The Mathematical Framework of Adjoint Equations for Illumination Computation

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

February 28, 1996

Abstract

There are two basic strategies used for carrying out the 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 provides the mathematical basis for the gathering methods and the potential equation provides the mathematical basis for the shooting methods. They together form an adjoint system of equations. In this paper, using the mathematical framework of the adjoint equations we review illumination computation methods, categorising them as using the gathering or shooting strategy or both. Another basis for categorisation is the basic equation solution strategy used, namely deterministic or nondeterministic.

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[1, 2, 3, 4] based on this approach. The other approach is based on simulating the propagation of light starting from the light sources of which progressive radiosity[5] and particle tracing[6] are representative methods. While illumination computation methods for computing the flux reaching the surface or the eye are primarily based on the radiance equation, the potential equation provides the basis for methods which simulate light propagation starting from the light sources. In this paper we first derive the radiance and potential equations and then review the existing methods within the framework of these two equations, particularly by considering the basic equation solving strategy adopted in the methods.

2 The Radiance Equation

The radiance from any surface point of an opaque object in a nonparticipating environment is due to emission from that point and due to reflection of incident radiation from the incoming hemisphere around the point. So the expression for the outgoing radiance L_o from surface point x along direction Θ_x can be given as:

$$L_o(x, \Theta_x) = L_e(x, \Theta_x) + L_r(x, \Theta_x) = L_e(x, \Theta_x) + \int_{\Omega_x} f_r(x, \Theta_x, \Theta_i) L_i(x, \Theta_i) \cos\theta_i d\omega_i$$

where

> $L_o(x, \Theta_x)$, $L_e(x, \Theta_x)$ and $L_r(x, \Theta_x)$ are respectively the outgoing, emitted and reflected radiance from point x along Θ_x direction,
> $L_i(x, \Theta_i)$ is the incident radiance at x from the incoming Θ_i direction,
> θ_i is the cone angle of the incoming direction,
> $d\omega_i$ is the differential solid angle around the incoming direction,
> Ω_x is the incoming hemisphere around x and f_r is the surface *brdf*.

This equation is one of the fundamental equations used in illumination computation. In an environment, incoming radiance at x is due to the outgoing radiance at some point y visible to x along that direction. So we may rewrite the above equation as follows:

$$L_o(x, \Theta_x) = L_e(x, \Theta_x) + \int_{\Omega_x} f_r(x, \Theta_x, \Theta_y) L_o(y, \Theta_y) \cos\theta_x d\omega_x \tag{1}$$

where $L_o(y,\Theta_y)$ is the outgoing radiance at point y visible to x along the direction Θ_y. We shall hence forth refer to Eq.1 as the *Radiance Equation*[1].

3 Potential Equation

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:

$$W(x,\Theta_x) = g(x,\Theta_x) + \int_{\Omega_y} f_r(y,\Theta_y,\Theta_x)W(y,\Theta_y)cos\theta_y d\omega_y \qquad (2)$$

4 Illumination Computation

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 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.

[1]Kajiya's rendering equation[3] is basically a slightly different form of this equation.

An expression for this flux using the radiance function will therefore be an integral of the form:

$$\Phi = \int_{position\ spread} \int_{direction\ spread} L_o(x, \Theta_x) cos\theta_x d\omega_x dx$$

If we assume that this flux represents the outgoing flux from $\mathcal{S}$ then we can use the earlier defined function $g(x, \Theta_x)$ which evaluates to 1 in the limits of the integration and 0 everywhere else, and rewrite the above equation as follows:

$$\Phi = \int_A \int_{\Omega_x} L_o(x, \Theta_x) g(x, \Theta_x) cos\theta_x d\omega_x dx \tag{3}$$

where A represents all the surface points and Ω_x represents the incoming hemispherical directions around all such points.

On the other hand an expression for the same flux using the potential function and radiance of the emitting points in the environment is as follows:

$$\Phi = \int_A \int_{\Omega_x} W(x, \Theta_x) L_e(x, \Theta_x) cos\theta_x d\omega_x dx \tag{4}$$

To sum up

- We have two different equations, Eq.3 and Eq.4, to express the same quantity Φ using two different functions L and W.
- Eq.3 and Eq.4 are similar in form and so also are Eq.1 and Eq.2 for L and W.

The equations thus form an adjoint system of equations. The flux computation equations Eq.3 and Eq.4 may be seen respectively as mathematically representing the process of gathering illumination coming in from all directions of the incoming hemisphere, and the process of shooting the illumination in all directions into the outgoing hemisphere[2]. In this paper we show that all the illumination computation methods either follow a gathering strategy or a shooting strategy or a combination of two.

Both radiance equation and the potential equation are linear integral equations of the second kind. A closed form solution for a general equation of this kind does not exist. One has to resort to various numerical quadrature

[2] The terms gathering and shooting were first introduced by Cohen *et al* in the context of explaining the difference in the strategy used by the full radiosity solution and the progressive radiosity algorithms[5].

techniques to get a reasonably accurate solution for this integral equation. Such methods may be considered as being deterministic or nondeterministic. Nondeterministic methods are generally based on principles of Monte Carlo quadrature. A method specific to the solution of integral equation of the second kind is the random walk. Monte Carlo quadrature and random walk techniques are briefly discussed in later sections. The choice of any particular method depends on its ability to handle a range of surface geometries, surface *brdf*s and surface emission properties.

5 Gathering by Deterministic Methods

5.1 Local Illumination Model and Ray Tracing

For the solution purpose it is often convenient to break up the incoming illumination in the integral part of Eq.1 into two components: illumination due to emission and due to reflection. Then we can have the following form for the radiance equation:

$$L_o(x,\Theta_x) = L_e(x,\Theta_x) + \int_\Omega f_r(x,\Theta_x,\Theta_y)\left[L_e(y,\Theta_y) + L_r(y,\Theta)\right]\cos\theta_x d\omega_x \quad (5)$$

The simplest approach towards solving the radiance equation is to approximate the integration over the hemisphere by a summation. If we assume that the light sources are differential elements and the indirect illumination from the incoming hemisphere is not very significant and hence may be considered as uniform then from Eq.5 we arrive at the following simplified form of the radiance equation:

$$L_o(x,\Theta_x) = L_e(x,\Theta_x) + L_a\rho_a(x) + \sum_{j=1}^{ns} f_r(x,\Theta_y(j),\Theta_x)L_e(j)\cos\theta(j)\Delta\omega(j) \quad (6)$$

where

> L_a is the uniform inter-reflection radiance (ambient term), ρ_a is the average hemispherical reflectance of point x,
> ns denotes the number of visible light sources,
> $L_e(j)$, $\Theta(j)$, $\Delta\omega(j)$ and $\theta(j)$ are all defined with respect to the differential j-th light source.

The earliest illumination computation methods use equations similar to Eq.6 and are popularly known as being based on *local illumination models*. The term local is because the illumination predicted by the equation is due to the direct or local effects of the light sources, with little consideration of effects such as inter-reflection amongst objects in the environment. Because of the computational simplicity these methods have been used very widely for many years, to produce shaded pictures of 3D objects. The very first equation of this kind is due to Bouknight[7] who gives the expression for the radiance[3] from diffuse surfaces as follows:

$$L_o(x) = \rho_a(x) L_a + k_d(x) \sum_{j=1}^{ns} L_e(j) \cos\theta(j) \quad (7)$$

As the outgoing radiance is considered from a perfectly diffuse surface L_o is independent of the outgoing direction,
k_d is the diffuse reflection coefficient, which take values from 0 to 1.

Phong[8] subsequently introduced an important improvement to this model for supporting shiny surfaces. Reflection from every surface is assumed to have a diffuse reflection component and an imperfect specular reflection component which is modelled by Phong's reflection model. With this extension Eq.7 takes the following form:

$$L_o(x, \Theta_x) = \rho_a(x) L_a + \sum_{j=1}^{ns} L_e(j) \left[k_{diffuse}(x) \cos\theta(j) + k_{specular}(x) \cos^n \alpha(j) \right] \quad (8)$$

where

$\alpha(j)$ is the angle between the mirror reflection direction of the j-th light source and the outgoing direction Θ_x,
$k_{specular}$ is the specular reflection coefficient, which take values from 0 to 1 subject to the condition that $k_{diffuse} + k_{specular} < 1$,
n is a an empirical surface roughness parameter.

Cook *et al*[9] proposed another equation using Phong's reflection model which was very much like Eq.6 and had the following form:

$$L_o(x, \Theta_x) = \rho_a(x) L_a + \sum_{j=1}^{ns} L_e(j) \left[k_{diffuse}(x) + k_{specular}(x) \cos^n \alpha(j) \right] \cos\theta(j) d\omega(j)$$

[3]Intensity (I) was the actual term used instead of radiance (L), but with the same meaning. The term intensity is still used by many for radiance. However they are not the same according to their definitions.

We see in the above that light is gathered accurately only from point light sources and for gathering from everywhere else in the hemisphere a very approximate term called ambient illumination has been used. For a more accurate solution of the radiance equation it is essential that more precise methods of gathering illumination from everywhere else be used. The basic ray tracing method attempts to do this by providing a mechanism to gather information from some other dominant directions also. In particular, for calculating illumination from a shiny surface it probes along the mirror reflection direction by sending a reflected ray. So if a significant amount of indirect light is incident from a mirror reflection direction then it is accounted for. Thus for computing the interreflection component, a ray is traced[1] along Θ_y^{-1}, the mirror reflection direction of Θ_x, where y is the visible point when viewed from x along Θ_y^{-1}. The radiance from y, $L_o(y, \Theta_y)$, is added to $L_o(x, \Theta_x)$ after taking into account the losses due to absorption at x. By incorporating the modification suggested above into Eq.6 we arrive at the following expression for computing illumination using ray tracing.

$$L_o(x, \Theta_x) = \rho_a(x)L_a + \rho_s(x)L_o(y, \Theta_y) + \sum_{j=1}^{N} f_r(x, \Theta_y(j), \Theta_x)\Delta\omega(j)L_e(j)\cos\theta(j) \quad (9)$$

where $\rho_s(x)$ is the specular reflectance of the surface at x.
The difficulty with evaluating this expression is that $L_o(y, \Theta_y)$ is also an unknown and hence needs further evaluation. If this calculation is carried out again using Eq.9 then it is called as recursive ray tracing.

The ray tracing method geometrically speaking is basically a visibility computation method. If the ray is intersected with all the objects in the environment then the nearest of all the intersections points is the visible point from the origin of the ray along the direction in which the ray is traced. Ray-object intersections play a major role in ray-tracing based methods. Fortunately, a ray is a simple 1D linear geometric primitive. Efficient algorithms for computing the intersection of a ray with a large number of object shapes have been devised. That is why a ray tracing method in general does not impose any restriction on the type of object shapes that it can handle. In ray tracing most of the time is spent in computing ray-object intersection. A very large number of extensions are basically acceleration methods for speeding up the ray object intersection computations. A good description of various ray-object intersection and acceleration techniques may be found in [10].

5.2 Radiosity

The basic radiosity method[4] provides a solution for an environment with surfaces exhibiting diffuse behaviour. The improvements to the basic radiosity method are in the form of extensions to support specular surface behaviour and general surface reflectance property.

In an environment every object surface can be considered as being completely surrounded by an envelope of surfaces of other solid objects or open areas. This envelope is the enclosure for the surface and it accounts for all directions surrounding the surface. By considering the radiation from the given surface to all parts of the enclosure and the radiation arriving at the surface from all parts of the enclosure, all the radiative contributions are accounted for. Because of this enclosure assumption every solid angle in the incoming or outgoing hemisphere around a point will be covered by a surface. If the environment is discretised to a number of small surface patches then the hemisphere around any surface point can be represented as a sum of solid angles occupied by each visible surface patch on the hemisphere. In general we may represent it as a sum of solid angles due to each of the surface patches, $\sum_N \Delta\omega(j)$ where N is the total number of surface patches. A surface patch, j, completely hidden to a point will have $\Delta\omega(j) = 0$. With the assumptions of diffuse reflection behaviour and uniformity of radiance over a patch we get the following simplified form for Eq.1.

$$L_o(i) = L_e(i)+f_{diffuse}(i)\sum_{j=1}^{N} L_o(j)\int_{\Delta\omega(j)} \cos\theta d\omega = L_e(i)+f_{diffuse}(i)\sum_{j=1}^{N} L_o(j)\pi F(ij) \quad (10)$$

where $F(ij) = \frac{1}{\pi}\int_{\Delta\omega(j)} cos\theta d\omega$.
We may interpret $F(ij)$ as the fraction of the total outgoing flux from the i-th patch reaching the j-th patch. So $\sum_{j=1}^{N} F(ij) = 1$. $F(ij)$ contains only geometry related terms and hence is termed as *geometric factor*, or *form-factor*. Eq.10 is a simple linear equation, containing geometry dependent and geometry independent terms for obtaining the radiance value at a point on the i-th patch of a diffuse 3D environment. The original equations used in radiosity methods contain radiosity, B, in the equation instead of the outgoing radiance L_o. The term radiosity means the rate of radiant energy outgoing per unit area from a surface. For a diffuse surface radiosity and radiance are related by the expression: $B = \pi L_o$. So we can easily convert

the above equation to the radiosity equation by multiplying π on both sides to obtain:

$$B(i) = E(i) + \rho_d(i) \sum_j B(j)F(ij) \tag{11}$$

where

> $B(i)$ and $E(i)$ are respectively the total radiosity and the radiosity due to emission from the i-th surface patch,
> $\rho_d(i) = \pi f_{diffuse}(i)$ is the diffuse surface reflectance of the i-th patch.

Writing down one equation for each of the surface patches in the environment results in a set of linear equations which can be solved for obtaining the equilibrium radiosity values.

If the environment consists of N patches, then we get a system of N linear equations with N unknowns of the form:

$$\begin{bmatrix} 1-\rho_d(1)F(11) & \dots & -\rho_d(1)F(1i) & \dots & -\rho_d(1)F(1N) \\ . & \dots & . & \dots & . \\ . & \dots & . & \dots & . \\ -\rho_d(i)F(i1) & \dots & 1-\rho_d(i)F(ii) & \dots & -\rho_d(i)F(iN) \\ . & \dots & . & \dots & . \\ . & \dots & . & \dots & . \\ -\rho_d(N)F(N1) & \dots & -\rho_d(N)F(Ni) & \dots & 1-\rho_d(N)F(NN) \end{bmatrix} \begin{bmatrix} B(1) \\ . \\ . \\ B(i) \\ . \\ . \\ B(N) \end{bmatrix} = \begin{bmatrix} E(1) \\ . \\ . \\ E(i) \\ . \\ . \\ E(N) \end{bmatrix}$$

or in short $A\,B = E$.
This set of equations has the unique characteristic of being diagonally dominant and hence is amenable to efficient solution by iterative methods such as Gauss-Seidel's. The major problem in using this method is the setting up of the system of equations. This amounts to calculating all the $F(ij)$ values. As shown in the derivation, $F(ij)$ is related to the solid angle subtended by the visible portion of the j-th patch over the points of the i-th patch. This turns out to be the most expensive step in radiosity methods.

The simplification of Eq.1 to a system of linear set of equations with geometry dependent and geometry independent terms was possible only due to the assumption of the diffuse reflecting nature of the surface patches. Such a simplification is no more possible if we assume a more general form for the surface *brdf* and the f_r term can not be taken out from inside the integration. Further, because of the directional independence of the radiance from diffuse surfaces, a single radiosity value was good enough to represent the outgoing radiance in any direction. However such simplification is not possible for surface patches with general reflectance behaviour because the complex *brdf* gives rise to complex directional distribution of radiance over the surface patches. We shall briefly discuss below a few of the computational strategies that have been suggested for extending the radiosity method to environments with non-diffuse surfaces.

5.2.1 Extensions to Non-Diffuse environment

(i) **Two-Pass Solutions:** Two-pass solutions[11, 12, 13] are based on the assumption
that in an environment majority of the surfaces are diffuse and only a few of the surfaces are non-diffuse. With this assumption it is possible to formulate a set of linear radiosity equations for the diffuse surfaces of the environment in which the non-diffuse surfaces play a role in light propagation by providing indirect light transport paths. The radiosity equations must now include geometry terms, called as *extended form-factors*. The extended form-factor is the fraction of total outgoing flux from the i-th surface patch reaching j-th surface patch directly and indirectly due to one or multiple reflections of this light by non-diffuse surfaces in their propagation path. Though not computationally simple it is possible to compute these extended form factors if we assume that the non-diffuse surfaces in the environment exhibit perfect specular behaviour.

The solution proceeds in two passes. In the first pass the equilibrium radiosity values for the diffuse surfaces of the environment are computed by solving the linear set of equations corresponding to the diffuse patches. In the second pass the radiance values for the non-diffuse surfaces in specific directions are computed by sampling the *brdf* of the non-diffuse surfaces. As most of the surfaces in the environment are diffuse, the sampled directions largely lead to diffuse surfaces whose radiance values are already known as a result of the first pass.

(ii) **Direction Discretisation:** This method[14] attempts to solve the directional distribution problem by discretising the hemispherical directions around a non-diffuse patch into a finite number of solid angles, in which the brightness is assumed uniform. Surfaces are also divided into small discrete patches. For each discrete direction around the patch, a radiosity like linear equation is formulated. Energy leaving every discrete solid angle is obtained by solving the linear set of equations.

For an accurate solution by this method one has to resort to very fine discretisation of the surfaces and the directions around the surface patches. Memory and computational expenses then become very high making it practically impossible to carry out the solution even for moderately complex environments. The discrete representation of directions also gives rise to severe aliasing problems.

6 Gathering by Nondeterministic Methods

Monte Carlo quadrature and Random Walk are the two main non-deterministic methods that have been used for gathering illumination from the incoming hemisphere. These methods are called nondeterministic or probabilistic because repeated applications of a solution method to the same problem are not guaranteed to give identical results. The methods include steps that depend not only on the input but also on results of some random events.

6.1 Monte Carlo Solution of Radiance Equation

The main principle behind a Monte Carlo quadrature for computing the integral $\int F(x)dx$ is as follows:

1. Rewrite *F(x)* as a product $f_1(x)f_2(x)$ such that $\int f_1(x)dx = 1$, *i.e.* $f_1(x)$ is a *pdf*.
2. Sample f_1 for a x_i.
3. For each such sample x_i evaluate $f_2(x_i)$.
4. Carry out the steps (2) and (3) for a large number, say n, of times. The average, $\frac{1}{n}\sum_{i=1}^{n} f_2(x_i)$, is the estimate of the integral.

Using Monte Carlo quadrature techniques, estimates for the integral part of the radiance equation can be arrived at by simply averaging the radiance

from a number of sampled directions from the incoming hemisphere. However, there are problems in this approach similar to the ones observed in ray tracing. Radiance at the visible point along the sampled direction is also not known and can only be obtained by a similar integration of the hemispherical contributions at that point. Along any sample direction, contribution due to emission is only known. The light sources are often localised and contribution from any point on the source is almost always significantly more as compared to the contribution from a reflector. Cook *et al*[2, 15] have presented a Monte Carlo solution to estimate the integral term in two stages. In the first stage, the source term, is estimated by sampling the light source surfaces and in the second stage, the inter-reflection term, is estimated by sampling the surface *brdf*. This technique is widely known today as distribution ray tracing. Thus the *distribution* ray tracing method is a modified recursive ray tracing method, where:
(i) For gathering the contribution from each light source illumination rays are not traced towards a single light direction, but are distributed according to the emission distribution function of the light source.
(ii) Again for the inter-reflection component, reflected rays are not traced in a single mirror direction but are distributed in the incoming hemisphere according to the bidirectional reflectance distribution function of the surface point.

Distribution ray tracing results in a very accurate solution to the radiance equation *albeit*, at a very high cost due to the excessively large number of rays that need to be traced. A number of attempts have been made to increase the efficiency of distribution ray tracing. One such attempt is the caching mechanism proposed by Ward *et al*[16] for diffuse environments. In a diffuse environment the *brdf* over a point is uniform in all the hemispherical directions. So the inter-reflection integration component of the Eq.9 can be written as a product of the surface reflectance and the incident hemispherical irradiance given below:

$$\int_{\Omega_x} f_r(x,\Theta_x,\Theta_y) L_r(y,\Theta_y)\cos\theta_x d\omega_x = f_d(x) \times \int_{\Omega_x} L_r(y,\Theta_y)\cos\theta_x d\omega_x = f_d(x) \times irradiance$$

Computation of the inter-reflection component at any point requires the evaluation of the incident hemispherical irradiance by distribution ray tracing. A cache is used to store previously computed irradiance values at various points in the environment. When calculating radiance at any point the stored irradiance of the nearby cached point(s) is used. Wherever possible, pixel to pixel illumination coherence is used to estimate the irradiance at any point from the stored irradiance.

6.2 Random Walk Solution of Radiance Equation

The radiance equation is an integral equation of the second kind. Estimates of the solutions of such equations can be carried out by *random walk*. A random walk is basically a sequence of steps. Each step is a random sample of the *pdf* defined over its previous step. Given an integral equation of the second kind, say $\phi(s) = f(s) + \int K(s,t)\phi(t)dt$, if $\int K(s,t)dt = 1$, then $K(s,t)$ at s can be used as a *pdf* . Given a starting point s, the random walk can proceed by sampling the *pdf* to arrive at a random t, and at the point t sample its associated *pdf*, $K(t,u)$ to arrive at a random point u and so on. Based on this we can provide an estimate for $\phi(s)$ as follows:

$$\begin{aligned}
\phi(s) &= f(s) + \int K(s,t)\phi(t)dt = f(s) + f(t) + \int K(t,u)\phi(u)du \\
&= f(s) + f(t) + f(u) + \ldots
\end{aligned}$$

If $\int K(s,t)dt < 1$ then also one can use the same method by introducing an additional event of absorption into the *pdf*. That means at every step either a next step is chosen according to the probability $K(s,t)$ or absorption is chosen according to the probability $1 - \int K(s,t)dt$. The random walk is bound to terminate and hence the sequence of steps is always finite. So the sum of these finite number of f terms will provide an estimate of $\phi(s)$. Applying this technique to the solution of radiance equation where $\int_\Omega f_r(x, \Theta_x, \Theta_y) \cos\theta_x d\omega_x < 1$ would mean estimating the radiance as a sum of finite number of emission radiance values as follows:

$$< L_o(x,\Theta_x) >= L_e(x,\Theta_x) + L_e(x',\Theta_{x'}) + L_e(x'',\Theta_{x''}) + \ldots \tag{12}$$

where

> $\Theta_{x'}$ is the direction chosen from the incoming hemisphere by sampling the *brdf* at x and x' is the surface position visible to x along $\Theta_{x'}$,
> $\Theta_{x''}$ is chosen by sampling *pdf* at x' and x'' is the surface position visible to x' along $\Theta_{x''}$, and so on.

The average of such estimates computed over a large number of paths provides a more accurate estimate of the actual solution. Kajiya proposed such a solution method for the *rendering equation* and the method is widely known as *path tracing*[3].

Path tracing differs from distribution ray tracing in that in path tracing a single ray emerges from each point where as in distribution ray tracing a large number of rays emerge from a single point. However, the requirement of tracing a large number of paths to get a reasonable estimate of the solution makes it as expensive as distribution ray tracing. Though path tracing in general cannot be considered as being very efficient, Kajiya[3] suggests careful use of various variance reduction techniques such as hierarchical and nonuniform sampling to make path tracing as an efficient and acceptable alternative for accurate illumination computations.

7 Deterministic Shooting Methods

As we said earlier the shooting process simulates the physical process of light distribution. As the sources are the ones responsible for illumination in the environment, intuitively a method based on the shooting strategy would be natural for computing illumination of the whole environment. However, the very first use of such a method in illumination computation, widely known as progressive radiosity[5], was proposed only in 1988. The method came as an extension to the standard radiosity method for increasing efficiency and is based on the observation that:

Radiosity computation of any patch requires the gathering of radiosity from every other patch in the environment. However, only a few of these gathered values are significant enough to contribute towards the brightness of the patch of interest. The significant contributions are mainly due to light sources and bright reflector surfaces (often highly reflecting surfaces which receive light directly from the light sources). So a method which can consider only those significant patches and ignore the rest is likely to be more efficient.

This makes finding the set of major contributors important. Light sources are undoubtedly included in this set. The other contributors are the ones receiving maximum emitted light directly or indirectly and hence can be found by shooting the light from the source(s) and keeping track of the quantity of light reaching every other surface patch of the environment.

The distinguishing feature of the progressive radiosity method is that while one surface is shooting light the outgoing flux of all other surfaces are simultaneously updated. The surfaces are processed in sorted order according to their flux contribution to the environment. The sorted list of surfaces initially contains only the emitters. As the shooting progresses, the receivers with acquired flux are added into the list. From the radiosity equation, Eq.11, the amount of light received by the i-th patch after a single shooting operation from a bright patch, say j, is given by

$$\Delta B(i) = \rho_d(i) F(ij) B(j)$$

where F is the geometric factor. Thus the total radiosity of the i-th surface patch is computed as follows:

$$B(i) = E(i) + \sum_{j=1}^{ns} E(j) \left[\rho_d(i) F(ji) + \sum_{k=1}^{N} \rho_d(k) F(jk) \rho_d(i) F(ki) + \ldots \right] \quad (13)$$

To derive the analytical approximation of Φ using the potential function we introduce a *hemispherical potential function* over any point of patch i, $\mathcal{W}(i)$, as the average potential of the surface points 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 will be:

$$\begin{aligned}
\mathcal{W}(i) &= \frac{1}{\pi} \int_{\Omega_{x_i}} W(x_i, \Theta_{x_i}) \cos\theta_{x_i} d\omega_{x_i} \\
&= \frac{1}{\pi} \int_{\Omega_{x_i}} \left[g(x_i, \Theta_{x_i}) + \int_{\Omega_y} f_r(y, \Theta_y, \Theta_{x_i}) W(y, \Theta_y) \cos\theta_y d\omega_y \right] \cos\theta_{x_i} d\omega_{x_i} \\
&= \frac{1}{\pi} \int_{\Omega_{x_i}} [g(i) + \pi f_r(j) \mathcal{W}(j)] \cos\theta_{x_i} d\omega_{x_i} \\
&= \frac{1}{\pi} g(i) \int_{\Omega_{x_i}} \cos\theta_{x_i} d\omega_{x_i} + \int_{\Omega_{x_i}} f_r(j) \mathcal{W}(j) \cos\theta_{x_i} d\omega_{x_i} \\
&= g(i) + \sum_{j=1}^{N} f_r(j) \mathcal{W}(j) \int_{\omega_{ij}} \cos\theta_{x_i} d\omega_{x_i} \\
&= g(i) + \sum_{j=1}^{N} f_r(j) \mathcal{W}(j) \pi F(ij) = g(i) + \sum_{j=1}^{N} \rho_d(j) F(ij) \left[g(j) + \sum_{l=1}^{N} \rho_d(l) \mathcal{W}_l F(jl) \right]
\end{aligned}$$

$$= \quad g(i) + \sum_{j=1}^{N} \rho_d(j)F(ij)g_j + \sum_{j=1}^{N} \rho_d(j)F(ij) \sum_{l=1}^{N} \rho_d(l)\mathcal{W}(l)F(jl)$$

If we assume that the set $\mathcal{S}$ represents the points of the k-th surface patch and all the directions over those points then $g(i)$ is 1 only for $i = k$ and zero otherwise. So the expression for $\mathcal{W}(i)$ further simplifies to

$$\mathcal{W}(i) = g(i) + \rho_d(k)F(ik) + \sum_{j=1}^{N} \rho_d(j)F(ij)\rho_d(k)F(jk) + \ldots$$

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

$$\begin{aligned}
\Phi &= \int_A \int_{\Omega_x} W(x,\Theta_x)L_e(x,\Theta_x)\cos\theta_x d\omega_x dx = \sum_{i=1}^{ns} \int_{A_i} \int_{\Omega_x} W(x,\Theta_x)L_e(x,\Theta_x)\cos\theta_x d\omega_x dx \\
&= \pi \sum_{i=1}^{ns} L_e(i)\mathcal{W}(i) \int_{A_i} dx = \pi \sum_{i=1}^{ns} L_e(i)A(i)\mathcal{W}(i) \\
&= \pi \sum_{i=1}^{ns} L_e(i)A(i) \left[g(i) + \rho_d(k)F(ik) + \sum_{j=1}^{N} \rho_d(j)F(ij)\rho_d(k)F(jk) + \ldots \right] \qquad (14)
\end{aligned}$$

And this is same as Eq.13.

Extensions to Nondiffuse Environment The shooting strategy can also be used for two pass methods and the directional discretisation method discussed in the earlier section. An important extension[17] that has been proposed for dealing with non-diffuse surface reflectance behaviour is the use of spherical harmonics for the representation of the directional variation in the outgoing radiance of a point on a non-diffuse surface. The extension is important because it maintains a continuous representation of the radiance distribution around a point.

8 Nondeterministic Shooting Methods

Particle tracing[6], the Monte Carlo simulation of the particle model of light, is a nondeterministic shooting method. In this method, particles carrying a finite amount of energy are emitted in random directions from random positions on the surfaces of the light sources. The emitted particles move in a straight path and hit other objects, hence forward termed receivers. At the surface of the receiver a particle is randomly absorbed or reflected. If reflected, the particle is assigned a random direction. The particle continues its flight until it is eventually absorbed. All the random choices are made by sampling the *pdf*s associated with the behaviour which is being simulated. Shooting a reasonable number of particles will result in a simulated particle flux on all the surfaces in the environment. This simulated flux is an estimate of the actual flux in the real environment. This solution method may be seen as the random walk solution of the potential equation in which random walk starts in a state $(x_{i_0}, \Theta_{x_{i_0}})$ drawn from the emission function, $L_e(x, \Theta_x)$. A random walk may terminate at the state $(x_{i_k}, \Theta_{x_{i_k}})$ with probability

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

or 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 is evaluated by carrying out the random walk. The potential estimated from this walk is given by

$$< W(x_i, \Theta_i) >= g(x_{i_0}, \Theta_{x_{i_0}}) + \sum_{k=1}^{m_i} g(x_{i_k}, \Theta_{x_{i_k}})$$

Φ is estimated from n such walks as:

$$\begin{aligned} \Phi &= \int_A \int_{\Omega_x} W(x, \Theta_x) L_e(x, \Theta_x) \cos\theta_x d\omega_x dx = \mathcal{E} \times \int_A \int_{\Omega_x} W(x, \Theta_x) \mathcal{L}(x, \Theta_x) \cos\theta_x d\omega_x dx \\ &= \mathcal{E} \times \frac{1}{n} \sum_{i=1}^{n} \left[g(x_{i_0}, \Theta_{x_{i_0}}) + \sum_{k=1}^{m_i} g(x_{i_k}, \Theta_{x_{i_k}}) \right] \end{aligned} \qquad (15)$$

where $\mathcal{E}$ is the normalising factor which makes $(1/\mathcal{E}) \int_A \int_{\Omega_x} L_e(x, \Theta_x) cos\theta_x d\omega_x dx = 1$. Thus the method of evaluating Φ by first shooting from the light source is the essence of the *particle tracing* or light ray tracing.

We have come across two random walk based illumination computation methods, *viz.* the path tracing and the particle tracing. An important observation is that both the random walk processes attempt to solve the same problem and are subject to similar statistical errors which in MonteCarlo studies is known as *variance*. But one thing which makes the *particle tracing* more attractive is that the simulation proceeds by sampling the source function. If we partition the position and directions in the environments into a finite number of sets $S_1, S_2, S_3, \ldots$, and formulate equal number of g functions $g_1, g_2, g_3, \ldots$, such that g_i is nonzero in the respective set S_i and zero otherwise, then each random walk originating from the emitting surfaces contributes towards the estimation of the Φ_i for each of the S_i. At the end of the simulation we have the estimates for Φ_i for all the sets. 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 towards 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 and to compute the illumination of all the subregions visible through a pixel by particle tracing are of equal magnitude. 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. All the same, for view independent illumination computations particle tracing does have its distinctive advantages.

9 Conclusion

Illumination computation has been one of the most extensively researched subjects in the field of computer graphics. The total amount of published work is enormously large and it would not have been feasible or beneficial to attempt to accommodate all of these in our review. However the treatment of these methods as being algorithmic solutions to the radiance equation, and the categorisation of strategies into gathering vs shooting in one dimension and deterministic vs nondeterministic along another is the first of its kind and has ensured that the more significant contributions have all been adequately covered.

The gathering strategy based methods can all be treated as methods providing solutions to the basic radiance equation and shooting strategy based methods as methods providing solutions to the basic potential equation. While the deterministic methods are efficient but restrictive solutions for these equations, the nondeterministic methods are general solutions and hence in principle can deal with all kinds of general environments in the same manner.

10 Acknowledgement

I am grateful to Dr S P Mudur, my thesis guide and Head Graphics & CAD Division, NCST, for his support, guidance and encouragement.

References

[1] Turner Whitted. An improved illumination model for shaded display. *Communications of the ACM*, 23(6):343–349, 1980.

[2] R. L. Cook, Thomas Porter, and Loren Carpenter. Distributed ray tracing. *Computer Graphics (SIGGRAPH '84 Proceedings)*, 18(3):137–145, July 1984.

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

[4] Cindy M. Goral, Kenneth E. Torrance, Donald P. Greenberg, and Bennett Battaile. Modelling the interaction of light between diffuse surfaces. *Computer Graphics (SIGGRAPH '84 Proceedings)*, 18(3):212–222, July 1984.

[5] 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.

[6] S. N. Pattanaik and S. P. Mudur. Computation of global illumination by monte carlo simulation of the particle model of light. In *Proceedings of the Third Eurographics Workshop on Rendering*, pages 71–83, Bristol, UK, May 1992.

[7] W.J. Bouknight. A procedure for generation of 3d half-toned computer graphics presentations. *Communications of ACM*, 13(9):527–536, 1970.

[8] Bui Tuo Phong. Illumination for computer generated pictures. *Communications of the ACM*, 8(6):311–317, 1975.

[9] Robert L. Cook and Kenneth E. Torrance. A reflectance model for computer graphics. *ACM Transactions on Graphics*, 1(1):7–24, 1982.

[10] Andrew Glassner. *An Introduction to Ray Tracing.* Academic Press, 1989.

[11] Shenchang Eric Chen, Holly E. Rushmeier, Gavin Miller, and Douglass Turner. A progressive multi-pass method for global illumination. *Computer Graphics (SIGGRAPH '91 Proceedings)*, 25(4):164–174, July 1991.

[12] Francois Sillion and Claude Puech. A general two-pass method integrating specular and diffuse reflection. *Computer Graphics (SIGGRAPH '89 Proceedings)*, 23(3):335–344, July 1989.

[13] John R. Wallace, Michael F. Cohen, and Donald P. Greenberg. A two-pass solution to the rendering equation: A synthesis of ray tracing and radiosity methods. *Computer Graphics (SIGGRAPH '87 Proceedings)*, 21(4):311–320, July 1987.

[14] Dave S. Immel, Michael Cohen, and Donald P. Greenberg. A radiosity method for non-diffuse environments. *Computer Graphics (SIGGRAPH '86 Proceedings)*, 20(4):133–142, Aug. 1986.

[15] R. L. Cook. Stochastic sampling in computer graphics. *ACM Transactions on Graphics*, 5(1):51–72, Jan. 1986.

[16] Gregory J. Ward, Francis M. Rubinstein, and Robert D. Clear. A ray tracing solution for diffuse interreflection. *Computer Graphics (SIGGRAPH '88 Proceedings)*, 22(4):85–92, Aug. 1988.

[17] Francois Sillion, James R. Arvo, Stephen H. Westin, and Donald P. Greenberg. A global illumination solution for general reflectance distributions. *Computer Graphics (SIGGRAPH '91 Proceedings)*, 25(4):187–196, July 1991.