一计算十分复杂,并且要求测试在光体中的所有动态几何体。

(3) 后期处理阶段。因为前两个阶段产生的图像作为一张纹理图存储起来,可以在显示前做一些后期处理增强图像效果。延迟光照技术的一个缺点是,当直接绘制到主帧缓存时不支持自动的硬件反走样技术。在这一阶段,需要加入反走样和另外一种很好的后期处理效果Bloom。Bloom近年来被大量使用在一些高端的数字娱乐产品(如游戏和电影特效)中,其方法是让图像中比较亮的点将能量泄漏到四周,使得该点周围本来比较暗的点也变亮,因此出现了比较大的光晕。经过朦胧处理后的图像显得比实际情况更加明亮。这种特效的优点是:计算量很小,只需要对被显示图像作一次过滤,并将过滤后的图像与原始图像进行叠加,使图像显得更明亮;朦胧的图像使得走样行为更不容易被察觉。缺点是会使图像不够清晰。

为了实现反走样过滤,只需在法向量或深度值不连续的像素处,使用一个shader程序来模糊最后的

图像。将边缘检测过滤法用于处理法向量或深度值纹理图,叠加在目标像素周围的八个相邻像素的差,从而得到模糊因子。模糊因子计算公式如下:

$$\text{Factor}_{\text{blur}}(u, v) = \sum_{i=1}^8 (t_{(u+\text{offset}[i].x)(v+\text{offset}[i].y)} - t_{uv})^2 \quad (2)$$

式中 t_{uv} 为位于 (u, v) 点处的目标像素; $\text{offset}[\cdot]$ 为查找表中的偏移量,用于区分像素位置和它周围的像素。

在定义了模糊因子后,可以用模糊因子乘以偏移量对延迟光照产生的图像进行取样。在没有边缘的地方,将没有偏移,当然也没有模糊。

(4) 最后阶段。这一阶段是展示最后图像给用户的阶段。把存储在P-Buffer0中的最后生成的图像作为纹理图送到帧缓存中。在这一阶段不使用任何shader程序,只需固定功能管线就可以实现。

3 基于HDR的延迟光照技术

HVS(Human Visibility System)是人眼的视觉系统,在计算机真实感绘制技术中,HVS是一个重要的概念。所谓真实就是模拟HVS的能力,在自然环境下,HVS能辨别的光强度为 $10^{-2} \sim 10^3$,分为五个级别。而普通显示器只能显示 $0 \sim 10^2$ 的光强度。普通电脑上的图像颜色是由红、绿、蓝三原色组合成的。但每个像素通道只有 $0 \sim 255$ 的灰度,根本无法表现自然界中所有的光线颜色和强度。目前的256级是综合考虑电脑硬件性能、照片级真实图片需要以及电脑二进制方案的结果, $256 \times 256 \times 256$ 带来的1670万色已经可以满足常规需要,也有利于电脑进行处理。

HDR(High Dynamic Range)为表现自然界中所有光线的颜色和强度提供了一个全新的模型。HDR是与低动态范围LDR(Low Dynamic Range)相对的,LDR采用八位纹理格式(24/32位颜色每像素),HDR将每个像素的RGB以及亮度值用实际物理参数或是线性函数表示,参数不再限于整数,可以达到更大的范围和更高的精度。

一幅图像亮度级的最大值和最小值之比被称为动态范围,定义为:

$$\alpha = I_{\text{max}}/I_{\text{min}} \quad (3)$$

通过眼睛瞳孔的自动调节,从明亮的日光到星光,人眼能分辨物体的动态范围可以达到 $10^8:1$ 亮度范围,即使在同一个场景内,不需调节,人眼也能分辨 $10^4:1$ 的亮度范围。然而,常规显示设备能重建的亮度动态范围仅为 $100:1$ 。

HDR主要有两个研究方向:

(1) 图像获取。根据LDR的相机照出的相片,经过处理得到HDR图像^[6-7];或者用专业的HDR相机照出HDR图像。

(2) 色调映射(Tone Mapping)。在基于图像的光照中把高动态范围的光照映射到低动态范围,以便显示器能够显示。

色调映射被用于低动态范围的显示设备。在真实世界中,人类视觉系统通过视觉调节适应高动态范围的光照。因为显示器本身不能产生高动态范围来触发视觉适应,所以要在显示器上输出高动态范围的图像,需要手动实现视觉适应过程。

色调映射算法可以分为两种类型:(1) 空域不变(Spatially Uniform)或者叫做全局动态范围压缩^[8-10]。该类算法对图像进行动态范围变换时,在每个像素上使用同一条变换曲线,变换曲线可以预先指定或者根据图像的内容获取。不足之处在于同一条变换曲线不能自适应图像的不同区域,导致结果图像在细节、颜色、明亮程度上有损失。(2) 空域变化(Spatial Varying)或者叫做局部动态范围压缩^[11-12]。该类算法针对图像的不同区域进行不同的变换。根据HVS的不同模型,各种不同的算法在压缩动态范围的同时,都以保留图像质量的某一方面为标准。

本文使用的色调映射算法的基本思想是,在高动态范围计算场景的平均光照,用平均光照和最大亮度值(White Point)缩放像素值。

1) 计算平均光照 L_w :

在高动态范围中计算平均光照的公式为:

$$\bar{L}_w = \exp\left[\frac{1}{N} \sum_{x,y} \log_2(\delta + L_w(x, y))\right] \quad (4)$$

式中 N 是像素的数量; x, y 是每个像素的2D坐标; δ 是一个很小的常数。

在GPU上实现式(4)并不太容易,因为需要在帧缓存中读入每个像素。如果在一个绘制管道中读入所有像素,即使在PS3.0硬件上也不能实现(在很多情况下,即使两个管道也不一定能实现)。本文采用两个管道实现的方式:(1) 第一个管道创建一个小的目的缓存绘制目标,画一个长方形块覆盖整个绘制目标。在目的缓存中的每个像素上,像素shader程序累加源缓存中的光强值,并且存储计算结果。(2) 第二个管道叠加这些小缓存,然后运行最后部分的计算,产生的纹理中就包含 \bar{L}_w 。PS3.0保证了512条指令,但是总的执行指令数是变化的。当前的硬件允许在一个shader程序中循环执行至多65 535条指令。每个像素执行的指令数决定了有多少管道可以使用。

在读取了RGB值后,用式(5)把RGB值转换成光强值:

$$L_w = 0.27R + 0.67G + 0.06B \quad (5)$$

2) 使用 \bar{L}_w :

通过式(6)计算光强:

$$L_d(x, y) = \frac{L(x, y)[1 + \frac{L(x, y)}{L_{white}^2}]}{1 + L(x, y)} \quad (6)$$

式中 L_d 是被映射的像素的光强值; L_{white} 是最大光强值, 是一个HDR值, 可以引起全白。 L_{white} 能被设成一个无限值, 用于把产生的光强值映射为一个可以显示的值。 $L(x, y)$ 可由下式得到:

$$L(x, y) = \frac{\alpha L_w(x, y)}{\bar{L}_w} \quad (7)$$

式中 α 是一个关键值, 在0~1之间, 通常取0.18。

总之, 先从纹理中读入RGB值, 由式(5)转换成实际光强值 L_w 。然后用 L_w 和式(4)计算出平均光强值 \bar{L}_w 。通过式(7), 计算出 $L(x, y)$ 。最后用式(6)计算出最终的值, 该值即是在显示器上显示的对应的高动态光强的值。

光照阶段绘制光照管线, 在高动态范围求解出光强值, 存入缓存中; 后期处理阶段, 进行色调映射, 把高动态范围的值映射到显示器可以输出的范围。

4 实验结果与分析

本文的实验是先用HDR绘制场景, 再用HDR绘制同一场景, 实验结果如图1所示。从实验结果可知, 虽然使用高动态范围绘制场景的帧率比没有使用高动态范围绘制场景的帧率降低了近一半, 但仍然达到了每秒25帧, 并且效果改善了很多, 背景也变亮了很多。从实验的结果图中可以很明显地看到, 使用高动态范围技术的场景更加细致。可以概括地说, 高动态范围技术的效果是: 亮的地方可以非常亮, 暗的地方可以非常暗, 亮暗部的细节都很明显。

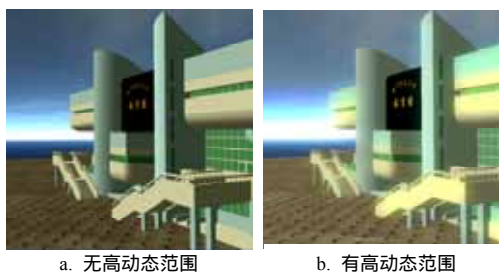


图1 有无HDR实验的结果对比图

5 结束语

本文研究了基于HDR的延迟光照技术, 把延迟光照和高动态范围技术结合起来, 相互取长补短, 从而产生出更好的效果。高动态范围能更好地表示场景细节, 高动态范围的计算, 主要是针对图像的色调映射计算。延迟光照技术是在多光源的情况下进行光照计算, 具有很好的效果和性能, 又由于延迟光照本身的特性是把几何计算和光照计算分开的, 所以, 把HDR应用于延迟光照技术, 可以实现实时多光源光照, 做出HDR光照特效, 使整个场景变亮。

参考文献

- [1] JAMES T K. The rendering equation[C]//Computer Graphics (Proceedings of SIGGRAPH 86). Dallas: ACM Press, 1986.
- [2] MICHAEL F C, JOHN R W. Radiosity and Realistic Image Synthesis[M]. Cambridge: Academic Press Professional, 1993.
- [3] GORAL C M, TORRANCE K E, GREENBERG D P, et al. Modelling the interaction of light between diffuse surfaces[C]//In: Proceedings SIGGRAPH 1984. Minnesota: [s.n.], 1984.
- [4] ANDREW S. Glassner, An introduction to ray tracing[M]. Cambridge: Academic Press Ltd., 1989.
- [5] PAUL S H. A minimal ray tracer in graphics gems IV[M]. Boston: Academic Press Professional Inc., 1994.
- [6] DEBEVEC P, MALIK J. Recovering high dynamic range radiance maps from photographs[C]//Proceedings of ACM SIGGRAPH 1997. Los Angeles: [s.n.], 1997.
- [7] ROBERTSON M A, BORMAN S, STEVENSON R L. Estimation-theoretic approach to dynamic range enhancement using multiple exposures[J]. Electronic Imaging, 2003, 12(2): 219-228.
- [8] WARD G L. A contrast-based scale factor for luminance display[M]. Boston: Academic Press Professional Inc., 1994.
- [9] WARD G L, RUSHMEIER H, PIATKO C. A visibility matching tone reproduction operator for high dynamic range scenes[J]. IEEE Transactions on Visualization and Computer Graphics, 1997, 3(4): 291-306.
- [10] TUMBLIN J, HODGINS J K, Guenter B K. Two methods for display of high contrast images[J]. ACM Trans. Graph, 1999, 18: 56-94.
- [11] PATTANAİK S N, JAMES A F, MARK D, et al. A multiscale model of adaptation and spatial vision for realistic image display[C]//Proceedings of SIGGRAPH'98. Orlando: [s.n.], 1998.
- [12] DANIEL J J, RAHMAN Z U, GIENN A W. A multiscale retinex for bridging the gap between color images and the human observation of scenes[J]. IEEE Transaction on Image Processing, 1997, 6(7): 965-976.