

基于 Harr 小波的动态场景全频阴影绘制算法*

赵沁平⁺, 车英慧

(北京航空航天大学 计算机学院 虚拟现实技术与系统国家重点实验室, 北京 100191)

All-Frequency Shadows Rendering Algorithm for Dynamic Scenes Based on Harr Wavelets

ZHAO Qin-Ping⁺, CHE Ying-Hui

(State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science, BeiHang University, Beijing 100191, China)

+ Corresponding author: E-mail: zhaomp@vrlab.buaa.edu.cn, http://vrlab.buaa.edu.cn

Zhao QP, Che YH. All-Frequency shadows rendering algorithm for dynamic scenes based on Harr wavelets. *Journal of Software*, 2011, 22(8): 1948-1959. <http://www.jos.org.cn/1000-9825/3865.htm>

Abstract: Current PRT (pre-computed radiance transfer) techniques are limited to having 3D static scene, or large low-frequency lights. In this paper, an all-frequency shadow rendering method for dynamic scenes is proposed. In the preprocessing phase, the blocking geometry is modeled as a set of spheres, according to complex 3D model. The lighting and BRDF (bidirectional reflectance distribution function) are projected onto the Harr wavelet basis. At runtime, with the advantage of different basis functions, the product of visibility vectors is computed for blocker spheres in the pixel basis, while the triple product integral of lighting, BRDF and visibility is computed in the Harr basis. CUDA (computed unified device architecture) is used to implement this method, which sufficiently utilizes the new features of GPU (graphics processing unit). Experiments show that the method greatly improves vision quality and satisfies real-time requirements.

Key words: pre-computed radiance transfer; Harr wavelet; triple product integral; CUDA (computed unified device architecture)

摘要: 针对现有的预计算辐射传递算法对三维场景限制严格、适合于低频光照环境等问题,提出了一种动态场景的全频阴影绘制算法。在预处理阶段使用球体对三维物体进行拟合,同时对光照函数和 BRDF(bidirectional reflectance distribution function)函数进行Harr小波变换;在运行时阶段利用不同基函数的优势,在像素基空间进行多个球体可见性函数的快速合并,在小波基空间进行光照函数、BRDF函数和可见性函数的三乘积分,得到最终的光照值。使用CUDA(computed unified device architecture)实现了该算法,充分利用了图形硬件的最新功能。实验结果表明,阴影绘制质量有很大的提高,可以基本达到实时绘制。

关键词: 预计算辐射传递;Harr小波;三乘积分;CUDA(computed unified device architecture)

中图法分类号: TP391 文献标识码: A

* 基金项目: 国家自然科学基金(60873159); 国家高技术研究发展计划(863)(2006AA01Z333)

收稿时间: 2008-04-01; 修改时间: 2008-08-07; 定稿时间: 2010-03-18

CNKI 网络优先出版: 2010-11-17 17:23, <http://www.cnki.net/kcms/detail/11.2560.tp.20101117.1723.002.html>

三维场景的实时真实感绘制是计算机图形学领域长期以来的一个重要研究方向.近年来,随着图形硬件计算能力的不断提高,以及对预计算辐射传递(pre-computed radiance transfer,简称 PRT)算法研究的不断深入,这一目标得到了一定程度的实现.在绘制质量上,基于不同基函数的预计算辐射传递算法可以得到很好的低频或全频阴影效果;在绘制速度上,一些算法也可以满足实时应用.然而,由于 PRT 算法将复杂的全局光照计算转化为预计算辐射传递矩阵,任何对三维场景的改变,都将使已经生成的预计算辐射传递矩阵无效.因此,早期的 PRT 算法只适用于完全静态的三维场景^[1,2].为了突破这一限制,众多研究者进行了深入的研究,如文献[3,4]允许物体在三维场景中运动,但仅限于刚体运动,物体本身形状不能发生变化.

上述工作使得 PRT 算法更为实用,然而在游戏、虚拟漫游等复杂应用,特别是动态的物体变形动画中,这些 PRT 算法全部失效.针对这一问题,文献[5]提出了 SHEXP(spherical harmonic exponentiation)算法,使用球体拟合三维物体,实时计算所有球体的辐射传递效果并进行累加,利用对数空间中高效的球面调和函数(spherical harmonic,简称 SH)多乘算法,在低频条件下取得了令人满意的实时绘制效果.

针对 SHEXP 算法只适用于低频光照环境的问题,本文在其基础上提出了一种全频环境下的动态场景阴影绘制算法.该算法充分发挥不同基函数的优势,首先利用像素基空间内可见性函数是二值函数的优点,在像素基空间进行多个物体遮挡效果的叠加计算;其次,为适应全频光照环境,在小波基空间进行光照函数、采样点 BRDF(bidirectional reflectance distribution function)函数和可见性函数三者的乘积积分,针对小波三乘积分计算耗时的问题,使用 CUDA(computed unified device architecture,统一计算设备架构)^[6]对本文算法进行加速.实验结果表明,在全频光照环境下,本文算法较好地实现了动态场景的阴影绘制效果;借助于最新的图形硬件支持,可以基本达到实时绘制.

1 相关工作

Sloan 等人^[1]在 2002 年首次提出预计算辐射传递算法.该算法在文献[7]的基础上,预计算时以球面调和函数系数的方式表示三维物体表面采样点的可见性及 BRDF 函数,并将其存储到预计算辐射传递矩阵;绘制时将表示光照的球面调和函数系数向量与预计算辐射传递矩阵相乘以求出三维物体的光照结果.预计算辐射传递算法利用球面调和函数的性质,将复杂的积分运算转换为矩阵和向量的乘积,从而加快了绘制速度,一般来说可以满足实时的需要.文献[2]提出使用小波基函数来取代球面调和基函数的预计算辐射传递算法,由于小波基函数在空域和频域上均具有局部性,因此不仅可以表现低频光照信号,而且可以表现高频小范围光照信号.从绘制结果上看,该算法可以很好地绘制全频阴影.

由于球面调和函数基和小波基均属于正交基,因此两个函数的乘积积分可以由其基函数系数向量点乘得到.这一性质是 PRT 类方法的基础,但是仅使用两个函数表示光源对物体的作用并仅允许其中之一发生变化,极大地限制了 PRT 类方法的应用范围.文献[8]在文献[2]的基础上深入研究了三函数乘积积分问题,给出一种在小波基下的次线性时间复杂度算法.文献[9]进一步提出一种通用多函数乘积积分算法,不再受限于 3 个函数.

上述方法均要求所处理的三维场景是静态的,即三维物体本身、三维物体之间均不能发生相对位置的变化,以保证预计算数据的有效性.为了突破这一限制,文献[4]提出了球面光辐射传递映射方法,通过对动态物体的包围球进行采样来计算物体之间相互作用造成的阴影和反射光.文献[3]提出了阴影场(shadow field)的概念,对于场景中的物体,在其周围的若干采样点处进行可见性采样,采样结果存储为遮挡场;对光源也进行辐射亮度采样,采样结果存储为辐射亮度场.绘制时对遮挡场和辐射亮度场进行排序,通过三重积运算计算最终的亮度值.该算法能够实现复杂局部光源下的动态场景光照明计算,但无法支持物体的形变.文献[10]提出了局部可变形预计算辐射传递(local deformable PRT,简称 LDPRT)算法.该算法使用球面调和函数的特殊形式,即带状调和函数(zonal harmonic,简称 ZH),拟合物体表面的遮挡关系和 BRDF 函数.利用带状调和函数的快速旋转性质,改变采样点处函数方向.当三维物体发生形变时,只需利用新的法线方向,就可以计算预计算辐射传递向量.该算法将全局光照退化到局部光照,在绘制效果方面存在较大的缺陷.文献[5]提出了一种创新性的动态场景辐射传递算法.首先将三维物体做球体拟合,然后以这一组球体作为遮挡体,进行基于球面调和函数的 PRT 计算.为了

加快计算速度,该文献提出了在对数空间中以向量加法代替多函数乘积积分的算法,使得实时可变形三维物体辐射传递成为可能.

2 三维场景的全局光照问题

对于三维场景中的 n 个不同物体,采样点 x 处 θ 方向的出射辐照度 B 可以根据入射环境光 L 、采样点材质 ρ 以及场景中所有物体的可见性函数 O 在球面 S 上的乘积积分计算得到:

$$B(x, \theta) = \int_S L(\varphi) \prod_{i=1}^n O_i(x, \varphi) \rho(x, \varphi \leftrightarrow \theta) (N \cdot \varphi) d\varphi \quad (1)$$

其中, φ 为光线的入射方向, N 为采样点 x 处的法线向量方向, $O_i (1 \leq i \leq n)$ 为三维场景中第 i 个三维物体对采样点位置 x 的动态可见性函数. 如果只考虑纯漫反射材质, 则出射辐照度 B 与视角 θ 无关. 另外, 将公式(1)中的余弦项 $(N \cdot \varphi)$ 与 BRDF 函数 ρ 合并为一项, 记为 $\tilde{\rho}$, 则公式(1)可简化为

$$B(x) = \int_S L(\varphi) \prod_{i=1}^n O_i(x, \varphi) \tilde{\rho}(x, \varphi) d\varphi \quad (2)$$

将 n 项动态可见性函数合并, 并且只考虑特定采样点, 即将 x 取值固定, 可得:

$$B = \int_S L(\varphi) V(\varphi) \tilde{\rho}(\varphi) d\varphi \quad (3)$$

其中,

$$V(x, \varphi) = \prod_{i=1}^n O_i(x, \varphi) \quad (4)$$

公式(3)是一个典型的三乘积分(triple product integral)问题. 为了求解该问题, 首先使用正交基函数 $b_i(\varphi)$ 对 L, V 和 $\tilde{\rho}$ 分别进行分解:

$$L(\varphi) = \sum_i L_i b_i(\varphi), \quad V(\varphi) = \sum_j V_j b_j(\varphi), \quad \tilde{\rho}(\varphi) = \sum_k \tilde{\rho}_k b_k(\varphi) \quad (5)$$

这样就得到了与函数 L, V 和 $\tilde{\rho}$ 对应的参数向量 L, T 和 $\tilde{\rho}$. 定义三乘系数(tripling coefficients)如下:

$$C_{ijk} = \int_S b_i(\varphi) b_j(\varphi) b_k(\varphi) d\varphi \quad (6)$$

三乘系数没有通用解析解. 特别地, 在使用球面调和函数基时, 三乘系数对应于 Clebsch-Gordan 系数, 可以通过解析方法求得^[8].

使用三乘系数, 公式(3)可以进一步求解:

$$B = \int_S \left(\sum_i L_i b_i(\varphi) \right) \left(\sum_j V_j b_j(\varphi) \right) \left(\sum_k \tilde{\rho}_k b_k(\varphi) \right) d\varphi = \sum_i \sum_j \sum_k L_i V_j \tilde{\rho}_k \int_S b_i(\varphi) b_j(\varphi) b_k(\varphi) d\varphi = \sum_i \sum_j \sum_k C_{ijk} L_i V_j \tilde{\rho}_k \quad (7)$$

可以看出, 对于多函数(>3)的预计算辐射传递, 如果仅使用以上的三乘积分方法, 计算时间复杂度较高. 因此, 早期的 PRT 类算法将公式(3)中的 V 和 $\tilde{\rho}$ 进一步合并, 将其简化为二乘积分进行计算^[1,2], 然而这种简化限制了算法的应用范围. 文献[5,9]给出了各自对于三乘积分的解决方法. 本文采用像素/小波混合基函数进行多乘积分的计算, 也是在一定程度上解决多乘积分问题, 较之前两种方法, 在绘制效果上有显著改进.

3 动态场景的全频阴影绘制算法

3.1 算法概述

本文采用像素/小波混合基函数来进行三乘积分的计算, 从而实现动态场景的全频阴影效果. 首先, 在像素基空间中以位运算的形式完成公式(4)给出的可见性函数之间的乘积; 随后, 将像素基空间中的可见性函数转入 Harr 小波基, 在 Harr 基空间中进行入射光函数、BRDF 函数以及可见性函数之间的一次三乘积分计算, 以获得采样点的颜色值. 本文算法的具体步骤如下:

- (1) 对原始三维模型进行球体拟合, 使外部空间体积值达到最小.
- (2) 预计算并存储像素基空间中的单个球体的可见性函数.

- (3) 对入射光函数 L 和物体 BRDF 函数 ρ 进行 Harr 小波变换,求得相应的小波基向量.
- (4) 更新三维模型顶点坐标,并以此为依据更新拟合球体的位置.
- (5) 在像素基空间中利用位操作进行多个球体可见性函数的乘积计算,得到完整的可见性函数 V .
- (6) 对可见性函数 V 进行 Harr 小波变换,求得相应的小波基向量.
- (7) 在小波基空间下进行 L , ρ 和 V 的三乘积计算,得到采样点颜色值,进行真实感绘制.若模拟尚未完成,则返回步骤(4);否则,算法结束.

3.2 三维模型的球体拟合

类似于文献[5],本文算法也需预先对原始三维模型进行球体拟合.由于算法目标相近,本文采用了文献[5,11]提出的球体集合拟合算法.使用一组包含 n_s 个球体的球体集合 $S_i(1 \leq i \leq n_s)$ 覆盖三维模型 T 所占用的三维空间,并试图最小化外部体积 E .外部体积 E 定义为

$$E(\{S_i\}, T) = \sum_{i=1}^{n_s} V(S_i - T) \quad (8)$$

其中, V 为体积函数,其具体计算方法可参考文献[11],本文不再详细讨论.

为了使球体集合 $\{S_i\}$ 完整覆盖三维模型 T 所占用的三维空间,算法在初始化阶段将 T 离散化为一组三维点集合,该集合包含了 T 的表面点(三角面片中心点)和内部点(网格体素化);同时将 S_i 的球心位置设置为随机,半径设置为 0.然后就开始迭代计算,直到误差足够小为止.迭代的主要步骤如下:

(1) 点分簇.点分簇使用基于栈的三维空间 Flood Fill 算法.该算法以某一个球体的球心为起始点,从栈中将三维点逐一弹出,计算其所对应的目标簇,并加入其中,随后将该三维点的邻居逐一压入栈中,以完成 Flood Fill 算法.算法尝试将该点加入到每个簇之中,比较所得到的新球体集合的外部体积 E 的增量,并最终将其加入导致外部体积增量最小的簇.在点分簇过程中,球体集合中所有球体的球心位置不变,仅改变球的半径.

(2) 根据分簇更新包围球体.完成点分簇后,每个球体所对应的簇就已确定.此时,可以根据确定的分簇调整包围球体的球心和半径.调整球体的目标仍然是找到最小的外部体积 E .即对于球体 S_i ,在三维空间中寻找函数 $V(S_i - T)$ 的最小值.具体实现时,求解这一最小值利用了 Powell 算法^[12].然后,返回步骤 1.

(3) 球体远程传输.当上述两个步骤无法进一步减小外部体积 E 时,球体远程传输算法^[13]将被启动.该算法将具有最大外部体积的球体一分为二,该球体所对应的簇内距离最远的两个点被选中作为新生成球体的球心,同时将原有球体集合中重合部分占球体体积比例最大的球体删除.球体远程传输有利于找到全局最优解,而不是陷入某一局部最优解之中.图 1 给出了部分模型的球体拟合结果.

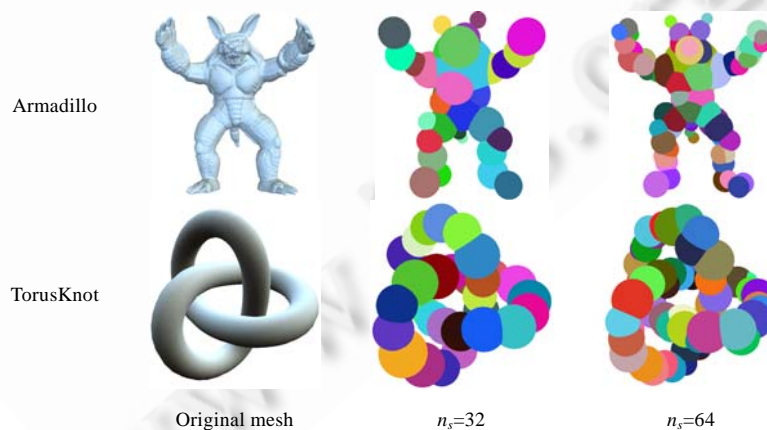


Fig.1 Sphere set approximation for some models

图 1 部分模型的球体拟合结果

3.3 像素基空间的可见性函数叠加

在利用球体拟合三维模型后,如何得到可见性函数是本文在运行阶段所面对的第 1 个问题.为此,预计算阶段本文预先计算并存储了单个球体的可见性函数,运行阶段则利用位操作完成多个球体可见性函数之间的乘积计算,上述操作均在像素基空间内进行.

在像素基空间中,可见性函数具有明显的特点:它是一个二值函数,即对于每一个表示可见性的像素,取值非 0 即 1.利用这一特点,使用单个比特便可以表示一个可见性像素.比如,本文采用立方体贴图(cube map)来存储单个采样点的可见性函数.设置 cube map 的分辨率为 $6 \times 32 \times 32$,因此单个采样点的函数只需 $6 \times 32 \times 32$ 比特=768 字节即可完整地表示,这在一定程度上实现了压缩存储.同时,在进行多个可见性函数的乘积计算时,只需进行位计算即可,简化了计算复杂度.如本文的例子中,两个球体的可见性函数叠加可以利用 $32 \times 6 = 192$ 次 32 位无符号整数之间的按位或运算得到.

球体作为基本的遮挡体,具有良好的对称性质,但由于本文使用 cube map 来存储单个采样点的可见性函数,cube map 缺乏快速旋转的能力.因此,对于球心在 P 处、半径为 R 的球体来说,在采样点 p 处产生的可见性函数必须使用参数 (\mathbf{v}, r) 表示.其中,归一化方向向量 $\mathbf{v} = (P - p) / \|P - p\|$,相对半径 $r = R / \|P - p\|$.具体实现时,我们也采用类似立方体贴图的方法来描述方向向量 \mathbf{v} ,每个面的分辨率为 32×32 ,即方向向量 \mathbf{v} 有 $6 \times 32 \times 32$ 种取值可能,相对半径 r 则有 32 种不同的取值.在此条件下,完整的单个球体的可见性立方图共占用 144M 内存空间.另外,由于球体的相似性和球体的可见性立方图是在球体的局部坐标系下表示的,我们只需要预计算 1 个球体的可见性立方图,相同半径的球体的可见性立方图完全相同,不同半径球体的可见性立方图只需要另外做一次缩放(scale)运算.因此,总的内存空间就是 144M.

有了单个球体的可见性立方图后,我们就可以对我们的绘制元素(几何顶点或者屏幕像素)进行可见性函数的计算.这些绘制元素也称为采样点.对采样点进行可见性计算需要叠加所有的球体的可见性立方图,最终的可见性函数也采用立方图的形式表示.立方图有 6 个面,我们采用十字架的表示形式,如图 2 所示(立方图采用十字架表示,其中,黑色 0 值表示该方向未被遮挡,白色 1 值表示该方向被遮挡),黑色代表未被遮挡,值为 0;白色代表遮挡,值为 1.球体可见性立方图的叠加是遮挡关系的叠加,只要被 1 个球体遮挡,结果就是被遮挡,因此是或运算.从图 2 可以看出,经过 64 个球体的叠加后,白色区域逐渐增多.同时可以看出,在多数情况下,可见性立方图的大部分都被黑色 0 值占据.因此,可以使用额外的数据结构表示某一块或某一行图像是否有需要合并的 1 值,以加速多个可见性函数之间的乘积计算.

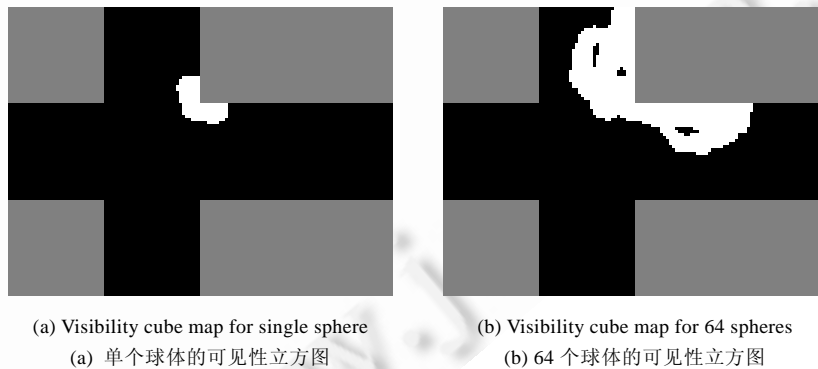


Fig.2 Cube map of visibility function for some sampler

图 2 某采样点的可见性函数立方图

3.4 二值位图的快速Harr小波变换

经过像素空间的可见性函数叠加,可以获得完整的可见性函数 V .但此时,光照函数 L 和 BRDF 函数 ρ 都处于 Harr 小波函数空间内.因此,在计算三乘积公式(3)之前,必须将像素空间的可见性函数转换至小波空间,即

进行一次 Harr 小波变换.

虽然 Harr 小波变换算法的时间复杂度为 $O(n)$,但在采样点数目较多的情况下,该算法仍然比较费时.因此,在可见性立方图是二值位图的前提下,本文采用查表法加速 Harr 小波变换,从而大幅降低了这一变换的时间开销.这一步骤的具体算法如图 3 所示.在预处理阶段,本文首先引入一张记录 4×4 大小的二值位图的 Harr 小波变换结果的数据表,表中每一项对应于在 4×4 二值位图的一种可能情况下,其 Harr 小波变换系数.由于 4×4 二值位图的不同组合情况共有 64K 种,故该表共包含 64K 项.在运行阶段,首先将输入的可见性立方图分割为 4×4 的位图块,每一块包含 16bit,将这 16bit 展开为一个整数,以此作为索引在表中查找,即可得到所对应的 Harr 小波系数.将其中的直流分量提出,存入第 2 级变换矩阵中,并使用通用 Harr 小波变换算法对第 2 级图像进行 Harr 小波变换,这样即可得到原输入位图的 Harr 小波变换结果.整个算法所使用时间约为原始算法的 1/16.

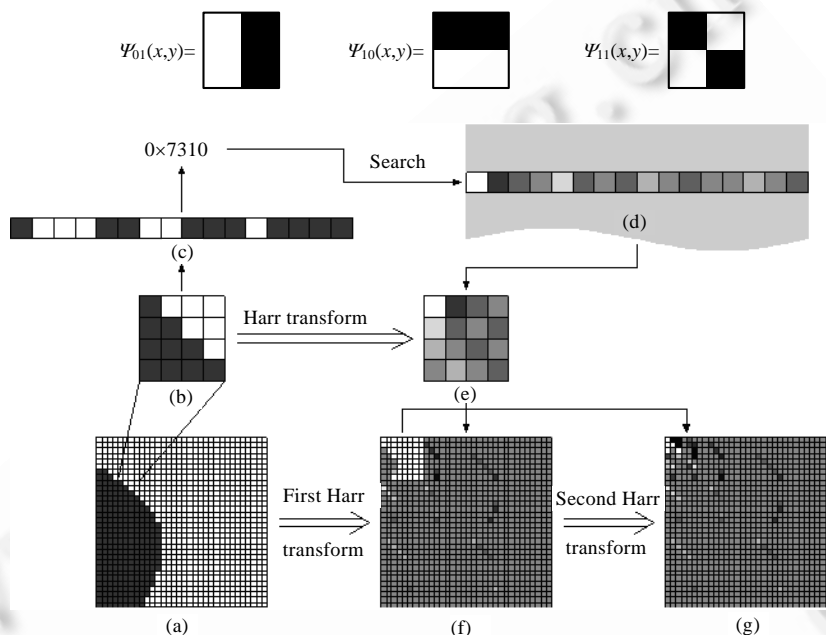


Fig.3 Fast Harr wavelet transform for binary bitmap

图 3 二值位图的快速 Harr 小波变换

3.5 小波空间的三乘积分

本文使用文献[8]提出的方法来计算小波空间的三乘积分.文献[8]分析并证明了采用二维 Harr 小波作为基函数的三乘积分中的所有非零三乘系数由 3 种不同情况组成,其算法复杂度仅为 $O(M \log N)$.

Harr 小波基是已知小波中最简单的小波基,其在二维空间中的定义如下^[14]:

- (1) 尺度基函数是单位面积:

$$\Phi(x,y)=1 \tag{9}$$

其中, $(x,y) \in [0,1]^2$.

- (2) 第 l 级的小波基函数由 3 种 Harr 母小波函数尺度扩张而成:

$$\Psi_M^{lij}(x,y) = 2^l \Psi_M(2^l x - i, 2^l y - j) \tag{10}$$

其中,偏移量 i 和 j 是位于 $[0,2^l]$ 范围内的整数, M 代表母小波函数的类型,取值为 01,10 和 11,分别对应于水平、垂直以及对角线差分.图 3 为 3 种母小波函数的定义,其中白色值为 +1,黑色值为 -1,边界外函数值为 0.图 3(a)是二值可见性立方图 6 个面中的一个面,大小为 32×32 ,将图 3(a)划分为 4×4 大小的子块,共有 8×8 个;图 3(b)为左数第 2 个、上数第 3 个子块,将图 3(b)表示为二维矩阵,则图 3(c)为图 3(b)的一维表示,黑色代表 0,白色代表 1,

图 3(c)的值为 0111 0011 0001 0000(十六进制为 7310);图 3(e)为图 3(b)的 Harr 小波变换结果,图 3(b)~图 3(e)的双线箭头代表此次 Harr 小波变换;图 3(d)表示记录 4×4 大小的二值位图的 Harr 小波变换结果的数据表,其中特别表示的一行为图 3(c)的 Harr 小波变换结果,图 3(d)和图 3(e)的关系与图 3(b)和图 3(c)的关系类似,都是二维矩阵的一维表示,但图 3(d)和图 3(e)不是二值,因此采用灰阶表示;图 3(f)为图 3(a)的每一个 4×4 子块完成 Harr 小波变换后的结果,此次 Harr 小波变换采用查表法,这时图 3(f)中的大部分区域都完成了 Harr 小波变换,只剩下左上角的 8×8 子块还需要进行进一步的 Harr 小波变换;图 3(g)为图 3(a)最终的 Harr 小波变换结果。

在定义了 Harr 小波基函数后,就可以给出 Harr 三乘系数定理:对于 3 个二维 Harr 小波函数的乘积积分,其三乘系数 C_{uvw} 取值非 0,当且仅当以下 3 项条件之一成立:

- (1) 3 个小波函数均为尺度基函数,在这种情况下, $C_{uvw}=1$.
- (2) 3 个小波函数作用区域相同,并且小波函数类型各不相同.在这种情况下, $C_{uvw}=2^l$, l 是小波级数.
- (3) 两个小波函数相同,其余一个小波函数是尺度基函数,或者是级别更低,作用区域与前两个小波函数重叠的小波函数.在这种情况下, $C_{uvw}=\pm 2^l$, l 是第 3 个小波函数所处的级数.

这个定理意味着,绝大部分 Harr 三乘系数的值都是 0.文献[8]指出,对于具有 N 个 Harr 小波基函数的二维图像来说,其精确的非零三乘系数为 $2-N+3\log_4 N$ 个.因此,Harr 小波三乘积分算法的时间复杂度上界为 $O(M\log N)$.下面给出 Harr 小波三乘积分的具体算法:

在下面的伪码中,分别使用 L,p 以及 V 代表公式(3)中的光照函数 L 、BRDF 函数 ρ 以及可见性函数 V ;对于小波基函数的访问,使用参数 $[s,M]$ 的形式, s 代表三元组 (l,i,j) , M 表示 Harr 母小波函数类型;另外,对于尺度基函数,直接使用 x_scale 的形式表示.

首先,使用条件(1)初始化三乘积分结果 $integral$:

```
 $integral=p\_scale*L\_scale*V\_scale;$ 
```

然后加入条件(2):

```
for each wavelet square  $s=(l,i,j)$ 
```

```
     $C_{uvw}=2^l;$ 
```

```
     $integral+=C_{uvw}*(p[s,01]*L[s,10]*V[s,11]+p[s,01]*L[s,11]*V[s,10]+p[s,10]*L[s,11]*V[s,01]+$   
         $p[s,10]*L[s,01]*V[s,11]+p[s,11]*L[s,01]*V[s,10]+p[s,11]*L[s,01]*V[s,10]);$ 
```

最后加入条件(3):

```
for each wavelet  $(s,M)$ 
```

```
     $integral+=(p[s,M]*L[s,M]*psum(V,s)+V[s,M]*p[s,M]*psum(L,s)+L[s,M]*V[s,M]*psum(p,s));$ 
```

其中,psum 函数定义为

```
 $psum(X,s)$  where  $s=(l,i,j)$ 
```

```
    if  $(l,i,j)=(0,0,0)$  return  $X\_scale;$ 
```

```
     $os=(ol,oi,oj)=(l-1,i/2,j/2);$ 
```

```
     $(qx,qy)=(i,j)-2*(oi,oj);$ 
```

```
    return  $psum(X,os)+2^{ol}*(X[os,01]*sign\_of\_quadrant(01,qx,qy)+$ 
```

```
         $X[os,10]*sign\_of\_quadrant(10,qx,qy)+X[os,11]*sign\_of\_quadrant(11,qx,qy));$ 
```

容易看出,psum 函数计算了与小波块 (l,i,j) 重叠并且面积大于它的所有小波系数之和.如果如伪码所示简单地令 psum 递归调用求值,则此函数的时间复杂度为 $O(M\log N)$.但如果采用动态规划算法,使用一个附加缓冲区记录已经计算得到的 psum 值,则可以实现优化的线性时间复杂度 $O(N)$ ^[8].

4 算法实现

本文算法的实现分为预处理与运行时计算两个主要部分.在预处理阶段,利用第 3.2 节给出的算法计算原始三维模型的球体拟合.良好的球体拟合是生成高质量全局渲染效果的先决条件.在运行时阶段,使用第 3.3 节~

第 3.5 节给出的算法,在像素基空间叠加多个球体可见性函数,然后将叠加后的可见性函数由像素基空间转为小波基空间,最后在小波基空间中将光照函数、BRDF 函数以及可见性函数进行三乘积分计算,得到最终的光照结果。

4.1 数据结构

在实现中,如何选择 L , $\tilde{\rho}$ 和 V 这 3 种主要函数系数的数据结构是一个重要的问题。对于光照函数 L 来说,由于在每帧渲染时光照函数是唯一且保持不变的;同时在大多数应用中,光照函数也不随时间发生改变,是完全静态的,因此对于光照函数,可以容易地获得它的完整小波存储(分辨率为 $6 \times 32 \times 32$)以及压缩后的稀疏存储。

由于本文仅考虑纯漫反射材质,因此合并了余弦项的 BRDF 函数 $\tilde{\rho}$ 仅与采样点本身的法向量方向相关。在本文实现中,BRDF 函数以分辨率为 $6 \times 8 \times 8$ 的法线方向为索引,同时以完整小波形式(分辨率为 $6 \times 32 \times 32$)以及压缩后的稀疏数组形式存储。

对于比较复杂的可见性函数 V ,同时存在像素基空间和小波基空间两种形式,两种形式下的分辨率均为 $6 \times 32 \times 32$ 。在像素基空间中,每个像素使用 1 个 bit,共用 768 字节。在转换至小波空间的过程中,使用第 3.4 节给出的二元位图的 Haar 小波快速变换算法,自然形成了两级 Haar 小波存储。第 1 级使用 $6 \times 8 \times 8$ 的分辨率存储 Haar 小波函数系数,同时使用指针指向第 2 级;第 2 级使用 4×4 的分辨率存储了较精细的 Haar 小波函数系数。两级存储共同实现了分辨率为 $6 \times 32 \times 32$ 的小波基可见性函数。第 2 级存储是预先计算并以表格形式给出的。

本文算法针对每个采样点分别计算,以完成场景的着色过程。在最终的小波空间三乘积分计算中,本文在具体实现时并没有直接使用第 3.5 节给出的线性时间复杂度版本,而是采用了具有次线性时间复杂度的非线性优化版本。在这一版本中,算法不再遍历所有的 (s, M) ,而是仅针对非线性优化后的 N 个小波项进行。由于 psum 函数的所有输出都包含了直流成分,不能被简单忽略,因此条件(3)的实现并非是完全对称的。光照函数和物体 BRDF 函数比较容易通过预计算的方法获得稀疏小波存储,所以循环针对这两个函数进行。

由于物体 BRDF 函数频率较低,因此采用该函数完成条件(3)中 3 项计算中的前两项。并且在这一情况下,仅使用第 1 级的可见性函数或可见性函数的 psum 函数。这也使整个算法实现过程中仅需要计算第 1 级 psum 函数,提高了算法速度。同时,使用条件(3)中的第 3 项计算 $(L[s, M] \times V[s, M] \times psum(p, s))$,可以选择是否使用第 2 级可见性函数。在实验结果中,给出了仅使用 1 级可见性函数与使用 2 级可见性函数的效果和速度对比。表 1 给出了本文中光照函数、物体 BRDF 函数以及可见性函数的采样分辨率、存储形式以及 psum 的计算策略。

Table 1 Parameters description of three functions
表 1 3 种函数的相关参数描述

	Resolution	Store method	Compute strategy of psum
Light (L)	$6 \times 32 \times 32$, 3 channels	Store both complete wavelet and sparse wavelet	Offline
BRDF ($\tilde{\rho}$)	$6 \times 32 \times 32$	Store both complete wavelet and sparse wavelet	Offline
Visibility (V)	First level $6 \times 8 \times 8$, second level 4×4 , totally $6 \times 32 \times 32$	Store 2 level complete wavelet	Real-Time, compute level 1 only

4.2 CUDA实现

为了获得更高的渲染速度,本文选择在 NVidia Geforce 8 系列 GPU(graphics processing unit)上基于 CUDA API 进行算法实现。与传统 GPU 编程相比,CUDA 提供了丰富的内存访问手段,使程序不再在写内存方面受限;并且 CUDA 提供了对整数处理的完整手段,包括本文算法严重依赖的位操作。虽然这是一个简单的改进,但对本文算法意义重大。本文的实现按照第 3.3 节~第 3.5 节自然划分为 3 个子程序。子程序之间使用显存进行数据传递,计算过程中不需要在图形总线上进行数据传输。在所有计算完成之后,利用一个结果缓冲区一次性将所有采样点渲染结果取回即可。

5 实验结果

本文所采用的实验平台为 Intel Core2 2.4GHz,主内存为 2GB,显卡为 NVidia Geforce 8800GTX 768MB,所有实验均在 Windows XP 下基于 DirectX API 和 CUDA 实现.实验场景由两部分组成:三维模型本身及模型所处平面.所采用的三维模型包括 TorusKnot 模型以及 Stanford 大学的 bunny 模型和 armadillo 模型等.模型所处平面采用 128×128 的分辨率,共包含 16 384 个采样点.

图 4 给出了本文算法的渲染结果,其中,bunny 模型使用 Grace Cathedral 光照立方图进行照明,armadillo 模型使用 St. Peters 光照立方图进行照明.图 4(a)为参考图像,使用原始三维模型作为遮挡体,光源、BRDF 以及可见性立方图精度与本文其他实验相同(均为 6×32×32),最终,使用像素基积分获得渲染结果.图 4(b)和图 4(c)采用本文算法渲染得到(Haar 小波压缩至非线性系数 200 项,2 级可见性函数渲染).在使用 Grace Cathedral 和 St. Peters 这类包含多个高频强光源的光照图进行照明时,本文算法产生了较为清晰的阴影,较好地体现了全频阴影的特点.

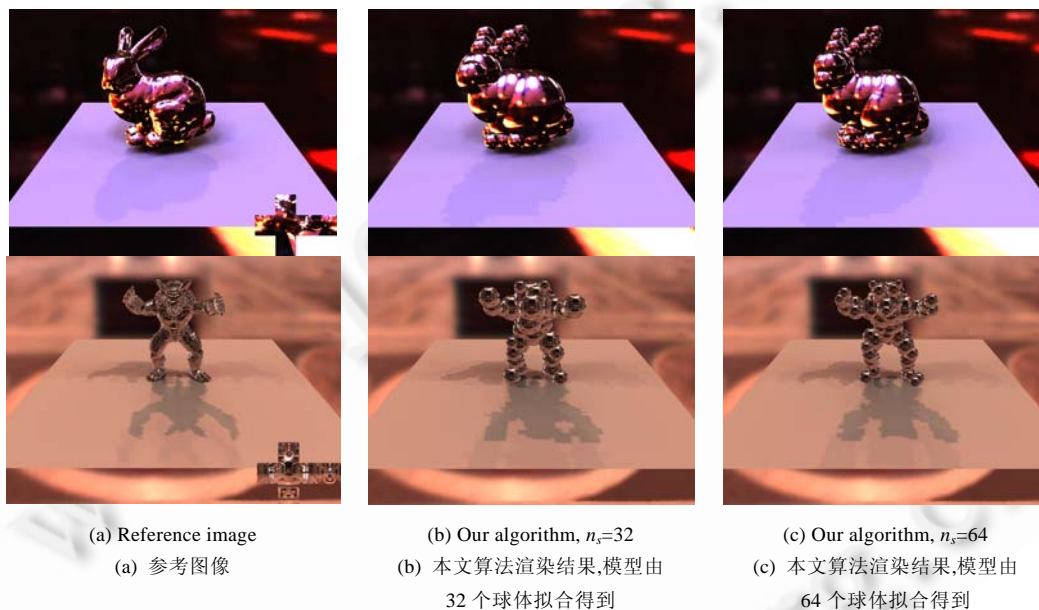


Fig.4 Rendering results of our algorithm

图 4 本文算法的渲染结果

图 5 给出了使用 1 级/2 级可见性函数的效果对比(遮挡球体数目为 64,Haar 小波压缩至非线性系数 200 项).在 RNL 光照图中,两种渲染效果差异很小;但在 Galileo 光照图中,可以看出使用 2 级可见性函数的渲染效果明显优于仅使用 1 级可见性函数的渲染效果.这是因为 RNL 光照图主要包含低频光照信号,而 Galileo 光照图同时包含高频和低频光照信号.如果在对光照函数完成条件(3)中的第 3 项计算中仅使用 1 级可见性函数,则导致这一计算中光照函数中的高频信号被忽略,进而导致最终渲染效果质量降低.图 6 给出了采用不同小波系数项数的渲染效果对比(遮挡体数目为 64,2 级可见性函数渲染,St. Peters 光照立方图照明).

图 7 对比了本文算法、SHEXP 算法以及像素基积分法的渲染效果,图中模型采用 RNL 光照立方图进行照明.图 7(a)采用原始 bunny 模型作为遮挡体,图 7(b)~图 7(d)采用 64 个球体作为遮挡体.可以看出,本文算法对于渲染质量的提升非常明显,即使采用 RNL 这类低频光源,本文算法也还是取得了远好于 SHEXP 算法的效果.

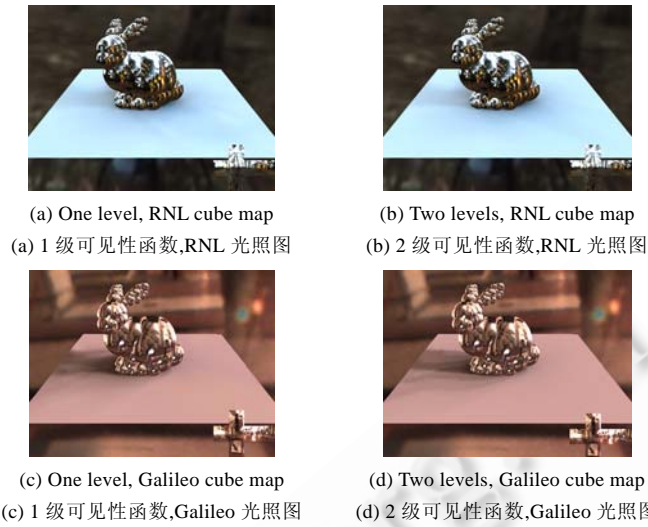


Fig.5 Comparison of accuracy with different level visibility function

图 5 不同级别可见性函数的渲染精度对比

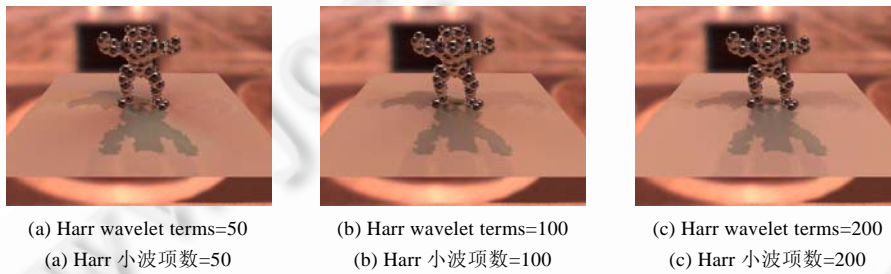


Fig.6 Comparison of accuracy with different numbers of Harr wavelet terms

图 6 不同 Harr 小波项数的渲染精度对比

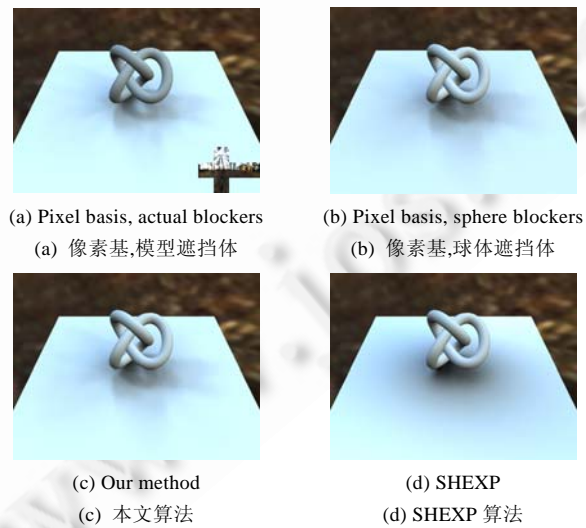


Fig.7 Rendering quality comparison

图 7 渲染质量对比

在性能方面,我们比较了本文算法在 CPU 和 GPU 上的性能,实验结果见表 2 和图 8.实验中,光源、BRDF 以及可见性函数的精度均为 $6 \times 32 \times 32$.表 2 中,CPU 表示算法在 CPU 上的实现,GPU-I 表示算法只使用 1 级可见性函数渲染,GPU-II 表示算法使用 2 级可见性函数进行渲染.从实验结果可以看出,基于 CUDA 的实现有效地利用了 GPU 高性能的计算能力.与 CPU 实现相比,GPU 实现的处理速度始终为 CPU 实现的 15 倍左右.

Table 2 Performance of our algorithm

表 2 本文算法的性能

Sphere number	Harr wavelet terms=50			Harr wavelet terms=200		
	CPU	GPU-I	GPU-II	CPU	GPU-I	GPU-II
32	2.91	34.45	31.53	1.52	27.16	22.64
64	1.91	24.83	23.62	1.20	21.01	17.87
96	1.42	18.72	18.15	0.98	16.75	14.78
128	1.13	15.50	15.07	0.83	14.10	12.80

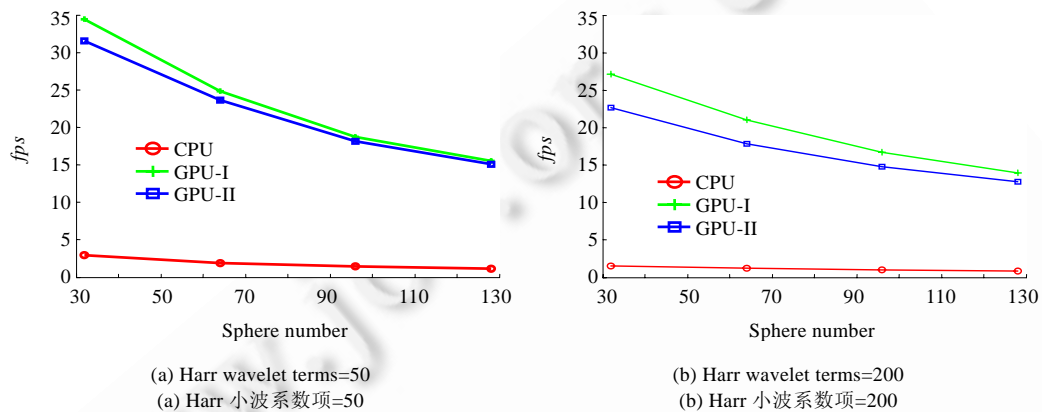


Fig.8 Algorithm performance comparison chart

图 8 算法性能比较图

6 结论及进一步的工作

本文提出了一种动态场景的全频阴影绘制算法,利用球体对三维物体进行拟合,利用小波变换获得全频阴影效果.本文算法较之以往算法有 3 个特点:首先,本文算法不仅限于低频光照环境,而且在具有高频信号的光照环境中也具有良好的表现;其次,本文算法使用像素/小波混合基函数进行三重乘积积分计算,利用不同基函数的优势,在像素基函数下进行多个球体可见性函数的快速合并,在小波基函数下进行光照函数、采样点的 BRDF 函数和可见性函数三者的乘积积分,得到最终的光照值;最后,使用 CUDA 实现了本文算法,充分利用了硬件特性.实验结果表明,在全频光照环境下,本文算法较好地实现了动态场景的全频阴影效果,可以基本达到实时绘制.

球体拟合毕竟是几何物体的近似表示,寻找更好的几何物体的表示方法是可以预见的进一步工作.另外,我们希望对遮挡球体建立层次结构、排除不可见采样点计算等方面进行进一步优化.然后,可以考虑加入遮挡体自反射计算.同时,对漫反射材质之外的镜面反射材质、各向异性材质的支持,也是我们未来的工作方向.

References:

- [1] Sloan PP, Kautz J, Snyder J. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. In: Proc. of the 29th Annual Conf. on Computer Graphics and Interactive Techniques. New York: ACM Press, 2002. 527-536. [doi: 10.1145/566570.566612]

- [2] Ng R, Ramamoorthi R, Hanrahan P. All-Frequency shadows using non-linear wavelet lighting approximation. In: Proc. of the ACM SIGGRAPH 2003 Papers. New York: ACM Press, 2003. 376–381. [doi: 10.1145/1201775.882280]
- [3] Zhou K, Hu YH, Lin S, Guo BN, Shum HY. Precomputed shadow fields for dynamic scenes. In: Marks J, ed. Proc. of the ACM SIGGRAPH 2005 Papers. New York: ACM Press, 2005. 1196–1201. [doi: 10.1145/1186822.1073332]
- [4] Mei CH, Shi JY, Wu FL. Rendering with spherical radiance transport maps. In: Proc. of the Eurographics. 2004. 281–290. [doi: 10.1111/j.1467-8659.2004.00759.x]
- [5] Ren Z, Wang R, Snyder J, Zhou K, Liu XG, Sun B, Sloan PP, Bao HJ, Peng QS, Guo BN. Real-Time soft shadows in dynamic scenes using spherical harmonic exponentiation. ACM Trans. on Graph, 2006,25(3):977–986. [doi: 10.1145/1179352.1141982]
- [6] <http://developer.nvidia.com/object/cuda.html>
- [7] Ramamoorthi R, Hanrahan P. An efficient representation for irradiance environment maps. In: Proc. of the 28th Annual Conf. on Computer Graphics and Interactive Techniques. New York: ACM Press, 2001. 497–500. [doi: 10.1145/383259.383317]
- [8] Ng R, Ramamoorthi R, Hanrahan P. Triple product wavelet integrals for all-frequency relighting. In: Marks J, ed. Proc. of the ACM SIGGRAPH 2004 Papers. New York: ACM Press, 2004. 477–487. [doi: 10.1145/1186562.1015749]
- [9] Sun WF, Mukherjee A. Generalized wavelet product integral for rendering dynamic glossy objects. ACM Trans. on Graph, 2006, 25(3):955–966. [doi: 10.1145/1179352.1141980]
- [10] Sloan P, Luna B, Snyder J. Local, deformable precomputed radiance transfer. ACM Trans. on Graph, 2005,24(3):1216–1224. [doi: 10.1145/1186822.1073335]
- [11] Wang R, Zhou K, Snyder J, Liu XG, Bao HJ, Peng QS, Guo BN. Variational sphere set approximation for solid objects. The Visual Computer: International Journal of Computer Graphics, 2006,22(9):612–621. [doi: 10.1007/s00371-006-0052-0]
- [12] Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical Recipes in C: The Art of Scientific Computing. New York: Cambridge University Press, 1992.
- [13] Cohen-Steiner D, Alliez P, Desbrun M. Variational shape approximation. In: Marks J, ed. Proc. of the ACM SIGGRAPH 2004 Papers. New York: ACM Press, 2004. 905–914. [doi: 10.1145/1186562.1015817]
- [14] Stollnitz EJ, Deroose TD, Salesin DH. Wavelets for Computer Graphics: Theory and Applications. Morgan Kaufmann Publishers, 1996.



赵沁平(1948—),男,山西介林人,博士,教授,博士生导师,CCF 高级会员,主要研究领域为虚拟现实,分布式应用系统,人工智能.



车英慧(1980—),女,博士生,主要研究领域为虚拟现实,计算机图形学.