



Monte Carlo Techniques for Direct Lighting Calculations

PETER SHIRLEY

Cornell University

CHANGYAW WANG

Alias Research Inc.

and

KURT ZIMMERMAN

Indiana University

In a distributed ray tracer, the sampling strategy is the crucial part of the direct lighting calculation. Monte Carlo integration with importance sampling is used to carry out this calculation. Importance sampling involves the design of integrand-specific probability density functions that are used to generate sample points for the numerical quadrature. Probability density functions are presented that aid in the direct lighting calculation from luminaires of various simple shapes. A method for defining a probability density function over a set of luminaires is presented that allows the direct lighting calculation to be carried out with a number of sample points that is independent of the number of luminaires.

Categories and Subject Descriptors: G.1.4 [Numerical Analysis]: Quadrature and Numerical Differentiation; I.3.0 [Computer Graphics]: General; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism

General Terms: Algorithms, Design

Additional Key Words and Phrases: Direct lighting, importance sampling, luminaires, Monte Carlo integration, ray tracing, realistic image synthesis

Much of this work was carried out while the authors were at Indiana University. This work was supported by Indiana University Faculty Start-Up Funds and the National Science Foundation under grant NSF-CCR-92-09457 and NSF-CCR-94-01961, and used equipment purchased with funds from CISE RI grant 92-23008 and NSF II grant 93-03189. The remainder of this work was carried out at the Cornell Program of Computer Graphics, was supported by the NSF/ARPA Science and Technology Center for Computer Graphics and Scientific Visualization (ASC-8920219), and was performed on workstations generously provided by the Hewlett-Packard Corporation.

Authors' addresses: P. Shirley, Program of Computer Graphics, Cornell University, Ithaca, NY 14853; C. Wang, Alias Research Inc., 110 Richmond Street East, Toronto, M5C 1P1, Canada; K. Zimmerman, Department of Computer Science, Indiana University, Bloomington, IN 47405. Permission to make digital/hard copy of all or part of this material without fee is granted provided that the copies are not made or distributed for profit or commercial advantage, the ACM copyright/server notice, the title of the publication, and its date appear, and notice is given that copying is by permission of ACM. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.

© 1996 ACM 0730-0301/96/0100-0001\$03.50

1. INTRODUCTION

Monte Carlo methods have been used extensively for image rendering since Cook et al.'s [1984] landmark paper on distributed ray tracing. Monte Carlo techniques have been used in classic ray-tracing (view-dependent) methods [Lee et al. 1985; Cook 1986; Kajiya 1986; Kirk and Arvo 1991b; Lange 1991; Kok and Jansen 1991; Rushmeier 1988], zonal (view-independent) methods [Malley 1988; Shirley et al. 1991], and hybrid techniques [Airey and Ouh-Young 1989; Sillion and Puech 1989; Chen et al. 1991; Kok and Jasen 1992; Rushmeier et al. 1993]. Most recent implementations have handled the direct lighting¹ using Monte Carlo integration in a view-dependent manner and have stored the indirect lighting within some kind of spatial lookup table [Ward et al. 1988; Rushmeier et al. 1988]. This paper extends traditional Monte Carlo methods for carrying out the direct lighting calculation so that they can be applied to scenes with hundreds or thousands of luminaires (light-emitting objects). Our method is designed to work with tens to thousands of samples for each numerical quadrature, and technically this is a "low sampling density," meaning that the results of traditional asymptotic analysis are not directly applicable. Instead, we follow the advice of Spanier and Maize [1994] and focus on the design of probability density functions that are well suited to the particular characteristics of the integrands that arise in direct lighting calculations.

As shown by Kajiya [1986], the radiance [American National Standard Institute 1986; Rea 1993] at a point in a scene can be written down as an integral equation over the set of all points in the environment.² The direct lighting component of this radiance is a two-dimensional integral, as opposed to an integral equation, over the set of all points in the scene. As shown by Cook et al. [1984], this direct lighting integral is a natural target for Monte Carlo integration [Hammersley and Handscomb 1964; Kalos and Whitlock 1986]. A difficult question in any application of Monte Carlo integration is what probability density function should be used to generate sample points on the domain of integration. This is especially difficult in the direct lighting calculation where the domain is typically all surfaces in a complex three-dimensional environment. The selection of a *good* density function, that is, one that produces a low-variance solution, is the key step in applying Monte Carlo integration. In this paper we attempt to find good probability density functions to be used in the direct lighting calculation. The main contributions of this paper are a careful discussion of the design of probability density functions for single luminaires, and a method for the construction of a single

¹ Lighting is often separated into three components in image synthesis programs: *emitted* (or *self-emitted*) lighting that originates at an object (called a *luminaire*), *direct* lighting that consists of light emitted from a luminaire and reflected from exactly one surface before reaching the viewer or camera, and *indirect* lighting that consists of light emitted from a luminaire and reflected from at least two surfaces before reaching the viewer or camera.

² This transport equation only applies to steady-state scenes made up of surfaces with a vacuum between them obeying geometrical optics. In practice, this is a reasonable approximation to the human-observable behavior of visible light for many scenes [Glassner 1995].

probability density function over a set of luminaires, so that only one shadow ray needs to be shot per viewing ray, regardless of the number of luminaires in the scene. This makes ray tracing feasible for scenes with many luminaires. This method is most suitable for images where multiple samples will be taken in each pixel. If the number of samples per pixel is very low, for example, four samples, then a culling method such as Ward's [1991] would be preferable to the method described in this paper.

In Section 2 we state the direct lighting integral and describe basic Monte Carlo integration. In Section 3 we design probability density functions for shapes commonly used to approximate luminaires. In Section 4 we describe how to design a probability density function over large sets of luminaires. In Section 6 we speculate on how the techniques presented in this paper will fit into future rendering systems. Table I gives a summary of terms and symbols.

2. MATHEMATICAL FORMULATION

In this section the basic Monte Carlo solution method for definite integrals is outlined, and Monte Carlo integration is applied to the direct lighting integral. Most of the basic Monte Carlo material of this section is also covered in several of the classic texts [Hammersley and Handscomb 1964; Shreider 1966; Halton 1970; Yakowitz 1977]. This section differs by being geared toward classes of problems that crop up in realistic image synthesis.

2.1 Monte Carlo Integration

Suppose we have a real-valued function $h: \mathcal{S} \rightarrow \mathbb{R}$, where \mathcal{S} is a possibly multidimensional space, and we wish to estimate the expected value of $h(\psi)$, where ψ is a random variable with probability density function $\varphi: \mathcal{S} \rightarrow \mathbb{R}^+$ (denoted $\psi \sim \varphi$). That φ is a probability density function on \mathcal{S} means that $\mu(\varphi) = 1$, where μ is a measure defined on \mathcal{S} .³ From the definition of expected value, we can approximate the expected value of $h(\psi)$, $E[h(\psi)]$, as a sum:

$$E[h(\psi)] = \int_{\mathcal{S}} h(\psi') \varphi(\psi') d\mu(\psi') \approx \frac{1}{N} \sum_{i=1}^N h(\psi_i). \quad (1)$$

The samples ψ_i are a set of N instantiations of the random variable ψ . We need this formulation with arbitrary dimension and measure, rather than the one-dimensional discussion found in most Monte Carlo texts, because we will integrate over surface areas and solid angles that do not have the simple

³ The notation $\mu(\varphi) = 1$ means that the measure of φ is one, and is equivalent to the notation $\int_{\mathcal{S}} \varphi(\psi) d\mu(\psi) = 1$. We occasionally use the symbol $d\mu(\psi)$ in isolation. Although this technically violates measure theoretic notation, we follow many authors by indulging in this violation for convenience.

Table I. Important Symbols and Terms

\mathbb{R}	Set of real numbers: $(-\infty, +\infty)$
\mathbb{R}^+	Set of nonnegative real numbers: $[0, +\infty)$
\mathcal{S}	Unspecified space used for examples
μ	Unspecified measure defined on \mathcal{S} used for examples
$\rho(\psi)$	Probability density function defined on \mathcal{S} using measure μ
ξ	Canonical random number: uniformly random in $[0, 1)$
$f: \mathcal{A} \rightarrow \mathcal{B}$	Function that maps elements of set \mathcal{A} to elements of set \mathcal{B}
$E(a)$	Expected value of random variable a
$V(a)$	Variance of random variable a
N	Number of samples taken in a Monte Carlo integration
\mathcal{X}	Set of all points on surfaces
\mathbf{x}	Point in \mathcal{X} being illuminated
\mathbf{x}'	Point in \mathcal{X} s.t. $\mathbf{x}' = \mathbf{x} + t(-\hat{\omega})$ with $t = \ \mathbf{x} - \mathbf{x}'\ $
A	Area measure
$\ \mathbf{x} - \mathbf{x}'\ $	Distance between points \mathbf{x} and \mathbf{x}'
\hat{n}	Unit surface normal direction at \mathbf{x}
\hat{n}'	Unit surface normal direction at \mathbf{x}'
Ω	Hemisphere of directions with pole \hat{n} (i.e., for all $\hat{\omega} \in \Omega$, $\hat{\omega} \cdot \hat{n} > 0$)
\mathcal{U}	Hemisphere of directions with pole $-\hat{n}$ (i.e., for all $\hat{\omega}' \in \mathcal{U}$, $\hat{\omega}' \cdot \hat{n} < 0$)
σ	Solid angle measure
$\hat{\omega}$	Direction (unit vector) in Ω from which \mathbf{x} is viewed
$\hat{\omega}'$	Direction (unit vector) in \mathcal{U} that is parallel to $\mathbf{x} - \mathbf{x}' / \ \mathbf{x} - \mathbf{x}'\ $
$\rho(\mathbf{x}, \hat{\omega}, \hat{\omega}')$	Spectral bidirectional reflectance distribution function (BRDF)
$g(\mathbf{x}, \mathbf{x}')$	Visibility function: 1 if no surface is between \mathbf{x} and \mathbf{x}' ; 0 otherwise
$L_s(\mathbf{x}, \hat{\omega})$	Surface spectral radiance at point \mathbf{x} in direction $\hat{\omega}$
$L_f(\mathbf{x}, \hat{\omega}')$	Field (incident) spectral radiance at point \mathbf{x} coming from direction $\hat{\omega}'$
$L_e(\mathbf{x}, \hat{\omega})$	Surface spectral radiance at point \mathbf{x} in direction $\hat{\omega}$ emitted from \mathbf{x}' (emitted lighting)
$L_d(\mathbf{x}, \hat{\omega})$	Surface spectral radiance at point \mathbf{x} in direction $\hat{\omega}$ caused by light coming directly from luminaire (direct lighting)
$p(\mathbf{x}')$	Probability density function defined on \mathcal{X} using measure A
$q(\hat{\omega}')$	Probability density function defined on Ω using measure σ
N_L	Number of luminaires in the environment
ℓ_i	i th luminaire
ρ_{\max}	Maximum BRDF of a point over all $\hat{\omega}$ and $\hat{\omega}'$
\mathcal{L}	Set of all luminaires ℓ_i
\mathcal{L}_{dim}	Set of "unimportant" luminaires
$\mathcal{L}_{\text{bright}}$	Set of "important" luminaires
\hat{L}_d	Estimate of average L_i for $\ell_i \in \mathcal{L}_{\text{dim}}$
N_d	Number of luminaires in \mathcal{L}_{dim}
N_b	Number of luminaires in $\mathcal{L}_{\text{bright}}$
L_i	Radiance contribution from ℓ_i at some $(\mathbf{x}, \hat{\omega})$
\hat{L}_i	Estimate of L_i
LO	Type of function that has "low variation" and so is roughly constant
HI	Type of function that has "high variation" and so is not roughly constant
UN	Type of function that is unknown, but can be evaluated at points

measure of length. The form of eq. (1) is a bit awkward; we would usually like to approximate an integral of a single function f rather than a product $h\varphi$. We can get around this by substituting $f = h\varphi$ as an integrand:

$$\int_{\mathcal{S}} f(\psi') d\mu(\psi') \approx \frac{1}{N} \sum_{i=1}^N \frac{f(\psi_i)}{\varphi(\psi_i)}. \quad (2)$$

For this formula to be valid, φ must be positive where f is nonzero. The variance of the sum in eq. (2) is

$$V\left[\frac{1}{N} \sum_{i=1}^N \frac{f(\psi_i)}{\varphi(\psi_i)}\right] = \frac{1}{N} V\left[\frac{f}{\varphi}\right]. \quad (3)$$

Equation (3) implies that to get a good (low-variance) estimate we want as many samples as possible (i.e., N is large), and we want the density f/φ to have a low variance. Choosing φ intelligently is called *importance sampling*, because if φ is relatively large where f is relatively large, there will be more samples in important regions. Eq. (3) also shows the fundamental problem with Monte Carlo integration: *diminishing return*. Because the variance of the estimate is proportional to $1/N$, the standard deviation is proportional to $1/\sqrt{N}$. Because the error in the estimate behaves similarly to the standard deviation, we need to approximately quadruple n to halve the error.

Another way to reduce variance is to partition \mathcal{S} , the domain of the integral, into N several smaller domains \mathcal{S}_i and to evaluate the integral as a sum of integrals over the \mathcal{S}_i . This is called *stratified sampling*. Normally, only one sample is taken in each \mathcal{S}_i (with density φ_i), and in this case, the variance of the estimate is

$$V\left[\sum_{i=1}^N \frac{f(\psi_i)}{\varphi_i(\psi_i)}\right] = \sum_{i=1}^N V\left[\frac{f(\psi_i)}{\varphi_i(\psi_i)}\right]. \quad (4)$$

Typically, a density φ on \mathcal{S} will be chosen, and φ_i will be made proportional to φ :

$$\varphi_i(\psi) = \frac{\varphi(\psi)}{\int_{\mathcal{S}_i} \varphi(\psi') d\mu(\psi')} . \quad (5)$$

There are some functions where stratification does no good. An example is a white noise function, where the variance is constant for all regions. Intuitively, stratification will pay off when f varies slowly relative to the distance between adjacent sample points. Stratification will never produce higher variance than unstratified sampling provided the integral of $\varphi(\psi)$ over each stratum is the same.

Although distributed ray tracing is usually phrased as an application of eq. (2), many researchers replace the ψ_i with more evenly distributed (quasi-random) samples (e.g., Cook [1986], Mitchell [1991], and Schlick [1991]). This approach can be shown to be sound by analyzing decreasing error in terms of

some discrepancy measure [Zeremba 1968; Wozniakowski 1991; Mitchell 1991; Shirley 1991], rather than in terms of variance. However, in practice it is often convenient to develop a sampling strategy using variance analysis on random samples, and then to turn around and use nonrandom but equidistributed samples in an implementation. This approach is almost certainly correct, but its justification and implications have yet to be fully explained. For the rest of this paper, we assume unstratified random sample points for our derivations, but in our implementation we have used stratified random samples.

2.2 Generating Random Samples

There are many ways to generate random samples with specified density functions. Most of these methods assume that we have the ability to generate a sequence of *canonical* random numbers ($\xi_1, \xi_2, \xi_3, \dots$), where each ξ_i is an identically distributed uniform random number between zero and one (more formally, $u:[0, 1] \rightarrow \mathbb{R}^+$ is the canonical probability density function, meaning for all $y \in [0, 1]$ that $u(y) = 1$, and ξ being a canonical random number means $\xi \sim u$). For example, to choose uniform random points in the ball with unit radius centered at the origin, we can pick a uniform point r in the cube with sidelength 2 centered at the origin, where $r = (2\xi_1 - 1, 2\xi_2 - 1, 2\xi_3 - 1)$, and check whether r is inside the ball. If it is not in the ball, we repeat the process with $r = (2\xi_4 - 1, 2\xi_5 - 1, 2\xi_6 - 1)$, and so on, until we find a point in the ball. This is a so-called *rejection* method and is very useful for picking random points in spaces with complex boundaries. Unfortunately, it is often not as obvious how to apply a rejection method for nonuniform densities, and a rejection method is not easy to combine with stratified sampling.

A method that does not suffer from the problems of the rejection method is to pass the canonical numbers through a function. To illustrate why this will work, notice that, for some function $f:[0, 1] \rightarrow \mathbb{R}$ and a canonical random number ξ , $f(\xi)$ is a random variable with a possibly nonuniform density function. For example, the quantity $f(\xi) = \xi^2$ is more likely to have a value near zero than near one, so it must have a nonuniform density function. We can reverse-engineer what $f(x)$ must be to get a sample with desired density function $\varphi:[\psi_0, \psi_1] \rightarrow \mathbb{R}$. This will be $\mathcal{P}^{-1}(\psi)$, where $\mathcal{P}:[\psi_0, \psi_1] \rightarrow [0, 1]$ is given by

$$\mathcal{P}(\psi) = \int_{\psi_0}^{\psi} \varphi(\psi') d\psi'. \quad (6)$$

The function \mathcal{P} is often called the *cumulative probability distribution function* associated with the probability density function φ because $\mathcal{P}(\psi)$ is the probability that a random variable with density φ has a value less than or equal to ψ . This basic method has been in use since at least the 1950s [Hammersley and Handscomb 1964], and was first used in image synthesis by Ward et al. [1988] to generate reflection rays with a cosine density. Note that \mathcal{P}^{-1} will always exist, but may not be analytical, in which case

numerical function inversion must be used. A nice graphical explanation of the role of \mathcal{P} in sample generation can be found in the book by Sillion and Puech [1994].

One nice thing about generating a sequence of nonuniform random numbers using the inverse of the cumulative distribution function is that the ordering of the ξ_i and the $\mathcal{P}^{-1}(\xi_i)$ will be the same, and thus, if we input stratified ξ_i instead of a completely random sequence, then $\mathcal{P}^{-1}(\xi_i)$ will also be stratified.

For a two-dimensional probability density function $\varphi : [\psi_0, \psi_1] \times [\tau_0, \tau_1] \rightarrow \mathbb{R}^+$, where φ is defined with respect to measure $d\mu(\psi, \tau) = m(\psi, \tau)d\psi d\tau$, we can do a similar transformation by using the marginal density function $\varphi_\psi : [\psi_0, \psi_1] \rightarrow \mathbb{R}^+$:

$$\varphi_\psi(\psi) = \int_{\tau_0}^{\tau_1} \varphi(\psi, \tau') m(\psi, \tau') d\tau'. \quad (7)$$

The function φ_ψ is a valid density function that represents the density of ψ values without reference to τ . Once a ψ value has been chosen with $\psi_i = \mathcal{P}_\psi^{-1}(\xi_i)$, the τ value can be chosen with density

$$\varphi_{T|\Psi}(\tau | \psi_i) = \frac{\varphi(\psi_i, \tau)}{\varphi_\psi(\psi_i)}, \quad (8)$$

where $\varphi_{T|\Psi}(\tau | \psi_i)$ is the conditional probability of τ given ψ_i . In the special case where φm is separable, meaning $\varphi(\psi, \tau)m(\psi, \tau)$ can be expressed as $b_1(\psi)b_2(\tau)$, where both $b_1 : [\psi_0, \psi_1] \rightarrow \mathbb{R}^+$ and $b_2 : [\tau_0, \tau_1] \rightarrow \mathbb{R}^+$ are probability density functions, then we can choose $\psi_i = B_1^{-1}(\xi_i)$ and $\tau_i = B_2^{-1}(\xi_{i+1})$, where B_1 is the cumulative probability distribution function associated with b_1 and where B_2 is the cumulative probability distribution function associated with b_2 .

2.3 Design Strategies for Importance Sampling

The key step in an implementation of Monte Carlo integration is the choice of the probability density function of the sample points. This is something of a “black art” because the goal is not to have the lowest variance for a given number of samples, but is instead to have the lowest variance for a given execution time. This is an important distinction because samples that produce low variance may take many times longer to generate than higher-variance samples [Kirk and Arvo 1991a]. In this section we outline the general rules we have used when choosing our density functions for sampling. We apply these rules when estimating the values of specific integrals in Section 3.

The integrals that arise in lighting calculations are often of the form

$$I = \int_{\mathcal{S}} f_1(\psi) f_2(\psi) \cdots f_r(\psi) d\mu(\psi), \quad (9)$$

where the f_i are strictly nonnegative functions. The f_i may vary a great deal in both shape and execution cost. They may be known a priori, or their

character may only be attained through point sampling. In this section we discuss several strategies for the design of density functions for Monte Carlo integration of integrals of the form of eq. (9).

First, we categorize the f_i into one of the following general behaviors:

- *Known, low variation (LO)*: f_i is known a priori and has low variation over \mathcal{S} ;
- *Known, high variation (HI)*: f_i is known a priori and has high variation over \mathcal{S} ; and
- *Unknown (UN)*: f_i is known only by sampling.

By grouping terms in eq. (9), we can rewrite it as a product of at most three functions of the types listed (note that grouping two terms of type **HI** could result in one term of type **LO**, so this grouping must be done carefully). The integrals in this paper will always contain an integral of type **UN** (usually evaluated by tracing a ray), and some weighting function (usually representing local properties of the surface) that is a combination of functions of type **LO** and **HI**. We consider several cases.

If the integrand is a product of a function f_{HI} of type **HI** and a function f_{UN} of type **UN**, then ideally we would set $\varphi(\psi) \propto f_{HI}(\psi)f_{UN}(\psi)$. Since f_{UN} is unknown, we might evaluate it at several locations and approximate it, but in image synthesis applications, this is not usually practical because evaluations are so expensive and f_{UN} may be quite complex. If we can assume nothing about f_{UN} , we might still want to make $\varphi(\psi) \propto f_{HI}(\psi)$, which is often done in practice. This is in some limited sense optimal, as is shown in the Appendix.

If the integrand is a product of a function f_{LO} of type **LO** and a function f_{UN} of type **UN**, we might follow the logic of the $f_{HI}(\psi)f_{UN}(\psi)$ case, but we might also just want to set $\varphi(\psi)$ to be a constant. This is because the improvements in variance caused by a choice of $\varphi(\psi) \propto f_{LO}(\psi)$ would probably be made inconsequential by the large variance that $f_{UN}(\psi)$ can cause. We certainly do not want to bother making $\varphi(\psi) \propto f_{LO}(\psi)$ if this would add much computation time to generating the random points with density φ .

In summary, when the integrand is a known weighting function times an unknown expensive function, we let φ be proportional to the components of the weighting function that are not roughly constant. If this is too expensive, we make $\varphi(\psi)$ proportional to an approximation to the components of the weighting function that are not roughly constant. More formally, we rewrite $f(\psi)$ as $f_{LO}(\psi)f_{HI}(\psi)f_{UN}(\psi)$ and try to find a way to generate samples $\psi_i \sim \varphi$ with $\varphi(\psi) \propto f_{HI}(\psi)$. If this is difficult, we use $\varphi(\psi) \propto \hat{f}_{HI}(\psi)$, where \hat{f}_{HI} is an approximation of $f_{HI}(\psi)$. This is the general strategy we use in later sections to choose density functions.

2.4 Direct Lighting

Here we review the rendering equation that governs our approximation to light transport and separate out the direct lighting integral that we later solve with Monte Carlo integration. The rendering equation is usually writ-

ten in one of two basic ways: one as an integral of radiometric quantities⁴ over solid angles, and one as an integral equation of radiometric quantities over all surfaces. It can be written down in terms of all directions $\hat{\omega}'$ visible to \mathbf{x} (as done by Immel et al. [1986]):

$$L_s(\mathbf{x}, \hat{\omega}) = L_e(\mathbf{x}, \hat{\omega}) + \int_{\mathcal{U}} \rho(\mathbf{x}, \hat{\omega}, \hat{\omega}') L_f(\mathbf{x}, \hat{\omega}') (-\hat{\omega}' \cdot \hat{n}) d\sigma(\hat{\omega}'), \quad (10)$$

where $L_s(\mathbf{x}, \hat{\omega})$ is the surface radiance⁵ of \mathbf{x} in direction $\hat{\omega}$, $L_e(\mathbf{x}, \hat{\omega})$ is the emitted surface radiance in direction $\hat{\omega}$ at \mathbf{x} , \mathcal{U} is the unit hemisphere of incoming directions oriented about \hat{n} , $\rho(\mathbf{x}, \hat{\omega}, \hat{\omega}')$ is the bidirectional reflectance distribution function (BRDF) at \mathbf{x} , $L_f(\mathbf{x}, \hat{\omega}')$ is the field radiance from direction $\hat{\omega}'$ incident at \mathbf{x} , and σ is the solid angle measure. The geometric quantities are illustrated in Figure 1.

Although eq. (10) has the form of an integral, the evaluation of the field radiance $L_f(\mathbf{x}, \hat{\omega}')$ typically requires the evaluation of $L_s(\mathbf{x}', \hat{\omega}')$ for some point \mathbf{x}' “seen” by \mathbf{x} in direction $-\hat{\omega}'$. This leads to another way the rendering equation is often written as an integral equation over all surfaces (as done by Kajiya [1986]⁶):

$$\begin{aligned} L_s(\mathbf{x}, \hat{\omega}) &= L_e(\mathbf{x}, \hat{\omega}) + \int_{\mathcal{X}} g(\mathbf{x}, \mathbf{x}') \rho(\mathbf{x}, \hat{\omega}, \hat{\omega}') \\ &\quad \times L_s(\mathbf{x}', \hat{\omega}') (-\hat{\omega}' \cdot \hat{n}) \frac{(\hat{\omega}' \cdot \hat{n}')}{\|\mathbf{x} - \mathbf{x}'\|^2} dA(\mathbf{x}'), \end{aligned} \quad (11)$$

where \mathcal{X} is the set of all points on surfaces; $g(\mathbf{x}, \mathbf{x}')$ is the *geometry term*, which is zero if there is an obstruction between \mathbf{x} and \mathbf{x}' , and is one otherwise; $\|\mathbf{x} - \mathbf{x}'\|$ is the distance between \mathbf{x} and \mathbf{x}' ; and A is the area measure.

⁴ In this paper we use a slight variant of the radiometric terms and symbols standardized by the Illumination Engineering Society (IES) [American National Standard Institute 1986; Rea 1993]. To simplify notation, we drop the spectral term even though all of the radiometric quantities we use are spectral quantities. So to denote spectral radiance we use L rather than the L_λ recommended by the IES.

⁵ Radiance is a quantity that is defined for all points in space, directions $\omega \in \mathcal{O}$, and wavelengths $\lambda \in \Lambda$ in a homogeneous material, and can be thought of as a function $L: \mathbb{R}^3 \times \mathcal{O} \times \Lambda \rightarrow \mathbb{R}$. This function is not defined at boundaries between materials (i.e., at surfaces) because the light abruptly changes direction [Shirley 1990]. To solve this problem, authors commonly refer to “incoming” and “outgoing” radiance. Because the outgoing radiance is what we perceive when we look at a surface, the outgoing radiance is sometimes called “surface radiance,” and the incoming radiance from the “field” is called “field radiance” [Arvo et al. 1994]. For transparent surfaces, both the surface and field radiances are defined on all directions in \mathcal{O} . For opaque surfaces, the surface radiance $L_s(\mathbf{x}, \hat{\omega})$ is defined on $\hat{\omega}$ in the outgoing hemisphere of directions Ω , and the field radiance $L_f(\mathbf{x}, \hat{\omega}')$ is defined on $\hat{\omega}'$ in the incoming hemisphere of directions \mathcal{U} .

⁶ Technically, Kajiya’s equation is not written in terms of radiance, but it can be mapped to eq. (11). An interesting discussion of the relationship of Kajiya’s equation and the radiance-based equation can be found in Arvo [1995, sect. 6.5].

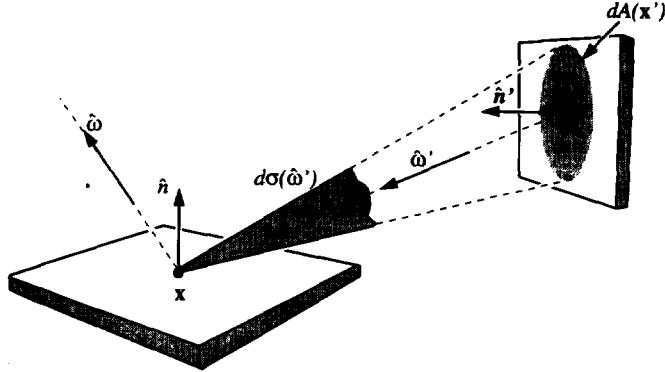


Fig. 1. Geometry for rendering equation.

The direct lighting, $L_d(\mathbf{x}, \hat{\omega})$, can be written as an integral over all directions by dropping the reflected light at \mathbf{x}' in eq. (10):

$$L_d(\mathbf{x}, \hat{\omega}) = \int_{\mathcal{U}} \rho(\mathbf{x}, \hat{\omega}, \hat{\omega}') L_e(\mathbf{x}', \hat{\omega}') (-\hat{\omega}' \cdot \hat{n}) d\sigma(\hat{\omega}'), \quad (12)$$

where \mathbf{x}' is the point seen by \mathbf{x} in direction $-\hat{\omega}'$. In a program a ray $(\mathbf{x} + t(-\hat{\omega}'))$ will be traced from \mathbf{x} in direction $-\hat{\omega}'$ to find \mathbf{x}' . The direct lighting can also be written as an integral over all points by dropping the reflected light at \mathbf{x}' from eq. (11):

$$L_d(\mathbf{x}, \hat{\omega}) = \int_{\mathcal{X}} g(\mathbf{x}, \mathbf{x}') \rho(\mathbf{x}, \hat{\omega}, \hat{\omega}') L_e(\mathbf{x}', \hat{\omega}') \frac{(\hat{\omega}' \cdot \hat{n}')}{\|\mathbf{x} - \mathbf{x}'\|^2} dA(\mathbf{x}'). \quad (13)$$

Here, the geometry term $g(\mathbf{x}, \mathbf{x}')$ is calculated by sending the “shadow” ray $(\mathbf{x} + t(\mathbf{x}' - \mathbf{x}))$ into the environment. The geometry term is one if $t \approx 1$ at the first surface the ray hits.

When Monte Carlo integration is used to approximate eq. (12), a probability density function $q(\hat{\omega}')$ (where q is defined using the solid angle measure σ) is built over \mathcal{U} . Applying eq. (2) with $N = 1$ (one sample) gives

$$L_d(\mathbf{x}, \hat{\omega}) \approx \frac{\rho(\mathbf{x}, \hat{\omega}, \hat{\omega}') L_e(\mathbf{x}', \hat{\omega}') (-\hat{\omega}' \cdot \hat{n})}{q(\hat{\omega}')}, \quad (14)$$

where $\hat{\omega}' \sim q$. Using Monte Carlo integration to approximate eq. (13) with a probability density function $p(\mathbf{x}')$, where p is defined on \mathcal{X} using the area measure A , yields

$$L_d(\mathbf{x}, \hat{\omega}) \approx g(\mathbf{x}, \mathbf{x}') \rho(\mathbf{x}, \hat{\omega}, \hat{\omega}') L_e(\mathbf{x}', \hat{\omega}') \frac{(-\hat{\omega}' \cdot \hat{n})}{p(\mathbf{x}') \|\mathbf{x} - \mathbf{x}'\|^2}, \quad (15)$$

where $\mathbf{x}' \sim p$.

All that must be done to implement the direct lighting calculation is to choose which integral we will evaluate (the directional integral or the integral over all points), to create a density function, and somehow to generate random samples according to that density. The details of these steps are described for direct lighting from a single luminaire in Section 3 and for multiple luminaires in Section 4.

3. DIRECT LIGHTING FROM ONE LUMINAIRE

To calculate $L_d(\mathbf{x}, \hat{\omega})$ for a single luminaire using Monte Carlo integration, we can use eq. (14), with a density that is positive for the directions subtended by the luminaire; or we can use eq. (15), with a probability density function that is positive at all visible points on the luminaire (in practice, it will be positive for all points on the luminaire because the visibility is hard to establish *a priori*). For large luminaires (such as the sky), sampling from solid angles might make sense, but for luminaires that are small in angle space as seen from the points where we are calculating direct lighting (e.g., the sun and most man-made luminaires), most of the integrand will be zero, and it will thus be hard to design good density functions. Therefore, we use eq. (15) because it is usually easier to design densities on the surfaces themselves. For reflection rays from specular or near-specular surfaces, it is usually more natural to use eq. (14), but we do not address reflection rays in this paper. The issues involved in choosing between eqs. (14) and (15) were discussed in more detail by Shirley and Wang [1992]. Several of the examples used in this section have been used by other authors (e.g., by Kirk and Arvo [1991b], Lange [1991], and Pattanaik [1993]), who used the same transformation strategies.

3.1 Designing Probability Density Functions

In designing a density function $p(\mathbf{x}')$ to use in eq. (15), we can immediately set $p(\mathbf{x}') = 0$ wherever $L_e(\mathbf{x}', \hat{\omega}') = 0$ (all points not on the luminaire). A perfect (zero-variance) estimate would result from

$$p(\mathbf{x}') = Cg(\mathbf{x}, \mathbf{x}')\rho(\mathbf{x}, \hat{\omega}, \hat{\omega}')L_e(\mathbf{x}', \hat{\omega}')(-\hat{\omega}' \cdot \hat{n}) \frac{(\hat{\omega}' \cdot \hat{n}')}{\|\mathbf{x} - \mathbf{x}'\|^2}, \quad (16)$$

where C is a normalization constant. As is usually pointed out in the Monte Carlo literature, the integral must be evaluated before we know C , so this is not very practical.

Instead, as suggested in Section 2.3, we attempt to partition the integrand into f_{LO} , f_{HI} , and f_{UN} . If we assume that the luminaire is diffuse or has a radiance that does not vary much across the surface as seen by most directions and if we assume that $\rho(\mathbf{x}, \hat{\omega}, \hat{\omega}')$ is roughly constant over the directions subtended by the luminaire, then our integrand can be classified as

$$\underbrace{\rho(\mathbf{x}, \hat{\omega}, \hat{\omega}')L_e(\mathbf{x}', \hat{\omega}')}_{\text{LO}} \underbrace{(-\hat{\omega}' \cdot \hat{n}) \frac{(\hat{\omega}' \cdot \hat{n}')}{\|\mathbf{x} - \mathbf{x}'\|^2}}_{\text{HI}} \underbrace{g(\mathbf{x}, \mathbf{x}')}_{\text{UN}}. \quad (17)$$

So following the strategy of using the **HI** component of the function guide for our design of p we get

$$p(\mathbf{x}') \propto (-\hat{\omega}' \cdot \hat{n}) \frac{(\hat{\omega}' \cdot \hat{n}')}{\|\mathbf{x} - \mathbf{x}'\|^2}. \quad (18)$$

Note that for some luminaires some of the components of eq. (18) might be roughly constant. For example, if the luminaire is small as seen from \mathbf{x} , then $(-\hat{\omega}' \cdot \hat{n})$ is roughly constant. If the luminaire is also planar, then $(\hat{\omega}' \cdot \hat{n}')$ is roughly constant. If the luminaire's extent is small compared to its distance from \mathbf{x} , then $\|\mathbf{x} - \mathbf{x}'\|^{-2}$ is roughly constant. If we can move any of these terms into the **LO** component of the integrand, then it will make p easier to design. For the remainder of this section, we discuss how to choose p for several simple luminaire shapes. In particular, we discuss probability density functions for luminaires that are approximated by spheres, polygons, disks, and cylinders. We use the case of spheres to illustrate the mechanics of choosing nonuniformly random points, and the case of triangles as an example of how to deal with nonseparable densities, and we then summarize results for the other shapes. More details on the derivations of the material in this section can be found in Wang [1993].

3.2 Sampling Spherical Luminaires

In this section we discuss three different density functions that could be used to choose samples on a spherical luminaire with radius r and center $\mathbf{c} = (c_x, c_y, c_z)$. To classify the components of $(-\hat{\omega}' \cdot \hat{n})(\hat{\omega}' \cdot \hat{n}')\|\mathbf{x} - \mathbf{x}'\|^{-2}$, we first notice that, even for small distant luminaires, $(\hat{\omega}' \cdot \hat{n}')$ varies from zero to one and so it is a **HI** component. The other two terms, $(-\hat{\omega}' \cdot \hat{n})$ and $\|\mathbf{x} - \mathbf{x}'\|^{-2}$, can be **LO** or **HI** depending on the parameters of the sphere and on the location of \mathbf{x} . For example, if the sphere almost touches \mathbf{x} , then both terms are **HI**, but for a distant source such as the sun, both terms are **LO**. So, in different circumstances, different density functions are useful.

3.2.1 Uniform Density. For the purposes of debugging, a constant density function is useful. Because the density function must have unit volume on the surface of the sphere, this density is just the inverse of the sphere area:

$$p_1(\mathbf{x}') = \frac{1}{4\pi r^2}.$$

To generate a random point on the sphere, we define the density in a spherical coordinate system with θ being the polar angle (the positive \hat{z} axis is $\theta = 0$) and with ϕ being the azimuthal angle (the positive \hat{x} axis is $\phi = 0$). This yields a density with a related differential measure $dA(\mathbf{x}') = r^2 \sin \theta d\theta d\phi$. Using the techniques of Section 2.2, we separate the density $1/(4\pi r^2)$ and $r^2 \sin \theta d\theta d\phi$ into two density functions: $p_a(\phi) = 1/(2\pi)$ and $p_b(\theta) = \frac{1}{2} \sin \theta$. This gives us $(\theta, \phi) = (\arccos(1 - 2\xi_1), 2\pi\xi_2)$. Converting to

Cartesian coordinates gives $\mathbf{x}' = [c_x + r \cos \phi \sin \theta, c_y + r \sin \phi \sin \theta, c_z + r \cos \theta]^T$, which can be simplified to

$$\mathbf{x}' = \begin{bmatrix} c_x + 2r \cos(2\pi\xi_2) \sqrt{\xi_1(1-\xi_1)} \\ c_y + 2r \sin(2\pi\xi_2) \sqrt{\xi_1(1-\xi_1)} \\ c_z + r(1-2\xi_1) \end{bmatrix}. \quad (19)$$

An immediate optimization can be to take samples only on the portion of the sphere visible to \mathbf{x} . However, because the uniform area case is ideal for debugging, we instead leave it as is and move on to nonuniform densities.

3.2.2 Sampling Uniformly in Directional Space. The first nonuniform density we might try is $p(\mathbf{x}') \propto (\hat{\omega}' \cdot \hat{n}')$. This turns out to be just as complicated as sampling with $p(\mathbf{x}') \propto (\hat{\omega}' \cdot \hat{n}') \|\mathbf{x} - \mathbf{x}'\|^{-2}$, so we instead discuss that here. We observe that sampling on the luminaire using $(\hat{\omega}' \cdot \hat{n}') \|\mathbf{x} - \mathbf{x}'\|^{-2}$ is the same as using a density constant function $q_2(\hat{\omega}')$ defined in the space of directions subtended by the luminaire as seen from \mathbf{x} . We now use a coordinate system defined with \mathbf{x} at the origin, and a right-handed orthonormal basis with $\hat{w} = \mathbf{c} - \mathbf{x}/\|\mathbf{c} - \mathbf{x}\|$ and $\hat{v} = (\hat{w} \times \hat{n})/\|(\hat{w} \times \hat{n})\|$ (see Figure 2). We also define (θ', ϕ') to be the azimuthal and polar angles with respect to the $\hat{u}\hat{v}\hat{w}$ coordinate system.

The maximum θ' that includes the spherical luminaire is given by

$$\theta'_{\max} = \arcsin\left(\frac{r}{\|\mathbf{x} - \mathbf{c}\|}\right) = \arccos\sqrt{1 - \left(\frac{r}{\|\mathbf{x} - \mathbf{c}\|}\right)^2}.$$

Thus, a uniform density (with respect to a solid angle) within the cone of directions subtended by the sphere is just the reciprocal of the solid angle $2\pi(1 - \cos \theta'_{\max})$ subtended by the sphere:

$$q_2(\hat{\omega}') = \frac{1}{2\pi\left(1 - \sqrt{1 - (r/\|\mathbf{x} - \mathbf{c}\|)^2}\right)}.$$

And we get

$$\begin{bmatrix} \theta' \\ \phi' \end{bmatrix} = \begin{bmatrix} \arccos\left(1 - \xi_1 + \xi_1\sqrt{1 - \left(\frac{r}{\|\mathbf{x} - \mathbf{c}\|}\right)^2}\right) \\ 2\pi\xi_2 \end{bmatrix}.$$

This gives us the direction to \mathbf{x}' . To find the actual point, we need to find the first point on the sphere in that direction. The ray in that direction is just $(\mathbf{x} + t\hat{a})$, where \hat{a} is given by

$$\hat{a} = \begin{bmatrix} u_x & v_x & w_x \\ u_y & v_y & w_y \\ u_z & v_z & w_z \end{bmatrix} \begin{bmatrix} \cos \phi' \sin \theta' \\ \sin \phi' \sin \theta' \\ \cos \theta' \end{bmatrix}.$$

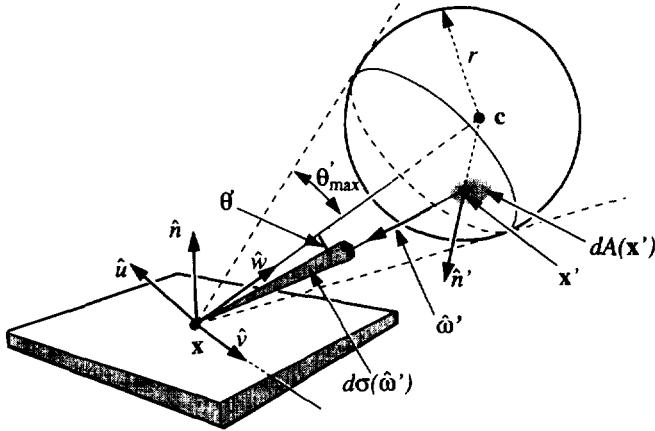


Fig. 2. Geometry for spherical luminaire.

We must also calculate $p_2(\mathbf{x}')$, the probability density function with respect to the area measure (recall that the density function q_2 is defined in a solid angle space). Since we know that q_2 is a valid probability density function using the σ measure and we know that $d\sigma(\hat{\omega}') = dA(\mathbf{x}')(\hat{\omega}' \cdot \hat{n}')/\|\mathbf{x} - \mathbf{x}'\|^2$, we can relate any probability density function $q(\hat{\omega}')$ with its associated probability density function $p(\mathbf{x}')$:

$$q(\hat{\omega}') = \frac{(\hat{\omega}' \cdot \hat{n}') p(\mathbf{x}')}{\|\mathbf{x} - \mathbf{x}'\|^2}. \quad (20)$$

So we can solve for $p_2(\mathbf{x}')$:

$$p_2(\mathbf{x}') = \frac{(\hat{\omega}' \cdot \hat{n}')}{2\pi\|\mathbf{x} - \mathbf{x}'\|^2 \left(1 - \sqrt{1 - (r/\|\mathbf{x} - \mathbf{c}\|)^2}\right)}.$$

For further details on sampling the sphere with p_2 , see Wang [1992].

3.2.3 Cosine-Weighted Solid Angle. To bring the $(-\hat{\omega}' \cdot \hat{n})$ term into the density function, we again use the space of directions: $q_3(\hat{\omega}') \propto (-\hat{\omega}' \cdot \hat{n})$. Integrating this density function gives the constant of proportionality for the case that the entire sphere is above the horizon as seen from \mathbf{x} (if this is not the case, we must use a different density function because q_3 as stated before would have regions with negative values), and thus, the density function is

$$q_1(\hat{\omega}') = \frac{(-\hat{\omega}' \cdot \hat{n})}{\pi(\hat{\omega}' \cdot \hat{n}) \sin^2 \theta_{\max}}.$$

The direction $\hat{\omega}'$ can be found by the following two formulas:

$$\xi_1 = \frac{\sin^2 \theta'}{\sin^2 \theta_{\max}},$$

and

$$\xi_2 = \frac{(\hat{u} \cdot \hat{n}) \sin \theta' \sin \phi' + (\hat{w} \cdot \hat{n}) \phi' \cos \theta'}{2\pi(\hat{u} \cdot \hat{n}) \cos \theta'}. \quad (21)$$

So we get $\theta' = \arcsin(\sqrt{\xi_1} \sin \theta_{\max})$, but we must numerically invert the second formula to solve for ϕ' . For our implementation, we use a simple binary search to perform this inversion, which requires tens of evaluations of eq. (21) and so is quite costly; in practice, we do not usually use p_3 to sample spherical luminaires. Getting the actual point \mathbf{x}' on the luminaire is handled in the same way as in Section 3.2.2, and the value of $p_3(\mathbf{x}')$ is found using eq. (20).

A spherical luminaire sampled with one sample per pixel for each of the three cases discussed in this section, along with a uniform sampling of the visible region of the sphere, is shown in Figure 3. Although the sphere sampled with p_3 generates the smoothest image, in practice we use p_2 because it is much less expensive than p_3 , and since we use multiple samples for edge and shadow antialiasing, we can tolerate the variance caused by p_2 .

3.3 Sampling Planar Luminaires

For any planar luminaire, the quantity $(-\hat{\omega}' \cdot \hat{n})(\hat{\omega}' \cdot \hat{n}')\|\mathbf{x} - \mathbf{x}'\|^{-2}$ stays roughly constant provided the luminaire subtends a relatively small solid angle at \mathbf{x} . So, for planar luminaires, we use a constant density function to generate most of our images. In our implementation we have disks, triangles, rectangles, and general polygons.

3.3.1 Sampling Polygonal Luminaires. For simple polygons (polygons that have no holes) we have resorted to picking uniformly from the bounding rectangle of the polygon and to using a rejection technique based on whether the sample is in the polygon. This test can be accomplished using a ray test in the plane of the polygon [Haines 1989], which works for all polygons, even those with holes. This probability density function is just the inverse of the polygon area, which can be found using Stoke's theorem [Rokne 1991]. Because we use a rejection method, much of the stratification of the sampling patterns is lost, so illumination calculations for polygonal luminaires produce noisy images unless the number of samples is relatively large.

3.3.2 Sampling Disk Luminaires. To choose a random sample from a disk, we suppose that its center is at the \mathbf{c} , the radius is r , and that $\hat{n}' = \hat{w}$ in its $\hat{u}\hat{v}\hat{w}$ coordinate system. So $p_4(\mathbf{x}') = 1/(\pi r^2)$, and

$$\mathbf{x}' = \mathbf{c}' + \begin{bmatrix} u_x & v_x & w_x \\ u_y & v_y & w_y \\ u_z & v_z & w_z \end{bmatrix} \begin{bmatrix} r\sqrt{\xi_1} \cos(2\pi\xi_2) \\ r\sqrt{\xi_1} \sin(2\pi\xi_2) \\ 0 \end{bmatrix}.$$

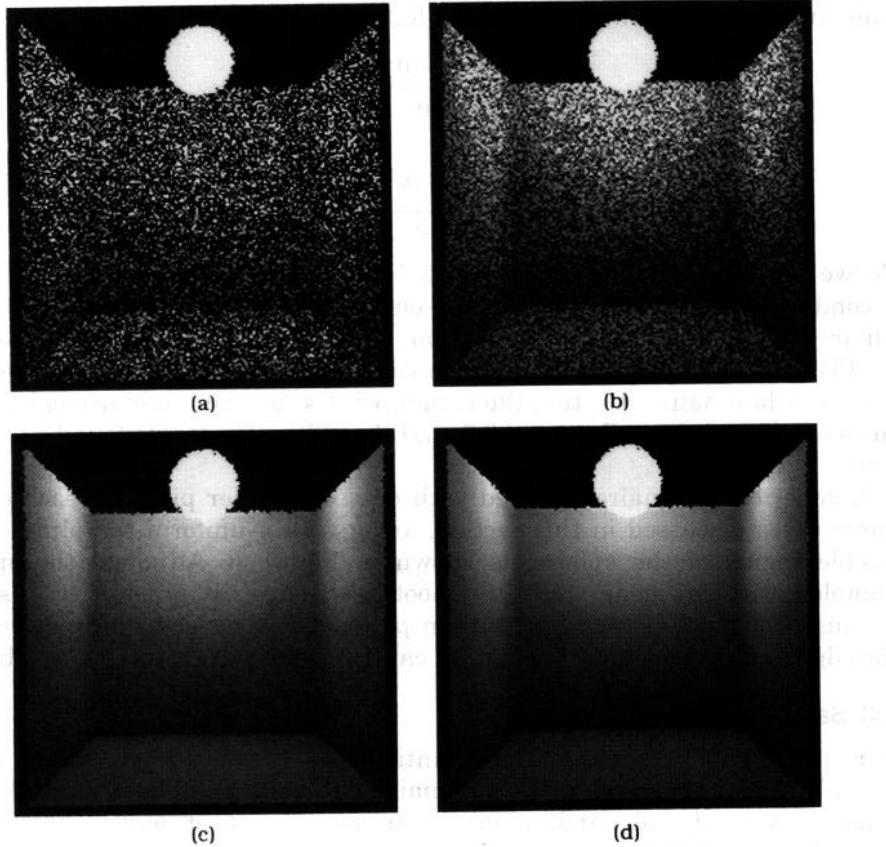


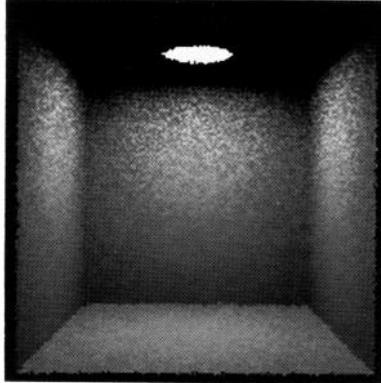
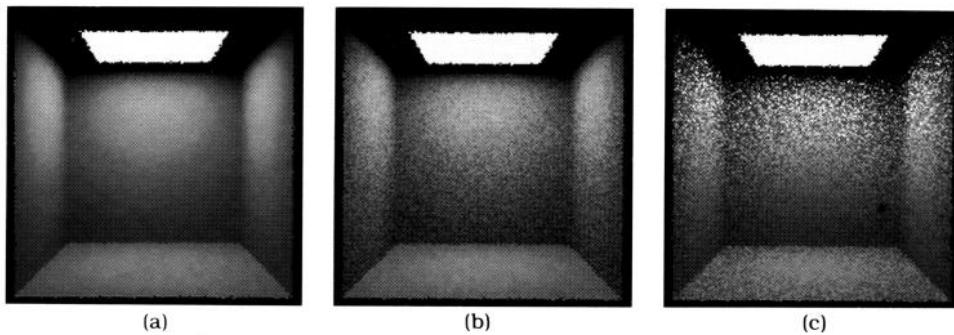
Fig. 3. Spherical luminaire sampled using (a) p_1 ; (b) p'_1 (uniform on visible portion); (c) p_2 ; and (d) p_3 .

An image with a disk luminaire sampled using p_4 with one sample per pixel is shown in Figure 4. The camera simulation work of Kolb et al. [1995] indicates that lower variances can be achieved by other stratification patterns on the disk.

3.3.3 Sampling Rectangular Luminaires. If a rectangle is defined with a point \mathbf{x}_0 and side vectors \mathbf{v}_1 and \mathbf{v}_2 , then a uniform random point is given by

$$\mathbf{x}' = \mathbf{x}_0 + \xi_1 \mathbf{v}_1 + \xi_2 \mathbf{v}_2$$

and has a density of $p_5(\mathbf{x}') = 2/\|\mathbf{v}_1 \times \mathbf{v}_2\|$. If the luminaire is entirely above the horizon, then it is also possible to choose a random point with density $p_6(\mathbf{x}') \propto (\hat{\omega}' \cdot \hat{n}') \|\mathbf{x} - \mathbf{x}'\|^{-2}$ and $p_7(\mathbf{x}') \propto (-\hat{\omega}' \cdot \hat{n})(\hat{\omega}' \cdot \hat{n}') \|\mathbf{x} - \mathbf{x}'\|^{-2}$; but the evaluation of p_6 and p_7 is costly, their implementation is difficult, and because of numerical inversion, they are also slow (see Wang [1993] for details). One sample per pixel image using p_5 , p_6 , and p_7 is shown in Figure 5. Note that the complex density functions perform better when the luminaire and \mathbf{x} are close to each other, as should be expected because all of the terms in $(-\hat{\omega}' \cdot \hat{n})(\hat{\omega}' \cdot \hat{n}') \|\mathbf{x} - \mathbf{x}'\|^{-2}$ vary most in those regions.

Fig. 4. Disk luminaire sampled using p_4 .Fig. 5. Rectangular luminaire sampled with (a) p_5 , (b) p_6 , and (c) p_7 .

3.3.4 Sampling Triangular Luminaires. To generate uniform random points on a triangle, we use the natural barycentric coordinates of the triangle and note that the determinant of the Jacobian matrix that takes a triangle defined with barycentric coordinates to a triangle with an unnormalized coordinate system is a constant. This means that, if we pick random barycentric coordinates for one triangle, we can use them for any other triangle, and the random point will still be uniform. We now go into some detail on the mechanics of choosing a uniform \mathbf{x}' , because when the uniform density function is written in terms of the two coordinates on the triangle, it is not separable. Using barycentric coordinates, every point on the plane containing a triangle with vertices \mathbf{p}_0 , \mathbf{p}_1 , and \mathbf{p}_2 can be described by $\mathbf{x}' = \mathbf{p}_0 + \beta(\mathbf{p}_1 - \mathbf{p}_0) + \gamma(\mathbf{p}_2 - \mathbf{p}_0)$,⁷ and the \mathbf{x}' is in the triangle if and only

⁷ Three barycentric coordinates are often used with $\mathbf{x}' = [\mathbf{o} + \alpha(\mathbf{p}_0 - \mathbf{o}) + \beta(\mathbf{p}_1 - \mathbf{o}) + \gamma(\mathbf{p}_2 - \mathbf{o})]$, where \mathbf{o} is the coordinate origin and $\alpha + \beta + \gamma = 1$. We just use the substitution $\alpha = (1 - \beta - \gamma)$.

if ($\beta > 0$), ($\gamma > 0$), and ($\beta + \gamma < 1$). Integrating the constant one across the triangle gives

$$\int_{\gamma=0}^1 \int_{\beta=0}^{1-\gamma} d\beta d\gamma = \frac{1}{2}.$$

This means that the value of our density function is 2 in barycentric coordinates (and $1/A$ in Cartesian space, where A is the area of the triangle, implying that the determinant of the Jacobian that transforms from barycentric to Cartesian space has the value $2A$).

To perform an inversion, we must first find densities that correspond to β and γ separately. If β and γ were independent, we could simply use the marginal density for each. In this case, β is dependent on γ . Therefore, we use the marginal density for γ , $f_G(\gamma)$, and the conditional density of β given γ , $f_{B|G}(\beta | \gamma)$. By definition,

$$f_G(\gamma) = \int_0^{1-\gamma} f(\beta, \gamma) d\beta = 2(1 - \gamma)$$

and

$$f_{B|G}(\beta | \gamma) = \frac{f(\beta, \gamma)}{f_G(\gamma)} = \frac{1}{1 - \gamma}.$$

Now we can perform the inversion by letting $\xi_1 = F_G(\gamma')$ and $\xi_2 = F_{B|G}(\beta' | \gamma')$, and by solving for β' and γ' . Here F_G and $F_{B|G}$ are the cumulative densities corresponding to the random variables γ' and $\beta' | \gamma'$, respectively. Again, by definition,

$$\xi_1 = F_G(\gamma') = \int_0^{\gamma'} f_G(\gamma) d\gamma = 2\gamma' - \gamma'^2.$$

Solving for γ' we get $\gamma' = 1 - \sqrt{1 - \xi_1}$. For β'

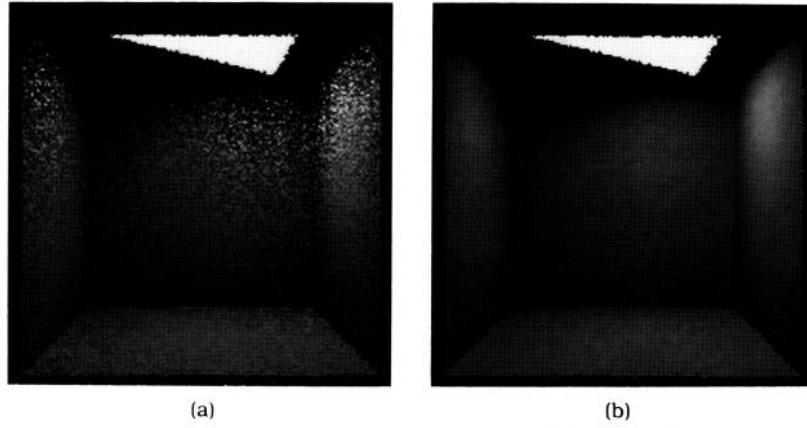
$$\xi_2 = F_{B|G}(\beta' | \gamma') = \int_0^{\beta'} f_{B|G}(\beta' | \gamma') d\beta = \frac{\beta'}{1 - \gamma'}.$$

Thus, $\beta' = \xi_2(1 - \gamma') = \xi_2\sqrt{1 - \xi_1}$. Therefore,

$$\mathbf{x}' = \mathbf{p}_0 + \xi_2\sqrt{1 - \xi_1}(\mathbf{p}_1 - \mathbf{p}_0) + 1 - \sqrt{1 - \xi_1}(\mathbf{p}_2 - \mathbf{p}_0). \quad (22)$$

This is the same method presented by Pattanaik [1993] in his dissertation. Turk [1990] presented a faster method for uniformly sampling triangles that makes stratification deteriorate slightly.

A triangular luminaire sampled using this constant p_8 , as well as using $p_9 \propto (-\hat{\omega}' \cdot \hat{n})(\hat{\omega}' \cdot \hat{n}')\|\mathbf{x} - \mathbf{x}'\|^{-2}$ (as with spheres and rectangles, this only works if the entire luminaire is above the horizon), is shown in Figure 6. Details on using p_9 (which requires a numerical inversion of an equation and so is slow) can be found in Wang [1993]. Arvo [1995a] recently developed a method to sample a triangle according to a solid angle $((\hat{\omega}' \cdot \hat{n}')\|\mathbf{x} - \mathbf{x}'\|^{-2})$; his method requires no numerical inversions, so it is fast. His approach to that problem has promise for dealing with p_9 without numerical inversion.

Fig. 6. Triangular luminaire sampled using (a) p_8 and (b) p_9 .

3.4 Sampling Cylindrical Luminaires

If we have a cylinder of radius r_c with its axis along the \hat{w} axis, and boundaries $w = 0$ and $w = w_{\max}$, then we can choose a point in the $\hat{u}\hat{v}\hat{w}$ coordinate system that is uniform on the cylinder: $[u, v, w]^T = [r_c \cos(2\pi\xi_1), r_c \sin(2\pi\xi_1), \xi_2 w_{\max}]^T$. If the base origin of the $\hat{u}\hat{v}\hat{w}$ coordinate system is \mathbf{o} , then a random point in Cartesian coordinates is

$$\mathbf{x}' = \mathbf{o} + \begin{bmatrix} u_x & v_x & w_x \\ u_y & v_y & w_y \\ u_z & v_z & w_z \end{bmatrix} \begin{bmatrix} r_c \cos(2\pi\xi_1) \\ r_c \sin(2\pi\xi_1) \\ \xi_2 w_{\max} \end{bmatrix}.$$

A more practical (lower-variance) cylinder sampling strategy that approximates a density function proportional to $(\hat{\omega}' \cdot \hat{n}')$ is presented in Zimmerman [1995].

3.5 Discussion

Because the density functions developed in this section assume that $\rho(\mathbf{x}, \hat{\omega}, \hat{\omega}')$ and $L_e(\mathbf{x}', \hat{\omega}')$ are roughly constant in the directions subtended by the luminaire, an increase in variance can be introduced by specular reflection or directional luminaires. If $\rho(\mathbf{x}, \hat{\omega}, \hat{\omega}')$ or $L_e(\mathbf{x}', \hat{\omega}')$ vary more than other terms in $L_e(\mathbf{x}, \hat{\omega})$, then all of the methods discussed in this section may have a high variance. For example, if we use a rectangle with a texture controlling emission to represent a TV screen, then the variance will be roughly proportional to the variance of the pixels in the texture. When $\rho(\mathbf{x}, \hat{\omega}, \hat{\omega}')$ is the most important term, we want $p \propto \rho(\mathbf{x}, \hat{\omega}, \hat{\omega}')$, as in Ward [1991] and Lange [1991], and when $L_e(\mathbf{x}', \hat{\omega}')$ is the key factor, we want to use $p \propto L_e(\mathbf{x}', \hat{\omega}')$. The paper by Veach and Guibas [1995] looked at this problem in a new way that combines directional and area sampling in an automatic way. Their method shows great promise for dealing with that tricky problem.

There are many luminaire shapes we have not discussed that might be of interest. For example, neon lights could be considered swept disks, and common light bulbs can be modeled as surfaces of revolution. However, in practice we can often use the simple shapes discussed in this section in preference to the complex shapes of the real luminaires. For example, Figure 7 shows a room lit by fluorescent fixtures that are approximated by rectangles for the lighting calculations at other surfaces. This idea is used extensively in Ward's *Radiance* program [1994], where the far-field photometry of the entire fixture is known (see Glassner [1995] for details on getting CIE or IES standard far-field photometry for real luminaires), and the luminaire is displayed with a complex geometry. This can be thought of as a special case of geometric simplification as used by Rushmeier et al. [1993].

4. DIRECT LIGHTING FROM MANY LUMINAIRES

Traditionally, when N_L luminaires are in a scene, the direct lighting integral is broken into N_L separate integrals [Cook et al. 1984]. This implies that at least N_L samples must be taken to approximate the direct lighting or that some bias must be introduced (as done by Ward where small-value samples are not calculated [Ward 1991]). Instead, we leave the direct lighting integral intact and design a probability density function over all N_L luminaires.

As an example, suppose we have two luminaires, ℓ_1 and ℓ_2 , and we devise two probability functions $p_1(\mathbf{x}')$ and $p_2(\mathbf{x}')$, where $p_i(\mathbf{x}') = 0$ for \mathbf{x}' not on ℓ_i and $p_i(\mathbf{x}')$ is found by a method such as one of those described in Section 3 for generating \mathbf{x}' on ℓ_i . These functions can be combined into a single density over both lights by applying a weighted average:

$$p(\mathbf{x}') = \alpha p_1(\mathbf{x}') + (1 - \alpha)p_2(\mathbf{x}'),$$

where $\alpha \in (0, 1)$. We can see that p is a probability density function because its integral over the two luminaires is one, and it is strictly positive at all points on the luminaires. Densities that are "mixed" from other densities are often called *mixture densities*, and the coefficients α and $(1 - \alpha)$ are called the *mixing weights* [Titterington et al. 1985].

To estimate $L = (L_1 + L_2)$, where L is the direct lighting and L_i is the lighting from luminaire ℓ_i , we first choose a random canonical pair (ξ_1, ξ_2) and use it to decide which luminaire will be sampled. If $0 \leq \xi_1 < \alpha$, we estimate L_1 with \hat{L}_1 using the methods described in Section 3 to choose \mathbf{x}' and to evaluate $p_1(\mathbf{x}')$, and we estimate L with \hat{L}_1/α . If $\xi_1 \geq \alpha$, then we estimate L with $\hat{L}_2/(1 - \alpha)$. In either case, once we decide which source to sample, we cannot use (ξ_1, ξ_2) directly because we have used some knowledge of ξ_1 . So, if we choose ℓ_1 (so $\xi_1 < \alpha$), then we choose a point on ℓ_1 using the random pair $(\xi_1/\alpha, \xi_2)$. If we sample ℓ_2 (so $\xi_1 \geq \alpha$), then we use the pair $((\xi_1 - \alpha)/(1 - \alpha), \xi_2)$. This way, a collection of stratified samples will remain stratified in some sense. Note that it is to our advantage to have ξ_1 stratified in one dimension, as well as having the pair (ξ_1, ξ_2) stratified in two dimensions, so that the ℓ_i we choose will be stratified over many (ξ_1, ξ_2) pairs, so some multijittered sampling method may be helpful (e.g., Chiu et al. [1994]).



Fig. 7. Rectangular luminaires used as Impostors

This basic idea used to estimate $L = (L_1 + L_2)$ can be extended to N_L luminaires by mixing N_L densities

$$p(\mathbf{x}') = \alpha_1 p_1(\mathbf{x}') + \alpha_2 p_2(\mathbf{x}') + \cdots + \alpha_{N_L} p_{N_L}(\mathbf{x}'), \quad (23)$$

where the α_i s sum to one, and where each α_i is positive if ℓ_i contributes to the direct lighting. The value of α_i is the probability of selecting a point on the ℓ_i , and p_i is then used to determine which point on ℓ_i is chosen. If ℓ_i is chosen, then we estimate L with \hat{L}_i/α_i . Given a pair (ξ_1, ξ_2) , we choose ℓ_i by enforcing the conditions

$$\sum_{j=1}^{i-1} \alpha_j < \xi_1 < \sum_{j=1}^i \alpha_j.$$

To sample the light, we can use the pair (ξ'_1, ξ'_2) , where

$$\xi'_1 = \frac{\xi_1 - \sum_{j=1}^{i-1} \alpha_j}{\alpha_i}.$$

This basic process is shown in Figure 8. It cannot be overstressed that it is important to “reuse” the random samples in this way to keep the variance low, in the same way we use stratified sampling (jittering) instead of random sampling in the space of the pixel.⁸ To choose the point on the luminaire ℓ_i given (ξ'_1, ξ'_2) , we can use the same types of p_i for luminaires as used in the last section. The question remaining is what to use for α_i .

4.1 Constant α_i

The simplest way to choose values for α_i was proposed by Lange [1991] (and this method is also implied in the figure in Kajiya [1986, p. 148]), where all weights are made equal: $\alpha_i = 1/N_L$ for all i . This would definitely make a valid estimator because the α_i sum to one and none of them is zero. Unfortunately, in many scenes this estimate would produce a high variance (when the L_i are very different, as occurs in most night “walk-throughs”).

4.2 Linear α_i

Suppose we had perfect p_i defined for all of the luminaires. A zero-variance solution would then result if we could set $\alpha_i \propto L_i$, where L_i is the contribution from the i th luminaire. If we can make α_i approximately proportional to L_i , then we should have a fairly good estimator. We call this the *linear method* of setting α_i because the time used to choose one sample is linearly proportional to N_L , the number of luminaires.

⁸ It is also important not to introduce any bias by correlated points in screen space and “luminaire space.” Typically, a set of N sample pairs is chosen in screen space, a different set of N sample pairs is chosen for luminaire space, and these are combined into N quadruples in screen-luminaire space. If there is some regularity in how these points are generated (e.g., the first point is in a particular corner of canonical space), then the mapping between pairs should be random [Cook 1986].

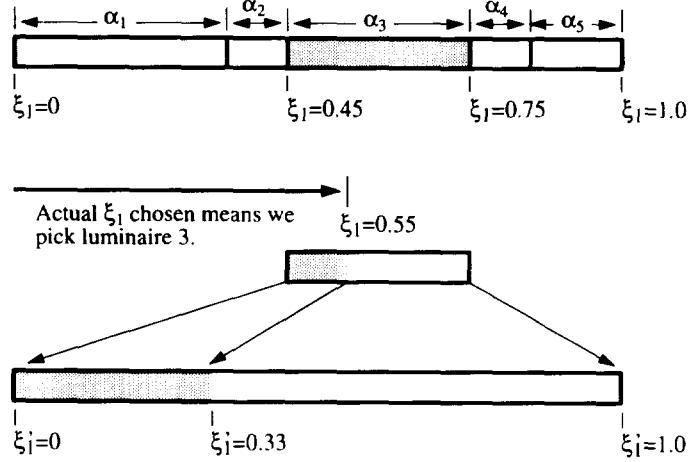


Fig. 8. Diagram of mapping ξ_l to choose α_i and resulting remapping to new canonical sample ξ'_l .

To obtain such α_i , we get an estimated contribution \hat{L}_i at \mathbf{x} by approximating eq. (13) for γ_i with the geometry term $g(\mathbf{x}, \mathbf{x}')$ set to one. These \hat{L}_i 's (from all luminaires) can be directly converted to α_i by scaling them so that their sum is one:

$$\alpha_i = \frac{\hat{L}_i}{\hat{L}_1 + \hat{L}_2 + \dots + \hat{L}_{N_L}}. \quad (24)$$

This method of choosing α_i is valid because all potentially visible luminaires will end up with positive α_i . We should expect the highest variance in areas where shadowing occurs, because this is where setting the geometry term to one causes α_i to be a poor estimate of α_i .

Figure 9 shows a scene with 100 rectangular luminaires sampled with 49 rays per pixel using the linear method. This method of setting α_i was first used by Shirley [1990a] and was first theoretically justified by Shirley and Wang [1991]. Figure 10 shows a scene with 100 rectangular luminaires sampled using the linear method and 49 rays per pixel.

Implementing the linear α_i method has several subtleties. We implemented a method for each type of luminaire that estimated L_i for a particular \mathbf{x} and $\rho(\mathbf{x}, \hat{\omega}, \hat{\omega}')$. If the entire luminaire is below the tangent plane at \mathbf{x} , then the estimate for \hat{L}_i should be zero. An easy mistake to make (which we made in our initial implementation) is to set \hat{L}_i to zero if the center of the luminaire is below the horizon. This will make α_i take the one value that is not allowed: an incorrect zero. Such a bug will become obvious in pictures of spheres illuminated by luminaires that subtend large solid angles, but for many scenes such errors are not noticeable (the figures in Shirley [1990a] had this bug, but it was not noticeable). To overcome this problem, we make sure that for a polygonal luminaire all of its vertices are below the horizon

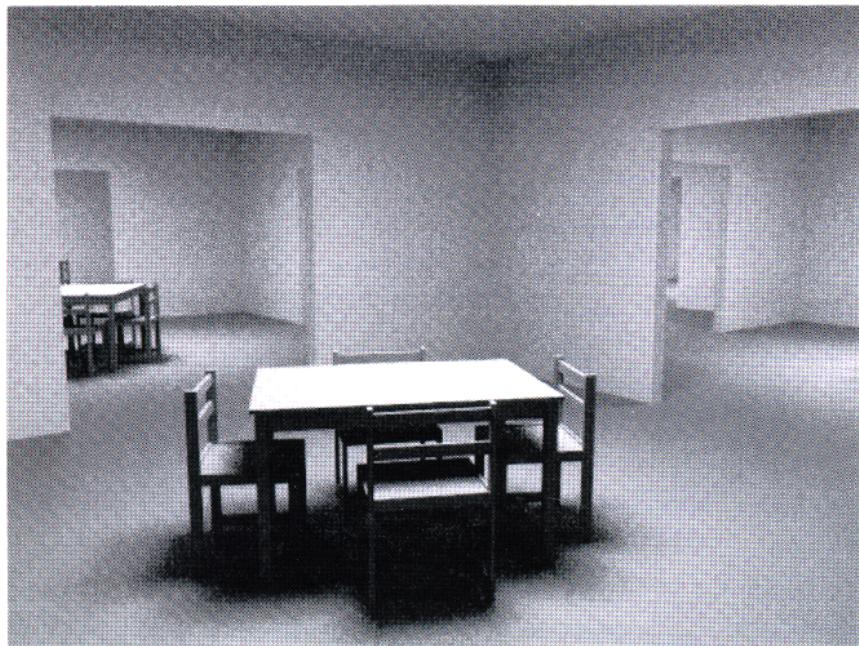


Fig. 9. Image that used the "linear" method with 100 luminaires.

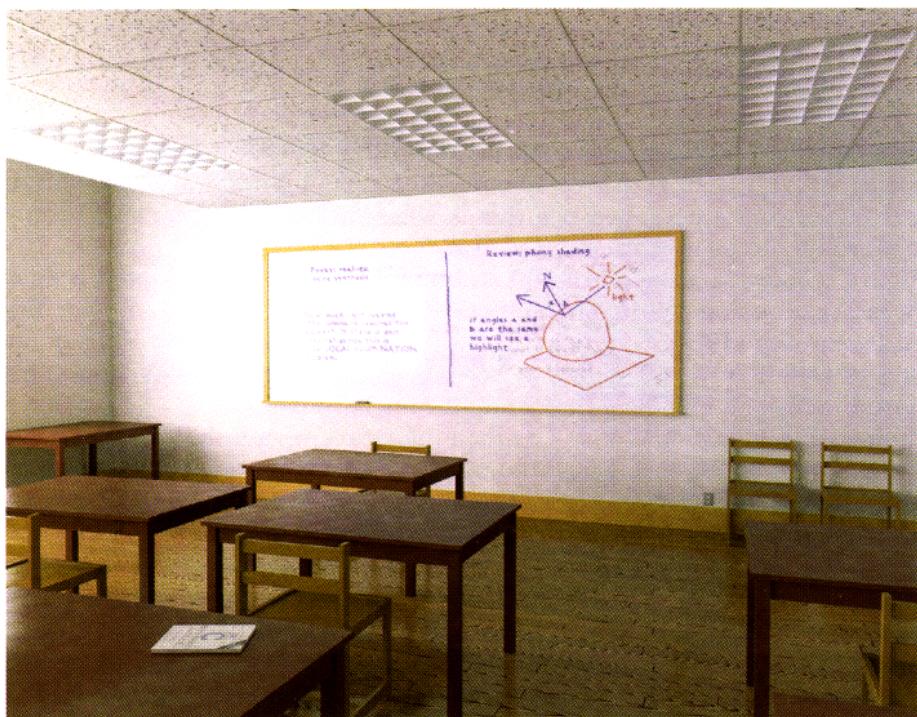


Fig. 10. Image that used the "linear" method with four luminaires.

ACM Transactions on Graphics, Vol. 15, No. 1, January 1996.

before it is given a zero probability of being sampled. For a spherical luminaire, we check that the center of the luminaire is a distance greater than the sphere radius under the horizon plane before it is given a zero probability of being sampled.

4.3 Spatial Subdivision

In the linear method, choosing α_i based on an estimated contribution requires querying every luminaire in the scene. This is acceptable for many scenes, but if N_L is large (thousands or millions), even that might be too slow. In such scenes at most a few hundred luminaires (and usually at most tens) will contribute significantly to the radiance at any particular point. Suppose we can partition $\mathcal{L} = \{\ell_1, \dots, \ell_{N_L}\}$ into two subsets $\mathcal{L}_{\text{bright}}$ and \mathcal{L}_{dim} , where $\mathcal{L}_{\text{bright}}$ is the set of luminaires that are “important” at \mathbf{x} (i.e., they contribute significantly to $L_s(\mathbf{x}, \hat{\omega})$ for some $\hat{\omega}$), and $\mathcal{L}_{\text{dim}} = (\mathcal{L} - \mathcal{L}_{\text{bright}})$. With these two subsets of \mathcal{L} , we can construct low-cost α_i .

Suppose that the size of $\mathcal{L}_{\text{bright}}$ is N_b , that $\mathcal{L}_{\text{bright}} = \{b_1, b_2, \dots, b_{N_b}\}$, and that $\mathcal{L}_{\text{dim}} = \{d_1, d_2, \dots, d_{N_d}\}$, where $N_d = N_L - N_b$. If N_L is large and we have partitioned \mathcal{L} correctly, then N_b should be much less than N_d . For each b_i in $\mathcal{L}_{\text{bright}}$, we estimate L_i in the same way we did in the linear method. We now assume that all members of \mathcal{L}_{dim} contribute approximately the same amount \hat{L}_d as a random element d_r of \mathcal{L}_{dim} . To pick a random element in \mathcal{L}_{dim} , we use a rejection method where we repeatedly choose a luminaire at random from \mathcal{L} (which we store in a linear array) until we find a light that is not in $\mathcal{L}_{\text{bright}}$ (that is faster than finding one that is in \mathcal{L}_{dim} because we expect that $N_d \gg N_b$). If we want a better estimate for the average contribution of a luminaire in \mathcal{L}_{dim} , we can take more than one sample. The appropriate number of such samples to take in \mathcal{L}_{dim} is scene dependent and is something that should be studied further. For our implementation we have taken one sample. This gives estimates \hat{L}_r for contributions from all luminaires without consulting more than one (or a few, if more than one sample is taken) element of \mathcal{L}_{dim} and will be reasonably accurate for the important luminaires. We can now construct a probability density function by setting α_i equal to the normalized value of \hat{L}_i :

$$\alpha_i = \frac{\hat{L}_i}{\hat{L}_1 + \hat{L}_2 + \dots + \hat{L}_{N_L}} = \frac{\hat{L}_i}{\sum_{j \in \mathcal{L}_{\text{bright}}} \hat{L}_j + N_d \hat{L}_r} \quad (25)$$

where \hat{L}_r is our estimate for the contribution of the random luminaire from \mathcal{L}_{dim} .

The difficult part of this method is deciding which luminaires are in $\mathcal{L}_{\text{bright}}$ for a particular \mathbf{x} . As pointed out by Kok and Jansen [1991], a luminaire that is responsible for a large fraction of the radiance of \mathbf{x} is likely to be responsible for a large fraction of the radiance of the neighboring points of \mathbf{x} . This implies that we can use a spatial subdivision scheme to precompute a $\mathcal{L}_{\text{bright}}$ list for each spatial cell in the spatial subdivision structure. For a particular cell, a luminaire is put in $\mathcal{L}_{\text{bright}}$ if it might contribute more than a threshold average spectral radiance to a diffuse surface within the cell.

A simple way to determine whether the maximum potential contribution of a luminaire is above the threshold is to evaluate eq. (13) with $g(\mathbf{x}, \mathbf{x}')$ and $(-\hat{\omega}' \cdot \hat{n})$ each set to one, and $\rho(\mathbf{x}, \hat{\omega}, \hat{\omega}')$ set to ρ_{\max} for all points (implementationally a large number of points) on the boundary of the spatial subdivision cell. Because specular surfaces are usually handled using reflection rays (and thus eq. (12); see Shirley and Wang [1992] for details) rather than shadow rays, we can expect ρ_{\max} not to be extremely large. In Figure 11 a set of luminaires and their regions of importance are shown. The stratification of the scene into boxes is arbitrary.

An easy way to choose the subdivision cells is to use the leaf cells of the conventional subdivision structure (e.g., the octree leaves of a Glassner-style octree used for ray intersection acceleration [Glassner 1994]) and to maintain a separate $\mathcal{L}_{\text{bright}}$ list at each leaf. This is a finer than needed subdivision for scenes where the number of objects is much greater than the number of luminaires. Advantages of this are that no luminaire lists need to be constructed for empty cells and that the characteristics of reflective objects can be used to construct the lists (because the objects in each cell are known in advance). Instead, we have implemented a separate *light octree* that recursively subdivides itself until each leaf is at a maximum allowed depth or when the size of $\mathcal{L}_{\text{bright}}$ for that cell is below a specified limit. The depth and size limits are similar to those in a conventional octree, and their values are even less well understood by us so far. To avoid excess subdivision, we check whether the minimum contribution of an important luminaire to a cell is above the threshold. If all members of $\mathcal{L}_{\text{bright}}$ are thus determined to be in $\mathcal{L}_{\text{bright}}$ for any possible descendant of that cell, we do not subdivide.

Another difficulty in building the light octree is what to use for the average radiance threshold. For our program we assume that we know what radiance L_w would map to white on the display device (white would be an rgb triple of [255, 255, 255] on a typical 24-bit frame-buffer-driven CRT). We set the radiance threshold T to be some fraction of L_w .⁹ Because our display device typically has 8 bits per channel, we usually make the threshold a few percent of L_w . Such a threshold will ensure that any luminaire that can change the pixel color more than a few steps will be included in $\mathcal{L}_{\text{bright}}$. If L_w is chosen correctly, then the sum of the contributions of luminaires should be no more than L_w , and thus, the number of important luminaires should be less than one hundred for most scenes.

To characterize important versus unimportant luminaires, we associate an axis-aligned *influence box* with each luminaire that includes all points that might include that luminaire in its $\mathcal{L}_{\text{bright}}$ list. When deciding whether a luminaire is important to a cell, we just check whether the cell and the influence box overlap, and if so, then the luminaire is treated as an important light source. To avoid an important luminaire being overlooked, the influence

⁹ L_w should be chosen according to a perceptual viewer model, such as the model implemented by Tumblin and Rushmeier [1993]. L_w may also vary from pixel to pixel if spatially varying mappings are used [Chiu et al. 1993]. Such models will become increasingly important as physically based rendering becomes more popular.

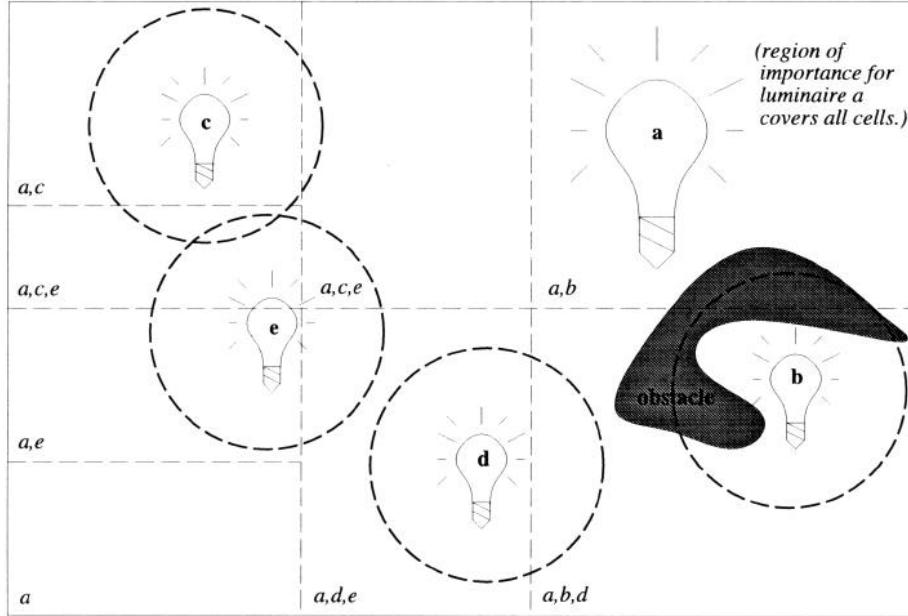


Fig. 11. Subdivision of space where luminaire is a member of $\mathcal{L}_{\text{bright}}$ for a cell if it has the potential to contribute significantly to the surface radiance of a point in that cell.

box must contain every point whose reflected radiance might be larger than a predefined threshold T when receiving the light only from that luminaire. Although this gives the chance of a nonimportant luminaire being selected as an important one, this is preferred to the risk of missing an important luminaire because our estimate is \hat{L}_i/α_i , which will blow up if \hat{L}_i is large and if α_i is small, but will just change a small component if \hat{L}_i is small, regardless of α_i .

To decide where a luminaire might cause an object to change its radiance more than T , we first examine a point source at point \mathbf{x}' with radiant intensity¹⁰ distribution $I(\hat{\omega}')$. The potential contribution due to this source can be bounded by replacing $\rho(\mathbf{x}, \hat{\omega}, \hat{\omega}')$ with ρ_{\max} and $(-\hat{\omega}' \cdot \hat{n})$ with one:

$$L_s(\mathbf{x}, \hat{\omega}) \leq \frac{\rho_{\max} I(\hat{\omega}')}{\|\mathbf{x} - \mathbf{x}'\|^2}. \quad (26)$$

To make sure this potential contribution is below T , we make sure the distance is above a certain quantity:

$$\|\mathbf{x} - \mathbf{x}'\| \geq \sqrt{\frac{\rho_{\max} I(\hat{\omega}')}{T}}. \quad (27)$$

¹⁰ The radiant intensity, $I(\hat{\omega}')$, is the power per unit solid angle in direction $\hat{\omega}'$.

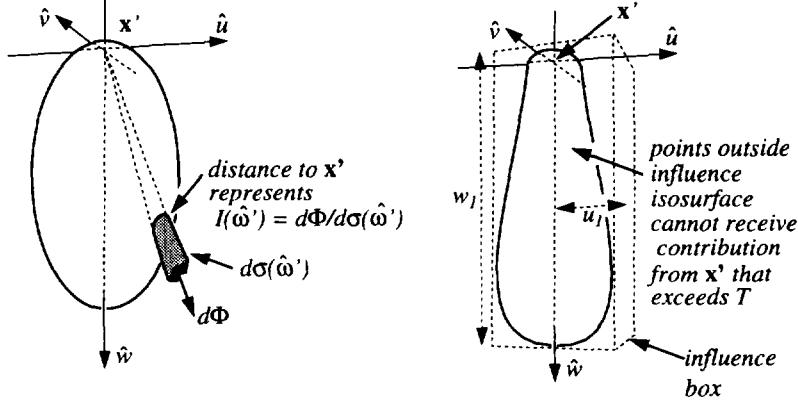


Fig. 12. (a) Goniometric diagram of radiant intensity of a point source. (b) Associated influence isosurface where the maximum radiance contribution is T .

Note that the set of points \mathbf{x} that have a potential above T is the inside of an *influence isosurface* that depends only on $I(\hat{\omega}')$ (see Figure 12). The box that bounds this influence isosurface is that luminaire's influence box.

As an example of finding an influence isosurface and the resulting influence box of a finite-sized luminaire, we now consider a spherical luminaire with a directionally isotropic emission pattern and emitted power Φ , and a resulting far-field radiant intensity of $I(\hat{\omega}') = \Phi/(4\pi)$. Allowing for a finite radius r just makes the radius r_i of the sphere's influence isosurface expand by r :

$$r_i = \sqrt{\frac{\rho_{\max}\Phi}{4\pi T}} + r. \quad (28)$$

If $I(\hat{\omega}')$ of the luminaire has \hat{w} as an axis of symmetry, then $I(\hat{\omega}')$ will depend only on $(\hat{w} \cdot \hat{\omega}')$. If we define ϑ to be the angle between \hat{w} and $\hat{\omega}'$, then we have a function $I(\vartheta)$ that fully describes the emission of the luminaire. To find the influence box that contains points $[u_0, u_1] \times [v_0, v_1] \times [w_0, w_1]$, we first note that because of symmetry $u_0 = v_0 = -u_1 = -v_1$. For simplicity, we assume that $I(\vartheta) = 0$ for $\vartheta \geq \pi/2$, so $w_0 = 0$ (when this is not true, the analysis for $\vartheta \geq \pi/2$ is analogous to the analysis for $\vartheta < \pi/2$). To find w_1 we note that on the potential isosurface at angle ϑ the value of w can be found from eq. (27):

$$w = \cos \vartheta \sqrt{\frac{\rho_{\max} I(\hat{\omega}')}{T}}.$$

To find w_1 we can differentiate w^2 (which will be maximum when w is maximum) with respect to ϑ and find where this derivative is zero. This occurs when

$$-2 \sin \vartheta I(\vartheta) + \cos \vartheta \frac{dI(\vartheta)}{d\vartheta} = 0. \quad (29)$$

Similar analysis for u_1 yields that the largest u inside the influence isosurface occurs at ϑ such that

$$2 \cos \vartheta I(\vartheta) + \sin \vartheta \frac{dI(\vartheta)}{d\vartheta} = 0. \quad (30)$$

Equations (29) and (30) can help us find the influence box of a point "phong" luminaire with power Φ located at \mathbf{x}' with

$$I(\vartheta) = \frac{(e+1)\Phi \cos^e \vartheta}{2\pi}.$$

Differentiating $I(\vartheta)$ yields

$$\frac{dI(\vartheta)}{d\vartheta} = -\frac{e(e+1)\Phi \cos^{e-1} \vartheta \sin \vartheta}{2\pi}.$$

This implies that w on the potential isosurface will be a maximum when $\vartheta = 0$, so

$$w_1 = \sqrt{\frac{\rho_{\max}(e+1)\Phi}{2\pi T}}.$$

Applying eq. (30) implies that u will be maximum on the potential isosurface when $(2 \cos^2 \vartheta = e \sin^2 \vartheta)$, so

$$u_1 = \left(\frac{e}{e+2} \right)^{e/2} \sqrt{\frac{\rho_{\max}\Phi}{\pi(e+2)T}}.$$

This gives us an influence box about a point phong source in the $\hat{u}\hat{v}\hat{w}$ coordinate system. To extend this result to polygonal luminaires, we can apply the point source analysis to a given vertex and add to u_1 the maximum u distance to any other vertex. To find the importance box in Cartesian coordinates, we apply the conservative strategy of finding the Cartesian axis-aligned bounding box of the $\hat{u}\hat{v}\hat{w}$ influence box.

Figure 13 shows two images of a pair of "sphereflakes" (generated using Eric Haines's procedural database software), with each sphereflake composed of 7381 spheres. In the image with 7381 luminaires, the light octree was used. In both cases, individual luminaires were sampled with a $(\hat{\omega}' \cdot \hat{n}') \|\mathbf{x} - \mathbf{x}'\|^2$ density function. The single-luminaire case used 10 samples per pixel, and the 7381-luminaire case used 40 samples per pixel and took about eight times as long as the 10-sample single-luminaire image. This means that the overhead of choosing the ray using the light octree does not swamp the cost of sending the shadow ray, even in this extreme case where half of the objects are luminaires. In less pathological scenes, the light octree should be smaller relative to the number of primitives, and performance should be even better.

5. DISCUSSION

Clearly, the linear and light octree methods are most useful for scenes with a large number of luminaires. Such scenes are becoming increasingly impor-

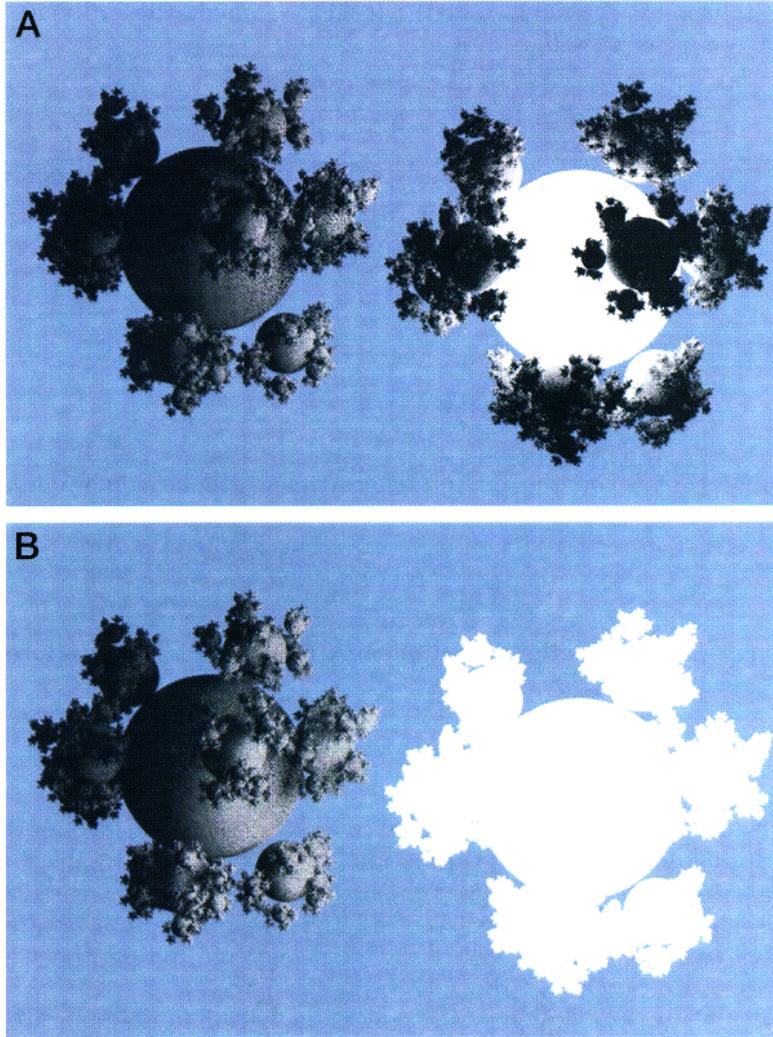


Fig. 13. (a) Single luminaire, 10 rays per pixel; (b) 7381 luminaires, 40 rays per pixel.

tant. In outdoor scenes, especially in urban settings, scenes with thousands and even hundreds of thousands of luminaires are commonplace. In opera and theater applications, hundreds or thousands of luminaires are common [Dorsey et al. 1991]. In infrared scenes, almost all surfaces are luminaires, so something such as the light octree is crucial. These methods also make it easy to use luminaires defined by many polygons or parametric patches. No matter how many patches define a light-bulb surface, it will receive only one shadow ray.

Recent work by Zimmerman and Shirley [1995] uses uniform spatial subdivision instead of a light octree. In addition, an estimate of visibility is

included for important light sources in each cell. This greatly reduces problems in environments where there is a very bright source that is blocked for the surface being rendered.

The recent work by Veach and Guibas [1995] can be combined with the linear method or light octree. The choice of the mixing weights α_i can be used to guide how many samples are directed toward directions covered by luminaires.

6. CONCLUSION

We have presented techniques for constructing probability spaces on luminaire surfaces for the direct lighting calculation. These techniques make the calculation of direct lighting from thousands of luminaires feasible. The chief limitation of the work is that it does not take into account the geometry term (visibility) or the BRDF. Extending the methods to include these terms is possible, but it will not be easy. The geometry term in particular should probably be attacked using probabilistic methods, because the visibility problem in high-complexity scenes is very difficult and exact solutions will take too long.

The basic rationale for the techniques presented in this paper is that direct lighting should not be calculated to a higher accuracy than necessary. This is very similar in concept to Kajiya's argument that we should not expend much work for deep parts of the ray tree [Kajiya 1986]. It is certainly not always true that one shadow ray per viewing ray is optimal, however. For the 100-luminaire case, one shadow ray is better than 100 shadow rays, but two or three might be better still. This issue requires further investigation.

Even in scenes with only a few luminaires, the linear or light octree techniques presented in Section 4 can be useful. Many of the recent rendering techniques can be viewed in terms of replacing surfaces with *impostors*. An example of an impostor is a luminaire with zero reflectivity that is shaped like a reflecting patch and that *emits* light in roughly the same distribution and intensity as the reflecting patch *reflects* light. This concept has been used by Kok and Jansen [1991], Chen et al. [1991], Rushmeier et al. [1993], and Ward [1994]. Adding an impostor reduces the number of patches whose reflected light needs to be calculated and increases the number of luminaires. Using impostors becomes more attractive if the cost of direct lighting does not increase linearly with N_L , so we believe the techniques in this paper will be used with both radiosity and ray-tracing programs.

APPENDIX: PROOF OF OPTIMAL SAMPLING FOR WEIGHTED INTEGRANDS

In image synthesis we often have integrands that take the form of a product of a unit-volume nonnegative function $w : \mathcal{S} \rightarrow \mathbb{R}^+$ and a nonnegative function $f : \mathcal{S} \rightarrow \mathbb{R}^+$ whose values are only attainable through point sampling:

$$I = \int_{\mathcal{S}} w(\psi) f(\psi) d\mu(\psi).$$

To solve this by eq. (2), the optimal choice for the probability function is $\varphi(\psi) \propto w(\psi)f(\psi)$, but as is often pointed out, this choice requires us already to know the value of I . Instead, people often either choose uniform φ or set $\varphi(\psi) = w(\psi)$ [Cook 1986; Purgathofer 1987; Lange 1991]. If f happens to be a constant function $f(\psi) = C$ for some real number C , then $\varphi(\psi) = w(\psi)$ is indeed optimal. In this Appendix we make some strict assumptions about f that will usually not be true, but that are closer to the truth than assuming that f is constant. This will not result in a proof that in a real implementation we should let $\varphi(\psi) = w(\psi)$, but it does increase our confidence that this is not a bad way to choose φ . This section extends arguments in Shirley and Wang [1992] that letting $\varphi(\psi) = w(\psi)$ is in some sense optimal. The proof uses techniques similar to those used by Glassner [1995, sect. 7.5.2].

In graphics we usually repeatedly perform an integral for every member of a set $F = \{f_1, f_2, \dots, f_n\}$. To decide which φ to use, we could try to minimize the average variance of our estimate across all $f \in F$. We cannot do this with such a vague definition of F , but we can approximate our situation in graphics by assuming that we are no more likely to have large values of f in a particular point in S than in any other. This makes sense when thinking of S as being the support of a filter w on a pixel on the image plane, and f as being the radiance hitting the film plane; we do not expect the exact location of features in the luminance function to be correlated with the details of how the pixels are arranged. More formally, since we will be doing a variance proof, we assume that the average value of $f_i^2(\psi)$ is the same for all ψ :

$$\sum_{i=1}^n f_i^2(\psi) = \text{constant}.$$

The average variance \bar{V} of the estimator $I' = w(\psi)f(\psi)/\varphi(\psi)$ is

$$\bar{V} = \frac{1}{n} \sum_{i=1}^n \left[\int_S \frac{w^2(\psi)f_i^2(\psi)}{\varphi(\psi)} d\mu(\psi) - \left(\int_S w(\psi)f_i(\psi) d\mu(\psi) \right)^2 \right]. \quad (31)$$

Because the right-hand (squared) term is $\sum I_i$ (a constant) and because we assume that average squared values of members of F are not correlated with the evaluation point in S , then as n becomes large we should expect that the average value of $f^2(\psi)$ at any given ψ is some constant \bar{f}^2 . The average variance of the estimator I' is then just

$$\bar{V} = \frac{1}{n} \sum_{i=1}^n \left[\int_S \frac{w^2(\psi)\bar{f}^2}{\varphi(\psi)} d\mu(\psi) - I^2 \right].$$

Because the I^2 term is a constant, the sum and the $1/n$ term cancel each other (there are no more f_i terms), and minimizing \bar{V} is equivalent to finding a function φ that minimizes

$$I_v = \int_S \frac{w^2(\psi)}{\varphi(\psi)} d\mu(\psi) \quad (32)$$

with the constraint that φ is nonnegative and with $\mu(\varphi) = 1$.

To find a constrained function that minimizes a quantity, we use the calculus of variations with Lagrange multipliers. We first replace $\varphi(\psi)$ with $\varphi(\psi) + \alpha\gamma(\psi)$, where α is a constant and $\gamma(\psi)$ is a perturbation function. Because $\varphi(\psi) + \alpha\gamma(\psi)$ must be a valid probability density function, we restrict γ to have zero volume: $\int \gamma d\mu = 0$. We now define a new integral $I_c(\alpha)$ to be

$$I_c(\alpha) = \int_{\mathcal{S}} \left[\frac{w^2(\psi)}{\varphi(\psi) + \alpha\gamma(\psi)} + \lambda(\varphi(\psi) + \alpha\gamma(\psi)) \right] d\mu(\psi), \quad (33)$$

where λ is the Lagrange multiplier (a constant). The calculus of variations tells us that I_c is stationary (an extremum) when the partial derivative of I_c evaluated at zero is zero. That is,

$$\frac{\partial I_c(\alpha)}{\partial \alpha} = \int_{\mathcal{S}} \left[\frac{-\gamma(\psi)w^2(\psi)}{(\varphi(\psi) + \alpha\gamma(\psi))^2} + \lambda\gamma(\psi) \right] d\mu(\psi). \quad (34)$$

Setting $\partial I_c(0)/\partial \alpha = 0$ yields

$$\frac{\partial I_c(0)}{\partial \alpha} = \int_{\mathcal{S}} \left[\frac{\gamma(\psi)w^2(\psi)}{\varphi^2(\psi)} - \lambda\gamma(\psi) \right] d\mu(\psi) = 0. \quad (35)$$

Because the integral must vanish for any valid $\gamma(\psi)$, we get

$$w^2(\psi) = \lambda\varphi^2(\psi). \quad (36)$$

Since both w and φ are nonnegative, this implies that the intuitively appealing choice of $\varphi = w$ has some theoretical justification as well. The second derivative $\partial^2 I_c(0)/\partial \alpha^2$ is always positive, so setting $\varphi = w$ ensures a minimum variance.

ACKNOWLEDGMENTS

The authors wish to thank Jim Arvo, Randy Bramley, Andrew Glassner, Steve Marschner, Holly Rushmeier, Ken Torrance, Bruce Walter, and Greg Ward for their help on algorithms and terminology; Ken Chiu for help on the code involved in the project; and Bill Kubitz, Dennis Gannon, Bob Skeel, and Don Hearn for initial suggestions and directions where the work should be pushed. Steve Marschner provided a proof (not included in the paper) that stratification will never produce higher variance than unstratified sampling for properly chosen strata. We are also grateful to the reviewers, who helped to improve a particularly difficult-to-read draft version of the paper.

REFERENCES

- AIREY, J. M., AND OUEH-YOUNG, M. 1989. Two adaptive techniques let progressive radiosity outperform the traditional radiosity algorithm. Tech. Rep. TR89-20, Computer Science, Dept., Univ. of North Carolina, Chapel Hill, Aug.
- AMERICAN NATIONAL STANDARDS INSTITUTE. 1986. Nomenclature and definitions for illumination engineering. ANSI Rep. ANSI/IES RP-16-1986, American National Standards Institute.

- ARVO, J. 1995a. Stratified sampling of spherical triangles. In ACM Siggraph 95 Conference Proceedings. *Comput. Graph.* (Aug.), 437-438.
- ARVO, J. 1995b. Analytic methods for simulated light transport. Ph.D. thesis, Yale Univ., New Haven, Conn., Dec.
- ARVO, J., TORRANCE, K., AND SMITS, B. 1994. A framework for the analysis of error in global illumination algorithms. In ACM Siggraph 94 Conference Proceedings. *Comput. Graph.* 28, 3 (July).
- CHEN, S. E., RUSHMEIER, H., MILLER, G., AND TURNER, D. 1991. A progressive multi-pass method for global illumination. In ACM Siggraph 91 Conference Proceedings. *Comput. Graph.* 25, 4 (July), 165-174.
- CHIU, K., SHIRLEY, P., AND WANG, C. 1994. Multi-jittered sampling. In *Graphics Gems IV*, P. Heckbert, Ed. Academic Press, Boston, Mass., 370-374.
- CHIU, K., HERF, M., SHIRLEY, P., SWAMY, S., WANG, C., AND ZIMMERMAN, K. 1993. Spatially nonuniform scaling functions for high contrast images. In *Graphics Interface 93* (May), 245-244.
- COOK, R. L. 1986. Stochastic sampling in computer graphics. *ACM Trans. Graph.* 5, 1 (Jan.), 51-72.
- COOK, R. L., PORTER, T., AND CARPENTER, L. 1984. Distributed ray tracing. In ACM Siggraph 84 Conference Proceedings. *Comput. Graph.* 18, 4 (July), 165-174.
- DORSEY, J. O. B., SILLION, F. X., AND GREENBERG, D. P. 1991. Design and simulation of opera lighting and projection effects. In ACM Siggraph 91 Conference Proceedings. *Comput. Graph.* 25, 4 (July), 41-50.
- GLASSNER, A. S. 1994. Space subdivision for fast ray tracing. *IEEE Comput. Graph. Appl.* 4, 10, 15-22.
- GLASSNER, A. S. 1995. *Principles of Digital Image Synthesis*. Morgan-Kauffman, San Francisco, Calif.
- HAINES, E. 1989. Essential ray tracing algorithms. In *An Introduction to Ray Tracing*, A. S. Glassner, Ed. Academic Press, San Diego, Calif., 33-77.
- HALTON, J. H. 1970. A retrospective and prospective of the Monte Carlo method. *SIAM Rev.* 12, 1 (Jan.), 1-63.
- HAMMERSLEY, J. M., AND HANDSCOMB, D. C. 1964. *Monte Carlo Methods*. Wiley, New York.
- IMMEL, D. S., COHEN, M. F., AND GREENBERG, D. P. 1986. A radiosity method for non-diffuse environments. In ACM Siggraph 86 Conference Proceedings. *Comput. Graph.* 20, 4 (Aug.), 133-142.
- KAJIYA, J. T. 1986. The rendering equation. In ACM Siggraph 86 Conference Proceedings. *Comput. Graph.* 20, 4 (Aug.), 143-150.
- KALOS, M. H., AND WHITLOCK, P. A. 1986. *Monte Carlo Methods*. Wiley, New York.
- KIRK, D., AND ARVO, J. 1991a. Unbiased sampling techniques for image synthesis. In ACM Siggraph 91 Conference Proceedings. *Comput. Graph.* 25, 4 (July), 153-156.
- KIRK, D., AND ARVO, J. 1991b. Unbiased variance reduction for global illumination. In *Proceedings of the 2nd Eurographics Workshop on Rendering* (Barcelona, May).
- KOK, A. J. F., AND JANSEN, F. W. 1991. Source selection for the direct lighting calculation in global illumination. In *Proceedings of the 2nd Eurographics Workshop on Rendering* (Barcelona, May), 75-82.
- KOK, A. J. F., AND JANSEN, F. W. 1992. Adaptive sampling of area light sources in ray tracing including diffuse interreflection. In *Eurographics 92. Comput. Graph. Forum* 11, 3, 289-298.
- KOLB, C., HANRAHAN, P., AND MITCHELL, D. 1995. A realistic camera model for computer graphics. In ACM Siggraph 95 Conference Proceedings. *Comput. Graph.* (Aug.), 317-324.
- LANGE, B. 1991. The simulation of radiant light transfer with stochastic ray-tracing. In *Proceedings of the 2nd Eurographics Workshop on Rendering*.
- LEE, M. E., REDNER, R. A., AND USELTON, S. P. 1985. Statistically optimized sampling for distributed ray tracing. In ACM Siggraph 85 Conference Proceedings. *Comput. Graph.* 19, 3 (July), 61-68.
- MALLEY, T. J. V. 1988. A shading method for computer generated images. Master's thesis, Computer Science Dept., Univ. of Utah, Salt Lake City, June.

- MITCHELL, D. P. 1991. Spectrally optimal sampling for distribution ray tracing. In ACM Siggraph 91 Conference Proceedings. *Comput. Graph.* 25, 4 (July).
- PATTANAIK, S. N. 1993. Computational methods for global illumination and visualization of complex 3D environments. Ph.D. thesis, Computer Science Dept., Birla Institute of Technology and Science, Pilani, India, Feb.
- PURGATHOFER, W. 1987. A statistical method for adaptive stochastic sampling. *Comput. Graph.* 11, 2, 157–162.
- REA, M. S., ED. 1993. *The Illumination Engineering Society Lighting Handbook*. 8th ed. Illumination Engineering Society, New York.
- ROKNE, J. 1991. The area of a simple polygon. In *Graphics Gems II*, J. Arvo, Ed. Academic Press, San Diego, Calif., 5–6.
- RUSHMEIER, H. E. 1988. Realistic image synthesis for scenes with radiatively participating media. Ph.D. thesis, Mechanical Eng. Dept., Cornell Univ., Ithaca, N.Y., May.
- RUSHMEIER, H. E., PATTERSON, C., AND VEERASAMY, A. 1993. Geometric simplification for indirect illumination calculations. In *Proceedings Graphics Interface 93* (May), 227–236.
- SCHLICK, C. 1991. The acne problem. *Ray Tracing News* 4, 1 (Mar.).
- SHIRLEY, P. 1990a. A ray tracing method for illumination calculation in diffuse-specular scenes. In *Proceeding of Graphics Interface 90* (May), 205–212.
- SHIRLEY, P. 1990b. Physically based lighting calculations for computer graphics. Ph.D. thesis, Univ. of Illinois, Urbana-Champaign, Nov.
- SHIRLEY, P. 1991. Discrepancy as a quality measure for sampling distributions. In *Eurographics 91* (Sept.), 183–193.
- SHIRLEY, P., AND WANG, C. 1991. Direct lighting by Monte Carlo integration. In *Proceedings of the 2nd Eurographics Workshop on Rendering*.
- SHIRLEY, P., AND WANG, C. 1992. Distribution ray tracing: Theory and practice. In *Proceedings of 3rd Eurographics Workshop on Rendering*, 200–209.
- SHIRLEY, P., SUNG, K., AND BROWN, W. 1991. A ray tracing framework for global illumination systems. In *Proceedings of Graphics Interface 91* (June), 117–128.
- SHREIDER, Y. A. 1966. *The Monte Carlo Method*. Pergamon Press, New York.
- SILLION, F. X., AND PUECH, C. 1989. A general two-pass method integrating specular and diffuse reflection. In ACM Siggraph 89 Conference Proceedings. *Comput. Graph.* 23, 3 (July), 335–344.
- SILLION, F. X., AND PUECH, C. 1994. *Radiosity and Global Illumination*. Morgan-Kaufman, San Francisco, Calif.
- SPANIER, J., AND MAIZE, E. H. 1994. Quasi-random methods for estimating integrals using relatively small samples. *SIAM Rev.* 36, 1 (Mar.), 18–44.
- TITTERINGTON, D. M., SMITH, A. F. M., AND MAKOV, U. E. 1985. *The Statistical Analysis of Finite Mixture Distributions*. Wiley, New York.
- TUMBLIN, J., AND RUSHMEIER, H. 1993. Tone reproduction for realistic computer generated images. *IEEE Comput. Graph. Appl.* 13, 6 (Nov.), 42–48.
- TURK, G. 1990. Generating random points in triangles. In *Graphics Gems*, A. Glassner, Ed. Academic Press, New York.
- VEACH, E., AND GUIBAS, L. 1995. Optimally combining sampling techniques for Monte Carlo rendering. In ACM Siggraph 95 Conference Proceedings. *Comput. Graph.* (Aug.), 419–428.
- WANG, C. 1992. Physically correct direct lighting for distribution ray tracing. In *Graphics Gems 3*, D. Kirk, Ed. Academic Press, New York.
- WANG, C. 1993. The direct lighting calculation in global illumination methods. Ph.D. thesis, Indiana Univ., Nov.
- WARD, G. 1991. Adaptive shadow testing for ray tracing. In *Proceedings of the 2nd Eurographics Workshop on Rendering*.
- WARD, G. J. 1994. The RADIANCE lighting simulation and rendering system. In ACM Siggraph 94 Conference Proceedings. *Comput. Graph.* 28, 2 (July), 459–472.
- WARD, G. J., RUBINSTEIN, F. M., AND CLEAR, R. D. 1988. A ray tracing solution for diffuse interreflection. In ACM Siggraph 88 Conference Proceedings. *Comput. Graph.* 22, 4 (Aug.), 85–92.

- WOZNIAKOWSKI, H. 1991. Average case complexity of multivariate integration. *Bulletin (New Ser.) Am. Math. Soc.* **24**, 1 (Jan.), 185–193.
- YAKOWITZ, S. J. 1977. *Computational Probability and Simulation*. Addison-Wesley, New York.
- ZEREMBA, S. K. 1968. The mathematical basis of Monte Carlo and quasi-Monte Carlo methods. *SIAM Rev.* **10**, 3 (July), 303–314.
- ZIMMERMAN, K. 1995. Direct lighting models for ray tracing with cylindrical lamps. In *Graphics Gems V*. Academic Press, New York.
- ZIMMERMAN, K., AND SHIRLEY, P. 1995. A two-pass solution to the rendering equation with a source visibility preprocess. In *Rendering Techniques 95*, P. Hanrahan and W. Purgathofer, Eds. Springer-Verlag, New York, 284–295.