Monte-Carlo Global Illumination Methods State of the Art and New Developments *

László Szirmay-Kalos
Department of Control Engineering and Information Technology, Technical University of Budapest
Budapest, Műegyetem rkp. 11, H-1111, HUNGARY
szirmay@fsz.bme.hu

Abstract

This paper presents the state of the art and recent developments of Monte-Carlo global illumination algorithms. First it surveys the basic tasks of global illumination, which can be formulated as the solution of either the rendering or the potential equation, then reviews the basic solution techniques, including inversion, expansion and iteration. The paper explains why stochastic approaches are good to solve these integral equations and highlights what kind of fundamental choices we have when designing such an algorithm. It compares, for example, finite-element and continuous methods, pure Monte-Carlo and quasi-Monte Carlo techniques, different versions of importance sampling, Russian roulette, etc. Then, a lot of methods are reviewed in a unified framework, that also allows to make comparisons.

Keywords: Rendering and potential equations, Monte-Carlo and quasi-Monte Carlo quadratures, finite-element techniques, importance sampling, Russian roulette, shooting and gathering random walks, stochastic iteration, Metropolis sampling, distributed ray-tracing, path tracing, photon tracing, light tracing, bi-directional path tracing, photon-map, instant radiosity, global ray-bundle tracing, stochastic ray-radiosity, transillumination method, first-shot.

1 Introduction

Generally, the *global illumination problem* is a quadruple[23]:

$$\langle S, f_r(\omega', \vec{x}, \omega), L^e(\vec{x}, \omega), \mathbf{W}^e(\vec{x}, \omega) \rangle$$

where S is the geometry of surfaces, f_r is the BRDF of surface points, L^e is the emitted radiance of surface points at different directions and $\mathbf{W}^e$ is a collection of measuring functions.

Global illumination algorithms aim at the modeling and simulation of multiple light-surface interactions to find out the power emitted by the surfaces and landing at the measuring devices after some reflections. A light-surface interaction can be formulated by the *rendering equation* or alternatively by its adjoint equation, called the *potential equation*.

The *rendering equation*[21] expresses the *radiance* $L(\vec{x}, \omega)$ $[W \cdot m^{-2} \cdot sr^{-1}]$ of a surface point $\vec{x}$ in direction ω, and has the following form:

$$L = L^e + \mathcal{T}L. \tag{1}$$

If only direct contribution is considered, then $L = L^e$. The light-surface interaction is described by integral operator $\mathcal{T}$, which has the following form

$$(\mathcal{T}L)(\vec{x}, \omega) = \int_\Omega L(h(\vec{x}, -\omega'), \omega') \cdot f_r(\omega', \vec{x}, \omega) \cdot \cos\theta' \, d\omega' \tag{2}$$

where $L(\vec{x}, \omega)$ and $L^e(\vec{x}, \omega)$ are the radiance and emission of the surface in point $\vec{x}$ at direction ω, Ω is the directional sphere, $h(\vec{x}, \omega')$ is the visibility function defining the point that is visible from point $\vec{x}$ at direction ω', $f_r(\omega', \vec{x}, \omega)$ is the bi-directional reflection/refraction function, and θ' is the angle between the surface normal and the incoming direction $-\omega'$ (figure 1).

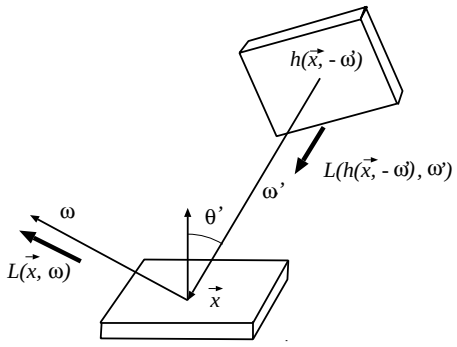

Figure 1: Geometry of the rendering equation

The *potential equation*[36], on the other hand, uses the *potential* $W(\vec{y}, \omega')$ as a fundamental measure, which expresses the effect of emitting unit power from $\vec{y}$ in direction ω' on a measuring device having sensitivity

*This work has been supported by the National Scientific Research Fund (OTKA ref.No.: F015884, T029135), the ÖAD (ref.No.: 32öu9) and the Spanish-Hungarian Fund (ref.No.: E9).

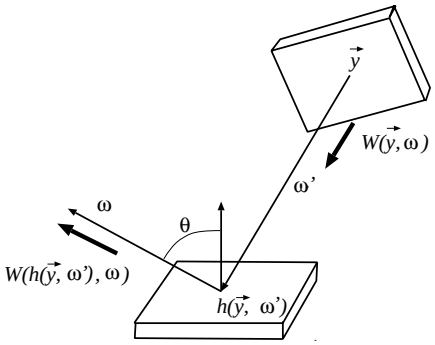

Figure 2: Geometry of the potential equation

$W^e(\vec{y},\omega')$ (for example, this device can measure the power going through a single pixel of the image, or leaving a surface element at any direction). If only direct contribution is considered, then $W(\vec{y},\omega') = W^e(\vec{y},\omega')$. To take into account light reflections, we can establish the potential equation

$$W = W^e + \mathcal{T}'W. \tag{3}$$

In this equation integral operator $\mathcal{T}'$ — which is the adjoint of $\mathcal{T}$ — describes the potential transport

$$(\mathcal{T}'W)(\vec{y},\omega') =$$

$$\int_\Omega W(h(\vec{y},\omega'),\omega)\cdot f_r(\omega',h(\vec{y},\omega'),\omega)\cdot\cos\theta\; d\omega, \tag{4}$$

where θ is the angle between the surface normal and the outgoing direction ω.

According to the definition of the radiance

$$L(\vec{y},\omega) = \frac{d\Phi(\vec{y},\omega)}{d\vec{y}\; d\omega\cos\theta},$$

the power detected by a measuring device can be computed by the measuring function from the radiance

$$\int_S\int_\Omega d\Phi(\vec{y},\omega)\cdot W^e(\vec{y},\omega) =$$

$$\int_S\int_\Omega L(\vec{y},\omega)\cos\theta\cdot W^e(\vec{y},\omega)\; d\vec{y}\; d\omega = \mathcal{M}L, \tag{5}$$

where $\mathcal{M}$ is the radiance measurement operator. Having introduced the scalar product $\langle u,v\rangle$

$$\langle u,v\rangle = \int_S\int_\Omega u(\vec{y},\omega)\cdot v(\vec{y},\omega)\; d\vec{y}\; d\omega,$$

and the cosine weighted scalar product $\langle u,v\rangle_{\cos}$

$$\langle u,v\rangle_{\cos} = \langle u\cdot\cos\theta, v\rangle = \langle u, v\cdot\cos\theta\rangle,$$

we can obtain an alternative form of the measurement operator

$$\mathcal{M}L = \langle L, W^e\rangle_{\cos}.$$

A simple measurement function for a pinhole camera is

$$W^e(\vec{y},\omega) = \frac{\delta(\omega-\omega_f)}{\cos\theta}\cdot\xi(h(\vec{y},\omega))$$

where ω_f is the focal point and $\cos\theta$ is the cosine angle between the normal of the visible surface and the viewing direction. With this measurement function, the power going through a pixel of area P can be obtained using equation (5):

$$\int_{S_P} L(h(\vec{p},-\omega_{\vec{p}}),\omega_{\vec{p}})\cdot\xi(\vec{p})\; d\vec{p}, \tag{6}$$

where S_P is the support of ξ. S_P is usually, but not necessarily, equal to the pixel surface.

Alternatively to the radiance, the power arriving at the measuring device can also be computed from the potential:

$$\int_S\int_\Omega d\Phi^e(\vec{y},\omega')\cdot W(\vec{y},\omega') =$$

$$\int_S\int_\Omega W(\vec{y},\omega')\cdot L^e(\vec{y},\omega')\cdot\cos\theta\; d\vec{y}\; d\omega' = \mathcal{M}'W, \tag{7}$$

where $\mathcal{M}'$ is the potential measuring operator. Note that unlike the radiance measuring operator, the potential measuring operator integrates on the lightsource.

This measuring operator can also be given in a scalar product form

$$\mathcal{M}'W = \langle L^e, W\rangle_{\cos}. \tag{8}$$

Since the rendering or the potential equation contain the unknown radiance function both inside and outside the integral, in order to express the solution, this coupling should be resolved. The possible solution techniques fall into one of the following three categories: *inversion*, *expansion* and *iteration*.

Operator $\mathcal{T}$ represents light-surface interaction, thus each of its application generates a higher-bounce estimate of the light transport (or alternatively $\mathcal{T}'$ represents potential-surface interaction). For physically plausible optical material models, a reflection or refraction always decreases the total energy, thus the integral operator is always a contraction. However, when the transport is evaluated numerically, computation errors may pose instability problems if the scene is highly reflective. As we shall see, expansion and iteration exploit the contractive property of the transport operator, but inversion does not.

1.1 Inversion

Inversion groups the terms that contain the unknown function on the same side of the equation and applies formally an inversion operation:

$$(1-\mathcal{T})L = L^e \implies L = (1-\mathcal{T})^{-1}L^e. \tag{9}$$

Thus the measured power is

$$\mathcal{M}L = \mathcal{M}(1-\mathcal{T})^{-1}L^e. \tag{10}$$

However, since $\mathcal{T}$ is infinite dimensional, it cannot be inverted in closed form. Thus it should be approximated by a finite dimensional mapping, that is usually given as a matrix. This kind of approximation is provided by finite-element techniques that project the problem into a finite dimensional function space, and approximate the solution here. This projection converts the original integral equation into a system of linear equations, which can be inverted, for example, by Gaussian elimination method. This approach was used in early radiosity methods, but have been ruled out due to the cubic time complexity and the numerical instability of the Gaussian elimination.

Since no stochastic alternative has been proposed yet for the deterministic inversion, we do not consider this option any further in this paper.

1.2 Expansion

Expansion techniques eliminate the coupling by obtaining the solution in the form of an infinite Neumann series.

1.2.1 Expansion of the rendering equation: gathering walks

Substituting the right side's L by $L^e + \mathcal{T}L$, which is obviously L according to the equation, we get:

$$L = L^e + \mathcal{T}L = L^e + \mathcal{T}(L^e + \mathcal{T}L) = L^e + \mathcal{T}L^e + \mathcal{T}^2L. \tag{11}$$

Repeating this step n times, the original equation can be expanded into a Neumann series:

$$L = \sum_{i=0}^{n} \mathcal{T}^i L^e + \mathcal{T}^{n+1}L. \tag{12}$$

If integral operator $\mathcal{T}$ is a contraction, then $\lim_{n\to\infty} \mathcal{T}^{n+1}L = 0$, thus

$$L = \sum_{i=0}^{\infty} \mathcal{T}^i L^e. \tag{13}$$

The measured power is

$$\mathcal{M}L = \sum_{i=0}^{\infty} \mathcal{M}\mathcal{T}^i L^e. \tag{14}$$

The terms of this infinite Neumann series have intuitive meaning as well: $\mathcal{M}\mathcal{T}^0L^e = L^e$ comes from the emission, $\mathcal{M}\mathcal{T}^1L^e$ comes from a single reflection, $\mathcal{M}\mathcal{T}^2L^e$ from two reflections, etc.

$\mathcal{M}\mathcal{T}^dL^e$ is a $2d+2$-dimensional integral where the integrand is a product of the radiance of a lightsource point and the probability that a patch of length d connects the eye to this lightsource point.

To obtain the integrand for a single point, a ray is emanated recursively from the visible point at direction ω'_1 then from the found surface at ω'_2, etc. until ω'_n. The emission intensity at the end of the walk is read and multiplied by the BRDFs and the cosine terms of the stages of the walk. These walks provide the value of the integrand at "point" $\vec{p}, \omega'_1, \omega'_2, \ldots, \omega'_n$.

Note that a single walk of length n can be used to estimate the 1-bounce, 2-bounce, etc. n-bounce transfer simultaneously, if the emission is transferred not only from the last visited point but from all visited points.

The presented walking technique starts at the eye and *gathers* the illumination encountered during the walk. The gathered illumination is attenuated according to the cosine weighted BRDFs of the path.

1.2.2 Expansion of the potential equation: shooting walks

The potential equation can also be expanded into a Neumann series similarly to the rendering equation.

$$W = \sum_{i=0}^{\infty} \mathcal{T}'^i W^e, \tag{15}$$

which results in the following measured power:

$$\mathcal{M}'W = \sum_{i=0}^{\infty} \mathcal{M}'\mathcal{T}'^i W^e. \tag{16}$$

$\mathcal{M}'W^e$ is the power measured by the device from direct emission. $\mathcal{M}'\mathcal{T}'W^e$ is the power after a single reflection, $\mathcal{M}'\mathcal{T}'^2W^e$ is after two reflections, etc.

This type of walk, called *shooting*, starts at a known point $\vec{y}_1$ of a lightsource and simulates the photon reflection for a few times and finally arrives at a pixel whose radiance this walk contributes to.

Note that in gathering walks the BRDF is multiplied with the cosine of the angle between the normal and the incoming direction, while in shooting walks with the cosine of the angle between the normal and the outgoing direction. On the other hand, in gathering walks, the cosine angle of the emitting surface is not used, while in shooting walks the cosine angle of the last visible surface is neglected.

1.2.3 Merits and disadvantages of expansion methods

The main problem of expansion techniques is that they require the evaluation of very high dimensional integrals that appear as terms in the infinite series. Practical implementations usually truncate the infinite Neumann series, which introduces some bias, or stop the walks randomly, which significantly reduces the samples of higher order interreflections. These can result in visible artifacts for highly reflective scenes.

On the other hand, expansion methods also have an important advantage. Namely, they do not require temporary representations of the complete radiance function, thus do not necessitate finite-element approximations. Consequently, these algorithms can work with the original geometry without tessellating the surfaces to planar polygons.

Expansion techniques generate random walks independently. It can be an advantage, since these algorithms can be suitable for parallel computing. However, it also means that these methods "forget" the previous history of walks, and they cannot reuse the visibility information gathered when computing the previous walks, thus they are not as fast as they could be.

1.3 Iteration

Iteration techniques realize that the solution of integral equation (1) is the fixed point of the following iteration scheme

$$L_n = L^e + \mathcal{T}L_{n-1}, \tag{17}$$

thus if operator $\mathcal{T}$ is a contraction, then this scheme will converge to the solution from any initial function L_0.

The measured power can be obtained as a limiting value

$$\mathcal{M}L = \lim_{n\to\infty} \mathcal{M}L_n, \tag{18}$$

In order to store the approximating functions L_n, usually finite-element techniques are applied, as for example, in *diffuse radiosity*[47], or in non-diffuse radiosity using *partitioned hemisphere*[16], *directional distributions*[49] or *illumination networks*[5].

There are two critical problems here. On the one hand, since the domain of L_n 4 dimensional, an accurate finite-element approximation usually requires very many basis functions, which, in turn, need a lot of storage space. Although, *hierarchical methods*[14, 3], *wavelet* or *multiresolution methods*[8, 41] and *clustering*[48, 7, 51] can help, the memory requirements are still prohibitive for complex scenes. This problem is less painful for the diffuse case since here the domain is only 2 dimensional.

On the other hand, when finite element techniques are applied, operator $\mathcal{T}$ is only approximated, which introduces some non-negligible error in each step. If the contraction ratio of the operator is λ, then the total accumulated error will be approximately $1/(1-\lambda)$ times the error of a single step[59]. For highly reflective scenes, the iteration is slow and the result is inaccurate if the approximation of the operator is not very precise. Very accurate approximations of the transport operator, however, require a lot of computation time and storage space.

Both the problem of prohibitive memory requirements and the problem of error accumulation can be successfully attacked by *stochastic iteration*.

Compared to expansion techniques, iteration has both advantages and disadvantages. Its important advantage is that it can potentially reuse all the information gained in previous computation steps, thus iteration is expected to be faster than expansion. Iteration can also be seen as a single infinite length random walk. If implemented carefully, iteration does not reduce the number of estimates for higher order interreflections, thus it is more robust when rendering highly reflective scenes than expansion.

2 Why should we use stochastic methods?

Expansion techniques require the evaluation of very high-dimensional — in fact, infinite dimensional — integrals. When using classical quadrature rules for multi-dimensional integrals [38], such as for example the trapezoidal rule, in order to provide a result with a given accuracy, the number of sample points is in the order of $O(M^D)$, where D is the dimension of the domain. This phenomenon is called the *dimensional core* or *dimensional explosion* and makes classical quadrature rules prohibitively expensive for higher dimensions. The reason of the dimensional explosion is that these rules are usually based on uniform grids — that are simple Cartesian products of the 1D grid in higher dimensions — in which different dimensions do not effectively interact.

However, Monte-Carlo or quasi-Monte Carlo techniques distribute the sample points simultaneously in all dimensions, thus they can avoid dimensional explosion. For example, the probabilistic error bound of Monte-Carlo integration is $O(M^{-0.5})$, independently of the dimension of the domain. D-dimensional low discrepancy series[35] can even achieve $O(\log^D M/M) = O(M^{-(1-\epsilon)})$ convergence rates for finite variation integrands.

Furthermore, classical quadrature cannot be used for infinite dimensional integrals, thus the Neumann series should be truncated after D terms. This truncation introduces a bias of order $\lambda^{D+1}\cdot||L_e||/(1-\lambda)$. Using a Russian roulette based technique, on the other hand, Monte-Carlo methods are appropriate for even infinite dimensional integrals.

Thus we can conclude that the stochastic approach is indispensable for expansion methods.

The application of randomized techniques in iteration is not so evident, but can also be justified. On the simplest level, these methods also use integration in each iteration step. The dimension of the domain is usually not very high. For example, iterative diffuse radiosity methods need to evaluate 4-dimensional integrals to obtain form factors. The dimension is often reduced to 2 by a brutal simplification, which computes one of the two surface integrals from a single value. For even 4-dimensional integrals Monte-Carlo methods are superior than classical quadratures thus in accurate algorithms they are highly recommended.

Furthermore, when stochastic iteration is applied, the operator should be like the real operator just in the average case. This allows us to use significantly simpler re-

alizations. For example, the integral part of the operator can also be approximated as an expectation value, thus in a single transfer usually no explicit integral is computed. As we shall see, it is relatively easy to apply random operators whose expected case behavior gives exactly back that of the real operator. Thus the error accumulation problem can also be avoided.

If the operator is highly simplified, it does not require the integrand everywhere in the domain, thus a lot of storage space can be saved. Compared to the astronomical storage requirements of non-diffuse radiosity methods, for example, with stochastic iteration we can achieve the same goal with one variable per patch[62]. This argument loses some of its importance when view-independent solution is also required, since the final solution should be stored anyway. This is not a problem if only the diffuse case is considered, since using a single radiosity value per patch the image can be generated from any viewpoint. For the non-diffuse case, the reduced storage gets particularly useful when the image is to be calculated in only a single, or in a few eye positions.

Summarizing, the advantages of stochastic iteration are the simplicity speed, affordable storage requirements and numerical stability even for very large systems containing highly reflective materials.

3 Options in stochastic rendering

3.1 Monte-Carlo versus quasi-Monte Carlo

The core of the computations of all methods is the evaluation of high-dimensional integrals (for inversion and iteration it means 4 dimensional integrals, for expansion, it means, at least theoretically, infinite-dimensional integrals). To evaluate an integral, we can use quadrature formulae, that have the following form in the simplest case:

$$\int\limits_{[0,1]^D} f(\mathbf{z})\, d\mathbf{z} \approx \frac{1}{M} \cdot \sum_{i=1}^{M} f(\mathbf{z}_i). \tag{19}$$

Those sets of sample points that provide an exact integral value in the asymptotic sense are called *uniform sequences*.

Well known examples for uniform sequences are the uniform grid, the uniformly distributed random samples and the family of low-discrepancy sequences.

To find out which are those sample sets that can effectively be used in numerical integration, the *Koksma-Hlawka inequality*[35] gives us some hints (unfortunately, it is valid only for finite-variation functions, but the basic observations are still useful in more general circumstances):

$$\left| \int\limits_{\mathbf{z}\in[0,1]^D} f(\mathbf{z})\, d\mathbf{z} - \frac{1}{M} \sum_{i=1}^{M} f(\mathbf{z}_i) \right| \leq \mathcal{V}_{\mathrm{HK}} \cdot \mathcal{D}^*(\mathbf{z}_1, \ldots \mathbf{z}_N), \tag{20}$$

where $\mathcal{V}_{\mathrm{HK}}$ is the *variation* of f in the sense of Hardy and Krause, and $\mathcal{D}^*(\mathbf{z}_1, \ldots \mathbf{z}_N)$ is the *star-discrepancy* of the used sample set (for the bounds and computation of the discrepancy refer to [35, 43, 11]).

According to this inequality, the error can be upper-bounded by the product of two independent factors, the variation of the integrand and the discrepancy of the used samples set. The discrepancy shows how uniformly the set is distributed[43]. This immediately presents two orthogonal strategies to improve the quality of quadratures. Either we try to make the function flat by appropriate variable transformations, or use very uniformly distributed sample sets. The first technique is called *importance sampling*[50], while the second involves the *stratification*[50, 30, 1] of random points or the application of *low-discrepancy series*[35, 71, 38, 24, 50].

Low-discrepancy samples are deterministic point sets that are designed to be optimally uniform, thus replacing the random points by them improves the accuracy of the integral quadrature. Quadrature rules that use low-discrepancy series instead of random points are called *quasi-Monte Carlo methods*.

Quasi-Monte Carlo techniques have been first applied to solve the diffuse rendering equation by Keller[22], where the integrand was generally discontinuous and therefore of infinite variation, thus the superiority of quasi-Monte Carlo method could not been theoretically justified (note that the Koksma-Hlawka inequality is meaningless if the variation is infinite). However, the numerical evidence showed that quasi-Monte Carlo methods can slightly be better than Monte-Carlo techniques.

The efficiency of the quasi-Monte Carlo integration for the rendering equation has been theoretically analyzed in [63] and an empirical study was presented in [4]. These studies concluded that quasi-Monte Carlo methods are still better but lose their advantage in higher dimensions. The difference is significant when the integrand is relatively smooth, which is the case in finite-element approaches. In [4] the different low-discrepancy series have also been compared, and it was conlcuded that all have them provide similar performance.

The other important problem is that although a low-discrepancy series has almost linearly decreasing discrepancy in the asymptotic sense, this discrepancy can still be high for not very many points (in the solution of the rendering equation we rarely use more than 1000 samples for the estimation of a single pixel). In the case of the Halton series, for example, the *base* of the series strongly affects the initial behavior of the discrepancy. These base numbers are different prime numbers for different dimensions, thus for high-dimensional integrals the base numbers can

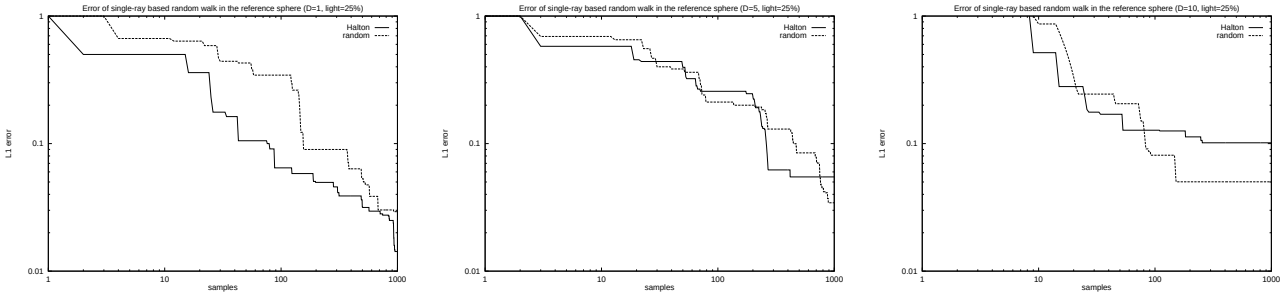

Figure 3: Error measurements for 1, 5 and 10 bounces

be quite high, which results in degraded performance.

To demonstrate this, in figure 3 the errors of different bounces generated by quasi-Monte Carlo and the Monte-Carlo quadratures have been compared for a spherical diffuse scene where only a part is lightsource. For this scene the analytical solution of the rendering equation is possible[15, 63].

3.2 Continuous versus finite-element based methods

Iteration requires the representation of the temporary radiance function L_n. So does expansion if view-independent solution is needed since the final radiance distribution must be represented in a continuous domain.

To represent a function over a continuous domain, finite element methods can be used which approximate the function in the following form:

$$L(\vec{x},\omega) \approx \sum_{j=1}^{n} \mathbf{L}_j \cdot b_j(\vec{x},\omega) = \mathbf{b}^T(\vec{x},\omega) \cdot \mathbf{L} \qquad (21)$$

where $b_j(\vec{x},\omega)$ is a system of predefined basis functions, and $\mathbf{L}_j$ factors are unknown coefficients.

This representation can also be seen as projecting the infinite dimensional space of the possible radiance functions into a finite-dimensional function space defined by the basis functions.

Substituting this approximation into the rendering equation we can obtain:

$$\mathbf{b}^T \cdot \mathbf{L} \approx \mathbf{b}^T \cdot \mathbf{L}^e + \mathcal{T}(\mathbf{b}^T \cdot \mathbf{L}). \qquad (22)$$

Note that equality cannot be guaranteed, since even if $\mathbf{b}^T(\vec{x},\omega) \cdot \mathbf{L}$ is in the subspace defined by the basis functions, the integral operator $\mathcal{T}$ may result in a function that is out of this space. This can be solved by projecting the result back to the subspace and using a projected integral operator $\mathcal{T}_F$ in the following way:

$$\mathcal{T}_F\mathbf{L} = \langle \mathcal{T}\mathbf{b}^T \cdot \mathbf{L}, \tilde{\mathbf{b}} \rangle. \qquad (23)$$

where $\langle \mathcal{T}L, \tilde{\mathbf{b}} \rangle$ is a vector of scalar products

$$\langle \mathcal{T}L, \tilde{b}_1 \rangle, \ldots \langle \mathcal{T}L, \tilde{b}_n \rangle$$

and $\tilde{b}_i$ is an adjoint basis of b_i, since we require that $\langle \tilde{b}_i, b_j \rangle = 1$ if $i = j$ and 0 otherwise.

Since $\mathbf{L}$ is constant, we can also obtain

$$\mathcal{T}_F\mathbf{L} = \langle \mathcal{T}\mathbf{b}^T, \tilde{\mathbf{b}} \rangle \cdot \mathbf{L} = \mathbf{F} \cdot \mathbf{L}, \qquad (24)$$

where $\mathbf{F} = \langle \mathcal{T}\mathbf{b}^T, \tilde{\mathbf{b}} \rangle$ is a matrix, where the i,j element is $\langle \mathcal{T}b_j, \tilde{b}_i \rangle$. Thus the projection converts the original integral to the following form:

$$\mathbf{L} = \mathbf{L}^e + \mathcal{T}_F\mathbf{L} = \mathbf{L}^e + \mathbf{F} \cdot \mathbf{L}. \qquad (25)$$

An adjoint of this linear equation can be derived by supposing that each basis function b_i is associated with a measurement device W_i^e that measures the power $\mathbf{P}_i$ leaving the support of the basis function. Thus we obtain

$$\langle W_i^e, \mathbf{b}^T \cdot \mathbf{L} \rangle_{\cos} = \langle W_i^e, b_i \rangle_{\cos} \cdot \mathbf{L}_i = \mathbf{P}_i.$$

Similarly, the measured emission power is

$$\langle W_i^e, \mathbf{b}^T \cdot \mathbf{L}^e \rangle_{\cos} = \langle W_i^e, b_i \rangle_{\cos} \cdot \mathbf{L}_i^e = \mathbf{P}_i^e.$$

Applying measurement operator W_i^e for equation (25), we can obtain the following equation:

$$\mathbf{P} = \mathbf{P}^e + \mathbf{H} \cdot \mathbf{P}, \qquad (26)$$

where

$$\mathbf{H}_{ij} = \mathbf{F}_{ij} \cdot \frac{\langle W_i^e, b_i \rangle_{\cos}}{\langle W_j^e, b_j \rangle_{\cos}}. \qquad (27)$$

When finite-element techniques are used together with expansion, finite-element representation can either be used to represent the final result[22], or even be involved in the random walk[36].

The latter case may correspond either to the random-walk solution of the linear equation derived by projecting the integral equation, or to the Monte-Carlo evaluation of the multi-dimensional integral containing both the transport and the projection operators. The second case is preferred, because it does not require matrix $\mathbf{F}$ to be explicitly computed and stored.

The main problem of finite-element representations is that they require a lot of basis functions to accurately

approximate high-variation, high-dimensional functions. Not surprisingly, finite-element methods become really popular only for the diffuse case, where the radiance depends on 2 scalars and is relatively smooth. For solving the non-diffuse case, they are good only if the surfaces are not very specular.

The property that certain methods require tessellation and finite-element representation is usually considered as a disadvantage. And indeed, sharp shadows and highlights on highly specular materials can be incorrectly rendered and light-leaks may appear, not to mention the unnecessary increase of the complexity of the scene description (think about, for example, the definition of the original and tessellated sphere). However, finite-element representation can also provide smoothing during all stages of rendering, which results in more visually pleasing and dot-noise free images.

3.3 Global versus local methods

Randomized transport operators transfer the radiance or the potential in the scene. The source and destination of the transfer can be points in the case of continuous methods or patches in the case of finite-element methods.

If the random operator is such that it always selects a single source for shooting or single destination for gathering, then the method is called *local method*. On the other hand, if many sources and destinations are taken into consideration simultaneously in each transfer, then the method is called *global method* or *multi-path method*[39].

Since global methods handle larger transfers in a single step, they can be expected to be more efficient than local methods. On the other hand, the single source or destination points of local methods directly correspond to the single "eye" of classical visibility algorithms. Thus, to exploit the capabilities of global methods, classical visibility algorithms should also be generalized for "moving" eye positions. These algorithms are called *global visibility algorithms*[37].

4 Stochastic expansion: random walks

In computer graphics the first Monte-Carlo random walk algorithm — called *distributed ray-tracing* — was proposed by Cook et al. [9], which spawned to a set of variations, including *path tracing*[21], *light-tracing*[12], *bi-directional path tracing*[25, 67], *Monte-Carlo radiosity*[44, 31, 36], and *two-pass methods* which combine radiosity and ray-tracing [42, 72, 69].

The problem of naive generation of walks is that the probability that a shooting path finds the eye is zero for a pin-hole camera or very small if a non-zero aperture camera model is used, while the probability that a gathering random path ends in a lightsource may be very little if the lightsources are small, thus the majority of the paths do not contribute to the image at all, and their computation is simply waste of time. Note that shooting is always superior for view-independent algorithms since they do not have to face the problem of small aperture.

Thus, on the one hand, random walk must be combined with a deterministic step that forces the walk to go to the eye and to find a lightsource. On the other hand, *importance sampling*[50] should be incorporated to prefer useful paths along which significant radiance is transferred. Note that although the contribution on the image is a function of the complete path, computer graphics applications usually assign estimated importance to individual steps of this path, which might be quite inaccurate. In a single step the importance is usually selected according to the BRDF [12, 25], or according to the direction of the direct lightsources [46]. Combined methods that find the important directions using both the BRDF and the incident illumination have been proposed in [66, 17, 26, 56]. Just recently, Veach and Guibas[68] proposed the Metropolis method to be used in the solution of the rendering equation. Unlike other approaches, Metropolis sampling[29] can assign importance to a complete walk not just to the steps of this walk, and it explores important regions of the domain adaptively while running the algorithm. Thus no a-priori knowledge is required about the important rays to construct a probability density function in advance. Instead, the algorithm converges to this probability density automatically.

4.1 Handling infinite-dimensional integrals

Expansion methods require the evaluation of infinite-dimensional integrals. One way of attacking the problem is truncating the Neumann series, but this introduces some bias, which can be quite high if the scene is highly reflective.

Fortunately, there is another approach that solves the infinite-dimensional integration problem through randomization. In the context of Monte-Carlo integration, this approach is called the *Russian roulette*[2], but here a somewhat more general treatment is given that can also justify this approach for quasi-Monte Carlo quadratures.

The basic idea is very simple. Higher order terms are included in the quadrature only randomly with probability decreasing with the order of the term. In order to compensate the missing terms in the expected value, the computed terms are multiplied by an appropriate factor. If the used probability goes to zero quickly, then the possibility of requiring very high dimensional integrals is rather low, which saves computation time but increases the variance. However, the expected value will still be correct, thus the integral quadrature will provide an asymptotically unbiased estimate.

A term of the Neumann series has generally the follow-

ing form

$$I_n = \int \dots \int W(\mathbf{z}_1, \dots \mathbf{z}_n) \cdot L^e(\mathbf{z}_1, \dots \mathbf{z}_n)\, d\mathbf{z}_1 \dots \mathbf{z}_n, \tag{28}$$

where $W(\mathbf{z}_1, \dots \mathbf{z}_n) = w_0 \cdot w_1 \cdot \dots \cdot w_n$ is the product of the weights including the cosine functions of the angles and the BRDFs.

Let us randomize this integral by introducing a random variable $C(\mathbf{z}_1, \dots \mathbf{z}_n)$, called the *contribution indicator*, that is 1 if a sample $\mathbf{z}_1, \dots \mathbf{z}_n$ should be taken into account in the integral quadrature and 0 if it should not. Using this, we can define the following random variable,

$$I_n^* = \int \dots \int C \cdot \tilde{W} \cdot \tilde{L}^e\, d\mathbf{z}_1 \dots \mathbf{z}_n, \tag{29}$$

where $\tilde{W}$ and $\tilde{L}^e$ are appropriate modifications of W and L^e, which can compensate the missing terms.

The expectation value of this random variable is

$$E[I_n^*] = \int \dots \int E[C(\mathbf{z}_1, \dots \mathbf{z}_n)] \cdot \tilde{W} \cdot \tilde{L}^e\, d\mathbf{z}_1 \dots \mathbf{z}_n =$$

$$\int \dots \int p(\mathbf{z}_1, \dots \mathbf{z}_n) \cdot \tilde{W} \cdot \tilde{L}^e\, d\mathbf{z}_1 \dots \mathbf{z}_n, \tag{30}$$

where $p(\mathbf{z}_1, \dots \mathbf{z}_n)$ is the probability of using sample $\mathbf{z}_1, \dots \mathbf{z}_n$ in the integral quadrature.

Obviously, this equals to the original integral I if

$$p(\mathbf{z}_1, \dots \mathbf{z}_n) \cdot \tilde{W} \cdot \tilde{L}^e = W \cdot L^e. \tag{31}$$

There are many possible selection of the contribution indicator and the $\tilde{W}$ and $\tilde{L}^e$ functions, that can satisfy this requirement, thus there are many different unbiased estimators.

A widely used selection is letting

$$\tilde{W} = 1, \quad \tilde{L}^e = L^e \text{ and } p(\mathbf{z}_1, \dots \mathbf{z}_n) = W(\mathbf{z}_1, \dots \mathbf{z}_n).$$

which corresponds to continuing the walk after step i with probability $w(\mathbf{z}_i)$.

4.2 Importance sampling

When solving the rendering equation, usually directional integrals (or surface integrals in other formulation) should be evaluated. Thus to allow the application of random or low-discrepancy point sets, the integration domain should be transformed to the unit cube or square.

For example, when dealing with directions, we have to find a mapping $\omega = T(\mathbf{z})$ that projects the unit square to the surface of the sphere (or hemisphere) and use the following integration rule

$$\int_\Omega f(\omega)\, d\omega = \int_{[0,1]^D} f(T^{-1}(\mathbf{z})) \cdot \left| \frac{dT^{-1}(\mathbf{z})}{d\mathbf{z}} \right| d\mathbf{z}, \tag{32}$$

where

$$\left| \frac{dT^{-1}(\mathbf{z})}{d\mathbf{z}} \right| = \frac{1}{t(\mathbf{z})}$$

is the Jacobi determinant of the inverse mapping.

If the Jacobi determinant is large, then a small portion of the unit square is mapped onto a large region. Thus sample points that are uniformly distributed in the unit square will be quite rare in these regions. Alternatively, where the Jacobi determinant is small, the sample points will be dense. Considering this, the meaning of $t(T^{-1}(\mathbf{z}))$ is the *density* of sample points in the neighborhood of $\omega = T^{-1}(\mathbf{z})$. This has an illustrative content for the random case. If $\mathbf{z}$ is uniformly distributed random variable, then the probability density of $\omega = T(\mathbf{z})$ will be $t(\mathbf{z})$.

Mathematically, the solution of either the rendering or the potential equation for a given point $(\vec{x}, \omega)$ requires the evaluation of the following multi-dimensional integral

$$L(\vec{x}, \omega) = L^e + \mathcal{T}L^e + \mathcal{T}^2 L^e + \dots =$$

$$\int \dots \int L^e + \frac{w_1}{t_1} \cdot L^e + \frac{w_1}{t_1} \cdot \frac{w_2}{t_2} \cdot L^e + \dots d\mathbf{z}_1 d\mathbf{z}_2 \dots \tag{33}$$

which can be estimated using formula (19) by evaluating the integrand in sample points and averaging the results.

An important design decision of such an algorithm is the selection of mappings T_i. Using probabilistic approach, it means the determination of the probability densities of finding new directions during the walks.

Following the directions concluded from the Koksma-Hlawka inequality, the mappings should make the integrand flat — that is of low variation, or constant in the ideal case. It means that the probability of selecting a walk is proportional to its contribution.

Looking at formula (33), which is the single multi-dimensional solution of the rendering equation, this decision seems to be hard to made, since there are too many free parameters to control simultaneously. Fortunately, this solution can also be presented in the following recursive form:

$$L^e + \int \frac{w_1}{t_1} \cdot [L^e + \int \frac{w_2}{t_2} \cdot [L^e + \dots] \dots]\, d\mathbf{z}_1 d\mathbf{z}_2 \dots \tag{34}$$

If we could ensure that each of the integrands of the form

$$\int \frac{w_i}{t_i} \cdot [L^e + \int \dots]\, d\mathbf{z}_i$$

is constant (at least approximately), then the integrand of the single multi-dimensional integral will also be constant.

An optimal importance sampling strategy thus requires density t_i to be proportional to the product of the incoming illumination $L^e + \int \dots$ and the cosine weighted BRDF w_i. Unfortunately, during random walks the incoming non-direct illumination is not known (the random walk is just being done to estimate it).

Thus, we have three alternatives. Information about the illumination in the space can be gathered in a preprocessing phase, then this information can be used to obtain probability densities for importance sampling. This is called the *global importance sampling*.

The second alternative is using the information gained during previous walks to approximate the illumination. This strategy is called *adaptive importance sampling*.

In the third alternative, the problem is simplified and the indirect illumination is not considered in importance sampling. When the directions are generated, we use only w_i depending on the local orientation, the BRDF and L^e representing the direct illumination of the actual point. This is called the *local importance sampling*.

It turns out that we have to encounter severe problems when we have to find a mapping which has density that is proportional to the product of the effects of the BRDF and the direct lighting. Consequently, local importance sampling strategies usually use only either w_i or L^e to identify important directions. The first alternative is called the *BRDF sampling*, while the second is called the *lightsource sampling*.

4.2.1 BRDF sampling

BRDF based importance sampling means that at step i the density t_i of the sample points is proportional to the weight w_i, that is

$$t_i \propto w_i = f_r(\omega_{\text{in}}, \vec{x}, \omega_{\text{out}}) \cdot \cos\theta \tag{35}$$

In gathering algorithms ω_{out} is known, θ is the angle between ω_{in} and the surface normal, and ω_{in} should be determined. In shooting algorithms, on the other hand, ω_{in} is known, θ is the angle between ω_{out} and the surface normal, and ω_{out} should be determined.

Due to the fact that t_i represents density (probability density for Monte-Carlo methods), its integral is 1. Thus for gathering walks, the ratio of proportionality in equation (35) is

$$\int w \, d\omega_{\text{in}} = \int f_r(\omega_{\text{in}}, \vec{x}, \omega_{\text{out}}) \cdot \cos\theta_{\text{in}} \, d\omega_{\text{in}} = a(\vec{x}, \omega_{\text{out}})$$

where $a(\vec{x}, \omega_{\text{out}})$ is the *albedo* of the surface at point $\vec{x}$ in the outgoing direction. Similarly, the proportionality ratio for shooting walks is

$$\int w \, d\omega_{\text{out}} = \int f_r(\omega_{\text{in}}, \vec{x}, \omega_{\text{out}}) \cdot \cos\theta_{\text{out}} \, d\omega_{\text{out}} = a(\vec{x}, \omega_{\text{in}}).$$

Thus the weights w_i/t_i are the albedos at the visited points.

When combining this with Russian roulette of type $\tilde{W} = 1, \tilde{L}^e = L^e$, the probability of continuing the walk will be equal to the albedo. This can also be interpreted in the following way. When the next direction is sampled, we use a subcritical density w_i which does not integrate to 1 but to a value $a(\vec{x}, \omega)$ and with the "missing" probability $1 - a(\vec{x}, \omega)$ it is decided whether or not the walk is stopped.

In order to use BRDF sampling, random samples should be generated with probability densities that are equal to $f_r(\omega_{\text{in}}, \vec{x}, \omega_{\text{out}}) \cdot \cos\theta$. This can be very difficult for complex BRDF models. Therefore, an important research direction focuses on the design of BRDF models that are physically realistic, simple and also allows for albedo computation and importance sampling [33].

4.2.2 Global importance sampling

Global importance sampling methods are two-phase procedures. In a preprocessing phase they build a data structure that guides the second phase to find important directions. These methods can be classified according to their incorporated data structure. Since the ray-space is 5-dimensional, it is straightforward to apply a *5D adaptive tree*[26] that is similar to the well-known octree to store radiance information. Jensen proposed the application of the *photon-map* as the basis of importance sampling[17]. We assigned the power computed in the preprocessing phase to *links* established between two interacting patches[56, 57].

4.2.3 Adaptive importance sampling

Adaptive importance sampling methods neither require the non-uniform probability densities to be constructed in advance, nor simplify them to take only into account local properties, but converge to a desired probability density using the knowledge of previous samples. Three techniques are particularly important, which have also been used in rendering: *genetic algorithms*[27] the Metropolis sampling[29, 68] and the VEGAS method[28, 52]. In this paper only the Metropolis sampling is discussed.

4.2.4 Metropolis sampling

The Metropolis algorithm[29] converges to the optimal probability density that is proportional to the importance, that is in the limiting case $\mathcal{I}(\mathbf{z}) = b \cdot p(\mathbf{z})$.

However, this probability density cannot be stored, thus in the Monte-Carlo formula the importance should be used instead, in the following way:

$$I = \int_V \frac{f(\mathbf{z})}{\mathcal{I}(\mathbf{z})} \cdot \mathcal{I}(\mathbf{z}) \, d\mathbf{z} = b \cdot \int_V \frac{f(\mathbf{z})}{\mathcal{I}(\mathbf{z})} \cdot p(\mathbf{z}) \, d\mathbf{z} =$$

$$b \cdot E\left[\frac{f(\mathbf{z})}{\mathcal{I}(\mathbf{z})}\right] \approx \frac{b}{M} \cdot \sum_{i=1}^{M} \frac{f(\mathbf{z}_i)}{\mathcal{I}(\mathbf{z}_i)} \tag{36}$$

In order to generate samples according to $p(\mathbf{z}) = 1/b \cdot \mathcal{I}(\mathbf{z})$, a Markovian process is constructed whose stationary distribution is just $p(\mathbf{z})$. Informally, the next state $\mathbf{z}_{i+1}$ of this process is found by letting an almost arbitrary *tentative transition function* $T(\mathbf{z}_i \to \mathbf{z}_t)$ generate a *tentative*

sample $\mathbf{z}_t$ which is either accepted as the real next state or rejected making the next state equal to the actual state using an "*acceptance probability*" $a(\mathbf{z}_i \to \mathbf{z}_t)$ that expresses the increase of the importance (if this "acceptance probability" is greater that 1, then the sample is accepted deterministically). The formal definition of this Markovian process $\{\mathbf{z}_1, \mathbf{z}_2, \ldots \mathbf{z}_i \ldots\}$ is as follows:

for $i = 1$ **to** M **do**
  Based on the actual state $\mathbf{z}_i$,
    choose another random, tentative point $\mathbf{z}_t$
  $a(\mathbf{z}_i \to \mathbf{z}_t) = \frac{\mathcal{I}(\mathbf{z}_t) \cdot T(\mathbf{z}_t \to \mathbf{z}_i)}{\mathcal{I}(\mathbf{z}_i) \cdot T(\mathbf{z}_i \to \mathbf{z}_t)}$
  if $a(\mathbf{z}_i \to \mathbf{z}_t) \geq 1$ **then** accept $(\mathbf{z}_{i+1} = \mathbf{z}_t)$
  else // *accept with probability* $a(\mathbf{z}_i \to \mathbf{z}_t)$
    Generate random number r in $[0, 1]$.
    if $r < a(\mathbf{z}_i \to \mathbf{z}_t)$ **then** $\mathbf{z}_{i+1} = \mathbf{z}_t$
    else $\mathbf{z}_{i+1} = \mathbf{z}_i$
  endif
endfor

Note that "acceptance probability" $a(\mathbf{x} \to \mathbf{y})$ has the following property: $a(\mathbf{x} \to \mathbf{y}) = 1/a(\mathbf{y} \to \mathbf{x})$.

The transition probability of this Markovian process is:

$$P(\mathbf{x} \to \mathbf{y}) = \begin{cases} T(\mathbf{x} \to \mathbf{y}) & \text{if } a(\mathbf{x} \to \mathbf{y}) \geq 1\,, \\ T(\mathbf{x} \to \mathbf{y}) \cdot a(\mathbf{x} \to \mathbf{y}) & \text{otherwise.} \end{cases} \tag{37}$$

In equilibrium state, the transitions between two states $\mathbf{x}$ and $\mathbf{y}$ are balanced, that is

$$p(\mathbf{x}) \cdot P(\mathbf{x} \to \mathbf{y}) = p(\mathbf{y}) \cdot P(\mathbf{y} \to \mathbf{x}).$$

Using this and equation (37), and assuming without the loss of generality that $a(\mathbf{x} \to \mathbf{y}) \geq 1$, we can prove that the stationary probability distribution is really proportional to the importance:

$$\frac{p(\mathbf{x})}{p(\mathbf{y})} = \frac{P(\mathbf{y} \to \mathbf{x})}{P(\mathbf{x} \to \mathbf{y})} = \frac{T(\mathbf{y} \to \mathbf{x})}{T(\mathbf{x} \to \mathbf{y})} \cdot a(\mathbf{y} \to \mathbf{x}) = \frac{\mathcal{I}(\mathbf{x})}{\mathcal{I}(\mathbf{y})}. \tag{38}$$

When we use Metropolis sampling in the solution of the global illumination problem, the "state" $\mathbf{z}$ corresponds to a complete walk. Mutation strategies are responsible for changing the walk a "little", by perturbing one or more directions or surface points, adding or deleting steps in the path, etc.

The first use of Metropolis sampling in rendering aimed at speeding up bi-directional path tracing[68]. The performance of the Metropolis sampling has been theoretically investigated in [58].

4.3 Random walk solution of the projected integral equation

Expansion expands the solution into a discrete Neumann series

$$\mathbf{L} = \mathbf{L}^e + \mathbf{F} \cdot \mathbf{L}^e + \mathbf{F}^2 \cdot \mathbf{L}^e + \mathbf{F}^3 \cdot \mathbf{L}^e + \ldots \tag{39}$$

A element i of the $\mathbf{F}^n \cdot \mathbf{L}^e$ term can be expressed as a multi-dimensional integral, where the integrand in a single point can be obtained by executing the following walk:

A point $\vec{x}_1$ is selected on patch i. Then a ray has to be traced from $\vec{x}_1$ in direction $-\omega_1'$ and the visible patch should be identified. Following this, another point on the identified patch j should be selected, which is denoted by $\vec{x}_2$, and a ray is traced in direction $-\omega_2'$. We keep doing this n times, and finally the emission of the last patch is propagated back on the walk. During propagation, the emission is multiplied by the BRDFs ($f_i, f_j, \ldots$) and the cosine ($\cos\theta_i', \cos\theta_j', \ldots$) factors of the visited patches (figure 4).

Note that this is basically the same walking scheme, as used to solve the original integral equation. The fundamental difference is that when a patch is hit by the ray, the walk is not continued from the found point but from another point of the patch.

In the section on comparing finite-element and discrete techniques, we conlcuded that discretization makes the integrand smoother, which improves the speed of the converges, especially for quasi-Monte Carlo techniques. Thus for such walks the quasi-Monte Carlo method should be selected [4].

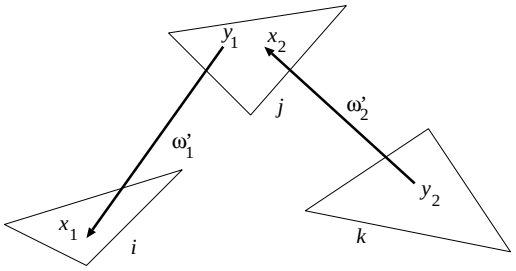

Figure 4: Random walk solution of linear equation

4.4 Gathering-type random walk algorithms

Gathering type random walks correspond to the Monte-Carlo solution of the rendering equations. They start at the eye position and gather the emission of the visited points. This approach is quite ineffective if the lightsources are small, since it has rather low probability that a walk visits a lightsource.

4.4.1 Distributed ray-tracing

Distributed ray tracing suggested by Cook[9] extends the classical ray-tracing method to model all the possible paths. In this method the ray tracing is not terminated when reaching a diffuse surface. After a ray has hit a diffuse surface, child rays are generated randomly according to the BRDF characterizing the surface. For the appropriate estimation of the diffuse interreflection, child rays have

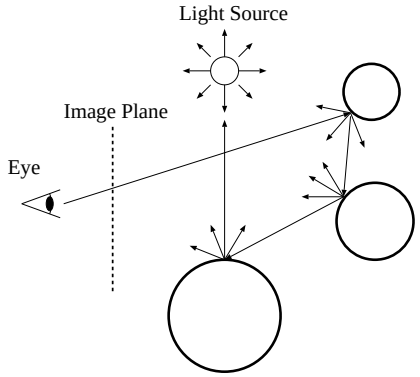

Figure 5: Distributed ray-tracing

to be traced and the average of their contributions have to be computed.

4.4.2 Path-tracing

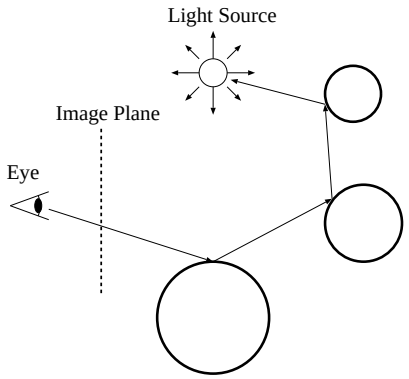

Figure 6: Path tracing

Path tracing, that was proposed by Kajiya, simply creates a path history for a single particle interacting with the environment until absorption. That is, instead of spawning new rays at an intersection, it simply chooses a random direction according to the BRDF for the ray to follow. The walk is continued with a probability equal to the albedo.

4.5 Shooting-type walks methods

Shooting walks are based on the Monte-Carlo solution of the potential equation.

4.5.1 Photon tracing

Photon tracing (forward ray-tracing) is the inverse of visibility ray-tracing and uses similar simplifying assumptions, thus they also stop tracing when hitting a surface that does not have coherent reflection or refraction. In photon tracing the rays are emitted from the light sources, and at each hit it is examined whether the surface has ideal reflection, refraction and incoherent reflection or refraction. In the directions of ideal reflection or refraction, the tracing is continued by starting new child rays. The effect of incoherent interactions, on the other hand, is stored in a map or is projected to the eye by tracing a ray towards the camera position.

4.5.2 Light Tracing

In *light tracing*[12] photons perform random walk through the scene starting at the lightsources. Whenever a surface is hit, a ray is traced from the intersection point to the eye and the contribution is added to the selected pixel (if any).

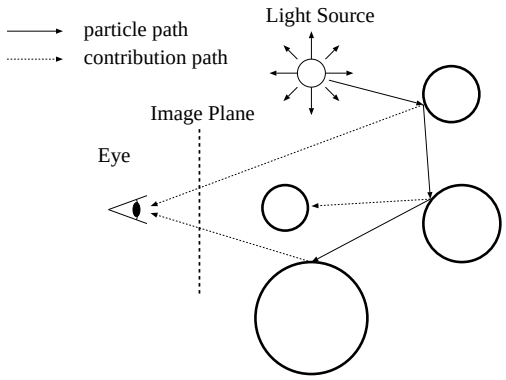

Figure 7: Light tracing

When the next direction is determined, the BRDF based importance sampling can be applied and combined with the random termination according to the albedo.

4.5.3 Bi-directional Path Tracing

Bi-directional path tracing[25, 67] is based on the combination of shooting and gathering walks thus it can combine the advantages of both techniques. Namely, it can effectively handle small lightsources and small aperture cameras.

Walks are initiated at the same time from a selected light source and from the viewpoint. After some steps, either a single deterministic shadow ray is used to connect the two types of walks[67], or all points of the gathering walk are connected to all points of the shooting walk using deterministic rays[25]. If the deterministic shadow ray detects that the two points are occluded from each other, then the contribution of this path is zero.

Note that gathering and shooting walks use different integration variables, namely a gathering walk is specified by a point on the pixel area and a sequence of incoming directions, while a shooting walk is defined by a point on the lightsource and a sequence of the outgoing directions. Thus when the two walks are connected, appropriate transformations should take place, which requires a multiplication of the radiance by

$$\frac{\cos\theta'_k \cdot \cos\theta_{n-k+1}}{r_k^2},$$

where θ'_k and θ_{n-k+1} are the angles between the surface normals and the direction of the connection at the last point of the gathering and shooting walks respectively, and r_k is the length of the connection.

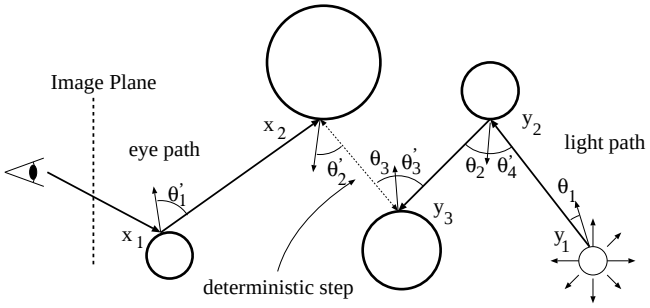

Figure 8: Bi-directional path tracing with a single deterministic step

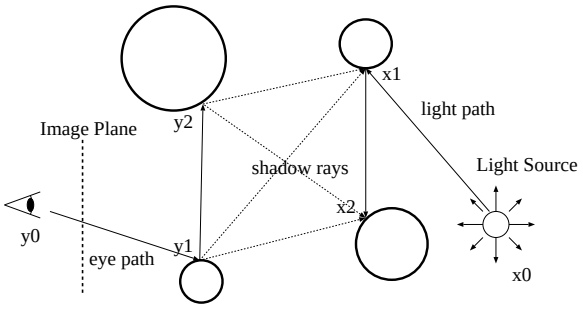

Figure 9: Bi-directional path tracing with multiple deterministic steps

In Lafortune's version of the bi-directional path tracing[25] not only the endpoints of the shooting and gathering walks are connected, but all intersection points are linked by shadow rays.

4.5.4 Photon-map

Bi-directional path tracing connects a single gathering walk to a single shooting walk. However, if the effects of a shooting walk, for instance, could be stored, then when a new gathering walk is computed, it could be connected to all of them simultaneously. This is exactly what Jensen[19, 18, 20] proposed, also giving the definition of a data structure, called the *photon-map* which can efficiently store the effects of many shooting walks.

A photon map is a collection of photon hits generated in the shooting phase of the algorithm. The photon-map is organized in a *kd-tree* to support efficient retrieval. A photon hit is stored with the power of the photon on different wavelengths, position, direction of arrival and with the surface normal.

The gathering phase is based on the following approximation of the transport operator:

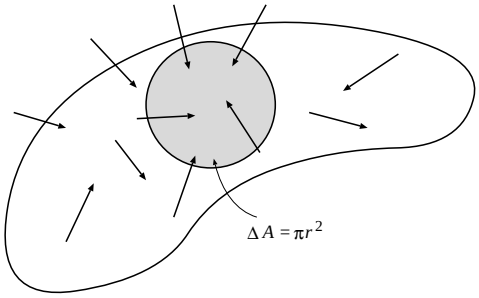

Figure 10: The information stored in the photon map

$$L(\vec{x},\omega') = \int_{\Omega} L(h(\vec{x},-\omega'),\omega')\cdot f_r(\omega',\vec{x},\omega)\cdot\cos\theta'\,d\omega' =$$

$$\int_{\Omega} \frac{d\Phi(\omega')}{dA\cos\theta' d\omega'}\cdot f_r(\omega',\vec{x},\omega)\cdot\cos\theta'\,d\omega' \approx$$

$$\sum_{i=1}^{n} \frac{\Delta\Phi(\omega'_i)}{\Delta A}\cdot f_r(\omega'_i,\vec{x},\omega), \qquad (40)$$

where $\Delta\Phi(\omega'_i)$ is the power of a photon landing at the surface ΔA from direction ω'_i. The $\Delta\Phi$ and ΔA quantities are approximated from the photons in the neighborhood of $\vec{x}$ in the following way. A sphere centered around $\vec{x}$ is extended until it contains n photons. If at this point the radius of the sphere is r, then the intersected surface area is $\Delta A = \pi r^2$.

4.5.5 Instant radiosity

Instant radiosity[23] elegantly subdivides the shooting walks into a view-independent walk and into the projection of the contribution to the eye. Let us call this last step with eye projection the *eye-step*. The view-independent walk is quite similar to the light-tracing algorithm, but the new directions are sampled from the Halton sequence instead of a random distribution.

When a surface hit is found, the eye-step is calculated taking advantage of the rendering hardware of advanced workstations. The reflection of this hit is assumed to be a point lightsource (in the radiosity setting the emission of the lightsource is also diffuse), and the rendering hardware is used to render the effect of this lightsource on the scene and also to compute shadows. The final image is the average of such estimates, which are computed using the hardware accumulation buffer.

Instant radiosity is quite similar to photon-map based techniques. However, instead of using ray-tracing for final gather, the photons in the photon map are used as lightsources and fast and hardware supported visibility and shadow algorithms are applied. The other fundamental difference is that instant radiosity allows just a relatively low number of photons which therefore should be very

well distributed. The optimal distribution is provided by quasi-Monte Carlo light walks.

4.5.6 Global ray-bundle tracing

Realizing that an accurate solution requires great many samples, *global ray-bundle tracing*[61, 62, 52] uses a bundle of very many (e.g. 1 million or even infinite) global parallel rays, which can be traced simultaneously using image coherence techniques. In order to represent the radiance that is transferred by a ray, finite-element techniques are applied that approximate the positional (but not the directional) dependence of the radiance by piece-wise continuous or piece-wise linear functions[60].

$$L(\vec{x},\omega) \approx \sum_{j=1}^{n} b_j(\vec{x}) \cdot L_j(\omega) = \mathbf{b}^T \cdot \mathbf{L}(\omega). \qquad (41)$$

Note that this is a mixed finite-element and continuous method, since the positional dependence of the radiance is approximated by finite-elements, while the directional dependence is not.

Substituting this into the rendering equation and projecting that into an adjoint base we obtain

$$\mathbf{L}(\omega) = \mathbf{L}^e(\omega) + \mathcal{T}_F \mathbf{L}(\omega), \qquad (42)$$

where $\mathcal{T}_F$ is a composition of the original transport operator and its projection to the adjoint base

$$\mathcal{T}_F \mathbf{L}(\omega) = \langle \mathcal{T}\mathbf{b}^T \cdot \mathbf{L}(\omega), \tilde{\mathbf{b}} \rangle. \qquad (43)$$

Let us use again piece-wise constant basis functions. Then the result of the application of the transport operator on patch i is

$$\mathcal{T}_F \mathbf{L}(\omega)|_i = \frac{1}{A_i} \cdot \int_\Omega \int_{A_i} L(h(\vec{x}, -\omega'), \omega') \cdot \cos\theta' \cdot \tilde{f}_i(\omega', \omega)\, d\vec{x}\, d\omega'. \qquad (44)$$

Taking into account that the integrand of the inner surface integral is piece-wise constant, it can also be presented in closed form:

$$\int_{A_i} L(h(\vec{x}, -\omega'), \omega') \cdot \cos\theta' \cdot \tilde{f}_i(\omega', \omega)\, d\vec{x} =$$

$$\sum_{j=1}^{n} \tilde{f}_i(\omega', \omega) \cdot A(i, j, \omega') \cdot L_j(\omega'), \qquad (45)$$

where $A(i, j, \omega')$ expresses the projected area of patch j that is visible from patch i in direction ω'. In the unoccluded case this is the intersection of the projections of patch i and patch j onto a plane perpendicular to ω'. If occlusion occurs, the projected areas of other patches that are in between patch i and patch j should be subtracted as shown in figure 11.

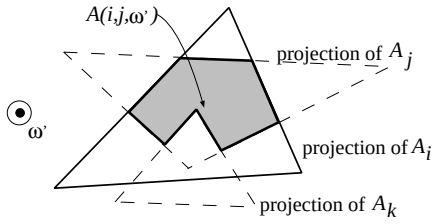

Figure 11: Interpretation of $A(i, j, \omega')$

This projected area can be efficiently calculated simultaneously for all patch pairs using global discrete or continuous visibility algorithms[52] and also exploiting the hardware z-buffer[62]. These algorithms can also have random nature, that is, they can result in $A(i, j, \omega') \cdot L_j(\omega')$ just as an the expected value[64, 53].

Using equation (45) the rendering equation can be obtained as:

$$\mathbf{L}(\omega) = \mathbf{L}^e(\omega) + \int_\Omega \mathbf{F}(\omega', \omega) \cdot \mathbf{A}(\omega') \cdot \mathbf{L}(\omega')\, d\omega', \quad (46)$$

where $\mathbf{L}(\omega)$ is the vector of radiance values, $\mathbf{F}(\omega', \omega)$ is a diagonal matrix of BRDFs, and *geometry matrix* $\mathbf{A}$ contains the relative visible areas: $\mathbf{A}(\omega')|_{ij} = A(i, j, \omega')/A_i$.

Note that equation (46) is highly intuitive as well. The radiance of a patch is the sum of the emission and the reflection of all incoming radiance. The role of the patch-direction-patch "form-factors" is played by $A(i, j, \omega')/A_i$.

This is also an integral equation but unlike the original rendering equation it provides the radiance of not only a single point but for all points at once. This integral equation is solved by random or quasi-random shooting type walks.

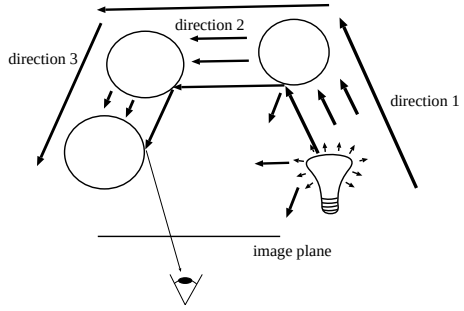

Figure 12: A path of ray-bundles

A single walk starts by selecting a direction either randomly or quasi-randomly, and the emission transfer of all patches is calculated into this direction (figure 12). Then a new direction is found, and the emission is transferred and the incoming radiance generated by the previous transfer is reflected from all patches into this new direction. The algorithm keeps doing this for a few times depending on how many bounces should be considered, then the emission is sent and the incoming radiance caused by the last

transfer is reflected towards the eye. Averaging these contributions results in the final image.

5 Stochastic iteration

The basic idea of stochastic iteration is that instead of approximating operator $\mathcal{T}$ in a deterministic way, a much simpler random operator is used during the iteration which "behaves" as the real operator just in the "average" case. The concept of stochastic iteration was proposed for the diffuse radiosity problem in [31], that is for the solution of finite-dimensional linear equations.

In this section we present a generalized formulation that is somewhat different from the original concepts to allow to attack also non-diffuse global illumination problems[53]. Suppose that we have a random linear operator $\mathcal{T}^*$ so that

$$E[\mathcal{T}^* L] = \mathcal{T} L \tag{47}$$

for any integrable function L.

In the case of finite-element representations, equation (47) should be true for the $\mathcal{T}_F \mathbf{L}$ operator that also involves the projection to the finite function space.

During stochastic iteration a random sequence of operators $\mathcal{T}_1^*, \mathcal{T}_2^*, \ldots \mathcal{T}_i^* \ldots$ is generated, which are instantiations of $\mathcal{T}^*$, and this sequence is used in the iteration formula:

$$L_n = L^e + \mathcal{T}_n^* L_{n-1}. \tag{48}$$

Since in computer implementations the calculation of a random operator may invoke finite number of random number generator calls, we are particularly interested in random operators having the following construction scheme:

1. Random "point" p_i is found from a finite dimensional set Π using probability density $\mathrm{prob}(p)$. This probability density may or may not depend on function L.

2. Using p_i a "deterministic" operator $\mathcal{T}^*(p_i)$ is applied to L.

Point p_i is called the *randomization point* since it is responsible for the random nature of operator $\mathcal{T}^*$.

Using a sequence of random transport operators, the measured power

$$P_n = \mathcal{M} L_n \tag{49}$$

will also be a random variable which does not converge but fluctuates around the real solution. Thus the solution can be found by averaging the estimates of the subsequent iteration steps.

Formally the sequence of the iteration is the following:

$$\begin{aligned}
P_1 &= \mathcal{M} L_1 = \mathcal{M}(L^e + \mathcal{T}_1^* L^e) \\
P_2 &= \mathcal{M} L_2 = \mathcal{M}(L^e + \mathcal{T}_2^* L^e + \mathcal{T}_2^* \mathcal{T}_1^* L^e) \\
&\vdots \\
P_M &= \mathcal{M} L_M = \mathcal{M}(L^e + \mathcal{T}_M^* L^e + \mathcal{T}_M^* \mathcal{T}_{M-1}^* L^e + \ldots)
\end{aligned}$$

Averaging the first M steps, we obtain:

$$\tilde{P} = \frac{1}{M} \sum_{i=1}^{M} \mathcal{M} L_i =$$

$$\mathcal{M}(L^e + \frac{1}{M} \sum_{i=1}^{M} \mathcal{T}_i^* L^e + \frac{1}{M} \sum_{i=1}^{M-1} \mathcal{T}_{i+1}^* \mathcal{T}_i^* L^e + \ldots) =$$

$$\mathcal{M}(L^e + \frac{1}{M} \sum_{i=1}^{M} \mathcal{T}_i^* L^e + \frac{M-1}{M} \cdot \frac{1}{M-1} \sum_{i=1}^{M-1} \mathcal{T}_{i+1}^* \mathcal{T}_i^* L^e + \ldots). \tag{50}$$

The expected value of the averaged solution $\tilde{P}$ is [55]:

$$E[\tilde{P}] =$$

$$\mathcal{M}(L^e + \mathcal{T} L^e + \frac{M-1}{M} \mathcal{T}^2 L^e + \frac{M-2}{M} \mathcal{T}^3 L^e + \ldots + \frac{1}{M} \mathcal{T}^M L^e), \tag{51}$$

which converges to the real solution

$$\mathcal{M}(L^e + \mathcal{T} L^e + \mathcal{T}^2 L^e + \mathcal{T}^3 L^e + \ldots)$$

if M goes to infinity.

5.1 Definition of random transport operators

In order to use this general stochastic iteration scheme in practice, the key problem is the definition of the random transport operator. This operator should meet the requirement of equation (47) and should be easy to compute.

For the continuous case, a single application of the transport operator contains a directional integral. For the finite element case, the transport operator also includes the projection to the adjoint basis which requires additional integration in the domain of basis functions. This additional integration means a surface integral for the diffuse radiosity setting and also for the ray-bundle tracing. For other non-diffuse finite-element methods a surface and a directional integrals need to be evaluated (note that directional integrals are sometimes "hidden" by integrals on the surfaces visible at different directions).

Following the general concepts of Monte-Carlo methods, we usually do not intend to compute the integrals *explicitly*, but want to get them as an expected value. Thus different random transport operators can be classified according to which integrals are evaluated explicitly using some deterministic quadrature and which integrals are computed *implicitly* as an expectation value.

5.2 Transport operator for the continuous, non-diffuse setting

The continuous formulation has just a single directional integral, thus a random transport operator can evaluate this single integral implicitly. This results in a method that uses a "single" random walk to obtain the solution.

An example of such single walk techniques is the following modification of the light tracing algorithm[53]:

In each step i a ray is obtained that has random origin $\vec{y}_i$ and direction ω_i with a probability that is proportional to the cosine weighted radiance of this point at the given direction. This ray is traced and the whole power

$$\Phi = \int_S \int_\Omega L(\vec{y}, \omega') \cos\theta_{\vec{y}} \, d\omega' \, d\vec{y}$$

is transported to that point $\vec{x}$ which is hit by the ray. Formally the random transport operator is

$$(\mathcal{T}^* L)(\vec{x}, \omega) = \Phi \cdot \delta(\vec{x} - h(\vec{y}, \omega_i)) \cdot f_r(\omega_i, \vec{x}, \omega). \quad (52)$$

Interestingly this iteration is a sequence of variable length random walks, since at each step the point that is last hit by the ray is only selected with a given probability as the starting point of the next ray. This probability depends on the albedo $a_{\vec{x}_i}(\omega_i)$ of the found point.

The algorithm selects initially a point from a lightsource and then starts a random walk. The walk finishes after each step with probability $1/(1 + a_{\vec{x}_i}(\omega_i))$ and also when the ray hits no object. If a walk finishes, another walk is initiated from the lightsource. When the walk is continued, the transferred power is weighted by $(1 + a_{\vec{x}_i}(\omega_i))$, which provides unbiased estimate even if less number of samples are used to simulate higher order bounces. This technique is called the *Russian roulette*[2, 47].

5.3 Random transport operators for the diffuse radiosity

In the gathering type radiosity algorithms the projected transport operator has the following form

$$\mathcal{T}_F \mathbf{L} = \mathbf{F} \cdot \mathbf{L}.$$

Alternatively, shooting radiosity algorithms are based on the projected potential equation $\mathcal{T}'_F \mathbf{P} = \mathbf{H} \cdot \mathbf{P}$.

According to the basic requirement of stochastic iteration we need to find random operators $\mathcal{T}^*_F$ or $\mathcal{T}'^*_F$ that behave as the real operator in average, that is

$$E[\mathcal{T}^*_F \mathbf{L}] = \mathbf{F} \cdot \mathbf{L}, \quad (53)$$

$$E[\mathcal{T}'^*_F \mathbf{P}] = \mathbf{H} \cdot \mathbf{P}. \quad (54)$$

The evaluation of $(\mathbf{F} \cdot \mathbf{L})_i$ or alternatively $(\mathbf{H} \cdot \mathbf{P})|_i$ requires a surface and a directional integration (or in other formulations two surface integrations).

The possible alternatives for a random transport operator are

1. Both integrals are explicitly computed but only for a randomly selected subset of the patches.
2. The surface integral explicitly computed but the directional integral implicitly.
3. Compute the surface integral implicitly but the directional integral explicitly. This method can, for example, use hemicubes for the directional integration but selects the center of the hemicube randomly on the patch.
4. Both integrals are computed implicitly.

5.3.1 Stochastic radiosity

In *stochastic radiosity*[32], the randomized operator is simplified in a sense that it first selects a single (or a few) patches with probability proportional to their power and then calculates the transfer only from this important patch as if it had all the power $\Phi = \sum_{k=1}^{n} \mathbf{P}_k$. Thus here both integrals are explicitly computed but only for a subset of patches.

To prove that it meets requirement stated by equation (54), let us examine the new power of patch i and suppose that patch j has been selected.

$$(\mathcal{T}'^*_F \mathbf{P})|_i = \mathbf{H}_{ij} \cdot \Phi \quad (55)$$

Since the probability of selecting patch j is $\mathbf{P}_j/\Phi$, the expectation of the new power is

$$E[(\mathcal{T}'^*_F \mathbf{P})|_i] = \sum_{j=1}^{n} \mathbf{H}_{ij} \cdot \Phi \cdot \frac{\mathbf{P}_j}{\Phi} = \sum_{j=1}^{n} \mathbf{H}_{ij} \cdot \mathbf{P}_j \quad (56)$$

which we wanted to prove.

5.3.2 Transillumination radiosity

The *transillumination radiosity method*[31, 59] has also a stochastic iteration version. It defines the random transport operator by uniformly selecting M *transillumination directions* $\omega'_1, \ldots \omega'_M$ and allowing patches to interact only in these transillumination directions. In order to calculate these interactions, a large discretized window is placed perpendicularly to each transillumination direction and the radiance transfer to a patch is approximated by elementary transfers going through the pixels covering the projection of the patch.

Let us consider a single transillumination direction. Projecting patch A_i onto a plane that is perpendicular to the transillumination direction and then approximating the integral of the incoming radiance here by a discrete sum, we get

$$\int_{A_i} L(h(\vec{x}, -\omega'_d)) \cdot \cos\theta'_d \, d\vec{x} =$$

$$\int_{A_i^p} L(h(\vec{x}', -\omega'_d)) \cdot \, d\vec{x}' \approx \sum_{P \in A_i^p} L_{\mathrm{buffer_d}[P]} \cdot \delta A. \quad (57)$$

where $\mathrm{buffer_d}[P]$ stores the index of that patch which is visible in pixel P in the transillumination direction ω_d' from patch i, and δA is the size of a pixel of the buffer (figure 13).

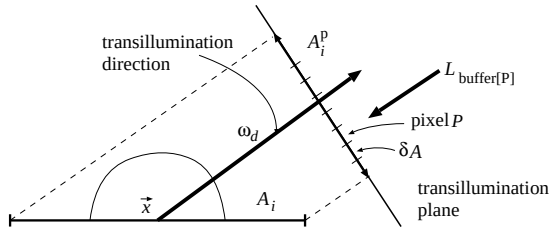

Figure 13: Integration on the transillumination plane

Thus the random transfer operator is

$$(\mathcal{T}_F^*\mathbf{L})|_i = \frac{4\pi \cdot f_i \cdot \delta A}{M} \sum_{d=1}^{M} \sum_{P \in A_i^p} L_{\mathrm{buffer_d}[P]}. \qquad (58)$$

If the transillumination directions are uniformly distributed and the buffer is uniformly jittered, then the expected value of this operator is equal to the real operator [54].

5.3.3 Stochastic ray-radiosity

Stochastic ray-radiosity[34] approximates the transport operator by M random rays that are sampled proportionally to the power of the patches. On a patch the starting point of the ray is sampled using a uniform distribution, while the direction follows a cosine distribution. A single ray carries Φ/M power. Thus this method approximates both integrals implicitly.

Let us examine the case when a single ray is selected (since different rays are sampled from the same distribution, the effect of M rays will be M times the effect of a single ray in the expected value). Suppose that patch j is selected as a shooting patch. The probability of the selection event is $\mathbf{P}_j/\Phi$. Thus the probability density of selecting a point $\vec{x}$ of a patch and a direction ω is

$$\frac{\mathbf{P}_j}{\Phi} \cdot \frac{1}{A_j} \cdot \cos\theta.$$

This transfers Φ/M power to the patch that is hit by the ray where the reflected power is computed. Thus the random transport operator for a single ray is

$$E[(\mathcal{T}_F^{'*}\mathbf{P})|_i] =$$

$$M \cdot f_i \cdot \sum_{j=1}^{n} \int_{A_j} \int_{\Omega} b_i(h(\vec{y},\omega)) \cdot \frac{\Phi}{M} \cdot \frac{1}{A_j} \cdot \cos\theta \, d\vec{y} d\omega \cdot \frac{\mathbf{P}_j}{\Phi} =$$

$$\sum_{j=1}^{n} \frac{f_i}{A_j} \cdot \int_{A_j} \int_{\Omega} b_i(h(\vec{y},\omega)) \cdot \cos\theta \, d\vec{y} d\omega \cdot \mathbf{P}_j = \sum_{j=1}^{n} \mathbf{H}_{ij} \cdot \mathbf{P}_j. \qquad (59)$$

5.4 Transport operators for the non-diffuse finite-element case

When moving towards the non-diffuse case, another requirement must be imposed upon the random transport operator. It must not only meet the requirement of equation (47), be easy to compute, but it must also allow the compact representation of the $\mathcal{T}_i^* L$ functions. This extra requirement is evident if we take into account that unlike in the diffuse case, the domain of L is a 4-dimensional continuous space, so is the domain of $\mathcal{T}_i^* L$ (for ray-bundle tracing only 2-dimensional continuous space). From the point of view of compact representation, what we have to avoid is the representation of these functions over the complete domain.

Thus those transport operators are preferred, which require the value of L just in a few "domain points" (e.g. in a single "domain point"). Note that the evaluation of $\mathcal{T}_i^* L$ now consists of the following steps: first a randomization point p_i is found to define random operator $\mathcal{T}_i^*$, which in turn determines at which domain point the value of L is required. Up to now, we have had complete freedom to define the set of randomization points. One straightforward way is defining this set to be the same as the domain of the radiance function and using random transport operators that require the value of the radiance function at their randomization points. Although this equivalence is not obligatory, it can significantly simplify the computations, since when the randomization point is generated, the required domain point is also known.

Using random operators that evaluate the radiance in a single point is not enough in itself, since even a single "point" can result in a continuous $\mathcal{T}_i^* L$ function, which must be stored and re-sampled in the subsequent iteration step and also by the measurement. The solution is the postponing of the complete calculation of $\mathcal{T}_i^* L$ until it is known where its value is needed in the next iteration step and by the measuring device. In this way, the random operator should be evaluated twice but just for two points. Once for the actual and the previous "points" resulting in $[\mathcal{T}^*(p_i)L(p_i)](p_{i+1})$, and once for p_{eye} which is needed by the measuring device and for previous point providing $[\mathcal{T}^*(p_i)L(p_i)](p_{\mathrm{eye}})$.

The complete iteration goes as follows:

$P = 0$
Find p_1 randomly
$L(p_1) = L^e(p_1)$
for $i = 1$ **to** M **do**
  $P^{\mathrm{new}} = L^e(p_{\mathrm{eye}}) + [\mathcal{T}^*(p_i)L(p_i)](p_{\mathrm{eye}})$
  $P = \mathcal{M}P^{\mathrm{new}} \cdot 1/i + (1 - 1/i) \cdot P$
  Find p_{i+1} randomly
  $L(p_{i+1}) = L^e(p_{i+1}) + [\mathcal{T}^*(p_i)L(p_i)](p_{i+1})$
endfor
Display final image

5.4.1 Global ray-bundle based iteration

Recall that the finite-element approximation applied by ray-bundle tracing converts the rendering equation to the following form (section 4.5.6):

$$\mathbf{L}(\omega) = \mathbf{L}^e(\omega) + \mathcal{T}_F\mathbf{L}(\omega), \tag{60}$$

where $\mathcal{T}_F$ is a composition of the original transport operator and its projection to the adjoint base

$$\mathcal{T}_F\mathbf{L}(\omega) = \int_\Omega \mathbf{F}(\omega', \omega) \cdot \mathbf{A}(\omega') \cdot \mathbf{L}(\omega')\, d\omega'. \tag{61}$$

Let the random approximation of the transport operator be the transfer of the radiance of all surface points of the scene in a single uniformly distributed random direction. This transfer can be effectively realized by sending a ray-bundle into this direction. Thus the random transport operator is ω'

$$(\mathcal{T}^*\mathbf{L})(\omega) = 4\pi \cdot \mathbf{F}(\omega', \omega) \cdot \mathbf{A}(\omega') \cdot \mathbf{L}(\omega'). \tag{62}$$

If the directions are sampled from a uniform distribution, then this obviously gives back the integral operator as an expected value:

$$E[(\mathcal{T}^*\mathbf{L})(\omega)] = \int_\Omega 4\pi \cdot \mathbf{F}(\omega', \omega) \cdot \mathbf{A}(\omega') \cdot \mathbf{L}(\omega') \frac{d\omega'}{4\pi} = \mathcal{T}_F\mathbf{L}(\omega). \tag{63}$$

In the definition of the random operator ω is the actually generated direction and ω' is the previously generated direction. Thus a "randomization point" is a global direction in this method.

The resulting algorithm is quite simple. In a step of the stochastic iteration a new direction is found and this direction together with the previous direction are used to evaluate the random transport operator. Then an image estimate is computed by reflecting the previously computed radiance estimate towards the eye. The complete algorithm is summarized in the following:

Generate the first random global direction ω_1
for each patch i **do** $L[i] = L_i^e(\omega_1)$
for $m = 1$ **to** M **do** *// iteration cycles*
  Calculate the image estimate relfecting
    the incoming radiance $L[1], L[2], \ldots L[n]$
    from ω_m towards the eye
  Average the estimate with the Image
  Generate random global direction ω_{m+1}
  for each patch i **do**
    $L^{\text{new}}[i] = L_i^e(\omega_{m+1}) + 4\pi \cdot$
    $\sum_{j=1}^n \tilde{f}_i(\omega_m, \omega_{m+1}) \cdot A(i, j, \omega_m)/A_i \cdot L[j]$
  endfor
endfor
Display Image

6 Handling point lightsources

Monte-Carlo integration is efficient if the integrand is relatively smooth and does not exhibit high variations. For gathering walks and for global methods, point lightsources may pose problems. Fortunately, these lightsources can be easily handled separately by deterministic techniques.

This algorithm can be applied either in a preprocessing step when it is called the *first-shot* or during the calculation of the random walks. For the diffuse radiosity problem, the preprocessing type first-shot algorithm has been first presented in [40], extended to multiple interreflections in [6] and has been generalized to non-diffuse environments in [62].

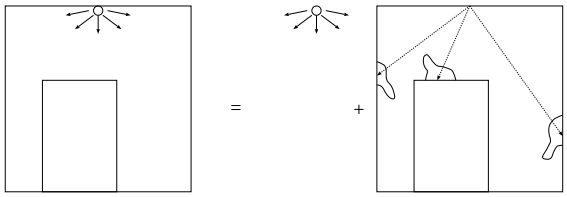

Figure 14: First shot technique

Formally, the unknown radiance L is decomposed into two terms:

$$L = L^{ep} + L^{np} \tag{64}$$

where L^{ep} is the emission of the small, point-like lightsources, L^{np} is the emission of the area lightsources and the reflected radiance. Substituting this into the rendering equation we have:

$$L^{ep} + L^{np} = L^e + \mathcal{T}(L^{ep} + L^{np}). \tag{65}$$

Introducing the new lightsource term

$$L^{e*} = L^e - L^{ep} + \mathcal{T}L^{ep} \tag{66}$$

which just replaces the point lightsources (L^{ep}) by their effect ($\mathcal{T}L^{ep}$), the equation for L^{np} is similar to the original rendering equation:

$$L^{np} = L^{e*} + \mathcal{T}L^{np}. \tag{67}$$

It means that first the direct illumination caused by the point lightsources must be computed, then they can be removed from the scene and added again at the end of the computation.

This idea can be combined with gathering type walks. In this case L^{e*} is computed on the fly when needed. To obtain L^{e*}, shadow rays are traced from each visited point of the walk towards every point lightsource and the direct lighting is deterministically computed. For finite-element-type methods, the computation of L^{e*} can be moved to a preprocessing phase. In diffuse environments the storage of the direction independent L^{e*} function requires just one extra variable per patch. In non-diffuse environments,

however, L^{e*} can be a non-constant function, which is difficult to represent and store. Instead, the incoming radiance received by the patches from each point lightsource should be stored (this requires l additional variables per patch, where l is the number of point lightsources). When L^{e*} is needed for a given direction, then it is computed on the fly from these incoming radiances.

7 Conclusions

This paper presented a review of stochastic global illumination algorithms. We concluded that stochastic methods are indispensible since they present the only reasonable alternative to compute the very high-dimensional integrals that emerge in global illumination methods.

Monte-Carlo algorithms take random or quasi-random samples and approximate the solution as an average. In order to make these methods fast, we either reduce the number of samples or reduce the cost of the samples. The number of samples depend on the variation of the integrand, that is how heterogeneous the light is in the space. Techniques, including importance sampling or the special treatment of point lightsources aim at the smoothing the integrand. The number of samples can also be reduced by distributing them very uniformly, which leads to the application of low-discrepancy series instead of random points. However, care should be taken, since these methods can be less efficient in higher dimensions, and can even provide wrong results in iteration. The other main direction, which aims at reducing the computational costs of the samples, includes global methods. This paper focused on ray-bundle tracing, which reduces the cost of tracing ray by tracing it with many other parallel rays and exploiting coherence properties.

References

[1] J. Arvo. Stratified sampling of spherical triangles. In *Computer Graphics (SIGGRAPH '95 Proceedings)*, pages 437–438, 1995.

[2] J. Arvo and D. Kirk. Particle transport and image synthesis. In *Computer Graphics (SIGGRAPH '90 Proceedings)*, pages 63–66, 1990.

[3] L. Aupperle and P. Hanrahan. A hierarchical illumination algorithms for surfaces with glossy reflection. *Computer Graphics (SIGGRAPH '93 Proceedings)*, pages 155–162, 1993.

[4] P. Bekaert, R. Cools, and D. Willems. An empirical comparison of monte carlo radiosity algorithms. In *Winter School of Computer Graphics '99*, pages 9–16, Plzen, Czech Republic, 1999.

[5] C. Buckalew and D. Fussell. Illumination networks: Fast realistic rendering with general reflectance functions. *Computer Graphics (SIGGRAPH '89 Proceedings)*, 23(3):89–98, July 1989.

[6] F. Castro, R. Martinez, and M. Sbert. Quasi Monte-Carlo and extended first-shot improvements to the multi-path method. In *Spring Conference on Computer Graphics '98*, pages 91–102, 1998.

[7] P. H. Christensen, D. Lischinski, E. J. Stollnitz, and D. H. Salesin. Clustering for glossy global illumination. *ACM Transactions on Graphics*, 16(1):3–33, 1997.

[8] P. H. Christensen, E. J. Stollnitz, D. H. Salesin, and T. D. DeRose. Global illumination of glossy environments using wavelets and importance. *ACM Transactions on Graphics*, 15(1):37–71, 1996.

[9] R. Cook, T. Porter, and L. Carpenter. Distributed ray tracing. In *Computer Graphics (SIGGRAPH '84 Proceedings)*, pages 137–145, 1984.

[10] B. Csebfalvi. A review of Monte-Carlo ray tracing algorithms. In *CESCG '97, Central European Seminar on Computer Graphics*, pages 87–103, 1997.

[11] D. P. Dobkin, D. Eppstein, and D. P. Mitchell. Computing the discrepancy with applications to supersampling patterns. *ACM Transactions on Graphics*, 15(4):354–376, 1996.

[12] P. Dutre, E. Lafortune, and Y. D. Willems. Monte Carlo light tracing with direct computation of pixel intensities. In *Compugraphics '93*, pages 128–137, Alvor, 1993.

[13] A. Glassner. *Principles of Digital Image Synthesis*. Morgan Kaufmann Publishers, Inc., San Francisco, 1995.

[14] P. Hanrahan, D. Salzman, and L. Aupperle. Rapid hierachical radiosity algorithm. *Computer Graphics (SIGGRAPH '91 Proceedings)*, 1991.

[15] M. Hyben, I. Martisovits, and A. Ferko. Scene complexity for rendering in flatland. In L. Szirmay-Kalos, editor, *Spring Conference on Computer Graphics*, pages 112–120, 1998.

[16] D. S. Immel, M. F. Cohen, and D. P. Greenberg. A radiosity method for non-diffuse environments. In *Computer Graphics (SIGGRAPH '86 Proceedings)*, pages 133–142, 1986.

[17] H. W. Jensen. Importance driven path tracing using the photon maps. In *Rendering Techniques '95*, pages 326–335, 1995.

[18] H. W. Jensen. Global illumination using photon maps. In *Rendering Techniques '96*, pages 21–30, 1996.

[19] H. W. Jensen and N. J. Christensen. Photon maps in bidirectional Monte Carlo ray tracing of complex objects. *Computers and Graphics*, 19(2):215–224, 1995.

[20] H. W. Jensen and P. H. Christensen. Efficient simulation of light transport in scenes with participating media using photon maps. *Computers and Graphics (SIGGRAPH '98 Proceedings)*, pages 311–320, 1998.

[21] J. T. Kajiya. The rendering equation. In *Computer Graphics (SIGGRAPH '86 Proceedings)*, pages 143–150, 1986.

[22] A. Keller. A quasi-Monte Carlo algorithm for the global illumination in the radiosity setting. In H. Niederreiter and P. Shiue, editors, *Monte-Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, pages 239–251. Springer, 1995.

[23] A. Keller. Instant radiosity. *Computer Graphics (SIGGRAPH '97 Proceedings)*, pages 49–55, 1997.

[24] D.E. Knuth. *The art of computer programming. Volume 2 (Seminumerical algorithms)*. Addison-Wesley, Reading, Mass. USA, 1981.

[25] E. Lafortune and Y. D. Willems. Bi-directional path-tracing. In *Compugraphics '93*, pages 145–153, Alvor, 1993.

[26] E. Lafortune and Y. D. Willems. A 5D tree to reduce the variance of Monte Carlo ray tracing. In *Rendering Techniques '96*, pages 11–19, 1996.

[27] B. Lange and B. Beyer. Rayvolution: An evolutionary ray tracing algorithm. In *Photorealistic Rendering Techniques*, pages 136–144, 1994.

[28] G. P. Lepage. An adaptive multidimensional integration program. Technical Report CLNS-80/447, Cornell University, 1980.

[29] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1091, 1953.

[30] D. P. Mitchell. Consequences of stratified sampling in graphics. *Computer Graphics (SIGGRAPH '96 Proceedings)*, pages 277–280, 1996.

[31] L. Neumann. Monte Carlo radiosity. *Computing*, 55:23–42, 1995.

[32] L. Neumann, M. Feda, M. Kopp, and W. Purgathofer. A new stochastic radiosity method for highly complex scenes. In *Proc. of the 5th. EG Workshop on Rendering*, 1994.

[33] L. Neumann, A. Neumann, and L. Szirmay-Kalos. Compact metallic reflectance models. *Computer Graphics Forum (Eurographics'99)*, 18(3), 1999.

[34] L. Neumann, W. Purgathofer, R. F. Tobler, A. Neumann, P. Elias, M. Feda, and X. Pueyo. The stochastic ray method for radiosity. In *Rendering Techniques '95*, pages 206–218, 1995.

[35] H. Niederreiter. *Random number generation and quasi-Monte Carlo methods*. SIAM, Pennsilvania, 1992.

[36] S. N. Pattanik and S. P. Mudur. Adjoint equations and random walks for illumination computation. *ACM Transactions on Graphics*, 14(1):77–102, 1995.

[37] M. Pellegrini. Monte Carlo approximation of form factors with error bounded a priori. *Discrete and Computational Geometry*, 1997. (to appear).

[38] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C (Second Edition)*. Cambridge University Press, Cambridge, USA, 1992.

[39] M. Sbert. *The Use of Global Directions to Compute Radiosity*. PhD thesis, Catalan Technical University, Barcelona, 1996.

[40] M. Sbert, X. Pueyo, L. Neumann, and W. Purgathofer. Global multipath Monte Carlo algorithms for radiosity. *Visual Computer*, pages 47–61, 1996.

[41] P. Schröder, S.J. Gortler, M.F. Cohen, and P. Hanrahan. Wavelet projections for radiosity. *Computer Graphics Forum*, 13(2):141–151, 1994.

[42] P. Shirley. A ray-tracing method for illumination calculation in diffuse-specular scenes. In *Proc. Graphics Interface*, pages 205–212, 1990.

[43] P. Shirley. Discrepancy as a quality measure for sampling distributions. In *Eurographics '91*, pages 183–194. Elsevier Science Publishers, 1991.

[44] P. Shirley. Time complexity of Monte-Carlo radiosity. In *Eurographics '91*, pages 459–466. Elsevier Science Publishers, 1991.

[45] P. Shirley. Monte Carlo simulation and integration. *SIGGRAPH '93 Global Illumination Course Notes (42)*, 1993.

[46] P. Shirley, C. Wang, and K. Zimmerman. Monte Carlo techniques for direct lighting calculations. *ACM Transactions on Graphics*, 15(1):1–36, 1996.

[47] F. Sillion and Puech C. *Radiosity and Global Illumination*. Morgan Kaufmann Publishers, Inc., San Francisco, 1994.

[48] F. Sillion, G. Drettakis, and C. Soler. Clustering algorithm for radiance calculation in general environments. In *Rendering Techniques '95*, pages 197–205, 1995.

[49] F. X. Sillion, J. R. Arvo, S. H. Westin, and D. P. Greenberg. A global illumination solution for general reflectance distributions. *Computer Graphics (SIGGRAPH '91 Proceedings)*, 25(4):187–198, 1991.

[50] I. Sobol. *Die Monte-Carlo Methode*. Deutscher Verlag der Wissenschaften, 1991.

[51] M. Stamminger, Slussalek P., and H-P. Seidel. Three point clustering for radiance computations. In *Rendering Techniques '98*, pages 211–222, 1998.

[52] L. Szirmay-Kalos. Global ray-bundle tracing. Technical Report TR-186-2-98-18, Institute of Computer Graphics, Vienna University of Technology, 1998. www.cg.tuwien.ac.at/.

[53] L. Szirmay-Kalos. Stochastic iteration for non-diffuse global illumination. Technical Report TR-186-2-98-21, Institute of Computer Graphics, Vienna University of Technology, 1998. www.cg.tuwien.ac.at/.

[54] L. Szirmay-Kalos. Stochastic methods in global illumination — state of the art report. Technical Report TR-186-2-98-23, Institute of Computer Graphics, Vienna University of Technology, 1998. www.cg.tuwien.ac.at/.

[55] L. Szirmay-Kalos. Stochastic iteration for non-diffuse global illumination. *Computer Graphics Forum (Eurographics'99)*, 18(3), 1999.

[56] L. Szirmay-Kalos, B. Csébfalvi, and W. Purgathofer. Importance-driven quasi-Monte Carlo solution of the rendering equation. In *Winter School of Computer Graphics '98*, pages 377–386, Plzen, Czech Republic, 1998.

[57] L. Szirmay-Kalos, B. Csébfalvi, and W. Purgathofer. Importance driben quasi-random walk solution of the rendering equation. *Computers and Graphics*, 23(2):203–211, 1999.

[58] L. Szirmay-Kalos, P. Dornbach, and W. Purgathofer. On the start-up bias problem of metropolis sampling. In *Winter School of Computer Graphics '99*, pages 273–280, Plzen, Czech Republic, 1999.

[59] L. Szirmay-Kalos, T. Fóris, L. Neumann, and B. Csébfalvi. An analysis to quasi-Monte Carlo integration applied to the transillumination radiosity method. *Computer Graphics Forum (Eurographics'97)*, 16(3):271–281, 1997.

[60] L. Szirmay-Kalos, T. Fóris, and W. Purgathofer. Non-diffuse, random-walk radiosity algorithm with linear basis functions. *Machine Graphics and Vision*, 7(1):475–484, 1998.

[61] L. Szirmay-Kalos, T. Fóris, and W. Purgathofer. Quasi-Monte Carlo global ray-bundle tracing with infinite number of rays. In *Winter School of Computer Graphics '98*, pages 386–393, Plzen, Czech Republic, 1998.

[62] L. Szirmay-Kalos and W. Purgathofer. Global ray-bundle tracing with hardware acceleration. In *Rendering Techniques '98*, pages 247–258, 1998.

[63] L. Szirmay-Kalos and W. Purgathofer. Analysis of the quasi-monte carlo integration of the rendering equation. In *Winter School of Computer Graphics '99*, pages 281–288, Plzen, Czech Republic, 1999.

[64] L. Szirmay-Kalos and W. Purgathofer. Global ray-bundle tracing with infinite number of rays. *Computers and Graphics*, 23(2), 1999.

[65] E. Veach. *Robust Monte Carlo Methods for Light Transport Simulation*. PhD thesis, Stanford University, http://graphics.stanford.edu/papers/veach_thesis, 1997.

[66] E. Veach and L. Guibas. Optimally combining sampling techniques for Monte Carlo rendering. In *Rendering Techniques '94*, pages 147–162, 1994.

[67] E. Veach and L. Guibas. Bidirectional estimators for light transport. In *Computer Graphics (SIGGRAPH '95 Proceedings)*, pages 419–428, 1995.

[68] E. Veach and L. Guibas. Metropolis light transport. *Computer Graphics (SIGGRAPH '97 Proceedings)*, pages 65–76, 1997.

[69] J. R. Wallace, M. F. Cohen, and D. P. Greenberg. A two-pass solution to the rendering equation: A synthesis of ray tracing and radiosity methods. In *Computer Graphics (SIGGRAPH '87 Proceedings)*, pages 311–324, 1987.

[70] G. J. Ward. The RADIANCE lighting simulation and rendering system. *Computer Graphics*, 28(4):459–472, 1994.

[71] T. Warnock. Computational investigations of low-discrepancy point sets. In H. Niederreiter and P. Shiue, editors, *Monte-Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, pages 354–361. Springer, 1995.

[72] K. Zimmerman and P. Shirley. A two-pass solution to the rendering equation with a source visibility preprocess. In *Rendering Techniques '95*, pages 284–295, 1995.