Global Illumination of Glossy Environments using Wavelets and Importance

Per H. Christensen Eric J. Stollnitz* David H. Salesin Tony D. DeRose

Department of Computer Science and Engineering
**Department of Applied Mathematics*
University of Washington
Seattle, Washington 98195

November 3, 1994

Abstract

We show how importance-driven refinement and a wavelet basis can be combined to provide an efficient solution to the global illumination problem with glossy and diffuse reflections. Importance is used to focus the computation on the interactions having the greatest impact on the visible solution. Wavelets are used to provide an efficient representation of radiance, importance, and the transport operator. We discuss a number of choices that must be made when constructing an algorithm for a finite-element solution method for glossy global illumination. The main contributions of this paper are a four-dimensional wavelet representation for spatially- and angularly-varying radiance distributions (which allows for sparse representation and fast evaluation), the use of the standard wavelet decomposition of the transport operator, and the formulation of a type of importance suited for light transports using a finite element radiance representation. Our implementation supports curved surfaces and spatially-varying anisotropic BRDFs. We use a final gathering step to improve the visual quality of the solution.

1 Introduction

Radiosity algorithms assume that all reflection in a scene is ideally diffuse. This assumption, while making the computation of global illumination more tractable, ignores important effects such as glossy highlights whose intensity varies smoothly with direction. Though more expensive, the simulation of directional reflection is essential for realistic image synthesis. In this paper, we consider the *glossy global illumination* problem, whose goal is to find the equilibrium distribution of light in a scene with surfaces that are glossy reflectors. The glossy global illumination problem includes radiosity as a special case.

One promising approach to solving the glossy global illumination problem is to extend the finite-element method used in radiosity algorithms. Designing a finite-element algorithm for glossy global illumination involves a number of choices, summarized below. The result is an efficient algorithm, which we demonstrate using both simple and complex scenes.

The first choice is in the parameterization of the unknown light distribution. One possibility is to use radiance distributions, which are functions of surface position and direction [23, 34]. The alternative is to use two-point transport intensities, which are functions of two surface positions [2, 32]. We describe our motivation for using radiance distributions.

A second area of choice in designing a glossy global illumination algorithm is that of basis functions. One can use a single fixed resolution or a hierarchy of multiresolution basis functions. The benefits of a multiresolution representation are apparent from the radiosity algorithms presented by Hanrahan *et al.* [21] and Gortler *et al.* [20, 31]. If we choose a multiresolution basis for glossy global illumination, there are further choices as well: we can use scaling functions or wavelets; we can choose from many types of wavelets; we can construct "standard" or "non-standard" tensor products of basis functions; and we can use the standard or the non-standard operator decomposition. We have chosen to represent radiance in a basis consisting of four-dimensional non-standard tensor products of Haar wavelets. These basis functions interact to simulate light transport through a standard decomposition of the light transport operator.

As a third area of choice, one must decide whether or not a view-independent solution is necessary. A view-dependent solution can be computed more efficiently using *importance*, as shown by Smits *et al.* [37] for radiosity. Assuming we are interested in accelerating our solution procedure using importance, we must choose between two types of importance. We introduce a type of importance whose angular distributions are in general continuous functions, and therefore can be efficiently represented with finite elements such as wavelets. We use this type of importance since it satisfies the same transport equation as radiance, and can be represented and transported identically.

The last area of choice is in the rendering of the solution. A complete solution to the global illumination problem should be both physically accurate and visually pleasing. However, certain good approximations may have artifacts that are very obvious to the human eye. We therefore use a final gathering step [25, 30, 36] to improve the visual quality of the solution.

We have implemented an algorithm based on these choices. Adaptive numerical integration is used to compute accurate light transports where needed. The implementation also supports curved surfaces and spatially-varying anisotropic bidirectional reflectance distribution functions (BRDFs). Both emission and reflection can be described by texture maps.

A preliminary version of this paper was presented at the Eurographics Workshop on Rendering [8]. In this article, we give motivations for the choices we made, as well as comparisons with alternative approaches. New theoretical results here include a proof concerning exitant directional

importance, a different adaptive numerical integration method, and a more efficient final gathering step. New practical contributions are descriptions of our data structures, tests of convergence and convergence rates, and a comparison of the standard and non-standard matrix decompositions of the light transport matrix.

The remainder of this paper is organized as follows: Section 2 motivates the use of radiance, gives a formal description of radiance and light transport, and shows how radiance and the light transport equation can be projected into a finite-element basis. Section 3 introduces a wavelet basis for radiance distributions. Section 4 introduces a type of importance that is convenient for a finite-element represention of radiance. Section 5 describes our importance-driven global illumination algorithm, using a wavelet basis for directional radiance and importance distributions, and Section 6 provides significant details of our implementation. Finally, Section 7 describes our results, and Section 8 contains a conclusion and suggestions for future work. Appendix A contains comparisons of operator decompositions in flatland.

2 Finite elements for radiance

To determine the unknown light distribution in the scene exactly, we would have to find the amount of light leaving all points in all directions. For non-trivial scenes, computing an exact solution is impossible, so we instead compute an approximate solution represented by a weighted sum of a finite number of basis functions.

In this section, we first contrast two-point transport intensities against radiance distributions, then give a formal description of radiance, and describe how the continuous radiance function and transport operator can be discretized to facilitate representation and efficient transport.

2.1 Radiance distributions vs. two-point transport intensities

Two fundamentally different representations of the transported light have been used for glossy global illumination. Immel *et al.* [23] and Sillion *et al.* [34] represent the light in the scene as radiance distributions, a function of two spatial and two angular variables on each surface patch. By contrast, Aupperle and Hanrahan [2], Pattanaik and Bouatouch [28], and Schröder and Hanrahan [32] use a "two-point transport intensity," a function of four spatial variables, which represents the amount of light traveling from a point on a patch to a point on another.

We choose to represent light as a radiance distribution, like Immel *et al.* and Sillion *et al.*, for the following reasons: Assume that the scene is initially split into n patches. The coarsest possible representation of radiance requires only one basis function per patch for radiance distributions, as opposed to n basis functions per patch for two-point transport intensities. Therefore, the initial, very rough, solution of the light transport equation requires $O(n^2)$ interactions between radiance distributions on patches, as opposed to $O(n^3)$ interactions for matching two-point transport intensities.[1]

[1] With a clustering technique to group patches (similar to Sillion [33] and Smits *et al.* [35] for purely diffuse reflection), the initial number of clusters could potentially be very small even in a complex scene, and the reduction to $O(n^2)$ might not make much difference. However, such clustering methods have yet to emerge for scenes with glossy reflections.

2.2 Radiance

We now give a formal description of radiance and light transport. Let x and y be points in space, and let ω and ω_{xy} be directions (ω_{xy} is the direction from x to y, thus $\omega_{xy} = -\omega_{yx}$). The *radiance* $L(y, \omega)$ is defined as the power emanating from y, per unit solid angle in the direction ω, per unit projected area perpendicular to that direction. Radiance L is measured in [watt $\cdot$ meter^{-2} $\cdot$ steradian^{-1}].

The equilibrium distribution of radiance satisfies the following *light transport equation* [12]:

$$L(y,\omega) \;=\; L_e(y,\omega) + \int_x f_r(\omega_{xy}, y, \omega)\, G(x,y)\, L(x, \omega_{xy})\, dx\,. \tag{1}$$

This equation states that the radiance L from a point y in direction ω is the sum of two terms: *emitted radiance* L_e and radiance reflected from all other points x. An infinitesimal area around point x is written dx. The term $f_r(\omega_{xy}, x, \omega)$ is the *bidirectional reflectance distribution function*, or BRDF, and describes the ratio of reflected radiance (in direction ω) to the differential irradiance (from direction ω_{xy}) that causes it. The BRDF has units [steradian^{-1}]. As a consequence of Helmholtz reciprocity, the BRDF satisfies $f_r(-\omega', x, \omega) = f_r(-\omega, x, \omega')$. Finally, the *geometric term* $G(x, y)$ is given by

$$G(x,y) \;\equiv\; V(x,y) \cdot \frac{\cos\theta_x \cos\theta_y}{||x-y||^2}\,,$$

where $V(x, y)$ is a *visibility term* that is 1 or 0, depending on whether or not x and y are visible to one another, and θ_x and θ_y are the angles between the line segment xy and the respective normals of differential areas at x and y. The geometric term describes how radiance leaving a differential area at x in direction towards y arrives at y. The geometric term has units [steradian $\cdot$ meter^{-2}], and is symmetric in its arguments, $G(x, y) = G(y, x)$. Some of these terms are illustrated in Figure 1.

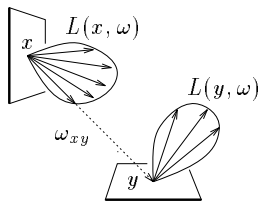


Figure 1 Light transport from point x to point y.

The light transport equation (1) can be rewritten in operator form as

$$L \;=\; L_e \;+\; \mathcal{T}L\,. \tag{2}$$

Here the *transport operator* $\mathcal{T}$ is defined by

$$(\mathcal{T}L)(y,\omega) \;\equiv\; \int_x f_r(\omega_{xy}, y, \omega)\, G(x,y)\, L(x, \omega_{xy})\, dx\,,$$

where $(\mathcal{T}L)(y, \omega)$ denotes the result of $\mathcal{T}$ operating on $L(x, \omega)$ to produce a function whose argument is (y, ω).

2.3 Discretization of radiance

Let $\mathbf{B}(x,\omega) = [b_1(x,\omega)\, b_2(x,\omega)\ \cdots\ ]$ be a basis for the space of radiance distributions. The unknown radiance distribution L can be projected onto the basis $\mathbf{B}$ by writing L as a series expansion,

$$L(x,\omega) \;=\; \sum_{i=1}^{\infty} \ell_i\, b_i(x,\omega)\,.$$

This equation can be written in matrix form as $L(x,\omega) = \mathbf{B}(x,\omega)\mathbf{L}$, where $\mathbf{L}$ is an infinite column matrix whose i-th entry is ℓ_i. When no confusion can arise, we suppress the arguments and simply write

$$L \;=\; \mathbf{BL}\,.$$

In the original formulation of radiosity, piecewise-constant functions were used as a basis for spatial variation [19]. In subsequent work on radiosity, Heckbert [22], Zatz [40], and Troutman and Max [38] used orthogonal polynomials, and Gortler *et al.* [20] used wavelets. In the more general context of radiance, the distribution of light leaving a patch has both spatial and angular variation. Immel *et al.* [23] used piecewise-constant basis functions for both spatial and angular variation. Later, Sillion *et al.* [34] used spherical harmonics for the angular variation and piecewise-constant basis functions for the spatial variation. In Section 3 we motivate and introduce our choice of basis, a hierarchical wavelet basis for both spatial and angular variation.

In order to project a radiance distribution onto the basis, we need an inner product and a dual basis. Let $\langle f \,|\, g \rangle$ denote the standard *inner product*, $\langle f \,|\, g \rangle \equiv \int_{\omega y} f(y,\omega)\, g(y,\omega)\, dy\, d\omega$. Let $[\langle \mathbf{F} \,|\, \mathbf{G} \rangle]$ be the outer product of $\mathbf{F}$ and $\mathbf{G}$, where each element of the outer product is the inner product of elements of $\mathbf{F}$ and $\mathbf{G}$. For example, if $\mathbf{F} = [f_1\ f_2\ \cdots\ ]$ and $\mathbf{G} = [g_1\ g_2\ \cdots\ ]$ are two row matrices of functions, then $[\langle \mathbf{F} \,|\, \mathbf{G} \rangle]$ is the matrix whose ij-th entry is $\langle f_i \,|\, g_j \rangle$. Likewise, $[\langle \mathbf{F} \,|\, g \rangle]$ is the row matrix consisting of elements $\langle f_1 \,|\, g \rangle$, $\langle f_2 \,|\, g \rangle$, ...

Let the *dual basis* associated with $\mathbf{B}$ be denoted $\overline{\mathbf{B}} = [\overline{b}_1(x,\omega)\, \overline{b}_2(x,\omega)\ \cdots\ ]$; each $\overline{b}$ is a linear combination of b's. The dual basis is characterized by the relation $\langle \overline{b}_i \,|\, b_j \rangle = \delta_{ij}$, or in matrix form $[\langle \overline{\mathbf{B}} \,|\, \mathbf{B} \rangle] = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Orthonormal bases are a special case: they are *self-dual*, meaning that $\overline{\mathbf{B}} = \mathbf{B}$.

2.4 Discrete light transport

Regardless of the choice of basis functions, we can obtain a system of equations for the unknown entries of $\mathbf{L}$ by substituting $L = \mathbf{BL}$ and $L_e = \mathbf{BL}_e$ into the light transport equation (2), and using linearity of the operator $\mathcal{T}$ to yield

$$\mathbf{BL} \;=\; \mathbf{BL}_e + \mathcal{T}(\mathbf{BL}) \;=\; \mathbf{BL}_e + (\mathcal{T}\mathbf{B})\mathbf{L}\,.$$

By applying the linear operator $[\langle \overline{\mathbf{B}} \,|\, \cdot \rangle]$ to both sides of this equation, we get

$$[\langle \overline{\mathbf{B}} \,|\, \mathbf{BL} \rangle] \;=\; [\langle \overline{\mathbf{B}} \,|\, \mathbf{BL}_e \rangle] + [\langle \overline{\mathbf{B}} \,|\, (\mathcal{T}\mathbf{B})\mathbf{L} \rangle]\,.$$

Using linearity and the duality relation, we arrive at the *discrete light transport equation*,

$$\mathbf{L} \;=\; \mathbf{L}_e + \mathbf{TL}\,. \tag{3}$$

In this infinite system of linear equations, $\mathbf{T} \equiv [\langle \overline{\mathbf{B}} \,|\, \mathcal{T}\mathbf{B} \rangle]$ is an infinite matrix representing the transport operator $\mathcal{T}$. The rs-th entry of $\mathbf{T}$ is a *transport coefficient*, representing the influence of the coefficient of b_s on the coefficient of b_r. It can be written explicitly as

$$\begin{aligned} T_{r \leftarrow s} &= \langle \overline{b}_r \,|\, \mathcal{T} b_s \rangle \\ &= \langle \overline{b}_r \,|\, \textstyle\int_x f_r(\omega_{xy}, y, \omega)\, G(x, y)\, b_s(x, \omega_{xy})\, dx \rangle \\ &= \displaystyle\int_{\omega y} \overline{b}_r(y, \omega) \int_x f_r(\omega_{xy}, y, \omega)\, G(x, y)\, b_s(x, \omega_{xy})\, dx\, dy\, d\omega\,, \end{aligned} \tag{4}$$

where the notation $r \leftarrow s$ is to emphasize that $T_{r \leftarrow s}$ represents the influence of the *sender* s on the *receiver* r. In this integral, the domain of x is the spatial support of the sending basis function b_s, the domain of y is the spatial support of the receiving basis function b_r, and the domain of ω is the angular support of b_r (directions on a hemisphere above y).

3 A wavelet basis for radiance

In this section we construct a multiresolution basis for efficiently representing radiance distributions. Results by Beylkin *et al.* [4, 5], Alpert [1], Gortler *et al.* [20], Hanrahan *et al.* [21] and others indicate that significant performance gains can be achieved by using a multiresolution basis.

By contrast, Immel *et al.* [23] used a single-resolution representation of radiance distributions using piecewise-constant basis functions. Sillion *et al.* [34] used spherical harmonics as basis functions for the angular variation of radiance. Their implementation also used a single-resolution representation: a fixed number of spherical harmonics interact at each light transport. One could envision multiresolution light transport using spherical harmonics, by only evaluating the most significant spherical harmonics (with largest coefficients). However, spherical harmonics have global support; that is, they are supported on the entire sphere, so all spherical harmonics interact throught light transports (except for occlusion). In contrast, the wavelets we will consider have compact support; a sending wavelet b_s will only interact with receiving wavelets b_r that have spatial support within the directional support of b_s. So for a wavelet basis, the compact directional support means that many of the transport coefficients are known a priori to be zero. Said differently, the transport matrix $\mathbf{T}$ is sparse in a wavelet basis, but dense in a spherical harmonics basis.

In the following, we first present some background on multiresolution analysis, and then describe one-dimensional wavelet bases and how they can be extended to the four-dimensional bases necessary for representing radiance distributions.

3.1 Multiresolution analysis

A straightforward method for solving the discrete light transport equation (3) approximately would represent the solution with a fixed, large number of basis functions, and transport light between all basis functions to compute the solution. If m is the number of basis functions, this method would require $O(m^2)$ interactions to compute a solution.

Instead, we use a hierarchical, or *multiresolution*, method, which results in only $O(m)$ or $O(m \log m)$ interactions (depending on the specific multiresolution method chosen). With this method, we first compute a very rough solution and then refine the representation and interactions based on that solution. After the refinement, an improved solution can be computed, new refinements can be performed, and so on. The multiresolution method exploits the facts that in some parts of the scene radiance distributions can be represented with sufficient accuracy using only a

few basis functions, and that even where the radiance distribution is represented with many basis functions, only few of these basis functions need to interact with most other basis functions.

Multiresolution analysis as formulated by Mallat [26] provides a convenient framework for studying multiresolution bases. More elaborate explanations of multiresolution bases are given by Chui [9], Daubechies [14], and DeRose *et al.* [15].

There are two basic ingredients for a multiresolution analysis: an infinite chain of nested linear function spaces $V^0 \subset V^1 \subset V^2 \subset \cdots$, and an inner product $\langle f \,|\, g \rangle$ defined on any pair of functions $f, g \in V^j$. The space V^j contains functions of resolution j, with resolution increasing as j increases. *Scaling functions* refer to bases for the spaces V^j.

A function can be approximated by a weighted sum of scaling functions. Alternatively, we can represent the same approximation as coarse scaling functions in V^0 along with detail at finer and finer resolution. Detail is accounted for by functions in the *orthogonal complement spaces* W^j defined by

$$W^j \equiv \{f \in V^{j+1} \mid \langle f \,|\, g \rangle = 0 \quad \forall g \in V^j\}.$$

Wavelets refer to bases for the orthogonal complement spaces W^j, and the spaces W^j are sometimes called *wavelet spaces*.

Orthogonal complements are often written as $V^{j+1} = V^j \oplus W^j$ since, intuitively, wavelet space W^j includes the functions that are in V^{j+1} but "missing" from V^j. More formally, any function $f^{j+1} \in V^{j+1}$ can be written uniquely as an orthogonal decomposition $f^{j+1} = f^j + f^j_\perp$, where $f^j \in V^j$ and $f^j_\perp \in W^j$. The space V^j can be fully decomposed as

$$V^j = V^0 \oplus W^0 \oplus \cdots \oplus W^{j-1}.$$

Therefore, a multiresolution basis for V^j can be formed by selecting a scaling function basis for V^0 together with wavelet bases for the spaces $W^0, ..., W^{j-1}$. The scaling functions spanning V^0 represent coarse variation, while the wavelets provide detail at increasing resolutions.

3.2 Choice of wavelet basis

The simplest multiresolution basis is the *Haar basis* in one dimension. The space V^j consists of piecewise-constant functions on $[0, 1]$ with discontinuities at $\{0, 1/2^j, 2/2^j, ..., 1\}$. The space V^j is spanned by piecewise-constant scaling functions $\phi^j_i(u)$, known as the Haar functions or "box" functions. The wavelet spaces W^j are spanned by piecewise-constant wavelets $\psi^j_i(u)$. A few Haar scaling functions and wavelets are shown in Figure 2.

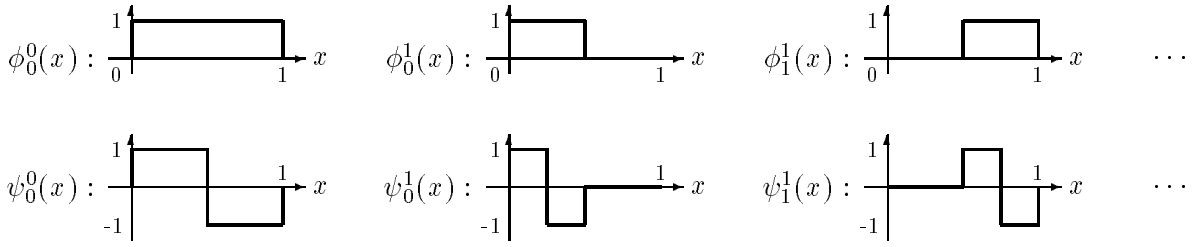


Figure 2 Haar scaling functions ϕ^j_i and wavelets ψ^j_i.

There are many alternatives to the Haar basis, each with advantages and disadvantages. Schröder and Hanrahan [32] have presented a comparison of wavelet bases for radiance, in which they exam-

ined rates of convergence, integration expense and accuracy, and the amount of work required to obtain a solution of a given accuracy.

The following list summarizes the properties desirable in a wavelet basis for global illumination:

- *Bounded domain.* Multiresolution analysis is usually formulated for functions on the unbounded real line, while we wish to represent radiance distributions defined on bounded domains. Unbounded wavelets could be used on a bounded interval by artificially extending the functions of interest [14], but it is unclear how radiance would be extended beyond the geometric extent of surface patches without introducing a number of artifacts. We have found it more convenient to use wavelets defined on a bounded interval, such as flatlets, multiwavelets [20], or B-spline wavelets adapted to a bounded interval [10, 15].

- *A single scaling function in V^0.* Some wavelet bases have more than one scaling function in space V^0. The high dimensionality of the radiance transport problem makes it advantageous to have only one scaling function in space V^0: A single scaling function leads to a single interaction between two patches at the coarsest level, while (as shown in Section 3.4) having two one-dimensional scaling functions leads to 16 four-dimensional scaling functions, requiring 256 interactions between two patches at the coarsest level.

- *Efficient numerical integration.* The finite-element method requires that we integrate the transport operator against scaling functions and wavelets (and their duals, if the basis is not orthonormal) to compute transport coefficients. It is advantageous to use basis functions for which we can develop inexpensive numerical integration formulas.

Among the wavelet bases that have bounded domains, there are both continuous and discontinuous choices. There are two families of bounded-interval continuous wavelets in the literature: Daubechies wavelets adapted to the bounded interval [11], and bounded-interval B-spline wavelets [10]. Having continuous basis functions on each patch is not sufficient to ensure a continuous solution; continuity must also be enforced across boundaries between adjacent patches, or the basis functions must be defined over complex shapes with arbitrary topology (for example, the floor in Figure 21). Both approaches seem unnecessarily complicated compared to the alternative, a discontinuous basis. Discontinuous wavelet bases that have been used for radiosity include the Haar basis, multiwavelets, and flatlets [20]. A final gathering step [25, 30, 36] can be used to smooth out discontinuities in the basis and at patch boundaries.

We have experimented with bounded-interval B-spline wavelets [7], Daubechies wavelets, and the Haar basis. As a result of these experiments, we have chosen to implement our glossy global illumination algorithm using the Haar basis. The Haar basis has many advantages, including a single scaling function in V^0, orthonormality (and therefore self-duality), compact support, and simple integration formulas. Flatlets, as an alternative, have more vanishing moments than the Haar basis, which increases the sparsity of a discretized smooth operator [1, 20]. But at the same time, flatlets have wider support and more costly integration formulas. Similarly, multiwavelets (which are constructed from higher-order polynomials) also require costly integration formulas. The main disadvantage of the Haar basis, its discontinuities, can be ameliorated by performing a final gathering step during rendering.

3.3 A convenient domain for radiance

Four-dimensional basis functions are required to represent radiance distributions: two variables describe spatial variation across a surface, and two variables describe angular variation. As is common, we split the surfaces into patches such that the spatial variables on each patch can be parameterized on the unit square $[0,1]^2$; see Figure 3. The domain of the radiance distribution on each path is then $[0,1]^2 \times H^2$, where H^2 is the unit hemisphere. By mapping H^2 onto $[0,1]^2$, we can use tensor products of one-dimensional basis functions for both angular and spatial variations. This process is described below.

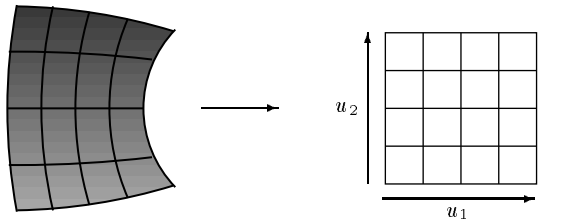


Figure 3 Spatial projection: mapping between 3D surface patch and 2D parameter space.

We use *gnomonic projection* to map between points in H^2 and points on a disc with radius $\pi/2$. As shown in Figure 4, gnomonic projection maps great circles through the pole of H^2 to radial lines, and preserves arc length along these curves. We use this map because it is easily computed and introduces only mild distortion. An alternative would be "flat" projection of H^2 onto a unit disc by simply ignoring the height component, but this projection results in points near the equator being mapped very densely near the circumference of the circle.

The gnomonic projection is followed by a radial "stretch" of the disc to exactly cover the unit square, also shown in Figure 4. The composition of these mappings is an invertible mapping between H^2 and the unit square.

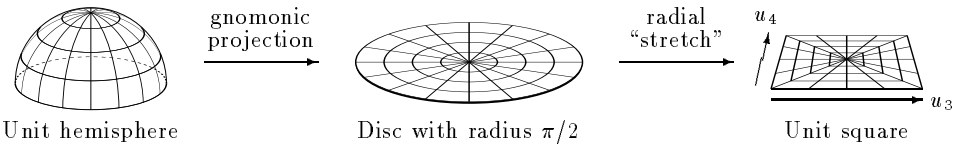


Figure 4 Angular projection: gnomonic projection and radial "stretch."

Figure 5 shows a typical radiance distribution (resulting from glossy reflection of light from a single point) before and after the angular projection. After the projection, the distribution is still continuous, but has a first-derivative discontinuity along the diagonals of the unit square.

3.4 A four-dimensional wavelet basis

We want to construct basis functions on the four-dimensional hypercube $[0,1]^4$ using tensor products of univariate basis functions. There are two alternative methods, the so-called "standard" and "non-standard" constructions [4]. The standard construction forms a basis from all possible tensor products of univariate basis functions. In the non-standard construction, on the other hand, each tensor product consists of univariate basis functions in the same space j (restricting the supports of

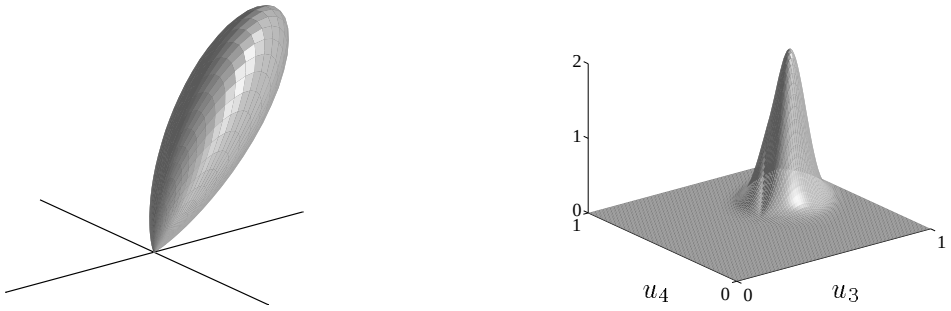


Figure 5 Radiance distribution before and after angular projection.

multivariate basis functions to be square for the Haar basis). The standard and non-standard basis constructions are illustrated in Figure 6 for the simpler case of two-dimensional basis functions. We choose the non-standard basis construction primarily because the required data structures are simpler (see Section 6.4).

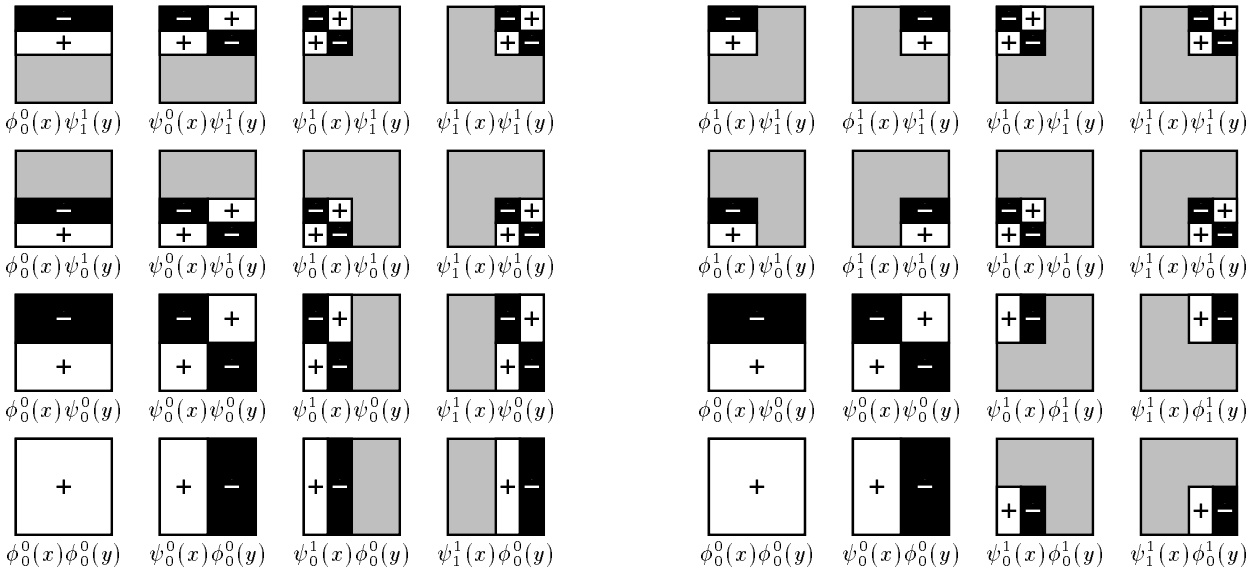


Figure 6 The standard and non-standard construction of a two-dimensional Haar wavelet basis for V^2. In the unnormalized case, functions are $+1$ where plus signs appear, -1 where minus signs appear, and 0 in gray regions.

Let $\mathbf{u} = (u_1, u_2, u_3, u_4)$ denote a point in $[0,1]^4$, and let $\mathbf{i} = (i_1, i_2, i_3, i_4)$ denote a four-component multi-index of integers. The four-dimensional scaling functions for V^j take the form

$$\phi\phi\phi\phi_{\mathbf{i}}^j(\mathbf{u}) \;\equiv\; \phi_{i_1}^j(u_1)\,\phi_{i_2}^j(u_2)\,\phi_{i_3}^j(u_3)\,\phi_{i_4}^j(u_4)\,.$$

That is, the scaling functions for resolution j consist of all possible products of the one-dimensional scaling functions for resolution j. The four-dimensional wavelets spanning the orthogonal complement W^j are formed by taking all other products of scaling functions and wavelets for resolution j. These wavelets consist of 15 types:

$$\phi\phi\phi\psi_{\mathbf{i}}^j(\mathbf{u}),\ \phi\phi\psi\phi_{\mathbf{i}}^j(\mathbf{u}),\ \phi\phi\psi\psi_{\mathbf{i}}^j(\mathbf{u}),\ \ldots\ ,\ \psi\psi\psi\psi_{\mathbf{i}}^j(\mathbf{u})\,.$$

We take as our basis $\mathbf{B}$ the set of basis functions spanning $V^0, W^0, W^1, \ldots$ for each patch in the scene.

The duals to each of the scaling functions and wavelets follow from the univariate duals since duals of products are products of duals. For example, $\overline{\phi\phi\psi\psi}^j_{\mathbf{i}}(\mathbf{u}) = \overline{\phi}^j_{i_1}(u_1)\,\overline{\phi}^j_{i_2}(u_2)\,\overline{\psi}^j_{i_3}(u_3)\,\overline{\psi}^j_{i_4}(u_4)$.

Gortler *et al.* [20] use a scaling function representation at all levels for radiosity. Their approach therefore requires "pushing" and "pulling" to distribute transported radiance to other levels of the hierarchy before the next transport iteration. By contrast, we use a wavelet representation, in which all basis functions are orthogonal to basis functions at other levels.

3.5 Transport matrix decomposition

Just as there are two different tensor-product constructions for multidimensional bases, so are there two ways to decompose a matrix: the so-called "standard" and "non-standard" decompositions [4]. Gortler *et al.* [20] and Schröder and Hanrahan [32] use the non-standard decomposition of the transport matrix, in which each transport coefficient relates two basis functions at the same level of the hierarchy. These transports are followed by "pushing" and "pulling" operations [20]. By contrast, we use the standard decomposition of the operator matrix, in which transport coefficients include interactions between basis functions at different levels of the hierarchy. Note that a single such interaction would require many transports in the non-standard decomposition, all of which are between basis functions at a single level of the hierarchy. Since we use a wavelet basis and a standard transport matrix decomposition, "pushing" and "pulling" procedures are unnecessary.

For a smooth operator, the non-standard operator decomposition has a theoretical growth of $O(m)$, while a standard operator decomposition has theoretical growth $O(m \log m)$, where m is the number of basis functions. However, because of occlusion, the radiance transport operator is only piecewise-smooth. Our experiments (described in Appendix A) show that the flexibility of the standard decomposition can in some cases cause a significant reduction in the number of interactions. These results indicate that in practice the standard decomposition is often more sparse than the non-standard decomposition. Similar conclusions were reached by Schröder *et al.* [31] for radiosity, and by Jaffard and Laurençot [24] for more general operators. For these reasons, we use the standard decomposition of the operator matrix.

4 Importance for glossy scenes

In order to maintain a tractably small problem for complex scenes, we use importance-driven refinement to compute a view-dependent solution. In this section, we introduce a type of importance that satisfies the same equilibrium equation as radiance. Radiance and importance can be used together to compute a solution, as described in Section 5.

Smits *et al.* [37] showed that diffuse importance gives a substantial speed-up for a complex diffuse scene. For glossy reflections, the gain in speed is even greater, due to the directionality of radiance and importance: a directional interaction is refined only if the radiance *in that direction* is sufficiently large, inaccurate, and important.

4.1 Incident and exitant directional importance

Smits *et al.* [37] define the "importance" at a point to be the fraction of light leaving that point that reaches the eye. Here we show that a slightly different definition is advantageous for the finite-element solution of glossy global illumination, since it allows importance to be represented in

the same manner as radiance, and makes importance satisfy the same transport equation as light. The form of diffuse importance defined by Smits *et al.* [37] for radiosity is an incident quantity similar to irradiance. For radiosity, storing and transporting incident importance is equivalent (up to a constant) to storing and transporting exitant importance, since the radiosity leaving a point is proportional to the irradiance received at that point.

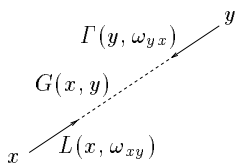
The most direct generalization of this quantity to radiance is an *incident* type of directional importance. Incident directional importance can be defined as the fraction of radiance that reaches the eye. This type of importance was used by Pattanaik and Mudur [27] for a Monte Carlo solution method and Aupperle *et al.* [3] for a finite-element solution method. Incident directional importance is represented and transported like differential irradiance.

One problem with incident directional importance is that it cannot always be represented efficiently, since its directional distribution at a point may have discontinuities, as in the case of light coming from a silhouette edge. Such discontinuities are difficult to represent with finite elements, resulting in slow convergence rates for adaptive refinement. For this reason, we define a type of *exitant* directional importance (which we will call *importance* for short). For glossy reflection, this exitant type of importance is a continuous function of direction at any surface point. In addition, as preliminary results in Christensen *et al.* [6] showed, using an exitant formulation for importance also makes it simple to develop an importance-driven glossy global illumination algorithm, since it allows light and importance to be transported identically.

4.2 Importance transport

Christensen *et al.* [6] showed that importance can be considered an exitant quantity like radiance and can be transported like light, thus simplifying a finite element representation. That presentation was based on adjoints. Here we present a simpler and more intuitive (but equivalent) definition and proof. With foresight, we define importance and emitted importance as follows:

Definition *Importance* $\Gamma(y, \omega_{yx})$ is the fraction of $G(x,y)\,L(x,\omega_{xy})$ that reaches the eye.



Light contributes directly to the image if it reaches the eye from one of the directions in the viewing pyramid. We can weight the light by a distribution of *emitted importance* Γ_e at the eye:

Definition

$$\Gamma_e(y,\omega) \equiv \begin{cases} 1, & \text{if } y \text{ is on the eye patch and } \omega \text{ is a direction within the viewing pyramid} \\ 0, & \text{elsewhere} \end{cases}$$

Lemma *$(\mathcal{T}^n\Gamma_e)(y,\omega_{yx})$ is the fraction of $G(x,y)\,L(x,\omega_{xy})$ that reaches the eye through exactly n bounces.*

Proof By induction over the number of bounces the radiance goes through before reaching the eye.

Basis: $\Gamma_e(y, \omega_{yx})$ is the fraction of $G(x,y)\,L(x,\omega_{xy})$ that reaches the eye directly; the fraction is zero if y is not on the eye patch or ω_{yx} is not within the viewing pyramid.

Inductive step: By the inductive hypothesis, $(\mathcal{T}^{n-1}\Gamma_e)(z,\omega_{zy})$ is the fraction of $G(y,z)\,L(y,\omega_{yz})$ that reaches the eye through exactly $n-1$ bounces. A single bounce of radiance $L(x,\omega_{xy})$ results in a radiance distribution $f_r(\omega_{xy},y,\cdot)\,G(x,y)\,L(x,\omega_{xy})$ at y; see the left illustration in Figure 7.

The amount of $L(x,\omega_{xy})$ that reaches the eye through exactly n bounces is the integral over all possible paths involving $(n-1)+1$ bounces:

$$\begin{aligned}
& \int_z \Big[(\mathcal{T}^{n-1}\Gamma_e)(z,\omega_{zy})\Big]\, G(y,z)\, \Big[f_r(\omega_{xy},y,\omega_{yz})\, G(x,y)\, L(x,\omega_{xy})\Big]\, dz \\
= & \left[\int_z (\mathcal{T}^{n-1}\Gamma_e)(z,\omega_{zy})\, G(y,z)\, f_r(\omega_{xy},y,\omega_{yz})\, dz\right] G(x,y)\, L(x,\omega_{xy}) \\
= & \left[\int_z f_r(\omega_{zy},y,\omega_{yx})\, G(z,y)\, (\mathcal{T}^{n-1}\Gamma_e)(z,\omega_{zy})\, dz\right] G(x,y)\, L(x,\omega_{xy}) \\
= & \; (\mathcal{T}\mathcal{T}^{n-1}\Gamma_e)(y,\omega_{yx})\, G(x,y)\, L(x,\omega_{xy}) \\
= & \; (\mathcal{T}^{n}\Gamma_e)(y,\omega_{yx})\, G(x,y)\, L(x,\omega_{xy})
\end{aligned}$$

The last expression is illustrated in the right part of Figure 7. The conclusion of these manipulations is that $(\mathcal{T}^n\Gamma_e)(y,\omega_{yx})$ is the fraction of $G(x,y)\,L(x,\omega_{xy})$ that reaches the eye through exactly n bounces.

□

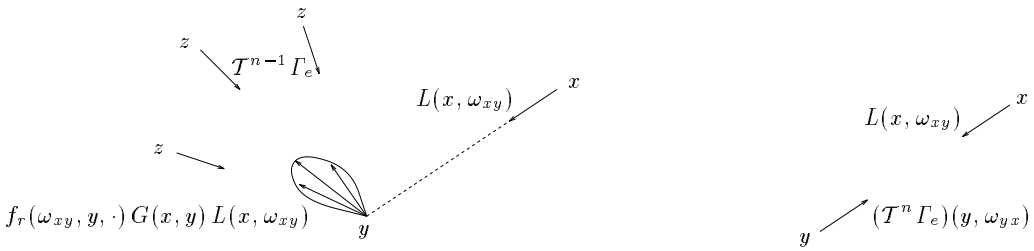


Figure 7 Two ways of computing the amount of radiance $L(x,\omega_{xy})$ that reaches the eye through exactly n bounces: The integral over all points z on the left is equivalent to the simple product on the right.

Lemma *The sum $\sum_{i=0}^{\infty}\mathcal{T}^i\Gamma_e$ satisfies our definition of importance Γ.*

Proof The total fraction of $G(x,y)\,L(x,\omega_{xy})$ that reaches the eye (through any number of bounces) is $\sum_{i=0}^{\infty}(\mathcal{T}^i\Gamma_e)(y,\omega_{yx})$.

□

Theorem *Importance Γ satisfies an equilibrium equation with the same transport operator as radiance, namely*

$$\Gamma \;=\; \Gamma_e + \mathcal{T}\Gamma\,.$$

Proof Assuming that reflections are energy dissipating, the norm of $\mathcal{T}$ is less than one and therefore $\mathcal{I}-\mathcal{T}$ is invertible (here $\mathcal{I}$ is the identity operator). The importance can then be rewritten using the Neumann series as $\Gamma = \sum_{i=0}^{\infty}\mathcal{T}^i\Gamma_e = (\mathcal{I}-\mathcal{T})^{-1}\Gamma_e$. Operating on both sides with $\mathcal{I}-\mathcal{T}$ gives $(\mathcal{I}-\mathcal{T})\Gamma = \Gamma_e$; the theorem follows directly.

□

Since (exitant directional) importance satisfies the same transport equation as radiance, it can be discretized like radiance and transported using the same transport coefficients. The *discrete importance transport equation* is

$$\mathbf{\Gamma} = \mathbf{\Gamma}_e + \mathbf{T}\mathbf{\Gamma} .$$

The only difference from radiance is that while radiance is emitted by light sources, importance is emitted by the eye.

5 Algorithm

Our solution method for radiance transport makes use of a wavelet representation and importance-driven refinement. The algorithm computes a view-dependent solution to the radiance equation; that is, the solution is refined most in the areas that contribute most to the image. In some respects, our algorithm is similar to the approach described by Gortler *et al.* [20] for wavelet radiosity. However, there are a number of areas, in addition to the higher dimensionality, in which our algorithm differs significantly from this previous work.

In this section, we first present the main algorithm and then discuss transport coefficients: how they are computed, which ones are computed as the refinement proceeds, and how their accuracy can be adaptively increased at small cost. Last, we describe our use of a final gathering step to generate smooth solutions with accurate shadows and textures.

5.1 Main algorithm

The primary task is to solve two systems of linear equations, one for radiance and one for importance:

$$\mathbf{L} = \mathbf{L}_e + \mathbf{T}\mathbf{L} \qquad \text{and} \qquad \mathbf{\Gamma} = \mathbf{\Gamma}_e + \mathbf{T}\mathbf{\Gamma} .$$

We first compute a small number of entries of the matrix $\mathbf{T}$ and solve the equations, then compute more entries of $\mathbf{T}$ and solve again, and so on. The high dimensionality of the global illumination problem makes the entries of $\mathbf{T}$ very expensive to compute, so we strive to compute as few of these entries as possible while generating a good approximation to the solution. Put briefly, only entries of $\mathbf{T}$ that are estimated to be large — and that connect large and important basis function coefficients — are computed.

The main part of the algorithm alternates between computing approximate radiance and importance solutions $\widetilde{\mathbf{L}}$ and $\widetilde{\mathbf{\Gamma}}$ and improving the finite representation of the transport operator $\widetilde{\mathbf{T}}$. Quantities with a tilde are approximate, both because they are computed numerically and because they are truncated versions of infinite matrices. Initially, we project L_e and Γ_e into space V^0, the space spanned by the coarsest-level scaling functions, to give $\widetilde{\mathbf{L}}_e$ and $\widetilde{\mathbf{\Gamma}}_e$. We also compute the entries of $\mathbf{T}$ corresponding to interactions of scaling functions in V^0 with one another (as described in Section 5.2), giving $\widetilde{\mathbf{T}}$. The algorithm is given in pseudocode below:

procedure *GlossyGlobalIllumination* ($\widetilde{\mathbf{T}}$, $\widetilde{\mathbf{L}}_e$, $\widetilde{\mathbf{\Gamma}}_e$):
    $\widetilde{\mathbf{L}} \leftarrow \widetilde{\mathbf{L}}_e$
    $\widetilde{\mathbf{\Gamma}} \leftarrow \widetilde{\mathbf{\Gamma}}_e$
    repeat
        $\widetilde{\mathbf{L}} \leftarrow \mathit{Solve}(\widetilde{\mathbf{T}}, \widetilde{\mathbf{L}}, \widetilde{\mathbf{L}}_e)$
        $\widetilde{\mathbf{\Gamma}} \leftarrow \mathit{Solve}(\widetilde{\mathbf{T}}, \widetilde{\mathbf{\Gamma}}, \widetilde{\mathbf{\Gamma}}_e)$
        $\widetilde{\mathbf{T}} \leftarrow \mathit{Refine}(\widetilde{\mathbf{T}}, \widetilde{\mathbf{L}}, \widetilde{\mathbf{\Gamma}})$
    until visual convergence of $\widetilde{\mathbf{L}}$
end procedure

The radiance and importance systems are solved simultaneously, with the solution in one system determining the refinements in the other system. Importance is used to refine the radiance solution only in areas that are significant to the final image. Likewise, radiance is used to refine the importance solution only in bright parts of the scene. The main loop iterates until *visual convergence* is achieved, that is, until further refinement does not significantly change the computed image. We use Gauss-Seidel iteration [18] to solve the approximate transport equations $\widetilde{\mathbf{L}} = \widetilde{\mathbf{L}}_e + \widetilde{\mathbf{T}}\widetilde{\mathbf{L}}$ and $\widetilde{\mathbf{\Gamma}} = \widetilde{\mathbf{\Gamma}}_e + \widetilde{\mathbf{T}}\widetilde{\mathbf{\Gamma}}$. Refinement is determined by an "oracle," described in Section 5.3.

5.2 Computing transport coefficients

The algorithm above requires computation of transport coefficients between basis functions. Each transport coefficient is defined in Equation (4) as a six-dimensional integral, which we approximate using numerical integration. Four-dimensional numerical integration formulas for wavelet radiosity are discussed by Gortler *et al.* [20].

The transport coefficients $T_{r \leftarrow s}$ are computed as inner products. For example, the influence of wavelet $\psi\phi\psi\phi^j_{\mathbf{i}}(\mathbf{u}_s)$ on wavelet $\psi\phi\phi\phi^{j'}_{\mathbf{i}'}(\mathbf{u}_r)$ is $T_{r \leftarrow s} = \langle \overline{\psi\phi\phi\phi}^{j'}_{\mathbf{i}'} \mid \mathcal{T}\psi\phi\psi\phi^j_{\mathbf{i}} \rangle$. The domain of radiance is position and direction, for example x and ω, while the domain of our tensor-product basis functions is the four-dimensional hypercube $[0,1]^4$. For convenience, we will make the spatial and angular transformations implicit, and write the basis functions as functions of points and directions: let the sending position x correspond to the two parameters u_1 and u_2, and let the direction ω_{xy} correspond to parameters u_3 and u_4 (and similarly for the parameters y and ω of the receiving basis function). Then the inner product in our example takes the form

$$\begin{aligned} T_{r \leftarrow s} &= \langle \overline{\psi\phi\phi\phi}^{j'}_{\mathbf{i}'} \mid \mathcal{T}\psi\phi\psi\phi^j_{\mathbf{i}} \rangle \\ &= \int_{\omega y} \overline{\psi\phi\phi\phi}^{j'}_{\mathbf{i}'}(y,\omega) \int_x f_r(\omega_{xy}, y, \omega)\, G(x,y)\, \psi\phi\psi\phi^j_{\mathbf{i}}(x, \omega_{xy})\, dx\, dy\, d\omega \\ &= \int_{xy} \left[\int_\omega \overline{\psi\phi\phi\phi}^{j'}_{\mathbf{i}'}(y,\omega)\, f_r(\omega_{xy}, y, \omega)\, d\omega \right] G(x,y)\, \psi\phi\psi\phi^j_{\mathbf{i}}(x, \omega_{xy})\, dy\, dx\,. \end{aligned} \tag{5}$$

Note that only the BRDF and the receiving basis function depend on ω. Our numerical integration routine samples these two functions in its innermost loop, while the remaining functions are evaluated only as the positional variables change.

We approximate integrals such as (5) using slightly jittered uniform sampling of the integrand. More accurate rules such as Gauss-Legendre or Gauss-Kronrod quadrature could be employed instead [20, 29, 40].

5.3 Refinement

In the algorithm, the approximate transport matrix $\widetilde{\mathbf{T}}$ is progressively refined. Here we describe how entries of $\widetilde{\mathbf{T}}$ are selected for computation.

In many applications of wavelets in numerical analysis [4], the goal is to obtain a sparse representation of a given matrix, thereby making repeated matrix–vector multiplications much faster. The wavelet decomposition of the matrix is done once and for all as a preprocess, so the cost of computing all the matrix elements is amortized by many fast matrix multiplications. In wavelet-based approaches to global illumination, the cost of explicitly constructing an entire transport matrix far outweighs the expense of any matrix–vector multiplications that follow. Therefore, it is essential to restrict the number of computed transport coefficients.

The goal of the *refinement oracle* is to determine which of the entries of $\mathbf{T}$ missing from $\widetilde{\mathbf{T}}$ should be computed to reduce the visible error in the current radiance solution. The two most important sources of error are:

- *truncation error* due to significant entries missing from $\widetilde{\mathbf{T}}$, and
- *integration error* in computing the entries of $\widetilde{\mathbf{T}}$.

In this section we describe how our oracle reduces truncation error. Section 5.4 outlines a method for simultaneously reducing integration errors.

The refinement oracle uses concepts from the brightness refinement criterion for hierarchical radiosity [21], the oracle used by Gortler *et al.* for wavelet radiosity [20], and the importance-based refinement strategy used by Smits *et al.* [37]. The idea is to estimate the influence on the visible image that would result if a new transport coefficient were to be added to $\widetilde{\mathbf{T}}$. If this quantity falls below some threshold, the expensive computation of the transport coefficient can be avoided without resulting in significant error in the solution.

Consider two basis functions b_s and b_r with no transport coefficient between them (yet), see Figure 8. We compute a new transport coefficient $\widetilde{T}_{r \leftarrow s}$ if a sufficiently large value results from the product of

- *radiance*: the magnitude of the sending basis function coefficient $\widetilde{\ell}_s$,
- *estimated transport coefficient*: the estimated new transport coefficient $\widetilde{T}_{r \leftarrow s}$ between the basis functions, and
- *importance*: the fraction $G\widetilde{I}$ of the receiving basis function that reaches the eye.

The product of the first two quantities estimates the amount of light transported between the two basis functions. Multiplying by the importance shining onto the receiving basis function (and the geometric term) gives the contribution to the final image. The sending basis function coefficient and approximate importance of the receiving basis function are known from the interim solution. Our estimate of the transport coefficient uses kernel variation, as explained shortly.

There are infinitely many new transport coefficients to be considered for computation. We need a scheme for considering only some of them in each iteration, while making it possible to eventually consider all. Associate with each basis function b (except for the basis functions in V^0 and W^0) a unique "parent" basis function b' that overlaps b and is in a space one level coarser. Also, let the parent of the wavelets in space W^0 be the scaling function in V^0. (For example, for the simple case

of the one-dimensional Haar wavelet basis, the parent of ψ_i^j is $\psi_{\lfloor i/2 \rfloor}^{j-1}$ and the parent of ψ_0^0 is ϕ_0^0.) In our implementation, we only consider computing a new transport coefficient $\widetilde{T}_{r \leftarrow s}$ if there is already a transport coefficient $\widetilde{T}_{r \leftarrow s'}$ or $\widetilde{T}_{r' \leftarrow s}$. The kernel variation encountered when computing $\widetilde{T}_{r \leftarrow s'}$ or $\widetilde{T}_{r' \leftarrow s}$ is used as a rough estimate of the transport coefficient $\widetilde{T}_{r \leftarrow s}$ for deciding whether to actually compute that transport coefficient. (The kernel variation is stored along with each transport coefficient.)

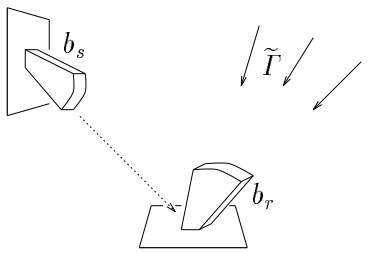


Figure 8 Sending and receiving basis functions

Since the kernel variation is the same for all fifteen transport coefficients from a given basis function to the fifteen wavelets sharing the same support, the estimated contribution to the image will be the same for all these fifteen wavelets. We therefore compute all fifteen transport coefficients at once. This approach saves many kernel-evaluations since the same kernel-evaluations can be re-used in computing all 15 transport coefficients. (At the same time, the transport coefficient between the two scaling functions that share support with the sending and receiving basis functions is computed. This transport coefficient is not used for light transports, but for adaptive improvements of other transport coefficients, as described in Section 5.4.)

For each call to the refinement procedure, the maximum product of radiance, kernel variation, and importance for all potential new transport coefficients is computed. Then all new transport coefficients with a product larger than some fraction (for example, 10%) of the maximum product are computed. This method requires two passes through all potential new transport coefficients.

5.4 Adaptive numerical integration

If we always use a numerical integration rule of high accuracy to compute transport coefficients, time is wasted evaluating the kernel for many interactions that have little effect on the final image. On the other hand, the significant transport coefficients must be computed to high precision; otherwise, the solution will not converge to the correct value. It is therefore advantageous to use an adaptive numerical integration technique that reduces error in transport coefficients, particularly on transport coefficients that are refined by the oracle. We have implemented such an adaptive integration as part of the refinement procedure.

At the time a transport coefficient is computed, a numerical integration method as described in Section 5.2 is used. Later, if more detailed transport coefficients are computed between the same supports, the kernel is sampled more densely. These samples are re-used to recompute the coarse transport coefficient more accurately.

As described in Section 5.3, the transport coefficient between two scaling functions is computed at the same time as the transport coefficients between other basis functions with the same sup-

port (at practically no extra cost, since the necessary kernel-evaluations have been done already). Since wavelets in a certain space can be expressed as a linear combination of scaling functions in higher spaces, coarse-level transport coefficients between wavelets can be recomputed by taking linear combinations of the transport coefficients between finer-level scaling functions. In this way, transport coefficients are adaptively recomputed where the kernel is sampled densely.

5.5 Final gather

For rendering the solution, we can evaluate the finite-element representation of the solution, or we can perform a rendering step that improves the visual quality of the solution. Following the ideas that Reichert [30], Lischinski *et al.* [25], and Smits [36] used for radiosity, we have implemented a *final radiance gathering* step. For each pixel in the final image, we perform a final gathering of light to the surface point that corresponds to the midpoint of the pixel. Call this point y.

We have tried three different final gathering methods. A complete final gather method gathers light from all basis functions in the solution. A faster final gather method gathers light from all basis functions that have transport coefficients to a basis function with support at the point y and in the direction of the eye e. This method only excludes light transports that have not been considered important in the solution. An even faster method gathers light only from basis functions that have a transport coefficient to the scaling function on the patch that point y is on. This fast method is nearly as accurate since if a transport coefficient is computed between two fine-level wavelets, there is a high probability that there is also a transport coefficient from the sending wavelet to the scaling function (since the scaling function has large support, and the kernel variation is usually high across large geometric extents). We have tried all three methods and have found no significant degradation in the final image from using the fastest final gather method.

For each basis function that we want to do a final gather from, we evaluate a simplified version of the integral in Equation (5). For example, the final gather from the wavelet $\psi\phi\psi\phi^j_{\mathbf{i}}(x,\omega)$ requires evaluating

$$\int_x f_r(\omega_{xy}, y, \omega_{ye})\, G(x,y)\, \psi\phi\psi\phi^j_{\mathbf{i}}(x,\omega_{xy})\, dx\ .$$

Since the receiving position y is fixed and the radiance is reflected towards the eye e, the integration is only over sending positions x.

Formally, this final gather corresponds to changing to a piecewise-constant basis, where the support of each basis function is the projection of a pixel onto a surface in the scene. Intuitively, this basis is tailored to be visually pleasing. The final gather smooths the discontinuities in the wavelet representation, and makes highlights, textures, and shadows crisper. The improvement brought about by the final gather can be seen by comparing Figures 21(e) and (f). The convergence improvement from final gather is described in Section 7.1.

Another way of thinking about the final gathering step is in the context of distribution ray tracing [13]. When a ray cast from the eye intersects a surface in the scene, a group of reflected rays are traced from the intersection point to points on other surfaces in the scene. A constant number of rays are cast to the support of each selected basis function in the radiance solution. In this way, the directions of the rays are guided by the solution. Thus, the most refined areas of the radiance solution are sampled the most by the distribution of reflected rays. Note that the costly "explosion" of the number of recursive bounces used in distribution ray tracing is avoided, since only a single bounce is followed. Once a finite element radiance solution has been computed, the final gather requires no additional memory.

6 Implementation

In this section, we describe features of our implementation, including the data structures for basis function coefficients and transport coefficients.

6.1 Surface geometry

To represent a given complex curved object, fewer spline patches are required than quadrilaterals. Since our method is independent of the geometry of the patches, as long as they can be parameterized on a unit square, it is natural to incorporate curved surfaces. In our implementation, a patch can be either a Bézier surface or a quadrilateral. It would also be straightforward to add non-uniform rational B-spline surfaces [16]. In fact, the only requirements are that we need to be able to quickly compute a position, surface normal, and differential area associated with a given parametric point (u_1, u_2), and be able to determine the intersection of a ray with the surface. The images in Figure 18 show a teapot consisting of 28 Bézier patches.

6.2 Reflection models and texture maps

We use the Ward isotropic and anisotropic reflection models [39] since they are physically valid and fast to evaluate. Examples of this reflection model can be seen in Figure 18. In addition to angular variation in reflectance, we use spatially varying reflectances to simulate details of the materials in the scene. We take the BRDF to be the tensor product of a spatially varying texture and the angular variation of the Ward model. Figure 21 demonstrates both texture (on the floor, walls, and pedestal) and an anisotropic reflectance function (on the teapot).

In the course of numerically approximating a transport coefficient, the geometric term and the BRDF are sampled at a number of quadrature points. The reflectance for each quadrature point is determined by a look-up in a texture map, multiplied by the angular variation given by Ward's model. Multiresolution textures could be incorporated in our method by using a pyramid of texture averages instead of sampling. This approach would eliminate sampling errors from sampling the texture. Gershbein *et al.* [17] present an alternative approach, using wavelet decompositions of textures for radiosity.

6.3 Light sources

By storing the wavelet decomposition of an image as the initial coefficients on a patch, we can model a light source that emits a spatially-varying radiance (such as a television screen). In general, not all coefficients of the emitting image will have links from them, but the coefficients are ready to be transported into the scene if the refinement procedure so decides. This technique allows a complex environment to be displayed using simple geometry.

A simple approach to angular variation is to let the emission depend on direction. For example, we model "spotlights" using a Phong-like function, in which emission depends on some power of the cosine of the angle between the emission direction and the surface normal of the patch. The spotlights appear dark from most directions because of the very narrow distribution of light they emit.

We demonstrate the use of spotlights and a spatially-varying emitter (the outdoor environment seen through the window) in Figure 21. More complex effects such as a slide projector or sunlight through a stained-glass window could be modeled by combining spatial and angular variations in an emitter.

6.4 Data structure for basis function coefficients

As in previous hierarchical radiosity algorithms [12], the matrices $\widetilde{\mathbf{T}}$, $\widetilde{\mathbf{L}}$, $\widetilde{\mathbf{L}}_e$, $\widetilde{\mathbf{\Gamma}}$, and $\widetilde{\mathbf{\Gamma}}_e$ are never formed explicitly. Entries of $\widetilde{\mathbf{L}}$, $\widetilde{\mathbf{L}}_e$, $\widetilde{\mathbf{\Gamma}}$, and $\widetilde{\mathbf{\Gamma}}_e$ are associated with the surface patches, while entries of $\widetilde{\mathbf{T}}$ are stored as "links" between radiance (and importance) coefficients. The coefficients and links are allocated dynamically as the solution is refined.

A hierarchy of basis function coefficients is associated with each patch. We have implemented the hierarchy as a tree where each node contains all coefficients ℓ with the same indices (space j and translations $i_1, \ldots, i_4$). Each node contains the 15 wavelet coefficients for each of six "color bands": red, green, and blue radiance and red, green, and blue importance. In addition, there is a scaling function coefficient within each root node (root nodes correspond to spaces V^0 and W^0). Each node contains 16 pointers to child nodes that contain the coefficients in the next (more refined) space, nodes corresponding to coefficients with all 16 combinations of indices $2i_1$ or $2i_1 + 1$, ..., $2i_4$ or $2i_4 + 1$. The pointers between nodes are illustrated in Figure 9. Initially, each patch only has a single node, containing a scaling function coefficient in space V^0 for each color band.

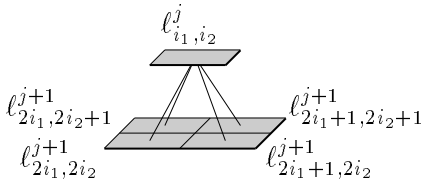


Figure 9 Tree of basis function coefficients on a patch (simplified to 2D flatland, where each node only has 4 children). Each node contains all coefficients of wavelets with the same support, and the root node also contains scaling function coefficients.

6.5 Data structure for transport coefficients

The transport coefficients that describe the interaction between radiance and importance basis functions at different patches are stored on *links*. As described in Section 5.3, the transport coefficients from a sending basis function to all 15 wavelets sharing support (and the single scaling function) are computed at the same time. In our implementation, these 16 transport coefficients are stored on the same link. Alternatively, each transport coefficient could be stored on a separate link, but the extra storage overhead makes this slightly infeasible, since each link needs to point to the two basis functions between which it is transporting light. As another alternative, the transport coefficients between all 15^2 possible combinations of wavelets on sender and receiver could be stored on the same link. This method also wastes memory, since it sets up links with room for many transport coefficients that are never computed, for example, because the sending coefficient is too low.

Each link contains a pointer to the node from which it is transporting radiance (and importance), information about what type of basis function it is transporting from, 15 entries of $\widetilde{\mathbf{T}}$ for each of the three color bands, the sample variation encountered while computing those transport coefficients (used for refinement as described in Section 5.3), a scaling function to scaling function transport coefficient for each of the three color bands (used for adaptive improvement of transport coefficients as described in Section 5.4), and a pointer to the next link to the same receiving basis functions. Note that there can be several links between the same pair of nodes, each connecting different basis functions with the same support, as illustrated in Figure 10. All links pointing to a given node

are organized in a linked list of links. Initially links are set up between root nodes for all pairs of patches that are mutually visible.

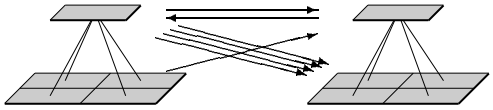


Figure 10 Example of links between basis functions (in flatland).

We only consider creating new links between basis functions b_s and b_r where there is a link from b_s to b_r's parent, or there is a link from b_s's parent to b_r; see Figure 11. This restriction reduces the number of new links that have to be considered for refinement at one time, while still allowing all possible links to be created eventually. The existing link contains information about the kernel variation encountered while computing that transport coefficient; this variation is used as an estimate of the (yet uncomputed) transport coefficient to or from a child basis function. The one exception to this scheme is root nodes, since they have no parent. Here the link between the two scaling functions is used for information about kernel variation for the wavelets in W^0.

Figure 11 New links to be considered (in flatland).

Links are never destroyed in our algorithm. By contrast, the approach described by Gortler *et al.* [20] removes a link at one level of the hierarchy and replaces it with multiple links at a finer level of detail (because they use a scaling function representation at all levels of detail).

The adaptive numerical integration takes place after new links have been set up. It is a bottom-up traversal of all links. For each link, it is checked whether the two nodes it connects have children, and if so, whether there are links between some of the children. If there are links between the children, the transport coefficients on the link are recomputed by linear combination of the scaling function to scaling function transport coefficients on links between the children.

7 Results

Here we present results from tests of our algorithm on two scenes, one simple and one complex. For the simple scene, a reference solution is easy to obtain, so convergence and convergence rates can be tested. For the complex scene, computing a reference solution is infeasible. However, the goal of our algorithm is to be able to generate realistic images of complex glossy scenes, so we have included an example of such a scene.

7.1 Simple scene

To test convergence and convergence rates, we tried a simple scene consisting of two tiny patches and a large patch. The geometry is shown in Figure 12. Patch 1 is emitting radiance L_e, and this light is reflected by patch 2 according to Ward's glossy reflection model [39] with $\alpha = 0.2$. This glossy reflection results in a directional radiance distribution on patch 2. The light from patch 2 is received at patch 3, which is a diffuse reflector.

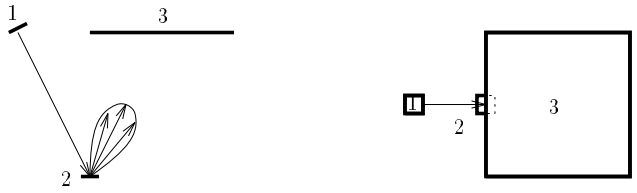


Figure 12 Simple scene geometry seen from the side and from above.

The angular variation of the radiance distribution on patch 2 is shown in the rightmost image in Figure 19. This reference solution was computed as

$$L_2(y,\omega) \;=\; f_r(\omega_{xy}, y, \omega)\, G(x,y)\, L_e(x,\omega_{xy})\, A_1 \,,$$

where x is the midpoint of patch 1, y is the midpoint of patch 2, and A_1 is the area of patch 1, for directions ω on the hemisphere (transformed from the unit square by the transformation described in Section 3.3). Converging finite element approximations of this angular variation are shown in Figure 19, along with difference images illustrating the difference between the approximations and the reference image. The corresponding convergence is shown in the graph in Figure 13(a).

The spatial variation of the radiance on the large diffuse receiver is shown in the rightmost image of Figure 20. This reference solution was computed as

$$L_3(z,\omega) \;=\; f_r(\omega_{yz}, z, \omega)\, G(y,z)\, L_2(y,\omega_{yz})\, A_2 \,,$$

for points z on patch 3, where A_2 is the area of patch 2. Here the direction ω is unimportant since patch 3 is a purely diffuse reflector. The top row of images shows the convergence of the wavelet representation. The first four images are identical, because all refinements take place between patches 1 and 2 (since larger radiance is involved in that transport, and importance is not taken into account). From the fifth image on, the interactions to patch 3 are also refined. Difference images are shown directly below each wavelet solution.

The middle pair of rows shows converging solutions (and difference images) when there is importance at the receiving patch. The interactions are refined more at the receiver than in the preceeding test.

The bottom pair of rows of images in Figure 20 shows the solution with a final gathering step, but without importance. Here the rendering takes advantage of the refinements of the interactions to patch 2 even before the interactions to patch 3 are refined. The convergence of the radiance distribution on patch 3, with and without importance, and with final gathering, is shown in the graph in Figure 13(b).

As these results show, a final gather improves the solution, and gives a better solution in the same amount of CPU time. However, the final gathering step does not appear to improve the convergence rate significantly. The final gather is only useful for display, since the result is an image and not a set of basis functions useful for further refinement and solution. Future research could examine how far the solution would have to proceed before the final gather is performed, if a given accuracy in the solution is required.

Faster convergence can be obtained by several means:

1. Selecting wavelets with more vanishing moments (but with acceptable numerical integration complexity) would make the transport matrix more sparse.

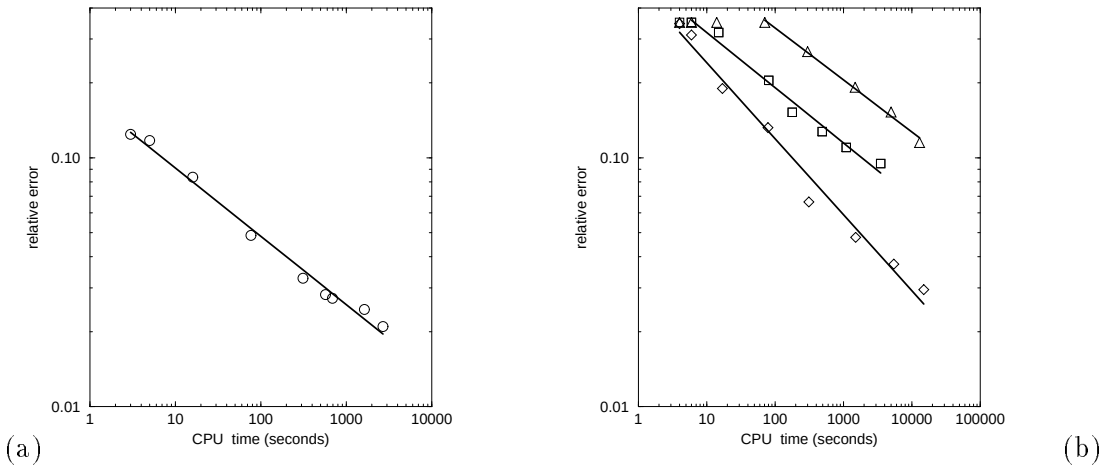


Figure 13 (a) Convergence of the radiance distribution on patch 2. (b) Convergence of the radiance distribution on patch 3: The top curve is the solution without importance, the middle curve is the solution with importance, and the bottom curve is the solution with final gather (without importance). The CPU times were measured on a DEC 3000/400 "Alpha" computer.

2. In this case, a standard construction of the basis would eliminate some basis functions and transports, since patch 2 has little spatial variation, and patch 3 has no angular variation.

7.2 Complex scene

As a more complex test scene, we used a maze of hallways with a glossy Bézier-patch teapot in the center (see Figure 21). The scene consists of 152 patches, including 28 Bézier patches, and has 12,603 mutually visible pairs of patches. The teapot's reflectance function uses Ward's reflection model [39], and is anisotropic with specularities $\alpha_u = 0.2$ and $\alpha_v = 0.5$, specular reflectivity $\rho_s = (0.1, 0.1, 0.1)$ and diffuse reflectivity $\rho_d = (0.2, 0.15, 0)$. The illumination consists of 24 "spotlights," patches that emit directional radiance. There is a patch outside the window that emits light according to a scanned image of an outdoor scene, giving the appearance of a full environment beyond the window. The radiance emitted by the lights and reflected in the scene is shown in Figure 21(a). The objective is to generate an image of this complex environment as seen from the eye, a small patch in the hallway in front of the teapot. All back faces, where no radiance is computed, are rendered gray.

Importance is emitted from the eye (just like a spotlight emits light) and reflected to the important parts of the scene, as shown in Figure 21(b). This picture demonstrates how small a fraction of the model significantly influences the solution visible from the eye. Figure 21(c) is a gray-scale encoding of the number of links between the basis functions on each surface patch. This "refinement image" verifies that most work is performed in areas that are bright and important. Note that we could get arbitrarily large speed-ups, compared to a solution obtained without using importance, by choosing a sufficiently complex scene where many parts do not contribute significantly to the final image.

The program begins by creating 12,603 links between scaling functions, and then solves for the equilibrium distribution. This initial solution can be seen in Figure 21(d). After six iterations of refinement and solution, there are 126 scaling functions in V^0, 1,518 wavelets in W^0, 18,852 wavelets

in W^1, 160,248 wavelets in W^2, 165,495 wavelets in W^3, and approximately 1.73 million links. This solution can be seen in Figure 21(e). In some refinements, new links are only created within already existing spaces, so the solution space reaches V^4 after six iterations. Running times on a DEC 3000/400 "Alpha" machine were approximately 5 minutes to compute the initial solution, then 100 minutes to iterate the main algorithm and refine as far as V^4 in important parts of the scene. Given this solution, an image of it needs to be created, either by evaluating the solution directly or using a final gather step. It takes 15 minutes to render a 600×600 image using ray casting and evaluation of the solution (see Figure 21(e)). Alternatively, using a final gathering step for the rendering takes approximately two hours, making it comparable to the solution process itself, and the result is shown in Figure 21(f). Note the significant color bleeding from the brick walls to the dim ceiling. Also note the glossy highlights on the teapot.

8 Conclusion

We have presented an efficient method for simulating light transport in an environment with diffuse and glossy reflections. The efficiency comes from using a wavelet representation of radiance along with importance-driven refinement for a view-dependent solution.

We use a finite-element representation of the four-dimensional radiance distributions associated with surfaces in a scene, since this representation has a lower initial cost than representation of two-point transport intensities. We outlined the criteria used for selecting a wavelet basis, and discussed reasons why the simplest wavelet basis, the Haar basis, is a good choice for our implementation. Wavelets adapt to the solution, so in areas with little spatial or angular variation a coarse solution is computed, and in areas with greater detail a more refined solution is found.

In contrast to previous algorithms for wavelet radiosity, we use a standard decomposition of the operator, which has the flexibility of being able to let detail at different resolutions interact directly. Since we use a wavelet representation rather than scaling functions at all levels, our algorithm does not require "pushing" and "pulling" procedures. In order to obtain accurate numerical integration without the expense of extraneous samples, we developed adaptive integration rules for the transport coefficients.

We use importance to focus the computations where their impact on the final image is highest. We showed that importance has a very intuitive meaning, and can be considered an exitant quantity similar to radiance.

Since light transport is formulated as a multidimensional Fredholm integral equation of the second kind, our approach may benefit other fields in which such equations arise — numerical analysis, finite-element analysis, computational heat transfer, and particle transport simulation, for example.

There are many possible extensions of the present algorithm. Wavelet bases, like all finite-element bases, are not suited to the representation of ideal specular reflections. Instead, a ray tracing step for ideal specular reflection could be incorporated in the same fashion as in Sillion *et al.* [34]. Surfaces that transmit light in addition to reflecting it could be incorporated into our algorithm by using wavelet basis functions defined for the entire sphere of directions. It may also be possible to extend the clustering algorithms developed for radiosity by Sillion [33] and Smits *et al.* [35] to handle glossy reflections, for example by storing radiance distributions representing radiance from entire clusters of patches.

Acknowledgements

This work was supported by an NSF Young Investigator award (CCR-9357790), an NSF Graduate Research Fellowship, and the University of Washington Graduate Research Fund (award 75-1721).

References

[1] Bradley K. Alpert. *Sparse Representations of Smooth Linear Operators.* PhD thesis, Yale University, 1990.

[2] Larry Aupperle and Pat Hanrahan. A hierarchical illumination algorithm for surfaces with glossy reflection. In *Proceedings of SIGGRAPH 93*, pages 155–162, August 1993.

[3] Larry Aupperle and Pat Hanrahan. Importance and discrete three point transport. In *Proceedings of the Fourth Eurographics Workshop on Rendering*, pages 85–94, June 1993.

[4] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms I. *Communications on Pure and Applied Mathematics*, 44:141–183, 1991.

[5] G. Beylkin, R. R. Coifman, and V. Rokhlin. Wavelets in numerical analysis. In Mary Beth Rushkai et al., editors, *Wavelets and Their Applications*, pages 181–210. Jones and Bartlett, 1992.

[6] Per H. Christensen, David H. Salesin, and Tony D. DeRose. A continuous adjoint formulation for radiance transport. In *Proceedings of the Fourth Eurographics Workshop on Rendering*, pages 95–104, June 1993.

[7] Per H. Christensen, Eric J. Stollnitz, David H. Salesin, and Tony D. DeRose. Importance-driven wavelet radiance. Technical Report 94-01-05, Department of Computer Science and Engineering, University of Washington, January 1994.

[8] Per H. Christensen, Eric J. Stollnitz, David H. Salesin, and Tony D. DeRose. Wavelet radiance. In *Proceedings of the Fifth Eurographics Workshop on Rendering*, pages 287–302, June 1994.

[9] Charles K. Chui. *An Introduction to Wavelets.* Academic Press, Boston, 1992.

[10] Charles K. Chui and Ewald Quak. Wavelets on a bounded interval. *Numerical Methods of Approximation Theory*, 9:53–75, 1992.

[11] A. Cohen, I. Daubechies, and J. C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 45:485–500, 1992.

[12] Michael F. Cohen and John R. Wallace. *Radiosity and Realistic Image Synthesis.* Academic Press Professional, Cambridge, Massachusets, 1993.

[13] Robert L. Cook. Distributed ray tracing. In *Proceedings of SIGGRAPH 84*, pages 137–146, July 1984.

[14] Ingrid Daubechies. *Ten Lectures on Wavelets.* SIAM, 1992.

[15] Tony D. DeRose, David H. Salesin, and Eric J. Stollnitz. Wavelets for computer graphics: A primer. In *SIGGRAPH 94 Computational Representations of Geometry, Course Notes 23*, pages 113–141, July 1994.

[16] Gerald Farin. *Curves and Surfaces for Computer Aided Geometric Design: A Practical Guide.* Academic Press, Boston, third edition, 1993.

[17] Reid Gershbein, Peter Schröder, and Pat Hanrahan. Textures and radiosity: Controlling emission and reflection with texture maps. In *Proceedings of SIGGRAPH 94*, pages 51–58, July 1994.

[18] Gene H. Golub and Charles F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, Baltimore, Maryland, second edition, 1989.

[19] Cindy M. Goral, Kenneth E. Torrance, Donald P. Greenberg, and Bennett Battaile. Modeling the interaction of light between diffuse surfaces. In *Proceedings of SIGGRAPH 84*, pages 213–222, July 1984.

[20] Steven J. Gortler, Peter Schröder, Micheal F. Cohen, and Pat Hanrahan. Wavelet radiosity. In *Proceedings of SIGGRAPH 93*, pages 221–230, August 1993.

[21] Pat Hanrahan, David Salzman, and Larry Aupperle. A rapid hierarchical radiosity algorithm. In *Proceedings of SIGGRAPH 91*, pages 197–206, July 1991.

[22] Paul S. Heckbert. *Simulating Global Illumination Using Adaptive Meshing.* PhD thesis, Department of EECS, UC Berkeley, California, June 1991.

[23] David S. Immel, Michael F. Cohen, and Donald P. Greenberg. A radiosity method for non-diffuse environments. In *Proceedings of SIGGRAPH 86*, pages 133–142, August 1986.

[24] S. Jaffard and Ph. Laurençot. Orthonormal wavelets, analysis of operators, and applications to numerical analysis. In Charles K. Chui, editor, *Wavelets: A Tutorial in Theory and Applications*, pages 543–602. Academic Press, 1992.

[25] Dani Lischinski, Filippo Tampieri, and Donald P. Greenberg. Combining hierarchical radiosity and discontinuity meshing. In *Proceedings of SIGGRAPH 93*, pages 199–208, August 1993.

[26] Stephane Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, July 1989.

[27] S. N. Pattanaik and S. P. Mudur. Efficient potential equation solutions for global illumination computation. *Computers and Graphics*, 17(4):387–396, 1993.

[28] Sumanta N. Pattanaik and Kadi Bouatouch. Haar wavelet: A solution to global illumination with general surface properties. In *Proceedings of the Fifth Eurographics Workshop on Rendering*, pages 273–286, June 1994.

[29] Robert Piessens, Elise de Doncker-Kapenga, Christoph W. Überhuber, and David K. Kahaner. *QUADPACK: A subroutine package for automatic integration.* Springer-Verlag, Berlin, 1983.

[30] Mark C. Reichert. A Two-Pass Radiosity Method Driven by Lights and Viewer Position. Master's thesis, Program of Computer Graphics, Cornell University, Ithaca, New York, January 1992.

[31] Peter Schröder, Steven J. Gortler, Micheal F. Cohen, and Pat Hanrahan. Wavelet projections for radiosity. In *Proceedings of the Fourth Eurographics Workshop on Rendering*, pages 95–104, June 1993.

[32] Peter Schröder and Pat Hanrahan. Wavelet methods for radiance computations. In *Proceedings of the Fifth Eurographics Workshop on Rendering*, pages 303–311, June 1994.

[33] François Sillion. Clustering and volume scattering for hierarchical radiosity calculations. In *Proceedings of the Fifth Eurographics Workshop on Rendering*, pages 105–117, June 1994.

[34] François X. Sillion, James R. Arvo, Stephen H. Westin, and Donald P. Greenberg. A global illumination solution for general reflectance distributions. In *Proceedings of SIGGRAPH 91*, pages 187–196, July 1991.

[35] Brian Smits, James Arvo, and Donald Greenberg. A clustering algorithm for radiosity in complex environments. In *Proceedings of SIGGRAPH 94*, pages 435–442, August 1994.

[36] Brian E. Smits. *Efficient Hierarchical Radiosity in Complex Environments.* PhD thesis, Department of Computer Science, Cornell University, August 1994. TR 94-1443.

[37] Brian E. Smits, James R. Arvo, and David H. Salesin. An importance-driven radiosity algorithm. In *Proceedings of SIGGRAPH 92*, pages 273–282, July 1992.

[38] Roy Troutman and Nelson L. Max. Radiosity algorithms using higher order finite elements. In *Proceedings of SIGGRAPH 93*, pages 209–212, August 1993.

[39] Gregory J. Ward. Measuring and modeling anisotropic reflection. In *Proceedings of SIGGRAPH 92*, pages 265–273, July 1992.

[40] Harold R. Zatz. Galerkin radiosity. In *Proceedings of SIGGRAPH 93*, pages 213–220, August 1993.

A Transport matrix decompositions

In this appendix, we compare different decompositions of the transport matrix **T** for a simple geometry. Our goal is to see how many interactions (transport coefficients relating sending and receiving basis function coefficients) are necessary in each decomposition.

Similar comparisons have been conducted for one-dimensional basis-functions and a smooth dense two-dimensional kernel by Jaffard and Laurençot [24] and Schröder *et al.* [31]. However, due to occlusion, the light transport operator is not smooth, and due to the directionality of radiance, the light transport matrix **T** is not dense. We have conducted tests in flatland, where the basis functions are two-dimensional and the kernel is three-dimensional. We have included experiments with occlusion to test a non-smooth kernel.

Theory indicates that for a smooth operator and a univariate wavelet basis with m basis functions, the standard decomposition results in a matrix with $O(m \log m)$ non-zero entries, while the non-standard decomposition has only $O(m)$ non-zero entries [24]. In practice, however, the standard decomposition has an advantage in some cases. For example, consider a uniformly emitting sender near one end of a receiving patch: the sender's distribution is uniform while the receiver's distribution is less regular. The standard decomposition allows an interaction between a coarse basis function at the sender and fine-scale basis functions at the receiver, while the non-standard decomposition would require interactions between fine-scale basis functions at both sender and receiver. In this example, the standard decomposition would generate fewer non-zero matrix entries than the non-standard decomposition.

We have conducted experiments in flatland to compare various matrix decompositions for the Haar basis. We consider two lines (corresponding to patches in 3D) of unit length. They are either parallel, with a separation of 1 unit, or perpendicular, meeting in a corner (see the illustration in Figure 14). The sending line s does not reflect light, so the radiance reaches equilibrium after a single reflection at the receiver r. The receiver reflects light according to Ward's model [39], with parameters $\rho_s = \rho_d = 0.5$ and $\alpha = 0.2$. The specular peak of this BRDF is about 10 times as bright as the diffuse portion.

Figure 14 Geometry in flatland: Parallel and perpendicular lines. The dashed lines indicate occluding patches in the last set of experiments.

Radiance distributions in flatland are two-dimensional; one parameter corresponds to position and the other to direction. We discretize each patch into 16 positional scaling functions, and

divide the semicircle of directions into 16 angular scaling functions. Radiance distributions can be represented using tensor products of these scaling functions, or using a standard or non-standard two-dimensional wavelet basis.

We consider four representations of the transport matrix in our experiments:

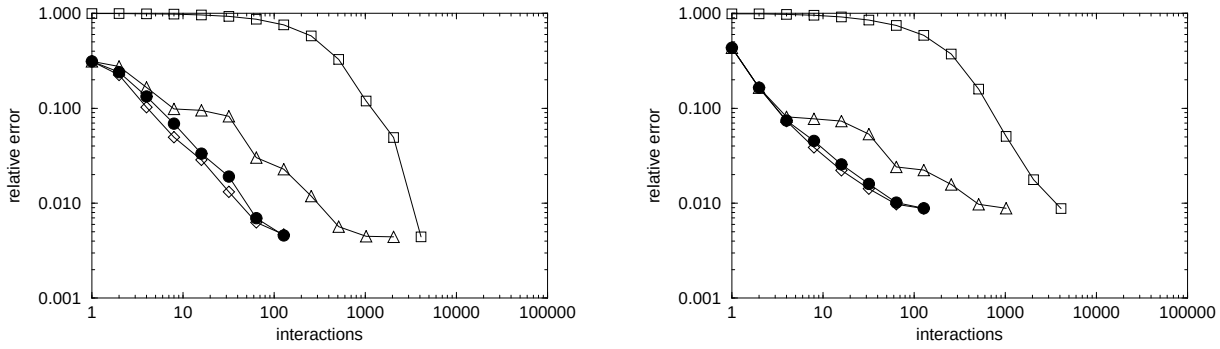
- □ the "direct" matrix, representing the influence of scaling functions on scaling functions;
- △ the non-standard matrix decomposition with the non-standard construction of a tensor product basis;
- • the standard matrix decomposition with the non-standard basis construction (this is what we have used in our implementation); and
- ◇ the standard matrix decomposition with the standard basis construction.

(It is not possible to use a non-standard matrix decomposition with a standard basis construction.) Standard and non-standard bases were discussed in Section 3.4 and standard and non-standard matrix decomposition were discussed in Section 3.5.

As a first test, we consider the special case when the sender has no detail at all: the sending distribution is set to have unit radiance at all points and in all directions. For each matrix decomposition, we evaluate the accuracy of the radiance distribution at the receiving patch r as the number of interactions increases. The interactions are included in order of decreasing magnitude of transport coefficient times sending basis coefficient. The error is computed as L^2 distance between the computed distribution and a reference solution (computed using a direct matrix and a larger number of scaling functions). Figure 15 shows the results. As expected, this experiment is significantly in favor of the standard matrix decomposition: if less than 1% error is desired, 10 times as many matrix entries are required for the non-standard matrix decomposition as for the standard. The reason for this difference is that there is detail at the receiving patch, but no detail at the sending patch. The standard matrix decomposition can represent this kind of interaction directly, whereas the non-standard decomposition forces the interactions to take place between fine-level basis-functions. In other words, the non-standard matrix decomposition forces the basis functions at the sender to be at the same fine level as at the receiver, even though a single basis function is sufficient to represent the sending radiance distribution. All these extra basis functions need to interact, creating many more interactions than for a standard matrix decomposition.

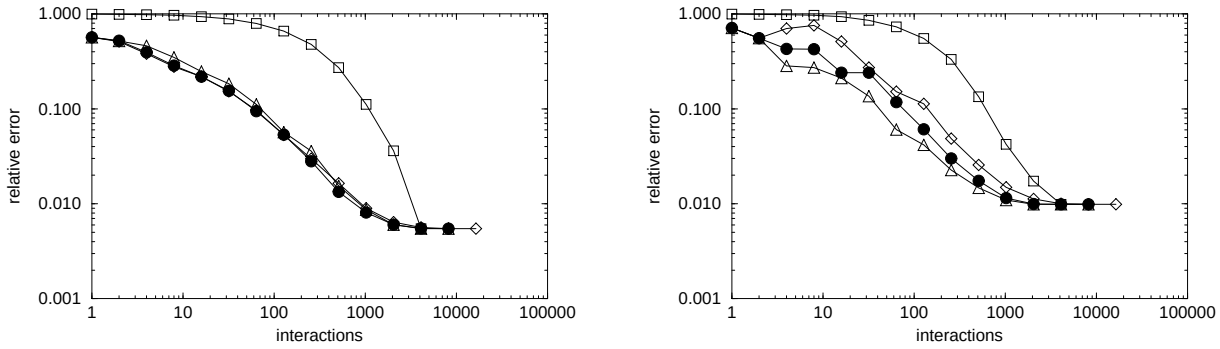
For a second comparison, the sending patch emits directional radiance. The emitted radiance distribution is computed according to Ward's reflection model (as if reflecting light from a single point). The results are shown in Figure 16. For both parallel and perpendicular geometries, the three wavelet matrix decompositions require a much smaller number of interactions than the direct matrix to achieve a given level of accuracy. The three decompositions are comparable in the parallel configuration, while the non-standard decomposition has an advantage in the perpendicular case (perhaps because the significant interactions near the singularity tend to be between detail at the same level).

When surfaces are partially occluded from one another, the transport operator is no longer smooth, and the theory for the cost of decomposing a smooth operator no longer applies. To test this case, a line of length 0.2 is inserted between the sending and receiving lines (the dashed lines in Figure 14), resulting in a non-smooth radiance distribution at the receiver. The results are shown in Figure 17. For parallel lines with occlusion, the standard decomposition requires the least number



parallel perpendicular

Figure 15 Error vs. number of coefficients for uniform emitted radiance and glossy reflection.



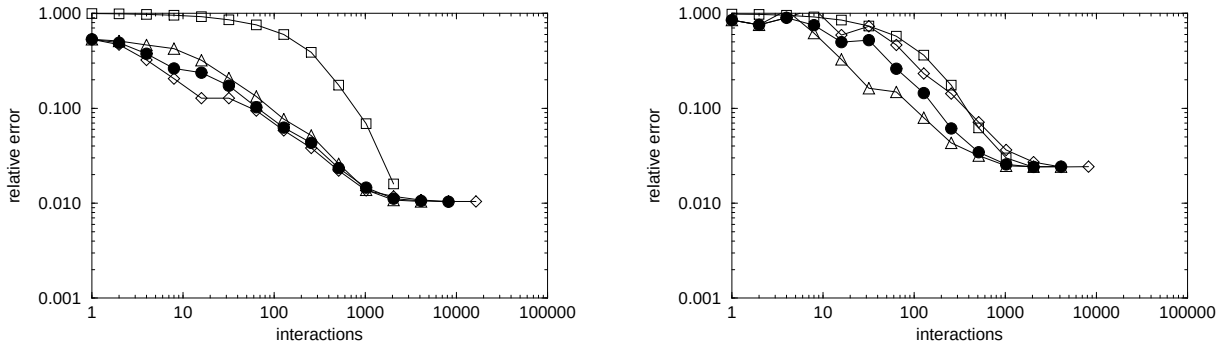
parallel perpendicular

Figure 16 Error vs. number of interactions for directional emitted radiance and glossy reflection.

of interactions for a given level of error. In the perpendicular case, the non-standard decomposition is still best.

Note that for most of our tests, the standard matrix decomposition performed equally well with both standard and non-standard basis function constructions. However, we can construct situations in which we expect the standard construction of basis functions to be advantageous. For example, a diffusely reflecting surface with spatially varying illumination would benefit from having separate hierarchy levels for spatial and angular detail. A glossy surface with much angular variation but little spatial variation is another example.

Comparisons similar to ours [24, 31] support the conclusion we draw for radiance transport: in practice, the standard operator decomposition often requires fewer interactions (matrix elements) than other decompositions to obtain a given accuracy.



parallel perpendicular

Figure 17 Error vs. number of coefficients for partial occlusion (with directional emitted radiance and glossy reflection).

Figure 18 teapots

Figure 19 refinement1

Figure 20 refinement2

Figure 21 scene

Contents