Adjoint Equations and Random Walks for Illumination Computation

S. N. PATTANAIK and S. P. MUDUR
National Centre for Software Technology

In this paper we introduce the potential equation that along with the rendering equation forms an adjoint system of equations and provides a mathematical framework for all known approaches to illumination computation based on geometric optics. The potential equation is more natural for illumination computations that simulate light propagation starting from the light sources, such as progressive radiosity and particle tracing. Using the mathematical handles provided by this framework and the random-walk solution model, we present a number of importance sampling schemes for improving the computation of flux estimation. Of particular significance is the use of approximately computed potential for directing a majority of the random walks through regions of importance in the environment, thus reducing the variance in the estimates of luminous flux in these regions. Finally, results from a simple implementation are presented to demonstrate the high-efficiency improvements made possible by the use of these techniques.

Categories and Subject Descriptors: I.3.3 [**Computer Graphics**]: Picture/Image Generation; I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism

General Terms: *Algorithms, Design, Experimentation*

Additional Key Words and Phrases: Adjoint equations, forward simulation, global illumination, *illumination potential equation, importance transformation, Monte Carlo, particle tracing,* path tracing, progressive radiosity, random walk

1. INTRODUCTION

There are basically two approaches used for the computation of global illumination. In one, the luminous flux *reaching* a surface or eye point is estimated by sampling its surroundings. Ray tracing, path tracing, and full matrix radiosity solution are extensively researched methods [Whitted 1980; Cook et al. 1984; Kajiya 1986; Goral et al. 1984] based on this approach. The other approach is based on simulating the propagation of light starting from the light sources, of which progressive radiosity and particle tracing [Arvo 1986; Cohen et al. 1988; Shirley, 1990; Pattanaik and Mudur 1992] are representative methods. So far, Kajiya's [1986] rendering equation has pro-

Authors' address: Graphics and CAD Division, National Centre for Software Technology, Juhu, Bombay 400 049, India.

vided the mathematical basis for these computations. The essence of the rendering equation is the balancing of point-to-point light transport through the use of a *transport-intensity* function that is closely related to the *radiance* function. In this paper we introduce an adjoint formulation based on what we call as the *potential* function. This function captures the potential that any point of the environment has toward the illumination of another point. The luminous flux at any point of a surface in the environment can be expressed using either the *radiance* (*transport intensity*) or the *potential*. Together they form an adjoint system of equations that provides a mathematical basis for illumination computation based on geometric optics. Although illumination-computation methods for computing the flux reaching the surface or the eye are primarily based on the rendering equation, we believe that the potential equation is more natural for describing methods that simulate light propagation starting from the light sources. Just as Kajiya [1986] has shown full radiosity to be an approximate analytical solution of the rendering equation, analogously we have also shown progressive radiosity [Cohen et al. 1988] to be an approximate analytical solution of the potential equation. The particle tracing [Pattanaik and Mudur 1992] (backward ray tracing [Arvo 1986] or light ray tracing [Shirley 1990]) method, mathematically speaking, is in fact a random-walk-based solution of the potential equation. Similarly, the absorption suppression scheme [Arvo and Kirk 1990; Pattanaik and Mudur 1992] used for increasing the efficiency of particle tracing automatically falls within the mathematical framework of the random walk.

In this paper we introduce the notion of the potential function intuitively, with the help of a simple hypothetical experiment placing light detectors in an environment, and then proceed with its mathematical formulation. We then discuss the adjoint system of equations for illumination computation and their corresponding approximate analytic solutions. Monte Carlo quadrature and random-walk techniques are then proposed for obtaining solutions to the general equations. Using the mathematical handles provided by the adjoint system of equations and the random-walk model, we present a number of importance-based sampling schemes for improving the computation of flux estimation. Of particular significance is the scheme to use approximately computed potential as the importance function for directing a majority of the random walks through regions of importance in the environment, thus reducing the variance in the estimates of flux in these regions. Finally, results from a simple straightforward implementation are presented illustrating the high-efficiency improvements made possible by the use of these techniques.

2. THE ADJOINT SYSTEM OF ILLUMINATION EQUATIONS

The illumination of any point of a surface in a complex 3-D environment is due to the emission of light from that point (if any) and/or due to the reflection from that point of the light received from incoming hemispherical directions around that point. This fundamental concept forms the basis for the derivation of the adjoint system of illumination equations. To simplify our

discussion, we have restricted our attention to environments containing only opaque solid objects. However, in no way should this assumption be considered as a limitation of the discussed framework. Illumination of environments containing transmitting surfaces can also be easily explained within the given framework.

2.1 Radiance Equation

The *radiance equation* is the expression for the outgoing radiance leaving a surface point x along the direction Θ_{out} (L_{out}) in terms of the radiance emitted (ϵ_{out}) and the incoming radiance at x (L_{in}) from (x, Θ_{out}). It has the following form [Cook et al. 1984]:

$$L_{out}(x, \Theta_{out}) = \epsilon_{out}(x, \Theta_{out}) + \int_{\Omega_x} f_r(x)(\Theta_{out}, \Theta_{in}) L_{in}(x, \Theta_{in}) \cos\theta_{in}\, d\omega_{in},$$

where f_r is the bidirectional reflectance distribution function, Ω_x is the hemisphere surrounding x, and Θ_{out}, Θ_{in}, θ_{in}, and $d\omega_{in}$ are as shown in Figure 1. For the kind of environment under discussion, the radiance in any incoming direction at x must be due to the outgoing radiance from some surface point y in an outgoing direction Θ_y, where Θ_y is defined by the vector joining the point x to y. If we now wish to rewrite the radiance equation in terms of outgoing radiance and outgoing directions only, then by representing the outgoing directions at x and y as Θ_x and Θ_y, we get

$$L(x, \Theta_x) = \epsilon(x, \Theta_x) + \int_{\Omega_x} f_r(x)(\Theta_x, \Theta_y) L(y, \Theta_y) \cos\theta_x\, d\omega_x. \qquad (1)$$

Note that there is an implicit assumption in this equation that y represents a surface point visible to x with Θ_y being the direction from y to x.

2.2 Potential Equation

Because of the optical properties of surfaces, in our case, primarily reflection, the light emitted from any surface in any direction can illuminate many other surfaces of an environment. Alternatively, we can say that a surface can be illuminated by lights placed anywhere in the environment. The placement of the lights will, of course, determine how brightly lit or how dimly lit that surface is. This phenomenon can be elegantly captured by the notion of a potential associated with every point and direction in the environment. We shall describe a simple experiment to make the concept of potential easier to understand.

To start with, consider an environment with all of its light sources turned off. Next, assume that hypothetical light detectors are located in the environment such that the outgoing illumination from a surface point and in a given direction gets registered in one of the detectors and only in that detector (see Figure 2). In other words, the surface region and the directions from which light is detected by these detectors are exclusive. The detectors are hypothetical and in no way affect the flow of light. Next, consider the location of a unidirectional unit point light source (hypothetical again) at a surface point

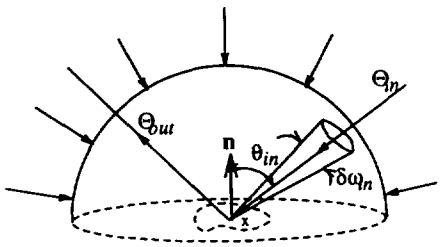

Fig. 1. Hemispherical directions for incoming illumination.

and in some orientation. That is, one unit of luminous flux is emitted from that surface point (x_0) in a particular direction (Θ_0). It is clear that due to this emission all of the detectors in the environment will register the receipt of some light. Of course, depending on their placement relative to the hypothetical point source, some will register more and some less. Concentrate only on one detector, say, the kth detector, and denote the light recorded by it as $W_k(x_0, \Theta_0)$. Similarly, the light recorded by this detector for a point source at (x_1, Θ_1) can be denoted as $W_k(x_1, \Theta_1)$. It is clear that the light recorded by this detector for all locations (x) and orientations (Θ) in the environment can be denoted by the function $W_k(x, \Theta)$. W_k captures the potential capability that every point and every orientation have toward illuminating the region in the directions in which the kth detector is focused. Other detectors would similarly define potential functions, say, W_i.

Next, we shall derive an expression for such a function. Let H_k denote the set of all points x over which the kth detector is focused. Similarly, Let $\mathscr{D}_k$ denote the set of all directions made by these points with the aperture of the kth detector. Then we define a function g_k as follows:

$$g_k(x, \Theta_x) = \begin{cases} 1 & \text{iff } x \in H_k \text{ and } \Theta_x \in \mathscr{D}_k, \\ 0 & \text{otherwise.} \end{cases}$$

Recall that the potential function W_k gives us the value of light detected by the kth detector for a hypothetical unit light source placed at any surface point and direction in the environment. Then the immediate contribution toward W_k by the unit light source placed at (x, Θ_x) in the environment is captured by the function $g_k(x, \Theta_x)$. We also have to account for an indirect contribution, which is the flux received by the detector due to any number of reflections of the light emitted from this unit light source at (x, Θ_x). For this component we will provide a recursive expression. The emission from the hypothetical source will reach the nearest surface point y and then possibly get reflected. If we take the probability of the whole amount of flux getting reflected in any one of the hemispherical directions Θ_y around y as $f_r(y)(\Theta_y, \Theta_x)\cos\theta_y \, d\omega_y$, where the terms used are as in Figure 3, then its contribution to the indirect component will be this probability times the potential of the point y along Θ_y, that is, $f_r(y)(\Theta_y, \Theta_x)\cos\theta_y \, d\omega_y W_k(y, \Theta_y)$. The indirect component is then the cumulative result of this expression obtained over the outgoing hemisphere around y, that is, $\int_{\Omega_y} f_r(y)(\Theta_y, \Theta_x)W_k(y, \Theta_y)\cos\theta_y \, d\omega_y$. The complete expression for the poten-

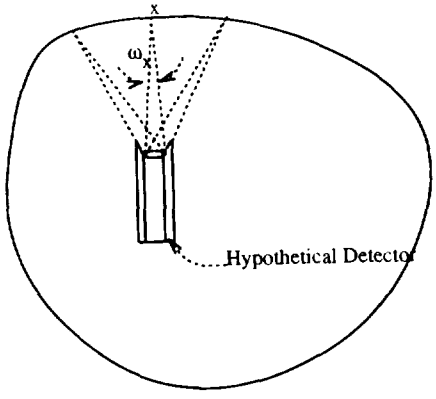

Fig. 2. Hypothetical detector focused on surface points of the environment.

tial function is therefore given by

$$W_k(x, \Theta_x) = g_k(x, \Theta_x) + \int_{\Omega_y} f_r(y)(\Theta_y, \Theta_x) W_k(y, \Theta_y) \cos\theta_y \, d\omega_y. \quad (2)$$

If we look back at eq. (1) for the radiance equation, we find a striking similarity in the forms of eq. (1) and eq. (2). However, note that in eq. (1) the integration is over the incoming hemisphere around x, whereas in eq. (2) the integration is over the outgoing hemisphere around y, where y is the surface point visible to x in the direction Θ_x. We now proceed to illustrate the relationship between these two equations.

2.3 Duality

In most of the illumination computations, one is interested in computing flux from a small region in a small spread of directions. For example, in image rendering the color of a pixel is assigned by computing the radiance from a few of the surface points visible to the eye through that pixel along the directions made by each such point with the eye point. Similarly, in the computation of view-independent global illumination of a diffuse environment by radiosity-based methods one is computing flux from a small surface patch in the hemispherical direction.

Expression of this flux using the radiance equation will therefore be an integral of the following form:

$$\Phi_k = \int_{\text{position spread}} \int_{\text{direction spread}} L(x, \Theta_x) \cos\theta_x \, d\omega_x \, dx.$$

If we assume that this flux represents the flux received by the kth hypothetical detector, then we can use the earlier defined function $g_k(x, \Theta_x)$, which evaluates to 1 in the limits of the integration and 0 everywhere else, and rewrite the above equation as follows:

$$\Phi_k = \int_A \int_{\Omega_x} L(x, \Theta_x) g_k(x, \Theta_x) \cos\theta_x \, d\omega_x \, dx. \quad (3)$$

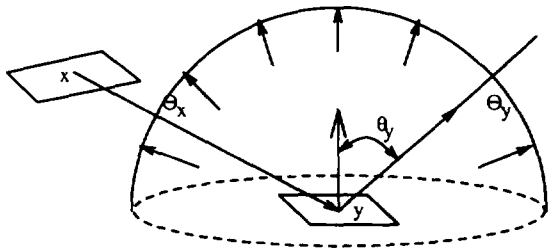

Fig. 3. Hemispherical directions for outgoing illumination.

Next, we get an expression for the above-discussed flux using the potential function. Let us represent the light sources by a function ϵ that is zero everywhere except at the positions belonging to emissive surfaces. By introducing ϵ we get the emission radiance at (x, Θ_x) to be $\epsilon(x, \Theta_x)$, and hence, the emission flux leaving (x, Θ_x) is $\epsilon(x, \Theta_x)\cos\theta_x\, d\omega_x\, dx$. The potential of (x, Θ_x) toward the kth detector is $W_k(x, \Theta_x)$. Then the flux received by the kth detector due to the actual emission at (x, Θ_x) will be $W_k(x, \Theta_x)\epsilon(x, \Theta_x)\cos\theta_x\, d\omega_x\, dx$. Thus, the expression for the total flux received by the kth detector is

$$\Phi_k = \int_A \int_{\Omega_x} W_k(x, \Theta_x)\epsilon(x, \Theta_x)\cos\theta_x\, d\omega_x\, dx. \tag{4}$$

To sum up,

—we have given two different equations, eqs. (3) and (4), to express the same quantity Φ_k using two different functions L and W_k;

—the eqs. (3) and (4) for Φ_k are similar in form, and so also are eqs. (1) and (2) for L and W_k, respectively; and

—eqs. (1)–(4) together form a closed system.

We will write again all of these four equations together to highlight the above-mentioned points:

$$\int_A \int_{\Omega_x} L(x, \Theta_x) g_k(x, \Theta_x)\cos\theta_x\, d\omega_x\, dx$$

$$= \Phi_k = \int_A \int_{\Omega_x} W_k(x, \Theta_x)\epsilon(x, \Theta_x)\cos\theta_x\, d\omega_x\, dx,$$

$$L(x, \Theta_x) = \epsilon(x, \Theta_x) + \int_{\Omega_x} f_r(x)(\Theta_x, \Theta_y) L(y, \Theta_y)\cos\theta_x\, d\omega_x,$$

$$W_k(x, \Theta_k) = g_k(x, \Theta_x) + \int_{\Omega_y} f_r(y)(\Theta_y, \Theta_x) W_k(y, \Theta_y)\cos\theta_y\, d\omega_y.$$

The equations satisfying the above-mentioned properties are said to form an adjoint system.[1] An adjoint system of linear equations for the discrete radiosity formulation referred to as importance equations has been independently derived [Smits et al. 1992]. In Pattanaik and Mudur 1993b, we showed how this adjoint radiosity system of linear equations can be derived as an analytical approximation to the adjoint system discussed in this paper. The discrete formulation of Smits et al. [1992] has just recently been extended to the continuous domain [Christensen et al. 1993] with emphasis on rigorously proving the adjointness.

The flux Φ_k can be computed using either eqs. (3) or (4). In the subsequent sections, we discuss solution methods for computing this flux. These methods can be used for any of the two equations.

3. ANALYTICAL SOLUTION FOR A DIFFUSE ENVIRONMENT

Because of their inherently complex nature, it is not possible to solve eq. (3) or (4) analytically. However, simplified forms under the following assumptions[2] are amenable to analytical solutions:

(1) The environment is a collection of a finite number, say, N, of small uniformly diffuse patches.

(2) As the radiance from any point of any such uniformly diffuse patch is $1/\pi$ times the flux per unit area, we shall compute this total flux leaving a patch in all of the hemispherical directions.

(3) The solution is carried out in an enclosure; that is, the hemispherical directions around any point in the environment are assumed to be covered by one or more of the patches of that environment, and every patch, j, may be assumed to occupy a solid angle, ω_j (which may be zero), in the hemisphere over any surface point.

Because of the diffuse nature of the surfaces, the radiance functions, $\epsilon(x, \Theta_x)$ and $L(x, \Theta_x)$, become independent of Θ_x, and because of the uniformity of the patch, they are also independent of the position x on any patch. Thus, they may be denoted as ϵ_i and L_i, respectively, for all of the x belonging to the patch i. Similarly, $f_r(x)(\Theta_x, \Theta_y)$ is independent of directions Θ_x and Θ_y and of the position x on any patch i and, hence, may be denoted as $f_{r,i}$.

Under these assumptions, eq. (1) simplifies to

$$L_i = \epsilon_i + f_{r,i} \sum_{j=1}^{N} L_j F_{ij}, \tag{5}$$

where L_j is the radiance of any point y belonging to the jth patch, ω_{ij} is the solid angle occupied by the jth patch in the visible hemisphere around the point x of the small patch i, and F_{ij} is the geometric factor between patch i

[1] Mathematically adjoint systems can be defined as follows: Given two systems of linear equations, $x = Ax + f$ and $y = By + h$, where A and B are linear operators, if $\langle By, x \rangle = \langle y, Ax \rangle$, where $\langle a, b \rangle$ denotes the inner product of a and b, then the pair of equations is said to form an adjoint system [Rubinstein 1981].

[2] These assumptions are typical of radiosity-based solutions [Goral et al. 1984].

and j. Recall that F_{ij} is similar to the form factor used in the radiosity computation [Goral et al. 1984] with the only difference that the summation of this factor over all the patches, that is, $\sum_{j=1}^{N} F_{ij}$, is equal to π. And the expression for the outgoing flux simplifies to

$$\Phi_k = E_k + A_k f_{r,k} \sum_{j=1}^{N} \frac{\Phi_j}{A_j} F_{kj}, \tag{6}$$

where E_k is the total emission flux leaving the kth patch and A_k denotes its area. These equations form the basis of the well-established radiosity method [Goral et al. 1984].

The analytical approximation of Φ_k using the potential function was derived by Pattanaik and Mudur [1993b] and is as follows:

$$\Phi_k = \pi \sum_{i=1}^{ns} \epsilon_i A_i \left[g_{k,i} + f_{r,k} F_{ik} + \sum_{j=1}^{N} f_{r,j} F_{ij} f_{r,k} F_{jk} + \cdots \right], \tag{7}$$

where ns is the total number of source patches in the environment. For the derivation the authors have introduced the notion of a hemispherical potential function defined over any point of patch i, say, $\mathscr{W}_k$. $\mathscr{W}_k$ is the average potential of the surface points in the patch in any hemispherical direction. If the patches are sufficiently small, this hemispherical potential function may be assumed to be independent of the position on each patch. The expression for this hemispherical potential function may be derived as follows [Pattanaik and Mudur 1993b]:

$$\begin{aligned}
\mathscr{W}_{k,i} &= \frac{1}{\pi} \int_{\Omega_{x_i}} W_k(x_i, \Theta_{x_i}) \cos\theta_{x_i} \, d\omega_{x_i} \\
&= \frac{1}{\pi} \int_{\Omega_{x_i}} \left[g_k(x, \Theta_x) + \int_{\Omega_y} f_r(y, \Theta_y, \Theta_x) W_y(y, \Theta_y) \cos\theta_y \, d\omega_y \right] \cos\theta_{x_i} \, d\omega_{x_i} \\
&= \frac{1}{\pi} \int_{\Omega_{x_i}} [g_{k,i} + \pi f_{r,j} \mathscr{W}_{k,j}] \cos\theta_{x_i} \, d\omega_{x_i} \\
&= \frac{1}{\pi} g_{k,i} \int_{\Omega_{x_i}} \cos\theta_{x_i} \, d\omega_{x_i} + \int_{\Omega_{x_i}} f_{r,j} \mathscr{W}_{k,j} \cos\theta_{x_i} \, d\omega_{x_i} \\
&= g_{k,i} + \sum_{j=1}^{N} f_{r,j} \int_{\omega_{ij}} \mathscr{W}_{k,j} \cos\theta_{x_i} \, d\omega_{x_i} \\
&= g_{k,i} + \sum_{j=1}^{N} f_{r,j} \mathscr{W}_{k,j} \int_{\omega_{ij}} \cos\theta_{x_i} \, d\omega_{x_i} \\
&= g_{k,i} + \sum_{j=1}^{N} f_{r,j} \mathscr{W}_{k,j} F_{ij} \qquad (8) \\
&= g_{k,i} + f_{r,k} F_{ik} + \sum_{j=1}^{N} f_{r,j} F_{ij} f_{r,k} F_{jk} + \cdots . \qquad (9)
\end{aligned}$$

Eq. (8) is the same as the *importance equation* proposed by Smits et al. [1992].

4. MONTE CARLO AND RANDOM WALK FOR GENERAL SOLUTIONS

Solutions of multidimensional integrations are best carried out by Monte Carlo quadrature techniques. The main principle behind a Monte Carlo quadrature technique for computing the integral $\int F(x)\,dx$ is as follows [Kalos and Whitlock 1986]:

(1) Write $F(x)$ as a product $f_1(x)f_2(x)$ such that $\int f_1(x) = 1$; that is, $f_1(x)$ is a *pdf*.
(2) Sample f_1 for an x_i.
(3) For each such sample x_i evaluate $f_2(x_i)$.
(4) Carry out steps (2) and (3) for some n times. The average, $1/n \sum_{i=1}^{n} f_2(x_i)$, is the estimate of the integral. Larger the n, better is the estimate.

We follow similar steps to evaluate eqs. (3) and (4). We have to find a *pdf* for this purpose. In both of the equations, we have a predefined known function each, ϵ and g_k, respectively. We can convert each of these functions to some constant times a normalized function. That means emission function $\epsilon(x, \Theta_x)$ may be converted to $\mathscr{E} \times S(x, \Theta_x)$, where $\mathscr{E} = \int_A \int_{\Omega_x} \epsilon(x, \Theta_x)\cos\theta_x\, d\omega_x\, dx$, and $g_k(x, \Theta_x)$ may be converted into $\mathscr{G}_k \times G_k(x, \Theta_x)$, where $\mathscr{G}_k = \int_A \int_{\Omega_x} g_k(x, \Theta_x)\cos\theta_x\, d\omega_x\, dx$. Then the quadrature process will start by sampling S and G_k using any standard sampling technique. For each such sample, the next task is to evaluate W_k and L. As said earlier, both W_k and L are integral equations of the second kind. The random-walk technique, a versatile mathematical method, is known to be useful in solving such integral equations [Rubinstein 1981]. We shall discuss in detail its use for the evaluation of radiance (L) and potential (W_k) values.

A random walk or a Markov chain is basically a sequence of states. Its formulation requires the definition of all possible states (discrete or continuous) of the system, a starting state, and the transition probability function (T) for transition from one state (s) to another (s') such that

$$\int T(s \to s')\, ds' \leq 1.$$

From a current state, the next state is chosen by sampling this transition probability function. The transition kernel is said to be normalized if $\int T(s \to s')\, ds' = 1$, whereas it is said to be subcritical when $\int T(s \to s')\, ds' < 1$. In a subcritical situation, the probability of $(1 - \int T(s \to s')\, ds')$ is taken as the probability of no transition (absorption) from any state s. Hence, a random walk with a subcritical transition kernel is bound to terminate in a finite number of steps, whereas any random walk with a normalized kernel can go on forever. So, in the latter cases, the walk has to be terminated by some external criterion [Arvo and Kirk 1990; Pattanaik and Mudur]. Any interaction of the light with the medium or the surface is always associated with some absorption. Thus, the environment for illumination computation is

always subcritical with $f_r(y)(\Theta_y, \Theta_x)\cos\theta_y$ as the transition kernel for potential equation solution and $f_r(x)(\Theta_x, \Theta_y)\cos\theta_x$ as the transition kernel for the radiance equation solution. Thus, straightforward evaluation of L or W_k using the random-walk technique leads to paths consisting of a finite number of steps and, hence, is computable. The states in our environment are the continuum of surface positions and hemispherical directions around each such surface position. The starting states are sampled from the respective *pdf*'s, that is, $S(x, \Theta_x)$ or $G_k(x, \Theta_x)$.

The evaluation of eq. (3) may be carried out by drawing n samples from the *pdf*, $G_k(x, \Theta_x)$, and evaluating L by the random walk for each sample (x_i, Θ_{x_i}). If the ith random walk starting from the state $(x_{i_0}, \Theta_{x_{i_0}})$ covers m_i steps, $(x_{i_1}, \Theta_{x_{i_1}}), \ldots, (x_{i_{m_i}}, \Theta_{x_{i_{m_i}}})$, then the radiance estimate from this walk will be given by

$$\langle L(x_i, \Theta_i)\rangle = \epsilon\left(x_{i_0}, \Theta_{x_{i_0}}\right) + \sum_{k=1}^{m_i} \epsilon\left(x_{i_k}, \Theta_{x_{i_k}}\right). \tag{10}$$

From this, the estimate of Φ_k is arrived at by averaging over n such random walks:

$$\begin{aligned}
\Phi_k &= \int_A \int_{\Omega_x} L(x, \Theta_x) g_k(x, \Theta_x)\cos\theta_x \, d\omega_x \, dx \\
&= \mathscr{G}_k \int_A \int_{\Omega_x} L(x, \Theta_x) G_k(x, \Theta_x)\cos\theta_x \, d\omega_x \, dx \\
&= \mathscr{G}_k \times \frac{1}{n} \sum_{i=1}^{n} \left[\epsilon\left(x_{i_0}, \Theta_{x_{i_0}}\right) + \sum_{k=1}^{m_i} \epsilon\left(x_{i_k}, \Theta_{x_{i_k}}\right) \right] \\
&= \mathscr{G}_k \times \frac{1}{n} \sum_{i=1}^{n} \sum_{k=0}^{m_i} \epsilon\left(x_{i_k}, \Theta_{x_{i_k}}\right).
\end{aligned} \tag{11}$$

This method of evaluating Φ_k by first sampling the G_k function is the essence of Kajiya's *path-tracing* method.

Similarly, the evaluation of eq. (4) may be carried out by drawing samples $(x_{i_0}, \Theta_{x_{i_0}})$ from the source function, $S(x, \Theta_x)$, and carrying out the random walk. A random walk may terminate at the state $(x_{i_k}, \Theta_{x_{i_k}})$ with probability

$$\sigma_{(x_{i_k}, \Theta_{x_{i_k}})} = 1 - \int_{\Omega_y} f_r(y)\left(\Theta_y, \Theta_{x_{i_k}}\right)\cos\theta_y \, d\omega_y \tag{12}$$

or may proceed to the next state $(x_{i_{k+1}}, \Theta_{x_{i_{k+1}}})$ chosen with probability $f_r(x_{i_{k+1}})(\Theta_{x_{i_{k+1}}}, \Theta_{x_{i_k}})\cos\theta_{x_{i_{k+1}}}$, and so on. For each such sample, W_k can be evaluated by carrying out the random walk. The potential estimate from this walk is given by

$$\langle W_k(x_i, \Theta_i)\rangle = g_k\left(x_{i_0}, \Theta_{x_{i_0}}\right) + \sum_{k=1}^{m_i} g_k\left(x_{i_k}, \Theta_{x_{i_k}}\right). \tag{13}$$

As before, Φ_k is estimated from n such walks as

$$\begin{aligned}\Phi_k &= \int_A \int_{\Omega_x} W_k(x, \Theta_x)\epsilon(x, \Theta_x)\cos\theta_x \, d\omega_x \, dx \\ &= \mathscr{E} \times \int_A \int_{\Omega_x} W_k(x, \Theta_x)S(x, \Theta_x)\cos\theta_x \, d\omega_x \, dx \\ &= \mathscr{E} \times \frac{1}{n}\sum_{i=1}^{n}\left[g_k\left(x_{i_0}, \Theta_{x_{i_0}}\right) + \sum_{k=1}^{m_i} g_k\left(x_{i_k}, \Theta_{x_{i_k}}\right)\right] \\ &= \mathscr{E} \times \frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{m_i} g_k\left(x_{i_k}, \Theta_{x_{i_k}}\right).\end{aligned} \tag{14}$$

This method of evaluating Φ_k by first sampling the source function is the essence of *particle tracing* or light ray tracing.

Of the two solution methods, *particle tracing* is highly intuitive, as it resembles the physical illumination process [Arvo 1986; Pattanaik and Mudur 1992]. Sampling of the source for a start state may be thought of as the emission of a photon from the source, and the transition for simulation of random walks may be thought of as the wandering of the photon in the environment as it gets reflected at the surface boundaries until it is absorbed. *Path tracing*, though not directly related to the physical process, is by now well known to the computer graphics community. The eye point, p_{eye}, and a random point on the pixel, p_{pixel}, define the direction, $\Theta_{out} = p_{eye} - p_{pixel}$. This direction, along with the nearest surface position along $-\Theta_{out}$, defines the starting state for the random walk. At that nearest surface, the ray is absorbed and the walk terminates, or the ray is reflected along one of the incoming hemispherical directions, Θ_{in}, obtained by sampling the *brdf*, and the walk continues.

What is more important in the discussion so far is that both of the random-walk processes attempt to solve the same problem and are subject to similar statistical error, which in Monte Carlo studies is known as *variance*. But one thing that makes *particle tracing* more attractive is that the simulation proceeds by sampling the source function. If we partition the space into a finite number of subregions, $sr_1, sr_2, sr_3, \ldots$, then we can locate detectors focused over each of these, that is, formulate an equal number of g functions $g_1, g_2, g_3, \ldots$, such that g_i is nonzero in the respective subregion sr_i and zero otherwise. Then each random walk originating from the source contributes toward the estimation of the Φ_i for each of the subregions. At the end of the simulation, we have the estimates for Φ_i for all of the subregions. Whereas in *path tracing* the random walk starts by sampling a particular g_i, for example, directions through a particular pixel. So each random walk contributes toward the estimation of only the Φ_i for that region, for which g_i is defined to be nonzero. This is not meant to be understood as saying that the computational efforts required to compute the brightness of a pixel by

path tracing are the same as the computational efforts required to compute the illumination of all of the subregions visible through a pixel by particle tracing. One may arrive at a low variance in the brightness estimate of the pixel by tracing a small number of paths, whereas it is possible that even after a large number of particle tracings the brightness estimates of a few of the subregions continue to show high variance. However, the difference is worth repeating:

—In particle tracing a single random walk contributes toward the estimation of many Φ_i's, as against many random walks contributing to the single Φ_i in path tracing.

An attempt at addressing this issue when tracing rays from the eye point was reported by Ward et al. [1988]. In this, the luminous flux gathered at the diffuse surfaces is cached for later use. Whenever flux has to be estimated at a diffuse surface, the cache is looked into first. If the surface under consideration is not found in the cache, then a Monte Carlo quadrature technique is invoked to estimate the flux.

There have been efforts to combine the two approaches of path tracing and particle tracing and thus derive benefits of both. These have typically come to be known as two-pass methods or, more generally, multipass methods [Wallace et al. 1987; Sillion and Puech 1989; Chen et al. 1991; Pattanaik and Mudur 1992]. In the initial passes, simulation proceeds starting from the light sources, and estimates are obtained for the flux in different subregions. For example, progressive radiosity [Sillion and Puech 1989; Chen et al. 1991] or particle tracing [Arvo 1986; Pattanaik and Mudur 1992] is used in the first pass to estimate the flux over diffuse surfaces. Chen et al. [1991] had an additional pass in which rays are traced from the light sources through nondiffuse surfaces to estimate caustics. In the case of multiple initial passes, care is taken to ensure that the flux computations are nonintrusive. The final rendering pass is always from the eye, which is based on the random-walk solution for eq. (1) with the slight difference from path tracing in that the walks are absorption suppressed and the walk terminates at a diffuse surface whose illumination computation has already been carried out in the earlier passes.

Knowing the basic solution processes, now we shall discuss some strategies for reduction in variance. Most of our discussions are based on the particle tracing method. However, it must be emphasized that both methods shall be equally benefited by these strategies.

5. IMPROVED ESTIMATION STRATEGIES

We discuss a few methods based on the observations that each random walk contributes either zero or nonzero values to the estimation of a Φ_i. In most of the situations of interest, more specifically in the problem of illumination computation of a reasonably complex environment, the fraction of random walks contributing nonzero values toward the estimate of any single Φ_i is

small.[3] A simple approach for improving the estimated result is to increase the number of random walks. Each random walk requires some amount of computational effort for sampling the initial state, sampling the transition probability function for moving to the next state, and computing the nearest surface along a given direction. So any increase in the number of random walks involves a proportionate increase in computation and must be contained. It can be seen that many random walks may in fact never visit the subregion(s) of interest or may visit subregions in which there have already been an adequate number of visits and, hence, may not substantially contribute any further to the accuracy of the flux estimates of those subregions. So an ideal strategy would be either to transform the basic underlying random-walk process, or to change the estimator (eqs. (11) and (14)), or to do both such that each random walk almost always contributes toward the subregions of our interest.

5.1 Next-Event Estimation

This technique [Coveyou et al. 1967] leaves the stochastic process under study unaltered, but modifies the form of the estimator. This modification involves the following:

(1) $g_k(x_{i_0}, \Theta_{x_{i_0}}) + \sum_{k=0}^{m_i} W_k^1(x_{i_k}, \Theta_{x_{i_k}})$ is used as the estimator of $W_k(x_i, \Theta_{x_i})$ so that

$$\Phi_k = \mathscr{E} \times \frac{1}{n} \sum_{i=1}^{n} \left[g_k\left(x_{i_0}, \Theta_{x_{i_0}}\right) + \sum_{k=0}^{m_i} W_k^1(x_{i_k}, \Theta_{i_k}) \right], \tag{15}$$

where

$$W_k^1(x, \Theta_x) = \int_{\Omega_y} f_r(y)(\Theta_y, \Theta_x) g_k(y, \Theta_y) \cos\theta_y \, d\omega_y.$$

(Note the difference between eq. (14) and (15).)

(2) Similarly, $\epsilon(x_{i_0}, \Theta_{x_{i_0}}) + \sum_{k=0}^{m_i} L^1(x_{i_k}, \Theta_{x_{i_k}})$ is used as the estimator of $L(x_i, \Theta_{x_i})$ so that

$$\Phi_k = \mathscr{G}_k \times \frac{1}{n} \sum_{i=1}^{n} \left[\epsilon\left(x_{i_0}, \Theta_{x_{i_0}}\right) + \sum_{k=0}^{m_i} L^1\left(x_{i_k}, \Theta_{x_{i_k}}\right) \right], \tag{16}$$

where

$$L^1(x, \Theta_x) = \int_{\Omega_x} f_r(x)(\Theta_x, \Theta_y) \epsilon(y, \Theta_y) \cos\theta_x \, d\omega_x.$$

(Note the difference between eq. (11) and eq. (16).

The choice of this estimator is based on intuition, as follows: $W_k^1(x, \Theta_x)$ is the expected direct potential contribution averaged over all possible next events.

[3] In *particle tracing* it rarely happens that every subregion of the space is visited in a single walk. Similarly, in *path tracing* it is also equally rare that every random walk starting from the eye will at all visit the light source during its walk.

Similarly, $L^1(x, \Theta_x)$ is the expected direct source contribution averaged over all possible sampling of the source. Replacing $g_k(x, \Theta_x)$ by $W_k^1(x, \Theta_x)$ and $\epsilon(x, \Theta_x)$ by $L^1(x, \Theta_x)$ may lead to convergence faster. The next-event estimation in *path tracing* would mean computing the local model at the points of ray-hit. Though not explicitly mentioned, we believe that Kajiya used this estimator in path tracing as he writes in Kajiya 1986, p. 146, "Calculating emittance... factors is simply a matter of consulting the ... lighting models." We would like also to point out that Chen et al. [1991, p. 167] used a variant of the next-event estimation principle in computing the final radiance $I(x, \Theta_x)$ by computing $I_{l,s}(x, \Theta_x)$, a part of $I(x, \Theta_x)$, by Monte Carlo sampling only the source contribution at x.

5.2 Transformation of the Stochastic Process

The methods discussed in this section attempt to transform the mathematical description of the stochastic process such that the modified form also gives an estimate of Φ and is likely to converge faster than the original process. The illumination process, as depicted in Section 2, is completely described by the source function and the surface *brdfs*. If we replace them by transformed functions satisfying some requirement and if we still wish to use them to estimate Φ in an unbiased fashion, then we must compensate for the transformation. In particle tracing the compensation required for the estimation may be derived as follows:

$$\Phi_k = \mathscr{E} \times \int_A \int_{\Omega_x} S(x, \Theta_x) W_k(x, \Theta_x) \cos\theta_x \, d\omega_x \, dx$$

$$= \mathscr{E} \times \int_A \int_{\Omega_x} S'(x, \Theta_x) \left(\frac{S(x, \Theta_x)}{S'(x, \Theta_x)} \right) W_k(x, \Theta_x) \cos\theta_x \, d\omega_x \, dx,$$

$$W_k(x, \Theta_x) = g_k(x, \Theta_x) + \int_{\Omega_y} f_r(y)(\Theta_y, \Theta_x) W_k(y, \Theta_y) \cos\theta_y \, d\omega_y$$

$$= g_k(x, \Theta_x) + \int_{\Omega_y} T'(\Theta_x \to \Theta_y) \left(\frac{T(\Theta_x \to \Theta_y)}{T'(\Theta_x \to \Theta_y)} \right) W_k(y, \Theta_y) \, d\omega_y,$$

where S' is the transformed normalized source function, $T(\Theta_x \to \Theta_y)$ is the transition function introduced only for notational convenience and is nothing but $f_r(y)(\Theta_y, \Theta_x)\cos\theta_y$, and T' is the transformed transition function. In order to make the equation more compact, we will define a multiplication factor f such that

$$f(x, y) = \frac{T(\Theta_x \to \Theta_y)}{T'(\Theta_x \to \Theta_y)}.$$

Then the transformed potential equation can simply be written as

$$W_k(x, \Theta_x) = g_k(x, \Theta_x) + \int_{\Omega_y} T'(\Theta_x \to \Theta_y) f(x, y) W_k(y, \Theta_y) \, d\omega_y. \quad (17)$$

An unbiased estimate is then given by

$$\Phi_k = \mathscr{E} \times \frac{1}{n} \sum_{i=1}^{n} \frac{S(x_{i_0}, \Theta_{x_{i_0}})}{S'(x_{i_0}, \Theta_{x_{i_0}})} \left[g_k(x_{i_0}, \Theta_{x_{i_0}}) + \sum_{k=1}^{m_i} \left(\prod_{l=0}^{k-1} f(x_{i_l}, x_{i_{l+1}}) \right) g_k(x_{i_k}, \Theta_{x_{i_k}}) \right]. \quad (18)$$

5.2.1 *Transition Transformation by Absorption Suppression.* As the name implies, in this method the absorption probability at the transition points is reduced (and may even be made zero). As a consequence, the random walk stretches to longer distances, and the probability of a nonzero contribution of each random walk to the estimation of Φ_i's is increased. The absorption probability σ at any state is given by eq. (12). Any reduction in this probability can be achieved by an appropriate increase in the reflection probability. A very convenient method is to scale the reflection probability simply by the factor $1/(1 - \sigma)$, consequently making the absorption probability at every state equal to zero. Thus, the unbiased estimate can be derived from eq. (17) to be

$$\Phi_k = \mathscr{E} \times \frac{1}{n} \sum_{i=1}^{n} \left[g_k(x_{i_0}, \Theta_{x_{i_0}}) + \sum_{k=1}^{m_i} \left(\prod_{l=0}^{k-1} (1 - \sigma_{x_{i_{l+1}}}) \right) g_k(x_{i_k}, \Theta_{x_{i_k}}) \right]. \quad (19)$$

A word of caution is needed here. If the transition probability is changed such that there is no absorption at any state, then any single random walk will go on forever without terminating. In practice, the walk is terminated when the product term in the above equations falls below some minimum threshold. However, this termination process introduces a bias into the estimation. An unbiased termination technique like Russian roulette may be used to overcome this [Avro and Kirk 1990; Pattanaik and Mudur 1992].

5.3 The Approximate Potential for Importance Transformation

Any transformation of the stochastic process needs to be carried out only if it results in an improvement in the estimation, either with an increase in computation speed or in a variance reduction. In the case of particle tracing, this means that the criterion based on which transformation may be carried out must be such that most of the emissions and transitions lead the random walk directly or indirectly to the region of interest, the hypothetical detector(s) of Section 2. Furthermore, the transformation computation must be straightforward. One such case could be a function, which can be easily computed and which can simply be multiplied with the emission function and/or transition probability function to give the appropriate transformed functions. Suppose we want to transform our random-walk process in this manner to improve the estimate of, say, some specified region of interest in the environment. Such a region of interest henceforward will be termed the *region of importance*. The potential function as defined in eq. (2) for this region of importance

may be seen as one such function. For source transformation, if we know the potential toward the region of importance of all the source points, that is, where $S(x, \Theta_x) > 0$, then we can transform the source function to $S'(x, \Theta_x)$ such that $S'(x, \Theta_x) \gg S(x, \Theta_x)$ for those points whose potential is higher and $S'(x, \Theta_x) \ll S(x, \Theta_x)$ for those points for which the potential is lower. Similarly, the transition probability that is, in our case, the reflection probability of the particle from a surface point in the hemispherical directions can be preferentially scaled up only for those directions for which the nearest surface points have larger potential toward the region of importance than those along the other directions. It is clear that if we can compute this function exactly then we also have the solution for the problem at hand and, hence, we will not require the simulation. However, if we can obtain an approximate value for this function easily, then our purpose is also well served. This approximate value can be used to transform the emission and transition probability functions and to increase the efficiency of the solution. Obviously, such transformation is not optimal.

A region of importance can be arrived at adaptively in a view-independent global illumination computation process, or can be predefined. For example, in rendering one or more views of a scene all of those surface points visible to the eye-point in the view(s) would form the region of importance. Our observation is that often it is possible to derive an approximate potential function $W_{imp}(x, \Theta_x)$ with respect to such a region of interest. This approximate function or one that can be easily derived from this is termed an *importance function*, $Imp(x, \Theta_x)$ [Coveyou et al. 1967]. We can see the usefulness of this importance function in transforming the mathematical description of the process:

$$\Phi_k = \mathscr{E} \times \int_A \int_{\Omega_x} S(x, \Theta_x) W_k(x, \Theta_x) \cos\theta_x \, d\omega_x \, dx$$

$$= \mathscr{E} \times \int_A \int_{\Omega_x} (S(x, \Theta_x) \mathrm{Imp}(x, \Theta_x)) \times \frac{W_k(x, \Theta_x)}{Imp(x, \Theta_x)} \cos\theta_x \, d\omega_x \, dx,$$

$$\frac{W_k(x, \Theta_x)}{Imp(x, \Theta_x)} = \frac{g_k(x, \Theta_k)}{Imp(x, \Theta_x)} + \frac{1}{Imp(x, \Theta_x)} \int_{\Omega_y} T(\Theta_x \to \Theta_y) W_k(y, \Theta_y) \, d\omega_y$$

$$= \frac{g_k(x, \Theta_x)}{Imp(x, \Theta_x)} + \int_{\Omega} T(\Theta_x \to \Theta_y) \frac{Imp(y, \Theta_y)}{Imp(x, \Theta_x)} \frac{W_k(y, \Theta_y)}{Imp(y, \Theta_y)} \, d\omega_y.$$

If we denote $S(x, \Theta_x) Imp(x, \Theta_x)$ as a transformed function $S'(x, \Theta_x)$,

$$\frac{W_k(x, \Theta_x)}{Imp(x, \Theta_x)}$$

as $W_k'(x, \Theta_x)$, and $T(\Theta_x \to \Theta_y)(Imp(y, \Theta_y))/(Imp(x, \Theta_x))$ as $T'(\Theta_x \to \Theta_y)$, then we have the following equations for the transformed potential function

and the flux using this transformed function:

$$\Phi = \mathscr{E} \times \int_A \int_{\Omega_x} S'(x, \Theta_x) W_k'(x, \Theta_x) \cos\theta_x \, d\omega_x \, dx,$$

$$W_k'(x, \Theta_x) = \frac{g_k(x, \Theta_x)}{Imp(x, \Theta_x)} + \int_{\Omega_y} T'(\Theta_x \to \Theta_y) W_k'(y, \Theta_y) \, d\omega_y.$$

As before, we can proceed to solve this quadrature by sampling S' followed by the simulation of random walks as discussed in Section 4, provided the following conditions are satisfied:

(1) $S'(x, \Theta_x)$ is normalized, and
(2) $\int_{\Omega_y} T'(\Theta_x \to \Theta_y) \, d\omega_y \leq 1$.

5.3.1 *Normalization of the Transformed Source Function.* To satisfy condition (1), we choose $(W_{imp}(x, \Theta_x))/(\int_A \int_{\Omega_x} S(x, \Theta_x) W_{imp}(x, \Theta_x) \cos\theta_x \, d\omega_x \, dx)$ for $Imp(x, \Theta_x)$, which trivially assures the normalization of $S'(x, \Theta_x)$. This is because

$$\int_A \int_{\Omega_x} S'(x, \Theta_x) \cos\theta_x \, d\omega_x \, dx$$

$$= \int_A \int_{\Omega_x} S(x, \Theta_x) \frac{W_{imp}(x, \Theta_x)}{\int_A \int_{\Omega_x} S(x, \Theta_x) W_{imp}(x, \Theta_x) \cos\theta_x \, d\omega_x \, dx} \times \cos\theta_x \, d\omega_x \, dx = 1.$$

5.3.2 *Subcritically of the Transformed Transition Function.* With the above definition of $Imp(x, \Theta_x)$,

$$T'(\Theta_x \to \Theta_y) = T(\Theta_x \to \Theta_y) \frac{Imp(y, \Theta_y)}{Imp(x, \Theta_x)} = T(\Theta_x \to \Theta_y) \frac{W_{imp}(y, \Theta_y)}{W_{imp}(x, \Theta_x)}.$$

Using eq. (2) we can write that

$$W_k(x, \Theta_x) = g_k(x, \Theta_x) + \int_{\Omega_y} T(\Theta_x \to \Theta_y) W_k(y, \Theta_y) \, d\omega_y,$$

$$W_k(x, \Theta_x) - g_k(x, \Theta_x) = \int_{\Omega_y} T(\Theta_x \to \Theta_y) W_k(y, \Theta_y) \, d\omega_y.$$

Thus,

$$\int_{\Omega_y} T'(\Theta_x \to \Theta_y) \, d\omega_y = \int_{\Omega_y} T(\Theta_x \to \Theta_y) \frac{W_{imp}(y, \Theta_y)}{W_{imp}(x, \Theta_x)} \, d\omega_y$$

$$= \frac{W_{imp}(x, \Theta_x) - g_{imp}(x, \Theta_x)}{W_{imp}(x, \Theta_x)}$$

$$= 1 - \frac{g_{imp}(x, \Theta_x)}{W_{imp}(x, \Theta_x)} \leq 1.$$

Thus, the second condition is also satisfied. Furthermore, it is established that the transformed transition probability function is subcritical, and the absorption probability at any (x, Θ_x) is

$$\frac{g_{imp}(x, \Theta_x)}{W_{imp}(x, \Theta_x)}.$$

An unbiased estimator for the transformed random walk is

$$\Phi_k = \mathscr{E} \times \frac{1}{n} \sum_{i=1}^{n} \sum_{k=0}^{m_i} \frac{g_k(x_{i_k}, \Theta_{x_{i_k}})}{Imp(x_{i_k}, \Theta_{x_{i_k}})}. \tag{20}$$

It is important to note here that by using the handles provided by the mathematical framework of the potential equation and carrying out adaptive sampling with the help of an importance transformed stochastic process we ensure that any statistical bias of the kind discussed by Arvo and Kirk [1990] is completely avoided.

6. IMPLEMENTATION

In its general form, the potential function is dependent on both positions and directions in the corresponding outgoing hemispheres (spheres) of the points of the surfaces (volumes) in the environment. Similarly, the region of importance is defined as a collection of points and corresponding directions. In order to illustrate the use of the potential, we shall make the following simplifying assumptions:

(1) The environment consists of ideal diffuse reflecting and emitting surfaces.
(2) The medium is nonparticipating.
(3) The region of importance is a set of patches with all of the corresponding hemispherical directions included.
(4) For transformation we shall only use the direction-independent hemispherical potential defined earlier.

With the above assumptions, the environment can be defined as being made up of patches, say, $E = \{P_1, P_2, \ldots, P_n\}$, such that the region of importance R is a subset of E and, for all $P_i \in E$, $\mathscr{W}_i$ denotes the hemispherical potential that patch P_i has toward illuminating patches of R.

The approximate potential values are easily computed from a particle-tracing simulation using a much smaller number of particles, say, 5%–10% of the total required for a complete unbiased simulation. For the purpose of computing hemispherical potential, the following additional information is kept track of:

—the number of particles leaving a patch P_i, that is, emitted/reflected, say, N_i; and
—the number of these particles reaching a patch belonging to the region of importance, say, M_i.

The ratio M_i/N_i gives us an estimate of the hemispherical potential of patch P_i.

6.1 Source Position Transformation Using Hemispherical Potential

If $\mathscr{W}_i$ is the hemispherical potential of patch P_i, then $\mathscr{W}(x)$ is also the hemispherical potential of point x, where $x \in P_i$. Using $\mathscr{W}(x)$ we can transform the normalized source function $S(x)$. Renormalizing the transformed source function then gives us the expression for the transformed source function:

$$S'(x) = \frac{S(x) \times \mathscr{W}(x)}{S0},$$

where $S0 = \int_A S(x)\mathscr{W}(x)\,dx \approx \sum_{i=1}^{ns} S_i \mathscr{W}_i A_i$, with ns being the number of source patches. This transformation results in an altered distribution of source strength, so that emissions take place more often on emitter patches from which the particles have a higher probability of reaching R, the region of importance. To compensate for this bias, for each particle the brightness contributing strength is multiplied by a factor, f_1, given below:

$$f_1 = \frac{S(x)}{S'(x)} = \frac{S0}{\mathscr{W}(x)}.$$

6.2 Direction Transformation Using Hemispherical Potential

In the normal simulation, the direction is chosen by sampling the diffuse distribution function. In the transformed case, both for emission and reflection, the idea is to look around the environment and decide on the direction that has a higher probability of leading the random walk to the region of importance. To understand this it is worthwhile to look at eq. (8) again, which gives a linear expression for the hemispherical potential, and eq. (7), which gives an expression for flux using the hemispherical potential:

$$\mathscr{W}_i = g_i + \sum_{j=1}^{N} f_r(j)\mathscr{W}_j F_{ij},$$

$$\Phi = \pi \sum_{i=1}^{ns} \epsilon_i A_i \mathscr{W}_i.$$

If we assume that the simulation is being carried out in an enclosure, then the outgoing hemisphere around any point is covered by other surface patches of the environment. Associated with each surface patch is its hemispherical potential. Now, using the above equations for particle tracing, the transition of a particle can be carried out by sampling the F_{ij} distribution to choose the patch, say, k, and by sampling the directions occupied by that patch on the hemisphere to arrive at the direction of flight. This F_{ik} times the approximate hemispherical potential now gives us a measure of the new relative importance, F'_{ik}, of each patch around the point p. Now, instead of the distribution of F_{ik}, the distribution of F'_{ik} is used for sampling and choosing

the appropriate range of directions. Furthermore, directions within that range are sampled to choose at the direction of flight for the particle. The resulting mathematical change of eq. (8) is as follows:

$$\mathscr{W}_i = g_i + \sum_{j=1}^{N} f_r(j) \frac{\mathscr{W}_j}{\mathscr{W}_{j,approx}/\mathscr{W}0} F'_{ij}, \tag{21}$$

where $\mathscr{W}0 = \Sigma \mathscr{W}_{j,approx} F_{ij}$ and $F'_{ij} = (\mathscr{W}_{j,approx}/\mathscr{W}0) F_{ij}$.

We have implemented the above idea and have applied it to a number of cases. The resulting improvements in efficiency are extremely encouraging, showing enormous efficiency gains by the use of this importance transformation of the underlying stochastic process in both cases:

(1) view-dependent illumination computation, that is, the situation in which region of importance (R) is predefined; and

(2) view-independent illumination computation, that is, the case in which the region of importance, which to begin with is the entire environment, is gradually pruned as the simulation progresses to smaller and smaller subsets of the environment, and transformation is done for each new subset of important regions.

We provide here the results of this implementation. Complete details of this implementation including an efficient method for computing approximate F_{ij}'s and the data structures required can be found in Pattanaik and Mudur 1993a. For the view-dependent global illumination computation, we have used a simple environment, a view of which is shown in Figure 4. The vertical wall on the left extreme has been defined as the region of importance. The wall has been divided into 32 × 16 patches. The simulation was carried out in two passes: The first pass was the normal untransformed one, and the second was a simulation of the transformed process. The results of the first pass were used to carry out the importance transformation. In the first pass, 300,000 particles were traced (i.e., 300,000 random walks were carried out). The average number of single-level ray-tracing operations[4] per random walk is 1–2. An equivalent number of direction-sampling operations are also required. At the end of this simulation, the left wall, that is, the region of interest, recorded a total of 46,462 particle hits. All of the results of this simulation were used to transform the source and transition functions for use in the second pass. Another 300,000 particles were traced in the second pass. This time, the left wall recorded a total of 352,922 particle hits. Figure 5a is the map of particle hits on the left wall for normal simulation, and Figure 5b is the map for transformed simulation. As one can see, visually there is appreciable improvement. Certainly, sampling the transformed functions for position and direction does involve more computation, particularly because

[4]Note that this average number will vary from environment to environment and will depend on the geometry and the average absorption coefficient of the environment. In our experience, for many typical environments this varies between 3 and 5. For this environment the number is low because the environment is open on all of its sides except the vertical wall on the extreme left.

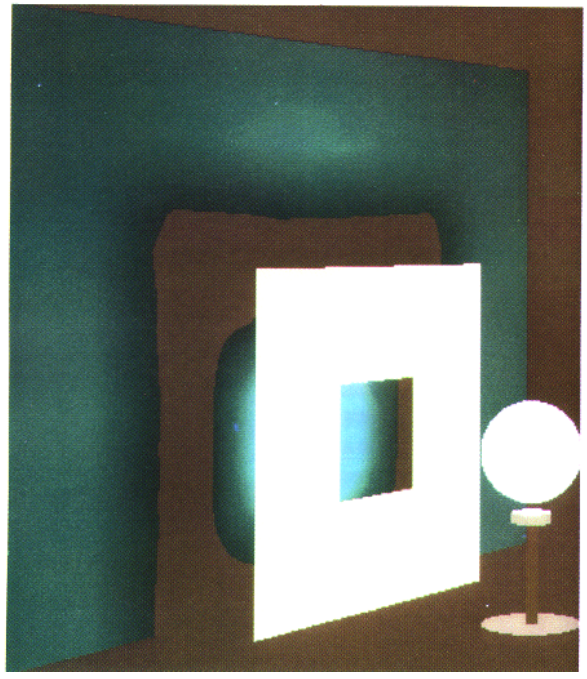

Fig. 4. Scene for importance biasing with predefined R

direction sampling now requires the use of the rejection sampling techniques [Pattanaik and Mudur 1993a]. However, the overall efficiency gains are considerable. We therefore compared the total time for global illumination computation for a total of 600,000 particles in the case of normal simulation and 600,000 particles in two passes as above in a transformed simulation. The speedup factor is a minimum of 4.

For the view-independent global illumination computation, we have used an environment like a maze, which we shall refer to as the ***Cornell Labyrinth***[*] (see Figures 6 and 7). It has a total of 523 patches, all of more or less the same area. In order to be able to quantify the effective gains obtained by the importance transformation techniques clearly, we carried out the simulations as follows: First, we chose a minimum number of particle hits that must take place on each patch so that the equilibrium state for the illumination can be assumed to have been reached. Note that because of indirect illumination every patch must eventually receive this minimum number of particle hits, assuming that the simulation can be carried on indefinitely. For the Cornell Labyrinth, we chose 100 particle hits as this minimum number. Then in the first phase a normal simulation was carried out until all of the patches

[*]The name has been chosen because a similar scene was chosen by the Cornell group in their recent publication on importance-driven radiosity [Smits et al. 1992].

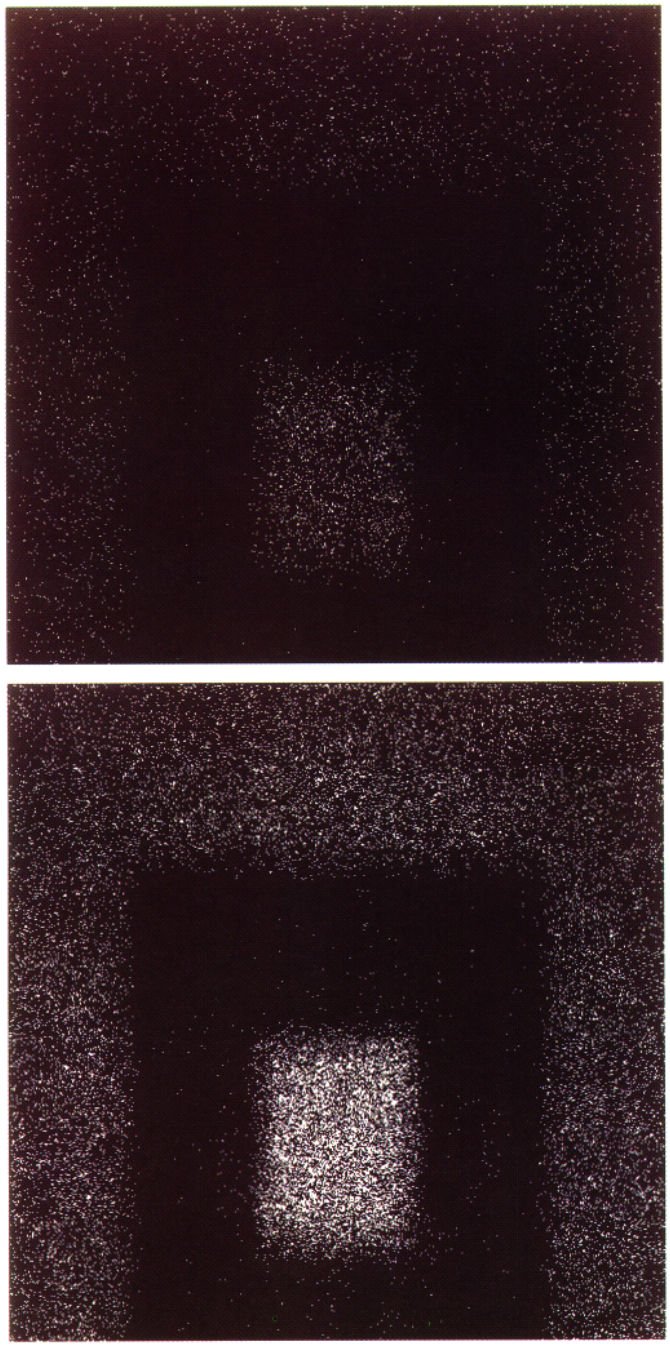

Fig. 5. Plot of particle incidences on region of importance.

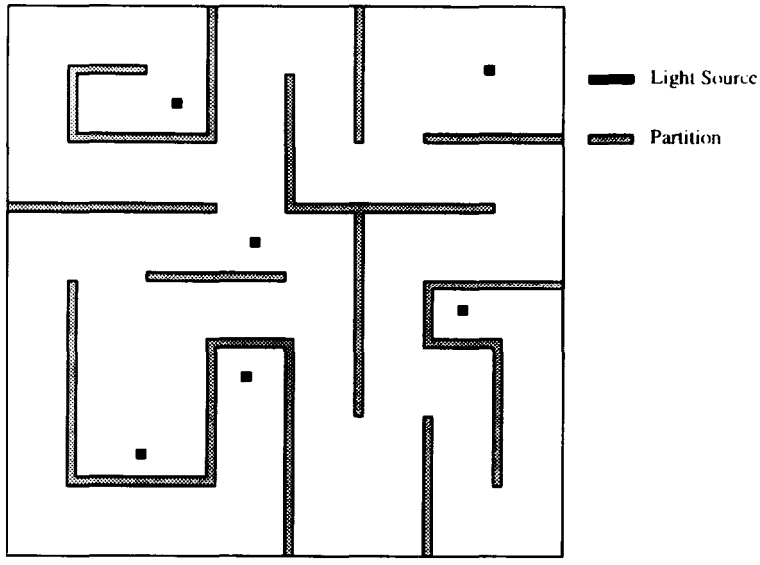

Fig. 6. Wire frame drawing showing top view of Cornell Labyrinth.

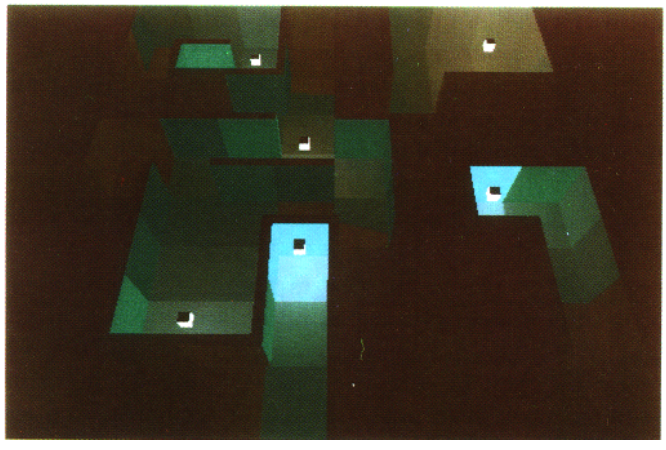

Fig. 7. Rendered view of Cornell Labyrinth.

registered a minimum of 100 particle hits each. For this, the total number o particles that needed to be traced was around 27 million.

The second phase of the simulation was as follows: To start with, all 523 patches were considered to constitute the region of interest, say, *R*. A normal simulation of 300,000 particles was carried out in the first pass, and the random-walk histories were stored for subsequent use. The region of interest was pruned by removing all patches that received a minimum of 100 particle hits. In the case of the Cornell Labyrinth, the resulting *R* had 205 patches.

Table I. Importance Transformation Improvements for Cornell Labyrinth

Simulation Volume (in number of particles)	Number of patches in region of interest		Particle hits in in region of interest
	Before	After	
300000(Normal)	523	205	658520
3000(Transformed)	205	131	10266
3000(Transformed)	131	72	10747
3000(Transformed)	72	26	10816
3000(Transformed)	26	13	10494
3000(Transformed)	13	8	10706
3000(Transformed)	8	6	10247
3000(Transformed)	6	4	10513
3000(Transformed)	4	0	11083

Using the results of the first pass, the stochastic process was transformed, and the second and subsequent passes were carried out in simulation volumes of 3,000 particles each. At the end of each pass, R is further pruned as before by removing all patches with a minimum number of 100 particle hits. Since R gets redefined after each pass, the hemispherical potential that any patch P_i has with respect to this newly defined R has also to be modified. This is done by scanning through all of the stored random-walk histories of the initial untransformed simulation. At the end of nine passes, R is empty (Table I). The total number of particles required to be traced in all of the passes was just 540,000, as compared to 27 million for the normal simulation. Even if some additional computational effort is required for maintaining and using the importance transformation, it is clear that the efficiency gains are enormous.

7. CONCLUSION

In this paper we have presented the mathematical framework of adjoint equations, consisting of the radiance equation and the potential equation for describing light transport in a general environment. Either of these or any combination may be used to compute global illumination values. Thus, all illumination-computation techniques based on geometric optics fall within the framework of this adjoint system of equations. Although the adjoint formulation of radiation transport has already been known in other disciplines such as Neutron Transport [Spanier and Gelbard 1969], one of the important contributions of this paper is an intuitive derivation of the potential equation and subsequent derivations, which go to show that the potential equation forms the mathematical basis for hitherto intuitively devised methods based on forward simulation starting from the light sources (progressive radiosity and particle transport). In the process we have shown [Pattanaik and Mudur 1993b] how the importance-driven radiosity method [Smits et al. 1992] is basically an analytical solution of a simplified (discretized) version of the potential equation. The principal contribution of this paper is an in-depth exploration of the use of the random walk, a powerful Monte Carlo technique,

for efficiently computing global illumination by solving either of the two integral equations of the adjoint system. In particular, the paper has formulated methods for importance-based transformation of the basic stochastic process of light transport, resulting in high-efficiency gains for random-walk solutions. The performance improvements that are possible by such random-walk solutions of the importance transformed process are demonstrated with the help of results from a simple and straightforward implementation. Certainly, there is a lot yet to be explored here. In particular, appropriate representations need to be evolved for the importance transformations, the transformed source and transition functions, and the patch illumination functions. In our simple implementation, we have only considered discretized sampled forms for these. Adaptive environmental discretization for complex environments is yet another topic that needs further exploration.

ACKNOWLEDGMENT

We are grateful to S. Ramani, director of the National Centre for Software Technology, for making the computational resources to carry out the simulation available.

REFERENCES

ARVO, J. 1986. Backward ray tracing. In *SIGGRAPH 86 Developments in Ray Tracing Course Notes* (Dallas, Tex., Aug. 18–22). ACM, New York.

ARVO, J., AND KIRK, D. 1990. Particle transport and image synthesis. In *SIGGRAPH 90 Proceedings. Comput. Graph. 24*, 4 (Aug.), 63–66.

CHEN, S. E., RUSHMEIER, H. E., MILLER, G., AND TURNER, D. 1991. A progressive multi-pass method for global illumination. In *SIGGRAPH 91 Proceedings. Comput. Graph. 25*, 4 (July), 164–174.

CHRISTENSEN, P., SALESIN, D., AND DE ROSE, T. 1993. A continuous adjoint formulation for radiance transport. In *Proceedings of 4th Eurographics Rendering Workshop* (Paris, June 14–16). Eurographics Assoc., 95–101.

COHEN, M., CHEN, S. E., WALLACE, J. R., AND GREENBERG, D. P. 1988. A progressive refinement approach to fast radiosity image generation. In *SIGGRAPH 88 Proceedings. Comput. Graph. 22*, 4 (Aug.), 75–84.

COOK, R. L., PORTER, T., AND CARPENTER, L. 1984. Distributed ray tracing. In *SIGGRAPH 84 Proceedings. Comput. Graph. 18*, 3 (July) 137–145.

COVEYOU, R. R., CAIN, V. R., AND YOST, K. J. 1967. Adjoint and importance in Monte Carlo application. *Nucl. Sci. Eng. 27*, 219–234.

GORAL, C. M., TORRANCE, K. E., GREENBERG, D. P., AND BATTAILE, B. 1984. Modelling the interaction of light between diffuse surfaces. In *SIGGRAPH 84 Proceedings. Comput. Graph. 18*, 3 (July), 212–222.

KAJIYA, J. T. 1986. The rendering equation. In *SIGGRAPH 86 Proceedings. Comput. Graph. 20*, 4 (Aug.), 143–150.

KALOS, M. H., AND WHITLOCK, P. A. 1986. *Monte Carlo Methods. Volume I.* Wiley, New York.

PATTANAIK, S. N., AND MUDUR, S. P. 1992. Computation of global illumination by Monte Carlo simulation of the particle model of light. In *Proceedings of the 3rd Eurographics Workshop on Rendering* (Bristol, U.K., May 17–20). Eurographics Assoc., 71–83.

PATTANAIK, S. N., AND MUDUR, S. P. 1993A. Efficient potential equation solutions for global illumination computation. *Comput. Graph. 17*, 4, 387–396.

PATTANAIK, S. N., AND MUDUR, S. P. 1993B. The potential equation and importance in illumination computations. *Comput. Graph. Forum 12*, 2 (June), 131–136.

RUBINSTEIN, R. Y. 1981. *Simulation and the Monte Carlo Method.* Wiley, New York.

SHIRLEY, P. 1990. Physically based lighting calculations for computer graphics. Ph.D. thesis, Dept. Computer Science, Univ. of Illinois, Urbana-Champaign.

SILLION, F., AND PUECH, C. 1989. A general two-pass method integrating specular and diffuse reflection. In *SIGGRAPH 89 Proceedings. Comput. Graph.* **23**, 3, 335–344.

SMITS, B. E., ARVO, J. R., AND SALESIN, D. H. 1992. An importance driven radiosity algorithm. In *SIGGRAPH 92 Proceedings. Comput. Graph. 26*, 2 (July), 273–282.

SPANIER, J., AND GELBARD, E. M. 1969. *Monte Carlo Principles and Neutron Transport Problems*. Addison-Wesley, Reading, Mass.

WALLACE, J. R., COHEN, M. F., AND GREENBERG, D. P. 1987. A two-pass solution to the rendering equation: A synthesis of ray tracing and radiosity methods. In *SIGGRAPH 87 Proceedings. Comput. Graph. 21*, 4 (July), 311–320.

WARD, G. J., RUBINSTEIN, F. M., AND CLEAR, R. D. 1988. A ray tracing solution for diffuse interreflection. In *SIGGRAPH 88 Proceedings. Comput. Graph. 22*, 4 (Aug.), 85–92.

WHITTED, T. 1980. An improved illumination model for shaded display. *Commun. ACM 23*, 6 (June), 343–349.

Received May 1992; revised October 1993; accepted June 1994