Efficient Potential Equation Solutions for Global Illumination Computation*

S.N.Pattanaik and S.P.Mudur
Graphics & CAD Division,
National Centre for Software Technology,
Juhu, Bombay 400 049, INDIA

Abstract

The *potential equation* provides a natural mathematical basis for all the illumination computation methods based on the strategy of shooting light. The particle tracing technique which physically simulates the particle model of light, can also be considered as a general, random walk based solution to the potential equation. As the complexity of the environment increases, the straight forward particle tracing method becomes inefficient. In this paper we discuss different biasing methods that could be used to increase the efficiency of the particle tracing process. In particular we present an importance driven scheme in which we use an approximate potential value as the biasing function for directing a majority of the random walks through regions of importance in the environment, thus efficiently estimating the light flux in these regions. A straight forward simple implementation of this scheme has been carried out and the results that we have obtained show that the scheme is very promising.

1 Introduction

There are two basic strategies used for carrying out global illumination computation – the gathering strategy in which light reaching a point from all directions is simulated and the shooting strategy in which light emitted from a point in all directions is simulated. Based on the strategy used, all the existing methods can be classified into two broad categories, namely gathering methods and shooting methods. The radiance equation, better known as rendering equation[1], provides the mathematical basis for the gathering methods and an adjoint equation termed as the potential equation provides the mathematical basis for the shooting methods. Though the shooting strategy methods are highly intuitive their mathematical basis, the potential equation, has only been recently introduced[2]. In this paper we shall briefly derive the potential equation and present the particle tracing technique, a general solution method by Monte Carlo and random walk. Then we discuss the efficiency

*Paper published in Computers & Graphics, Vol 17(4), Pergammon Press

issues involved and the different biasing techniques that can be used to improve the efficiency of particle tracing. We discuss the implementation of a particular biasing scheme in which an approximate potential value defined over each patch of the environment is used as the biasing function.

2 The Potential Equation

We know that the illumination of any point of a surface in a complex 3D 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 through all the incoming hemispherical directions around that point. To simplify our discussion, we shall restrict our attention to environments containing only opaque solid objects[1]. Because of the optical properties of surfaces, which for the present discussion is primarily reflection, the light emitted from any surface in any direction can illuminate many other surfaces of an environment. This phenomenon can be elegantly captured by the notion of a potential associated with every point and direction in the environment.

The potential equation provides an expression for the potential capability, W, of any (x, Θ_x) towards the illumination of $\mathcal{S}$, a set of points and directions around those points. Emission from a point x along Θ_x can directly and/or indirectly illuminate $\mathcal{S}$. The emitted radiance from (x, Θ_x) can directly account for the illumination of the set $\mathcal{S}$ if (x, Θ_x) itself belongs to the set. So to represent the direct component we shall use a function g defined over all the surface points and all the directions around those points such that $g(x, \Theta_x)$ is 1 if $(x, \Theta_x) \in \mathcal{S}$, and 0 otherwise.

The quantity of emitted light from (x, Θ_x) responsible for the outgoing component at $\mathcal{S}$ due to one or more reflections may be expressed recursively as follows. The emission from any (x, Θ_x) will reach the nearest surface point y and then possibly be reflected. The fraction of the incident flux getting reflected in any one of the hemispherical directions Θ_y around y is $f_r(y, \Theta_y, \Theta_x) cos\theta_y d\omega_y$. Then this fraction times the potential of the point y along Θ_y integrated over the outgoing hemisphere around y, i.e. $\int_{\Omega_y} f_r(y, \Theta_y, \Theta_x) W(y, \Theta_y) cos\theta_y d\omega_y$, will represent the indirect component. Thus the complete expression for the potential function is given by:

$$W(x, \Theta_x) = g(x, \Theta_x) + \int_{\Omega_y} f_r(y, \Theta_y, \Theta_x) W(y, \Theta_y) cos\theta_y d\omega_y \tag{1}$$

3 Illumination using Potential Equation

Most often in illumination computation one is interested in computing flux from a small region in a small spread of directions. For example: in rendering an image of a scene the colour of a pixel is assigned by computing the radiance from all the surface points visible to the eye through that pixel and in a spread of directions made by each such point with the aperture of the eye. 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.

[1] The potential equation for complex environments with participating volume is derived in [2].

We can derive the expression for the actual flux received by the detector by using the potential function and the actual emission behaviour of the environment. If we define a function L_e which is zero everywhere except at the positions belonging to emissive surfaces, then we get the emission radiance at (x, Θ_x) to be $L_e(x, \Theta_x)$ and hence the emission flux leaving (x, Θ_x) is $L_e(x, \Theta_x) \cos\theta_x d\omega_x dx$. The potential of (x, Θ_x) towards the k-th detector is $W_k(x, \Theta_x)$. Then the flux received by the k-th detector due to the actual emission at (x, Θ_x) will be $W_k(x, \Theta_x) \times L_e(x, \Theta_x) \cos\theta_x d\omega_x dx$. Thus the expression for the total flux received by the k-th detector will be

$$\Phi_k = \int_A \int_{\Omega_x} W_k(x, \Theta_x) L_e(x, \Theta_x) \cos\theta_x d\omega_x dx \tag{2}$$

4 Monte Carlo Methods and Random Walks for General Solution

Computation of light flux using Eq.2 is a twofold problem. The first one is solving a multidimensional quadrature, and the second one is solving the potential equation which is an integral equation of the second kind. The numerical solution of Eq.2 for a diffuse environment leads to the method of progressive radiosity[3, 4, 5]. A general solution by Monte Carlo quadrature and random walk leads to *particle tracing*. In the particle tracing technique the illumination in any environment is computed by simulating the flow of a large number of light particles originating at light sources, moving in different directions, getting absorbed, reflected and scattered depending on the object they hit. Its main advantage stems from the fact that it is easily extended to handle general environments. Particle tracing and its implementation has been described in detail earlier[6]. In Fig. 1 we reproduce some of the results of these implementations. In the rest of the section we will concentrate only on this method.

The solution of multidimensional integrations are best carried out by Monte Carlo quadrature technique[7, 8]. Monte Carlo quadrature of an integration of the form $\int F(x)dx$ involves sampling some probability distribution function[2](*pdf*) $f(x)$ and for each such sample evaluating $F(x)/f(x)$. The average of $F(x)/f(x)$ over a reasonably large number of samples is the estimate of $\int F(x)dx$. In Eq.2 we have an emission function, L_e, whose values are known. This function can be converted to some constant times the *pdf*. Let $L_e(x, \Theta_x)$ be converted to $\mathcal{E} \times S(x, \Theta_x)$ where $\mathcal{E} = \int_A \int_{\Omega_x} L_e(x, \Theta_x) \cos\theta_x d\omega_x dx$ and $\int_A \int_{\Omega_x} S(x, \Theta_x) d\omega_x dx = 1$. We can use S as the required *pdf* and carry out the the quadrature by sampling S. Any standard sampling technique may be used. For each such sample W_k is evaluated. As W_k is an integral equation of the second kind the random walk technique can be used for solving this integral equation.

A random walk is basically a sequence of steps. In order to formulate it we must define the set of all possible steps (discrete or continuous) of the system, a starting step and the transition probability function (T) for transition from one step (s) to another (s') such that $\int T(s \to s')ds' \leq 1$. The next step is chosen from the current step by sampling this transition

[2]f(x) is a probability distribution function or in short *pdf* if $f(x) > 0$ for all x in the domain of integration and $\int f(x)dx = 1$.

Figure 1: Rendered Views a Few Particle Traced Scenes.

probability function. The transition kernel is said to be normalised 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 from step s i.e. absorption at step s. Hence a random walk with a subcritical transition kernel is bound to terminate in a finite number of steps as every particle has eventually to be absorbed in some step. Whereas any random walk with a normalised kernel can go on for ever. In the latter case, the walk has to be terminated by some other external criterion. It is natural to expect that all environments would include some absorption. Thus the environment for illumination computation is always subcritical with $f_r(y, \Theta_y, \Theta_x)\cos\theta_y$ as the the subcritical transition kernel for solving the potential equation. A straight forward evaluation of W_k using a random walk results in paths consisting of a finite number of steps. The steps in our environment are the continuum of surface positions and hemispherical directions around each such surface position. The starting steps are sampled from the *pdf* $S(x, \Theta_x)$.

The evaluation of Eq.2 may be carried out by drawing n 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 step $(x_{i_k}, \Theta_{x_{i_k}})$ with probability

$$\sigma_{(x_{i_k}, \Theta_{x_{i_k}})} = 1 - \int_{\Omega_y} f_r(y, \Theta_y, \Theta_{x_{i_k}}) \cos\theta_y d\omega_y \tag{3}$$

or proceed to the next step $(x_{i_{k+1}}, \Theta_{x_{i_{k+1}}})$ 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 is estimated from the random walk by the formula given below:

$$< W_k(x_i, \Theta_i) >= g_k(x_{i_0}, \Theta_{x_{i_0}}) + \sum_{k=1}^{m_i} g_k(x_{i_k}, \Theta_{x_{i_k}}) \tag{4}$$

Then Φ_k can be estimated from n such walks as follows:

$$\begin{aligned}
\Phi_k &= \int_A \int_{\Omega_x} W_k(x, \Theta_x) L_e(x, \Theta_x) \cos\theta_x d\omega_x dx \\
&= \mathcal{E} \times \int_A \int_{\Omega_x} W_k(x, \Theta_x) S(x, \Theta_x) \cos\theta_x d\omega_x dx \\
&= \mathcal{E} \times \frac{1}{n} \sum_{i=1}^{n} \left[g_k(x_{i_0}, \Theta_{x_{i_0}}) + \sum_{k=1}^{m_i} g_k(x_{i_k}, \Theta_{x_{i_k}}) \right] \\
&= \mathcal{E} \times \frac{1}{n} \sum_{i=1}^{n} \sum_{k=0}^{m_i} g_k(x_{i_k}, \Theta_{x_{i_k}})
\end{aligned} \tag{5}$$

This method of evaluating Φ_k by first sampling the source function is the essence of *particle tracing*[2].

Particle tracing resembles the physical illumination process. Sampling of the source for a start step 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 and scattered by the objects in the environment until it is absorbed. One aspect which makes *particle tracing* an attractive method for global illumination purposes 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 i.e. formulate 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 towards the estimation of Φ_i for each of the subregions. At the end of the simulation we have the estimates for Φ_i for all the subregions.

However, the method is not without its disadvantages. The number of particles that need to be traced for computing the illumination accurately in a complex environment can be prohibitively large. A primary reason for this is that many particle traces are not necessarily effective when it comes to computing illumination within a reasonable but limited accuracy and could actually be wasteful tracing of the particles. A naive simulation of the physical model of light using light particles results in particle paths which are solely determined by the probability distribution functions that are used in various sampling steps of the simulation process. Many of these particle emissions and the paths traced may not in any way make a significant difference to illumination computations. For example during the course of simulation many particles may interact with an object even after the object's illumination has reasonably stabilised. Similarly many particles may be interacting with objects which are not very relevant to the illumination computation. For example surfaces which are never visible and/or do not illuminate other visible surfaces in any significant manner.

It is clear that if we have to reduce wasted particle emissions and wasteful particle tracings then we have to change the *pdf*'s used in the solution method so that all particles originate and get distributed in the most useful manner. In Monte Carlo parlance this is known as biasing.

5 Biasing

All the methods discussed under this topic transform the mathematical description of the stochastic process with an appropriately modified estimator for Φ in order to make the random walk process converge faster. The illumination process as described in Section 2 and 3 is completely described by the source function and the surface *brdf*s. If we replace them by suitably biased functions then when estimating Φ correctly we must remove the bias by properly compensating for the change. In particle tracing the compensation required is derived below:
Let $S'(x, \Theta_x)$ be the biased normalised source function.
If $T(\Theta_x \rightarrow \Theta_y) = f_r(y, \Theta_y, \Theta_x)\cos\theta_y$ is used to denote the transition function then
let $T'(\Theta_x \rightarrow \Theta_y)$ be the biased transition function.

$$\begin{aligned}
\Phi_k &= \mathcal{E} \times \int_A \int_{\Omega_x} S(x, \Theta_x) W_k(x, \Theta_x) \cos\theta_x d\omega_x dx \\
&= \mathcal{E} \times \int_A \int_{\Omega_x} S'(x, \Theta_x) \left(\frac{S(x, \Theta_x)}{S'(x, \Theta_x)} \right) W_k(x, \Theta_x) \cos\theta_x d\omega_x dx
\end{aligned}$$

$$
\begin{aligned}
W_k(x,\Theta_x) &= g_k(x,\Theta_x) + \int_{\Omega_y} f_r(y,\Theta_y,\Theta_x) W_k(y,\Theta_y)\cos\theta_y d\omega_y \\
&= g_k(x,\Theta_x) + \int_{\Omega_y} T'(\Theta_x \to \Theta_y)\left(\frac{T(\Theta_x \to \Theta_y)}{T'(\Theta_x \to \Theta_y)}\right) W_k(y,\Theta_y) d\omega_y
\end{aligned}
$$

The above equation can be written in a more compact form by defining a multiplication factor $\mathcal{F}$ where

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

The transformed potential equation can now be written as

$$
W_k(x,\Theta_x) = g_k(x,\Theta_x) + \int_{\Omega_y} T'(\Theta_x \to \Theta_y)\mathcal{F}(x,y) W_k(y,\Theta_y) d\omega_y \tag{6}
$$

With compensation the flux is now estimated by using the following equation:

$$
\Phi_k = \mathcal{E} \times \frac{1}{n}\sum_{i=1}^{n} \frac{S(x_{i_0},\Theta_{x_{i_0}})}{S'(x_{i_0},\Theta_{x_{i_0}})}\left[g_k(x_{i_0},\Theta_{x_{i_0}}) + \sum_{k=1}^{m_i}\left(\prod_{l=0}^{k-1}\mathcal{F}(x_{i_l},x_{i_{l+1}})\right) g_k(x_{i_k},\Theta_{x_{i_k}})\right] \tag{7}
$$

Below we first consider two special cases of this general biasing mechanism. The first is absorption suppression in which only the transition function is biased and not the source. The second is source biasing in which transitions are not biased. Later in Section 5.3 we discuss a more general method for biasing using the potential associated with surfaces.

5.1 Survival Biasing or Absorption Suppression

As the name implies, in this method the absorption probability at the transition points is reduced (may even be made zero) and as a consequence the random walk stretches to longer distances and the probability of each random walk making a nonzero contribution to the estimation of Φ_is is increased. The absorption probability σ at any step is given by Eq.3. Any reduction in this probability can be achieved by appropriate increases in the reflection probabilities. A very convenient method is to scale the reflection probabilities simply by the factor $\frac{1}{1-\sigma}$, consequently making the absorption probability at every step to zero. Thus the compensated estimate can be derived from Eq.7 as shown below:

$$
\Phi_k = \mathcal{E} \times \frac{1}{n}\sum_{i=1}^{n}\left[g_k(x_{i_0},\Theta_{x_{i_0}}) + \sum_{k=1}^{m_i}\left(\prod_{l=0}^{k-1}(1-\sigma_{x_{i_{l+1}}})\right) g_k(x_{i_k},\Theta_{x_{i_k}})\right] \tag{8}
$$

A word of caution is needed here; if the transition probability is changed such that there is no absorption at all then every single random walk will go on for ever 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[9, 2].

5.2 Source Biasing

In particle tracing, the emission function, $S(x, \Theta_x)$, plays an important role as every random walk originates at the light source. Any biasing of this function while still keeping the normalisation condition satisfied and the transition probability unaltered will change the form of flux estimation equation from Eq.7 to the one shown below:

$$\Phi_k = \mathcal{E} \times \frac{1}{n} \sum_{i=1}^{n} \frac{S(x_{i_0}, \Theta_{x_{i_0}})}{S'(x_{i_0}, \Theta_{x_{i_0}})} \left[g_k(x_{i_0}, \Theta_{x_{i_0}}) + \sum_{k=1}^{m_i} g_k(x_{i_k}, \Theta_{x_{i_k}}) \right] \tag{9}$$

where $S'(x, \Theta)$ is the biased source function for sampling.
As we shall see later source biasing provides a simple and convenient mechanism for improving the efficiency of particle tracing.

5.3 The Use of Approximate Potential for Biasing

Suppose we wish to bias our random walk process to improve the estimate of some specified region in the environment, say, the *region of importance*. This region of importance could be predefined. For example, in rendering a view of a 3D environment, the set of all visible surfaces could form the region of importance. Alternatively the region of importance could be adaptively defined as the solution progresses. This would imply that the importance of regions would change depending on the values computed from a partial simulation. Biasing of particle tracing must be such that the resulting emissions and transitions must lead most of the the random walks directly or indirectly to the region of importance. Further the computations required by the biasing scheme must be simple and straight forward. One possible scheme is to suitably weight the emission function and the transition probability functions.

The potential towards the region of importance provides an excellent basis for this weighting. Let $S(x, \Theta_x) > 0$. Then we can bias the source function to $S'(x, \Theta_x)$ such that the $S'(x, \Theta_x)$ is much greater than $S(x, \Theta_x)$ for those points, (x, Θ_x), whose potential is higher and $S'(x, \Theta_x)$ is much lower than $S(x, \Theta_x)$ for those points whose potential is lower. We can similarly bias the transition probability i.e. in our case the reflection probability as follows: Consider two directions Θ_1 and Θ_2 in the outgoing hemisphere at point x. Let S_1 and S_2 be the two surfaces nearest to point x along directions Θ_1 and Θ_2 respectively. Denote their potential towards the region of importance by W_1 and W_2 respectively. Without loss of generality assume that $W_1 > W_2$. Then the transition function T at x must be biased such that $T(x, \Theta_1)$ is much greater than $T(x, \Theta_2)$.

There is however one catch to the above biasing scheme. It will work provided we know the value of potential that all surfaces in the environment have towards the region of importance. It is clear that if we can derive the exact potential values then we can also derive the solution for the problem at hand and hence we do not require the simulation. Fortunately for biasing purposes we need not know the exact potential values. It is sufficient to obtain approximate values of this potential, hopefully, with much reduced computational costs. Provided these approximate values maintain their relative ordering they can be effectively used to bias the emission and transition probability functions.

6 Computation of Approximate Potential and Biasing

In its general form the potential function is dependent both on 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 for biasing we shall make the following simplifying assumptions:

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

With the above assumptions the environment can now 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 $P_i \in E$, $\mathcal{W}_i$ denotes the hemispherical potential that patch P_i has towards 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, i.e. emitted/reflected, say N_i,
- the number of these particles reaching a patch belonging to the region of importance, say M_i.

The ratio $\frac{M_i}{N_i}$ gives us an estimate of the hemispherical potential of patch P_i.

6.1 Hemispherical Potential

Here we introduce the notion of a *hemispherical potential* of patch i and denote it by $\mathcal{W}_{k,i}$. $\mathcal{W}_{k,i}$ is obtained as the average potential of the points of the patch in any hemispherical direction. If the patches are sufficiently small this hemispherical potential may be assumed to be independent of the position on each patch. The expression for this hemispherical potential is derived below:

$$\begin{aligned}
\mathcal{W}_{k,i} &= \frac{1}{\pi}\int_{\Omega_{x_i}} W_k(x_i, \Theta_{x_i}) \cos\theta_{x_i} d\omega_{x_i} \\
&= \frac{1}{\pi}\int_{\Omega_{x_i}} \left[g_k(x, \Theta_x) + \int_{\Omega_y} f_r(y, \Theta_y, \Theta_x) W_y(y, \Theta_y) \cos\theta_y d\omega_y \right] \cos\theta_{x_i} d\omega_{x_i} \\
&= \frac{1}{\pi}\int_{\Omega_{x_i}} [g_{k,i} + \pi f_r(j)\mathcal{W}_{k,j}] \cos\theta_{x_i} d\omega_{x_i}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\pi} g_{k,i} \int_{\Omega_{x_i}} \cos\theta_{x_i} d\omega_{x_i} + \int_{\Omega_{x_i}} f_r(j)\mathcal{W}_{k,j} \cos\theta_{x_i} d\omega_{x_i} \\
&= g_{k,i} + \sum_{j=1}^{N} f_r(j) \int_{\omega_{ij}} \mathcal{W}_{k,j} \cos\theta_{x_i} d\omega_{x_i} \\
&= g_{k,i} + \sum_{j=1}^{N} f_r(j)\mathcal{W}_{k,j} \int_{\omega_{ij}} \cos\theta_{x_i} d\omega_{x_i} \\
&= g_{k,i} + \sum_{j=1}^{N} f_r(j)\mathcal{W}_{k,j} F_{ij} \qquad (10)
\end{aligned}
$$

Using this hemispherical potential we can derive the simplified expression for the total flux over the k-th patch as follows:

$$
\begin{aligned}
\Phi_k &= \int_A \int_{\Omega_x} W_k(x,\Theta_x) L_e(x,\Theta_x) \cos\theta_x d\omega_x dx \\
&= \sum_{i=1}^{ns} \int_{A_i} \int_{\Omega_x} W_k(x,\Theta_x) L_e(x,\Theta_x) \cos\theta_x d\omega_x dx \\
&= \pi \sum_{i=1}^{ns} L_e(i)\mathcal{W}_{k,i} \int_{A_i} dx \\
&= \pi \sum_{i=1}^{ns} L_e(i) A_i \mathcal{W}_{k,i} \qquad (11)
\end{aligned}
$$

where ns is the total number of source patches in the environment.

6.2 Source Position Biasing using Hemispherical Potential

If $\mathcal{W}_i$ is the hemispherical potential of patch P_i then $\mathcal{W}(x)$ is also the hemispherical potential of point x, where $x \in P_i$. Using $\mathcal{W}(x)$ we can bias the normalised source function $S(x)$. Renormalising the biased source function then gives us the following definition:

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

Where $S0 = \int_A S(x)\mathcal{W}(x)dx \approx \sum_{i=1}^{ns} S_i \mathcal{W}_i A_i$ and ns is the number of source patches.
This biasing 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 biasing for each particle the brightness contributing strength is multiplied by a factor, f_1. The expression for f_1 is given below:

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

6.3 Direction Biasing using Hemispherical Potential

Direction biasing is used both for choosing the direction for emission and for reflection. In the normal simulation the direction is chosen by sampling the diffuse distribution function. In the biased 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 direction biasing using hemispherical potential it may be worth while to look at Eq.10 again, which gives a linear expression for the hemispherical potential, and Eq.11 which gives us an expression for flux using the hemispherical potential:

$$\begin{aligned}\mathcal{W}_i &= g_i + \sum_{j=1}^{N} f_r(j)\mathcal{W}_j F_{ij} \\ \Phi &= \pi \sum_{i=1}^{ns} A_e(i) A_i \mathcal{W}_i\end{aligned}$$

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 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. Further directions within that range are sampled to choose at the direction of flight for the particle. The resulting mathematical change to Eq.10 is as follows:

$$\mathcal{W}_i = g_i + \sum_{j=1}^{N} f_r(j) \frac{\mathcal{W}_j}{\mathcal{W}_{j,approx}/\mathcal{W}0} F'_{ij} \tag{12}$$

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

There is one major task to be carried out in implementing this idea which is finding the F'_{ij} distribution around a point p on patch P_i. We have already discussed a simple method of computing approximate potential earlier in this section. The task that remains is to determine the F_{ij} distribution in the outgoing hemisphere of the point p. Though it is possible to do this by carrying out hemicube projections coupled with depth sorting it is impractical to use this method for every transition of a particle. Of course the fact that we need only relative importance of the surrounding patches implies that the exact F_{ij} values are not necessary. Once again any suitable approximation which maintains this relative ordering of F_{ij} would do. We have devised a simple method of obtaining this information from the partial simulation used to compute approximate hemispherical potential. This method is based on the following observation:

> If particles are shot diffusely towards the outgoing hemisphere from a point p of the i-th patch, P_i, then the number of particles reaching the j-th patch, P_j,

visible to this point is proportional to F_{ij}. If N is the total number of particles shot from p of a patch P_i and M is the number directly reaching patch P_j then the ratio $\pi\frac{M}{N}$ approaches F_{ij} as N increases. Obviously if a patch is not visible to the point p then M would be zero and so would be F_{ij}.

To be able to capture this information from an unbiased particle tracing process we have used a simple data structure. The data-structure is a 2-D array of size $N \times N$. We shall name this data structure as *Vis*. Each row of *Vis* corresponds to an immediate emitter and each column corresponds to an immediate receiver. During the process of particle tracing, for every emission/transition, the array cell corresponding to the row of the source patch and the column of the target patch is incremented. $\pi Vis[i][j]/\sum_j Vis[i][j]$ is then used to obtain the approximation to F_{ij}. This value multiplied with the computed approximate hemispherical potential gives us F'_{ij}. The F'_{ij} of all the N patches results in a discrete distribution of patches for transition, or in other words, directions for emission/reflection. By sampling this distribution we get a patch for transition and by sampling the surface of the patch we get the point of transition. The current point and the chosen point of transition define the direction of flight for the particle. This method however has a problem. A patch may only be partially visible. Hence the chosen point of transition on the patch may give rise to a direction which is hidden from the source point. This problem is avoided by rejection sampling. That is the surface is resampled until we arrive at a proper direction. The biasing algorithm is now given below, assuming that at the point p on patch P_i a direction has to be chosen.

1. Compute the discrete distribution of F'_{ij}, i.e. $\frac{\mathcal{W}_{j,approx}F_{ij}}{\sum_j \mathcal{W}_{j,approx}F_{ij}}$ of patches around p.
2. Discrete sample the above distribution and choose a patch say k.
3. do{

   Sample the surface of the patch k and choose the transition point y on the k-th patch

   }while(transition point y on the k-th patch is not visible to point p).
4. Choose the interaction.
   If the interaction is not absorption then set $p = y$, $i = k$ and repeat from step 1.

For the proper computation of flux the compensation factor which appropriately modifies the brightness contribution of the particle is derived from equation (Eq.12) and is as follows:

$$f_2 = \frac{\mathcal{W}0}{\mathcal{W}_{j,approx}} = \frac{\sum \mathcal{W}_{j,approx}F_{ij}}{\mathcal{W}_{j,approx}}$$

We have implemented the above biasing mechanisms and have applied it to a number of cases. The resulting improvements in efficiency have been extremely encouraging. Below we discuss these results in a little more detail. Let us first consider the situation in which R is predefined. The simulation then proceeds in two distinct phases. In the first phase the approximate potential values are computed, while in the second phase the computed

Figure 2: Scene for importance biasing with predefined R.

potentials are used to bias the source and transition functions and a biased simulation is carried out to obtain global illumination information in the environment. We demonstrate the improvements due to biasing towards a predefined R by using a simple environment, a view of which is shown in Fig. 2. The vertical wall on the left extreme has been defined as the region of importance. The wall has been divided into 32×16 patches. Fig.3 shows the particle incidence map on the wall with a total of 100,000 particles traced in the simulation. Fig.3(a) is the map for normal simulation and Fig.3(b) is the map for biased simulation. As one can see visually there is appreciable improvement. The quantitative figures are as follows: 46,462 incidences in the unbiased simulation and 352,922 incidences in the biased simulation. The number of samples rejected during the transition biasing is 69,520 giving an over all improvement factor of 4 with equivalent computation effort.

In the above we have assumed that the region of importance, R, is predefined, and that R forms a small subset of the entire environment. The basic strategy has been based on the use of approximate potential values obtained from a small simulation run for biasing and thus directing most of the random walks to R. This situation is typical of view dependent illumination computation. On the other hand for view independent illumination computation clearly the whole environment is the region of importance. Biasing techniques that direct random walks to a region of importance therefore are not meaningful. However this biasing mechanism could still be used effectively to improve computational efficiency of the simulation provided we could devise a strategy like the one stated below:

> 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 biasing is done for each new subset of important regions.

In our work we have been able to devise one such strategy. This is based on the observation

Figure 3: The plot of particle incidences on the region of importance.

that, as the simulation progresses, some of the regions of the environment would have received enough particle incidences so that the illumination estimates due to even more incidences can be said to be reasonably invariant. That is, as far as these regions are concerned the simulation need not be continued. We make use of this in order to reduce the set of important regions, R, by saying that the regions which have received enough incidences are no more of importance. What we need is then the capability to bias again with this R and carry out a further simulation.

The simulation starts with the whole environment as R (equivalent to no biasing). After a reasonable number of particle traces, R is reduced by removing those regions having close to equilibrium illumination values. The simulation is continued after using the earlier simulation results to bias towards the new R. The process is continued until R is empty.

The two important tasks in realising the above strategy are:

1. A method of deciding on when a subregion of R has reached near equilibrium illumination.

2. A method of computing approximate hemispherical potential for all the regions of the environment for the reduced R.

For the present the first task has been carried out by taking a very simple approach. Regions receiving particle incidences above some predefined number are assumed to have reached equilibrium status. Of course in practice this strategy would have to be much more sophisticated and would have to depend on other criteria which enable one to decide that adequate incidences have been registered over a patch. So after each simulation R is scanned and pruned. As the simulation progresses R is redefined many times. It is therefore not possible to compute the hemispherical potential values only once at the start as was done earlier. Instead we store the complete history of the particle traces from the initial unbiased run. Every time R is redefined these traces are scanned and the new hemispherical potential is computed.

The expression for the biased transition function F'_{ij} is the same as before. Because we have successive biasing we shall use $F_{ij}^{(n)}$ instead of F'_{ij} and we get the recursive expression for it below:

$$F_{ij}^{(n)} = F_{ij}^{(n-1)} \frac{\mathcal{W}_{j,approx}}{\sum_j F_{ij}^{(n-1)} \mathcal{W}_{j,approx}}$$

For computation of $F_{ij}^{(n-1)}$ one approach is to use the recursive expansion till one reaches $F^(0)_{ij}$ (i.e. F_{ij} of our earlier experiment) and use the Vis data structure computed from the initial unbiased run. The other approach is to update Vis in every simulation and extract $F_{ij}^{(n-1)}$ directly from the updated Vis. This latter approach is what we have used. We believe it is more efficient due to the fact that the information in Vis is enriched in each simulation. The compensated strength of the particle at the n-th biasing step then becomes:

$$f_2^{(n)} = \frac{\sum_j F_{ij}^{(n-1)} \mathcal{W}_{j,approx}}{\mathcal{W}_{j,approx}}$$

We have used this scheme for computing the illumination in an environment like a maze similar to the *Cornell Labyrinth*[3] (see Fig.4 and Fig.5). It has a total of 523 patches, all of more or less the same area. We have chosen 100 as the minimum number of particle incidences on a patch after which we assume that the patch has reached equilibrium illumination. For an unbiased simulation if each and every patch had to receive at least 100 particle incidences then the total number of particles that had to be traced in the entire simulation was 27,000,000. In the case of a biased simulation, 300,000 particles without any bias were first traced. The results were used to prune R to result in 205 patches. The particle tracing history of all these 300,000 was stored for computing $\mathcal{W}$ as and when necessary. The subsequent simulation runs were carried out in batches of 3000 particles each. After each biasing run R was updated and a new $F^{(n)}$ computed. Table 1 gives some of the statistics from this experiment. In the table the column corresponding to "Rejected Samples" indicates the number of times the position sampling on the patches for choosing a transition direction resulted in hidden transition points. As can be seen efficiency improvement due to biasing is enormous.

7 Conclusion

The potential equation for illumination is a powerful mathematical tool for illumination computation by what have usually been called forward simulation techniques or light shooting strategy methods. The most popular of these, progressive radiosity, is an analytic solution

[3] The name has been chosen because a similar scene was chosen by the Cornell group in a recent SIGGRAPH presentation on *Importance Driven Radiosity*[10].

Figure 4: Wire frame drawing showing the top view of the Cornell Labyrinth.

Batch Size	Region of Importance		Rejected Samples	Hits
	Before	After		
300000(Normal)	523	205	0	5625
3000(Biased)	205	131	4142	10266
3000(Biased)	131	72	3758	10747
3000(Biased)	72	26	4048	10816
3000(Biased)	26	13	3642	10494
3000(Biased)	13	8	4017	10706
3000(Biased)	8	6	4224	10247
3000(Biased)	6	4	4324	10513
3000(Biased)	4	0	3888	11083

Table 1: Biasing improvements for Cornell Labyrinth.

Figure 5: A rendered view of the Cornell Labyrinth.

to this equation. Monte Carlo quadrature and random walk methods that can be devised for solving this equation are more general, in the sense that, the simplifying assumptions of ideal diffuse behaviour made for progressive radiosity are not any more necessary. The idea of using the potential for biasing and improving the efficiency of the Monte Carlo solution has been used in other disciplines like Neutron Transport. Its application to illumination computation is not only interesting but also very beneficial. The use of illumination computation in Computer Graphics is for imaging and this naturally defines visible regions as being more important. Using the potential for biasing random walks towards these regions of importance has resulted in very high efficiency improvement factors. Similarly the strategy of successively pruning the region of importance, recomputing the bias and carrying out continuously biased simulations has also proved to result in very high efficiency. This in spite of the fact that the biasing scheme that has been devised and implemented is rather simple and straightforward. Certainly one can expect more sophisticated biasing techniques resulting in even more efficient Monte Carlo solutions to the potential equation and its use for illumination flux computation.

8 Acknowledgements

We are grateful to Dr S Ramani, Director, NCST for his support and encouragement.

References

[1] James T. Kajiya. The rendering equation. *Computer Graphics (SIGGRAPH '86 Proceedings)*, 20(4):143–150, Aug. 1986.

[2] S. N. Pattanaik. *Computational Methods for Global Illumination and Visualisation of Complex 3D Environments*. PhD thesis, submitted to Birla Institute of Technology & Science, Pilani, from National Centre for Software Technology, Feb 1993.

[3] Michael Cohen, Shenchang Eric Chen, John R. Wallace, and Donald P. Greenberg. A progressive refinement approach to fast radiosity image generation. *Computer Graphics (SIGGRAPH '88 Proceedings)*, 22(4):75–84, Aug. 1988.

[4] S. N. Pattanaik. The mathematical framework of adjoint equations for illumination computation. In S. P. Mudur and S. N. Pattanaik, editors, *Proceedings of International Conference on Computer Graphics*. IFIP-CSI, Bombay, India, North-Holland, February 1993.

[5] S. N. Pattanaik and S. P. Mudur. The potential equation and importance in illumination computations. *Computer Graphics Forum*, 12(2):131–136, 1993.

[6] S. N. Pattanaik and S. P. Mudur. Computation of global illumination in a participating medium by monte carlo simulation. *The Journal of Visualisation and Computer Animation, To appear*, 4, 1993.

[7] Malvin H. Kalos and Paula A. Whitlock. *Monte Carlo Methods. Volume I.* John Wiley & Sons, 1986.

[8] Reuven Y. Rubinstein. *Simulation and the Monte Carlo Method.* John Wiley & Sons, 1981.

[9] James Arvo and David Kirk. Particle transport and image synthesis. *Computer Graphics (SIGGRAPH '90 Proceedings)*, 24(4):63–66, Aug. 1990.

[10] Brian E. Smits, James R. Arvo, and David H. Salesin. An importance driven radiosity algorithm. *Computer Graphics (SIGGRAPH '92 Proceedings)*, 26(2):273–282, July 1992.