Global Illumination of Glossy Environments Using Wavelets and Importance

PER H. CHRISTENSEN, ERIC J. STOLLNITZ, DAVID H. SALESIN
and
TONY D. DEROSE
University of Washington

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 a finite element algorithm for glossy global illumination. Our algorithm is based on the standard wavelet decomposition of the transport operator and makes use of a four-dimensional wavelet representation for spatially and angularly varying radiance distributions. We use a final gathering step to improve the visual quality of the solution. Features of our implementation include support for curved surfaces as well as texture-mapped anisotropic emission and reflection functions.

Categories and Subject Descriptors: G.1.9 [**Numerical Analysis**]: Integral Equations—*Fredholm equations*; I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism—*color, shading, shadowing, and texture*; *radiosity*; *raytracing*

General Terms: Algorithms

Additional Key Words and Phrases: Glossy global illumination, importance, radiance wavelets

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

[1]Alternatively, we could construct wavelets directly on the hemisphere, in a manner similar to that used by Schröder and Sweldens [1995] for spherical wavelets, but the use of different basis functions for spatial and angular variables would significantly complicate our implementation.

Authors' current addresses: P. H. Christensen, Mental Images GmbH and Co. KG, Fasanenstrasse 81, D-10623, Berlin, Germany; D. H. Salesin (Department of Computer Science and Engineering), and E. J. Stollnitz (Department of Mathematics), University of Washington, Seattle, WA 98195; T. D. DeRose, Pixar Animation Studios, 1001 West Cutting Blvd., Richmond, CA 94804.

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.

This article explores a promising approach to solving the glossy global illumination problem: extending the finite element method used in radiosity algorithms. Designing a finite element algorithm for glossy global illumination involves a number of choices, as summarized in the following.

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 [Immel et al. 1986; Sillion et al. 1991]. The alternative is to use two-point transport intensities, which are functions of two surface positions [Aupperle and Hanrahan 1993a; Schröder and Hanrahan 1995]. 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. [1993] and Gortler et al. [1993]; and also by Schröder et al. [1994]. 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 "nonstandard" tensor products of basis functions; and we can use the standard or the nonstandard operator decomposition. We have chosen to represent radiance in a basis consisting of four-dimensional nonstandard 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 a view-independent solution is necessary. A view-dependent solution can be computed more efficiently using *importance*, as shown by Smits et al. [1992] for radiosity. Assuming we are interested in accelerating our solution procedure using importance, we must choose between incident and exitant importance. We describe a formulation of exitant importance that satisfies the same transport equation as radiance, and that 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, many algorithms produce solutions that are numerically accurate yet still contain artifacts that are very obvious to the human eye. We therefore use a final gathering step [Lischinski et al. 1993; Reichert 1992; Smits 1994] to improve the visual quality of the solution.

We have implemented an algorithm based on these outlined choices. Other features of the implementation include support for curved surfaces and anisotropic bidirectional reflectance distribution functions (BRDFs). Texture maps can be used to describe the spatial variation of both emission and reflection.

A preliminary version of this article was presented at the Fifth Eurographics Workshop on Rendering [Christensen et al. 1995]. In the current article, we give motivations for the choices we made. We also present a proof concerning exitant directional importance. This article extends our previous algorithm by including an adaptive numerical integration method and a more efficient final gathering step. New practical contributions include descriptions of our data structures, as well as tests of convergence and convergence rates.

The remainder of this article is organized as follows: Section 2 motivates the use of radiance distributions, gives a formal description of radiance and light transport, and shows how radiance and the light transport equation can be discretized using a finite element basis. Section 3 introduces a wavelet basis for radiance distributions. Section 4 describes a type of importance that is convenient for a finite element representation of radiance. Section 5 presents our importance-driven glossy global illumination algorithm, which uses a wavelet basis for directional radiance and importance distributions. 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.

2. FINITE ELEMENTS FOR RADIANCE

To determine the exact solution to the glossy global illumination problem for a particular scene, we would have to find the amount of light leaving all points in all directions. To date, it has not been possible to derive analytical solutions for nontrivial scenes with glossy surfaces. Instead, we compute an approximate solution represented by a weighted sum of a finite number of basis functions.

In this section, we first discuss the domain of the basis functions, contrasting two-point transport intensities against radiance distributions. Then we give a formal description of radiance and discuss how the continuous radiance function and transport operator can be discretized to facilitate representation and efficient transport.

2.1 Radiance Distributions Versus Two-Point Transport Intensities

Two fundamentally different representations of light have been used for glossy global illumination. Immel et al. [1986] and Sillion et al. [1991] represent the light in a scene as radiance distributions, which are functions of two spatial and two angular variables on each surface patch. By contrast, Aupperle and Hanrahan [1993], Pattanaik and Bouatouch [1995], and Schröder and Hanrahan [1995] use "two-point transport intensities," which are functions of four spatial variables; these functions represent the amount of light traveling from a point on one patch to a point on another.

We choose to represent light as radiance distributions, as did Immel et al. and Sillion et al., for the following reasons. First, assuming that the scene is initially split into p patches, the coarsest possible representation of radiance requires only one basis function per patch when we use radiance distributions, but it requires p basis functions per patch when we use two-point transport intensities. Therefore, the initial, very coarse solution of the light

transport equation requires $O(p^2)$ interactions between radiance distributions on patches, as opposed to $O(p^3)$ interactions for matching two-point transport intensities. Second, clustering algorithms for radiosity are very effective at reducing the number of initial interactions by grouping nearby patches together [Sillion 1995; Smits et al. 1994], and the only clustering methods for glossy reflections that we are aware of use radiance distributions [Christensen et al. 1996; Sillion et al. 1995].

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, so $\omega_{xy} = -\omega_{yx}$). 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 · meter^{-2} · steradian^{-1}].

The equilibrium distribution of radiance satisfies the following *light transport equation* [Cohen and Wallace 1993]:

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

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 $f_r(\omega_{xy}, y, \omega)$ 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 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 the direction toward y arrives at y. The geometric term has units [steradian · meter^{-2}], and is symmetric in its arguments: $G(x, y) = G(y, x)$. Some of these terms are illustrated in Figure 1.

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).

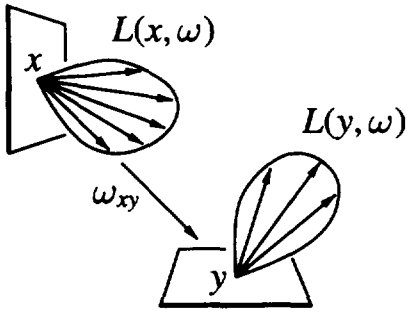

Fig. 1. Light transport from point x to point y.

2.3 Discretization of Radiance

In this section and the following one, we summarize the steps we use to convert the glossy global illumination problem into a system of linear equations. These steps amount to an application of the Galerkin method of finite elements [Zienkiewicz 1989].

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 expressed as a linear combination of the basis functions $b_i(x, \omega)$ with unknown coefficients ℓ_i:

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

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 ith entry if 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 [Gorral et al. 1984]. In subsequent work on radiosity, Heckbert [1991], Zatz [1993], and Troutman and Max [1993] used orthogonal polynomials, and Gortler et al. [1993] 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. [1986] used piecewise-constant basis functions for both spatial and angular variation. Later, Sillion et al. [1991] 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 ijth entry is $\langle f_i | g_j \rangle$.

Likewise, $[\langle \mathbf{F}|g\rangle]$ is the column matrix consisting of elements $\langle f_1|g\rangle$, $\langle f_2|g\rangle, \cdots$.

Let the *dual basis* associated with $\mathbf{B}$ be denoted $\overline{\mathbf{B}} = [\bar{b}_1(x, \omega)\bar{b}_2(x, \omega)\cdots]$. The dual basis is characterized by the relation $\langle \bar{b}_i|b_j\rangle = \delta_{ij}$, 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

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

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 rsth 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

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

where the notation $r \leftarrow s$ serves 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. [1991, 1992], Alpert [1990], Gortler et al. [1993], Hanrahan et al. [1993], and others indicate that significant performance gains can be achieved by using a multiresolution basis.

By contrast, Immel et al. [1986] used a single-resolution representation of radiance distributions using piecewise-constant basis functions. Sillion et al.

[1991] used spherical harmonics as basis functions for the angular variation of radiance. Their implementation also used a single-resolution representation: each distribution uses a fixed number of spherical harmonics. Unfortunately, there is no obvious way to make a useful multiresolution basis from spherical harmonics because they have global support (each spherical harmonic is nonzero over the entire sphere). Therefore it is possible for all spherical harmonics in one location to interact with all those in another location. In contrast, the wavelets we 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. The compact directional support of wavelet basis functions guarantees that many of the transport coefficients will be zero. In short, the transport matrix $\mathbf{T}$ is sparse in a wavelet basis, but dense in a spherical harmonics basis.

In what follows, 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 pairs of basis functions to compute the solution. If n is the number of basis functions, this method would require $O(n^2)$ interactions to compute a solution.

Instead, we use a hierarchical, or *multiresolution*, method that results in only $O(n)$ or $O(n \log n)$ interactions (depending on the specific multiresolution method chosen). With this method, we first compute a very coarse 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 fact that in some parts of the scene radiance distributions can be represented with sufficient accuracy using only a few basis functions. Furthermore, even where many basis functions are required, each basis function will interact with just a few others.

A convenient framework for studying multiresolution bases is provided by multiresolution analysis as formulated by Mallat [1989], which we describe briefly here. A more detailed exposition on the use of multiresolution bases in graphics is given by Stollnitz et al. [1995, 1996]; see the books by Chui [1992] and Daubechies [1992] for more mathematical treatments.

There are two basic ingredients for Mallat's 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

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

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

Orthogonal complements are often written as $V^{j+1} = V^j \oplus W^j$ because, 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 $W^0, \ldots, W^{j-1}$. The scaling functions spanning V^0 represent coarse variation, and 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 $\{0, 1/2^j, 2/2^j, \ldots, 1\}$. The space V^j is spanned by piecewise-constant scaling functions $\phi^j_i(u)$, known as *box functions.* The wavelet spaces W^j are spanned by piecewise-constant functions $\psi^j_i(u)$, known as *Haar wavelets.* A few box functions and Haar wavelets are shown in Figure 2. The Haar basis consists of the single coarsest scaling function $\phi^0_0(u)$ along with all the wavelets ψ^j_i.

There are many alternatives to the Haar basis, each with advantages and disadvantages. Like the Haar basis, flatlets and multiwavelets are suited to the bounded domains over which we define radiance distributions [Gortler et al. 1993]; B-spline wavelets can also be adapted to a bounded interval [Chui and Quak 1992; Stollnitz et al. 1995, 1996]. These higher-order basis functions are appealing because of their improved convergence properties, but they also require more costly numerical integration rules than the Haar basis functions. Our algorithm uses the Haar basis because of its simplicity and convenience, but further research may demonstrate that the benefits of other wavelet bases outweigh their costs. In fact, Schröder and Hanrahan [1995] compared a number of wavelet bases for radiance and, after testing convergence rates and integration expense, found that higher-order wavelets best suited their implementation.

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

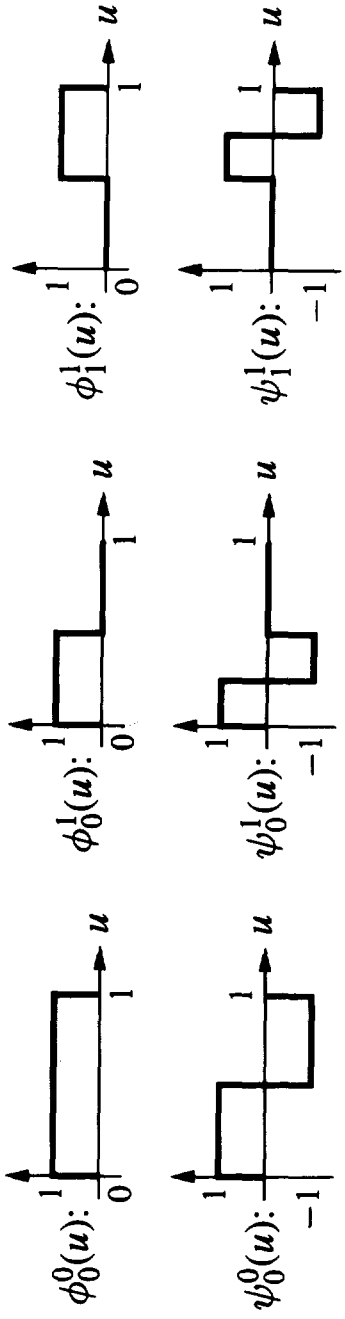

Fig. 2. Some box functions $\phi_i^j(u)$ and Haar wavelets $\psi_i^j(u)$.

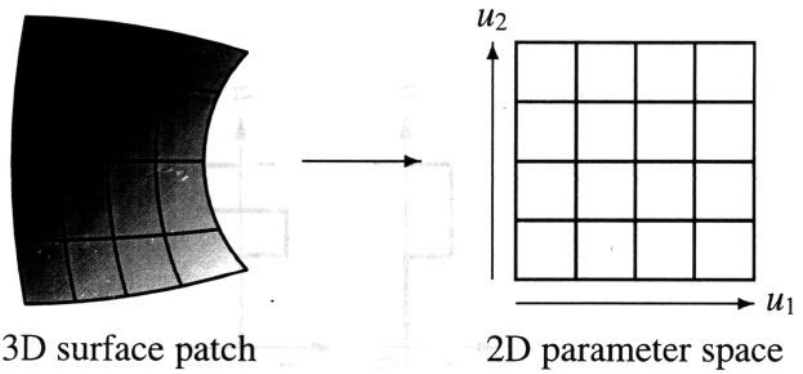

Fig. 3. Spatial projection mapping between 3D surface patch and 2D parameter space.

on the unit square $[0,1]^2$ as illustrated in Figure 3. The domain of the radiance distribution on each patch is then $[0,1]^2 \times H^2$, where H^2 is the unit hemisphere.

Next, for convenience, we transform the hemisphere of directions into another unit square, which allows us to use tensor products of one-dimensional basis functions for both angular and spatial variations.[1] To achieve this transformation, we first use *gnomonic projection* to map between points in H^2 and points on a disc with radius $\pi/2$. As shown on the left side of 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 it introduces less distortion than "flat" projection (flat projection maps H^2 to a unit disc by simply ignoring the height component, resulting in points near the equator being mapped very densely near the circumference of the circle). We then follow the gnomonic projection by a radial "stretch" of the disc to exactly cover the unit square, as shown on the right side of Figure 4. Note that the composition of the gnomonic projection and radial stretch is an invertible mapping between H^2 and the unit square. However, the radial stretch introduces a derivative discontinuity along the diagonals of the square. Figure 5 shows a typical radiance distribution (resulting from glossy reflection of light from a single point) before and after this transformation. 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 now need to construct basis functions on the four-dimensional hypercube $[0,1]^4$. There are two commonly used methods, which both employ tensor products of univariate basis functions: the so-called "standard" and "non-standard" constructions [Beylkin et al. 1991]. The standard construction forms a basis from all possible tensor products of univariate basis functions.

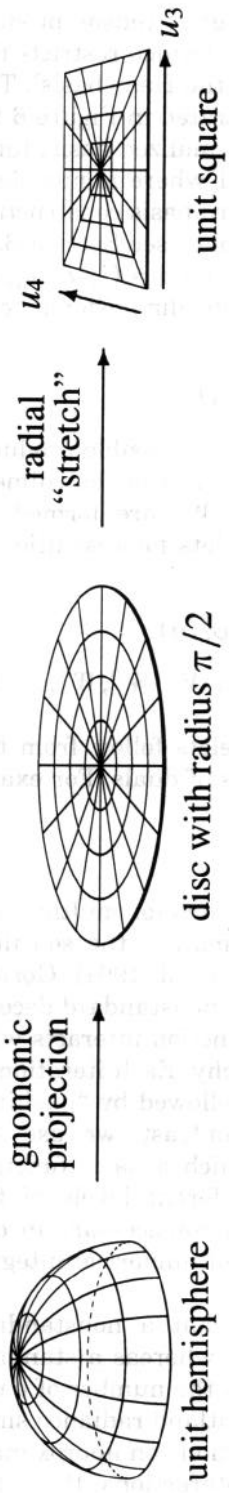

Fig. 4. Angular projection: gnomonic projection and radial "stretch."

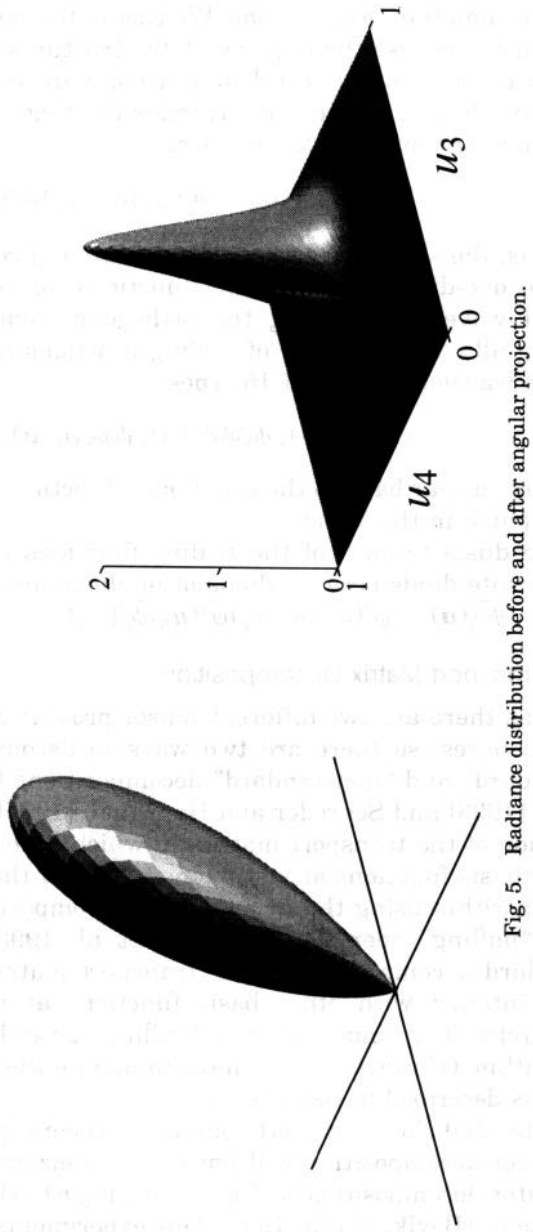

Fig. 5. Radiance distribution before and after angular projection.

In the nonstandard construction, on the other hand, each tensor product consists of univariate basis functions in the same space j (which restricts the supports of multivariate basis functions to be square in the Haar basis). The standard and nonstandard basis constructions are illustrated in Figure 6 for the case of two-dimensional basis functions. In an unnormalized basis, functions are $+1$ where plus signs appear in the figure, -1 where minus signs appear, and 0 in gray regions. We choose the nonstandard basis construction primarily because the required data structures are simpler (see Section 6.4).

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 because duals of products are products of duals. For example, $\overline{\phi\phi\psi\psi}_{\mathbf{i}}^j(\mathbf{u}) = \bar{\phi}_{i_1}^j(u_1)\bar{\phi}_{i_2}^j(u_2)\bar{\psi}_{i_3}^j(u_3)\bar{\psi}_{i_4}^j(u_4)$.

3.5 Transport Matrix Decomposition

Just as there are two different tensor-product constructions for multidimensional bases, so there are two ways to decompose a matrix: the so-called "standard" and "nonstandard" decompositions [Beylkin et al. 1991]. Gortler et al. [1993] and Schröder and Hanrahan [1995] use the nonstandard decomposition of the transport matrix, in which each basis function interacts with other basis functions at just a single level of the hierarchy. Each iteration of an algorithm using the nonstandard decomposition is followed by "pushing" and "pulling" operations [Gortler et al. 1993]. By contrast, we use the standard decomposition of the transport matrix, in which a basis function may interact with other basis functions at many different levels of the hierarchy. The "pushing" and "pulling" procedures are unnecessary in our algorithm. (However, we do have to update the results of numerical integration as described in Section 5.4.)

Note that for a smooth operator, theory predicts that a nonstandard operator decomposition will have $O(n)$ nonzero entries, whereas a standard operator decomposition will have $O(n \log n)$, where n is the number of basis functions [Beylkin et al. 1991]. Our experiments with flatland radiance show that for a given error tolerance, the standard decomposition can approximate some transport matrices using as few or fewer interactions than the

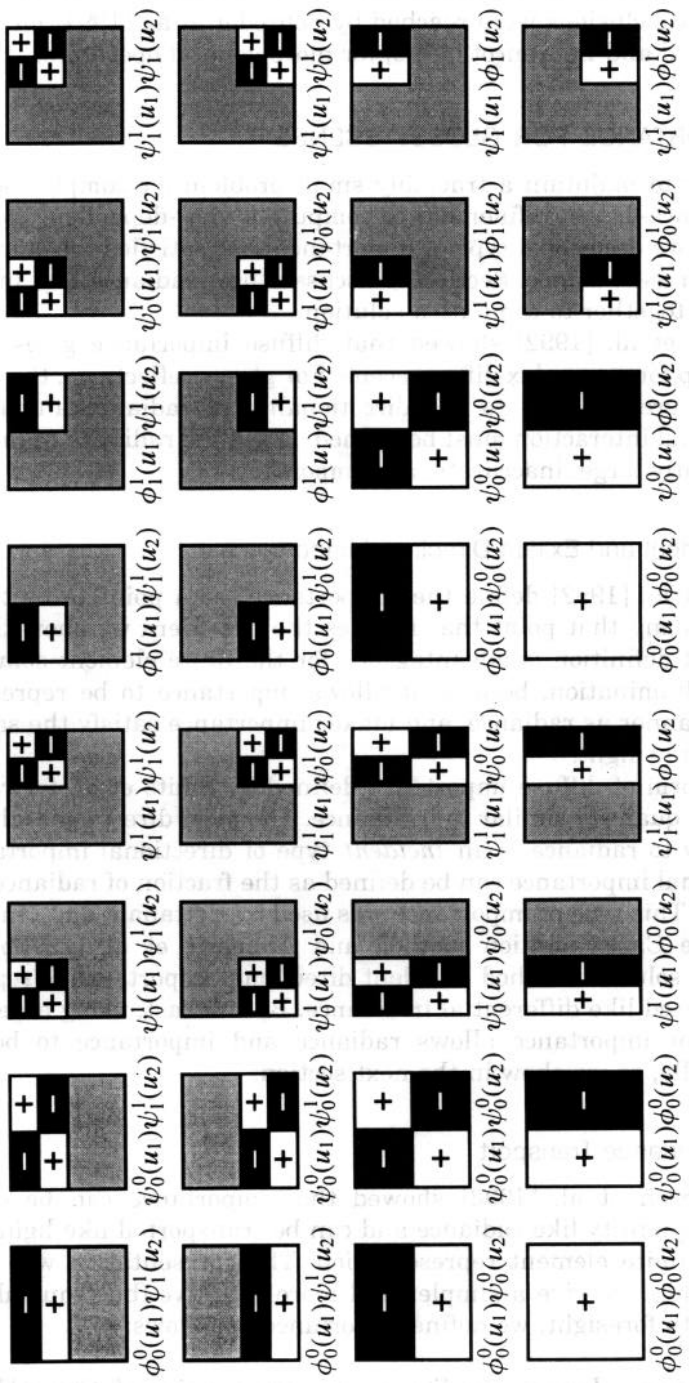

Fig. 6. Standard (left) and nonstandard (right) constructions of a two-dimensional Haar wavelet basis for V^2.

nonstandard decomposition [Christensen 1995]. Thus in practice the nonstandard decomposition is often no more sparse then the standard decomposition. Similar conclusions were reached by Schröder et al. [1994] for radiosity, and by Jaffard and Laurençot [1992] for more general operators.

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 describe a type of importance that satisfies the same equilibrium equation as radiance. Section 5 discusses how radiance and importance can be used together to compute a solution.

Smits et al. [1992] showed that diffuse importance gives a substantial speed-up for a complex diffuse scene. For glossy reflections, the potential gain is even greater, due to the directionality of radiance and importance: a directional interaction must be refined only if the radiance *in that direction* is sufficiently large, inaccurate, and important.

4.1 Incident and Exitant Directional Importance

Smits et al. [1992] 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, because 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. for radiosity is an incident quantity similar to irradiance. 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 [1993] for a Monte Carlo solution method and Aupperle et al. [1993b] for a finite element solution method. Incident directional importance is represented and transported like differential irradiance. By contrast, using an exitant formulation for importance allows radiance and importance to be transported identically, as we show in the next section.

4.2 Importance Transport

Christensen et al. [1993] 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 give a simpler and more intuitive (but equivalent) explanation. With foresight, we define 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.

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

In order to show that importance as previously defined satisfies the same transport equation as radiance, we will need the following lemma.

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. We use induction over the number of bounces taken by the radiance before it reaches 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 the quantity $G(y, z)L(y, \omega_{yz})$ 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:

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

Arriving at the last expression (illustrated in the right side of Figure 7) proves the lemma. □

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

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

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

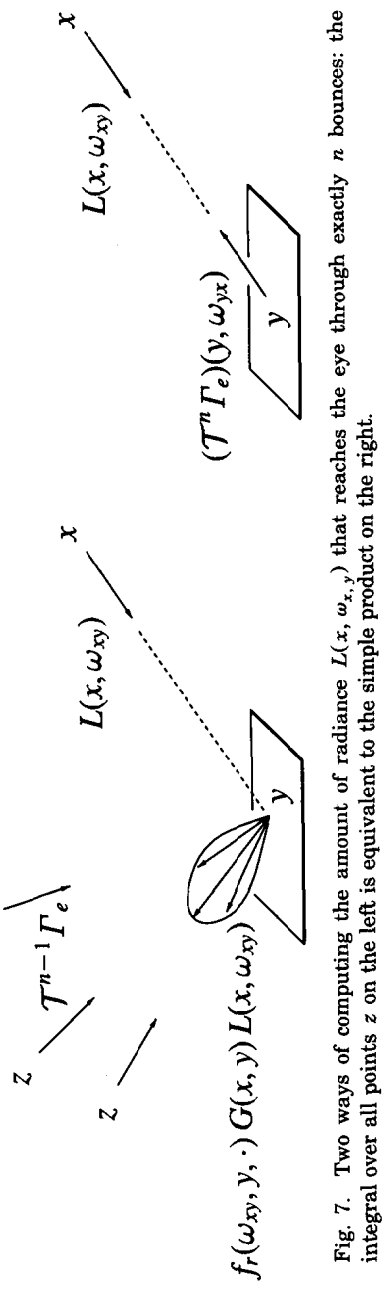

Fig. 7. Two ways of computing the amount of radiance $L(x, \omega_{x,y})$ that reaches the eye through exactly n bounces: the integral over all points z on the left is equivalent to the simple product on the right.

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. □

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

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

The only difference from radiance is that radiance is emitted by light sources, and 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. [1993] for wavelet radiosity. However, there are a number of areas aside from the higher dimensionality of radiance 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 refinement proceeds, and how their accuracy can be increased adaptively at little 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:

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

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 $\tilde{\mathbf{L}}$ and $\tilde{\boldsymbol{\Gamma}}$ and improving the finite representation of the transport operator $\tilde{\mathbf{T}}$. 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

$\tilde{\mathbf{L}}_e$ and $\tilde{\Gamma}_e$. 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 $\tilde{\mathbf{T}}$. The algorithm in pseudocode is as follows:

procedure *GlossyGlobalIllumination* ($\tilde{\mathbf{T}}, \tilde{\mathbf{L}}_e, \tilde{\Gamma}_e$):
  $\tilde{\mathbf{L}} \leftarrow \tilde{\mathbf{L}}_e$
  $\tilde{\Gamma} \leftarrow \tilde{\Gamma}_e$
  repeat
    $\tilde{\mathbf{L}} \leftarrow$ *Solve*($\tilde{\mathbf{T}}, \tilde{\mathbf{L}}, \tilde{\mathbf{L}}_e$)
    $\tilde{\Gamma} \leftarrow$ *Solve*($\tilde{\mathbf{T}}, \tilde{\Gamma}, \tilde{\Gamma}_e$)
    $\tilde{\mathbf{T}} \leftarrow$ *Refine*($\tilde{\mathbf{T}}, \tilde{\mathbf{L}}, \tilde{\Gamma}$)
  until visual convergence of $\tilde{\mathbf{L}}$
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 [Golub and Van Loan 1989] to solve the approximate transport equations $\tilde{\mathbf{L}} = \tilde{\mathbf{L}}_e + \tilde{\mathbf{T}}\tilde{\mathbf{L}}$ and $\tilde{\Gamma} = \tilde{\Gamma}_e + \tilde{\mathbf{T}}\tilde{\Gamma}$. Refinement is determined by an "oracle," described in Section 5.3.

5.2 Computing Transport Coefficients

The preceding algorithm 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. [1993].

The transport coefficients $T_{r \leftarrow s}$ are computed as inner products. For example, the influence of wavelet $\psi\phi\psi\phi_1^j(\mathbf{u}_s)$ on wavelet $\psi\phi\phi\phi_{1'}^{j'}(\mathbf{u}_r)$ is $T_{r \leftarrow s} = \langle \overline{\psi\phi\phi\phi_{1'}^{j'}} \mid \mathcal{T}\psi\phi\psi\phi_1^j \rangle$. Although radiance varies with position x and direction ω, the domain of our tensor-product basis functions is the four-dimensional hypercube $[0, 1]^4$. For convenience, we 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

$$
\begin{aligned}
T_{r \leftarrow s} &= \left\langle \overline{\psi\phi\phi\phi_{1'}^{j'}} \mid \mathcal{T}\psi\phi\psi\phi_1^j \right\rangle \\
&= \int_{\omega y} \overline{\psi\phi\phi\phi_{1'}^{j'}}(y, \omega) \int_x f_r(\omega_{xy}, y, \omega) G(x, y) \psi\phi\psi\phi_1^j(x, \omega_{xy})\, dx\, dy\, d\omega \\
&= \int_{xy} \left[\int_\omega \overline{\psi\phi\phi\phi_{1'}^{j'}}(y, \omega) f_r(\omega_{xy}, y, \omega)\, d\omega \right] G(x, y) \psi\phi\psi\phi_1^j(x, \omega_{xy})\, dy\, dx.
\end{aligned}
\tag{5}
$$

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

We approximate integrals such as the one in Equation (5) using slightly jittered uniform sampling of the integrand. One area for future research in glossy global illumination is the exploration of more accurate integration rules such as Gauss-Legendre or Gauss-Kronrod quadrature [Gortler et al. 1993; Piessens et al. 1983; Zatz 1993].

5.3 Refinement

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

In many applications of wavelets in numerical analysis [Beylkin et al. 1991], the goal is to obtain a sparse representation of a given matrix, thereby making repeated matrix–vector multiplications much faster. In such applications, 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, on the other hand, 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 $\tilde{\mathbf{T}}$ 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 $\tilde{\mathbf{T}}$, and
—*integration error* in computing the entries of $\tilde{\mathbf{T}}$.

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 [Hanrahan et al. 1993], the oracle used by Gortler et al. for wavelet radiosity [1993], and the importance-based refinement strategy used by Smits et al. [1992]. The idea is to estimate the influence on the visible image that would result if a new transport coefficient were to be included in $\tilde{\mathbf{T}}$. 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, as depicted in Figure 8. We compute a new transport coefficient $\tilde{T}_{r \leftarrow s}$ if a sufficiently large value results from the product of

(1) *radiance*: the magnitude of the sending basis function coefficient $\tilde{\ell}_s$;
(2) *estimated transport coefficient*: the estimated new transport coefficient $\tilde{T}_{r \leftarrow s}$ between the basis functions; and

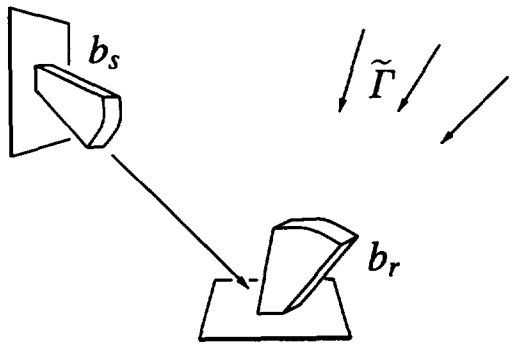

Fig. 8. Sending and receiving basis functions.

(3) *importance*: the integral of $G\tilde{\Gamma}$ over all patches whose importance reaches the receiving basis function.

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 gives the contribution of the transported light to the final image. The sending basis function coefficient is known from the interim solution. The integral of importance arriving at the receiving basis function can be computed from the interim solution as well. Our estimate of the transport coefficient uses kernel variation, as we explain 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. To do this, we associate with each wavelet ψ a unique "parent" wavelet ψ' that overlaps ψ and is in a space one level coarser. We define the parent of a wavelet in W^0 to be the scaling function in V^0 sharing the same support. (For example, in 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 then consider computing a new transport coefficient $\tilde{T}_{r \leftarrow s}$ only if there is already a transport coefficient $\tilde{T}_{r \leftarrow s'}$ or $\tilde{T}_{r' \leftarrow s}$, where r' is the parent of r and s' is the parent of s.

As an estimate for the transport coefficient $\tilde{T}_{r \leftarrow s}$ under consideration, we use the variation (maximum minus minimum) of the samples of the kernel that were obtained when computing either $\tilde{T}_{r \leftarrow s'}$ or $\tilde{T}_{r' \leftarrow s}$. This variation estimates the kernel's deviation from a constant function; the oracle described by Gortler et al. [1993] used a similar measure of deviation from an interpolating polynomial. (One might improve the estimate of $\tilde{T}_{r \leftarrow s}$ by using a global visibility algorithm [Teller and Hanrahan 1993] instead of computing an approximate visibility using sampling, although we have not explored this possibility ourselves.)

Because the kernel variation is the same for all 15 transport coefficients from a given basis function to the 15 wavelets sharing the same support, and because the radiance and importance reaching these basis functions are also

the same, their estimated contribution to the image will be identical. We therefore compute all 15 transport coefficients at once. This approach allows us to avoid reevaluating the kernel for each of these transport coefficients. (At the same time, we also compute the transport coefficient between the two scaling functions that share support with the sending and receiving basis functions. 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% in our implementation) of the maximum product are calculated and incorporated into $\tilde{\mathbf{T}}$. This method requires two passes through all potential new transport coefficients, but it allows refinement to proceed automatically without any user-specified tolerances.

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 necessary to use an adaptive numerical integration technique that reduces error in transport coefficients, particularly those 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 is used as described in Section 5.2. Later, if we compute transport coefficients that link narrower wavelets sharing supports with the original basis functions, the kernel is sampled more densely. These samples are reused to compute the coarse transport coefficient more accurately.

As we mentioned in the previous section, the transport coefficient between two scaling functions is computed at the same time as the transport coefficients between other basis functions with the same support (at practically no extra cost, as the necessary samples of the kernel have already been obtained). Because 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 wherever the kernel is sampled densely.

5.5 Final Gather

In order to render the solution, we can either evaluate the finite element representation of the solution directly, or we can perform an extra step that

improves its visual quality. Following the ideas that Reichert [1992], Lischinski et al. [1993], and Smits [1994] 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 y that corresponds to the midpoint of the pixel.

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 each of the basis functions that has a transport coefficient linking it to a basis function with support at the point y and in the direction of the eye e. This faster method only excludes light transports that were not considered to be significant in the solution process. The third alternative achieves still greater speed by gathering light only from each basis function that has a transport coefficient linking it to the single scaling function on the patch where point y lies. We have tried all three methods and found that the fastest final gather method causes no visible degradation in the final image; exploring the tradeoffs between the speed and accuracy of these (and other) final gather approaches remains an open area for research.

For each basis function from which we want to do a final gather, we evaluate a simplified version of the integral in Equation (5). For example, the final gather from the wavelet $\psi\phi\psi\phi_1^j(x, \omega)$ requires evaluating

$$\int_x f_r(\omega_{xy}, y, \omega_{ye}) G(x, y) \psi\phi\psi\phi_1^j(x, \omega_{xy})\, dx.$$

Because 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 Figure 16, parts (e) and (f).

Another way of thinking about the final gathering step is in the context of distribution raytracing [Cook et al. 1984]. 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 we avoid the costly "explosion" in the number of rays associated with the recursive bounces used in distribution raytracing because we only follow a single bounce. Also, once a finite element radiance solution has been computed, the final gather requires no additional memory.

6. IMPLEMENTATION

In this section, we describe specific features of our implementation, as well as the data structures used to represent basis-function and transport coefficients.

6.1 Surface Geometry

Curved objects are more compactly represented by splines than quadrilaterals. Our algorithm naturally applies to spline surfaces, as long as they are split into convex patches, so that no patch can interact with itself. With this restriction, we can use any surface representation for which it is possible to determine the intersection of a ray with a surface and compute a position, surface normal, and differential area associated with a given parametric point (u_1, u_2). Our implementation includes the Bézier surfaces and quadrilaterals.

6.2 Reflection Models and Texture Maps

We use the Ward isotropic and anisotropic reflection models [Ward 1992] as they are inexpensive to evaluate and consistent with physical observations. Ward's models account for angular variations in reflectance; we also allow reflectance to vary spatially to simulate the texture of materials. Figure 16 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 reduce the errors caused by point-sampling the texture. Gershbein et al. [1994] present a more rigorous mathematical approach for using textures in radiosity, which employs wavelet decompositions of the textures themselves.

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 interact with other parts of the scene; instead, the refinement procedure determines which coefficients affect the visible solution. This technique allows a complex environment to be displayed using simple geometry.

A simple approach to angular variation is to make emission an explicit function of 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 16. 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 [Cohen and Wallace 1993], the matrices $\tilde{\mathbf{T}}$, $\tilde{\mathbf{L}}$, $\tilde{\mathbf{L}}_e$, $\tilde{\Gamma}$, and $\tilde{\Gamma}_e$ are never formed explicitly. In our implementation, entries of $\tilde{\mathbf{L}}$, $\tilde{\mathbf{L}}_e$, $\tilde{\Gamma}$, and $\tilde{\Gamma}_e$ are associated with the surface patches, and entries of $\tilde{\mathbf{T}}$ 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 in which each node contains all coefficients ℓ with the same indices (space j and translation $i_1, \ldots, i_4$). Initially, each patch has a single root node associated with it, containing a scaling function coefficient in space V^0 for each of six "color bands": red, green, and blue radiance and red, green, and blue importance. Only root nodes store these scaling function coefficients, but all nodes contain storage for 15 wavelet coefficients for each color band, and 16 pointers to child nodes that contain the coefficients in the next (more refined) space. The pointers between nodes are illustrated in Figure 9.

6.5 Data Structure for Transport Coefficients

The transport coefficients that describe the interaction between radiance (or importance) basis functions on different patches are stored in *links*. As described in Section 5.3, the transport coefficients from a sending basis function to all 15 wavelets and the single scaling function sharing support are computed at the same time. In our implementation, all 16 of these transport coefficients are stored in the same link. (Alternatively, each transport coefficient could be stored in its own link, but the extra storage overhead would make this approach infeasible as scenes became complex. As another alternative, the transport coefficients between all 15^2 possible combinations of wavelets on sender and receiver could be stored in a single link. However, this method would also waste memory by creating links with room for many transport coefficients that might never be computed if, for example, a sending coefficient was too small.)

Thus each link contains a pointer to the node from which it is transporting radiance (or importance); information about what type of basis function it is transporting from; 15 entries of $\tilde{\mathbf{T}}$ 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 with the same receiving basis functions.

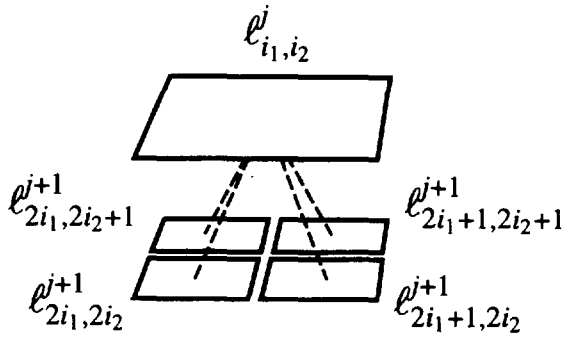

Fig. 9. Tree of basis function coefficients on a patch (simplified to two dimensions, where each node has only four children).

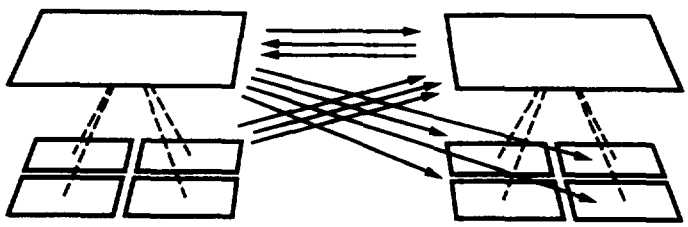

Fig. 10. Example of links between basis functions (in two dimensions).

Note that there can be several links between the same pair of nodes, each connecting different sending basis functions with the same support, as illustrated in Figure 10. All links pointing to a given node are organized in a dynamic list. Initially, links are set up between root nodes for all pairs of patches that are mutually visible.

We consider creating new links between basis functions b_s and b_r only if there is already a link from b_s to the parent of b_r, or from the parent of b_s 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 its transport coefficients; 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, because 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. Note that because a new link can be created in two different ways (by refining a link at the sending end or by refining a different link at the receiving end), we need to check for a link's existence to avoid creating a duplicate.

Links are never destroyed in our algorithm. By contrast, the approach described by Gortler et al. [1993] removes a link at one level of the hierarchy

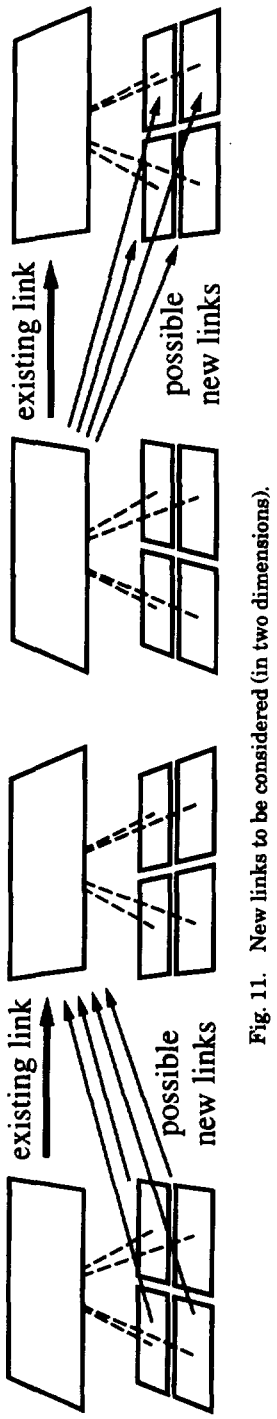

Fig. 11. New links to be considered (in two dimensions).

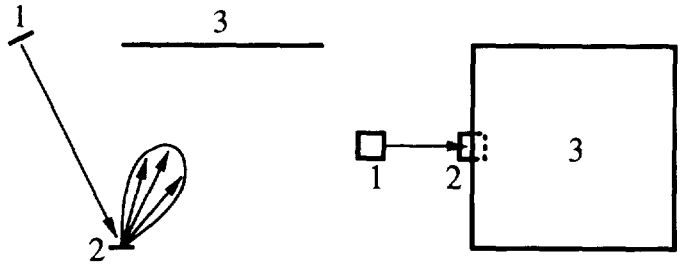

Fig. 12. Simple scene geometry seen from the side and from above.

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 described in Section 5.4 takes place after new links have been set up, as a bottom-up traversal of all links. During this traversal, we compute a new transport coefficient for a given link by taking a linear combination of the scaling-function-to-scaling-function transport coefficients on all links between child nodes. The particular linear combination that is used depends on the sending and receiving basis functions involved.

7. RESULTS

Here we present results from tests of our algorithm on two scenes, one very simple and one more complex. For the simple scene, a reference solution is easily obtained, so convergence and convergence rates can be tested. For the more complex scene, computing a reference solution is infeasible. However, we have included the scene to provide an example that is nontrivial.

7.1 Convergence Tests

To test convergence and convergence rates, we used 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 [1992] 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.

The angular variation of the radiance distribution on patch 2 is shown in the rightmost image in Figure 14 after being transformed from the hemisphere of directions to the unit square as described in Section 3.3. 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. Converging finite element approximations of this angular variation are shown in Figure 14, along with difference images illustrating the error in the approximations relative to the reference image. The graph in Figure 13(a) quantifies the

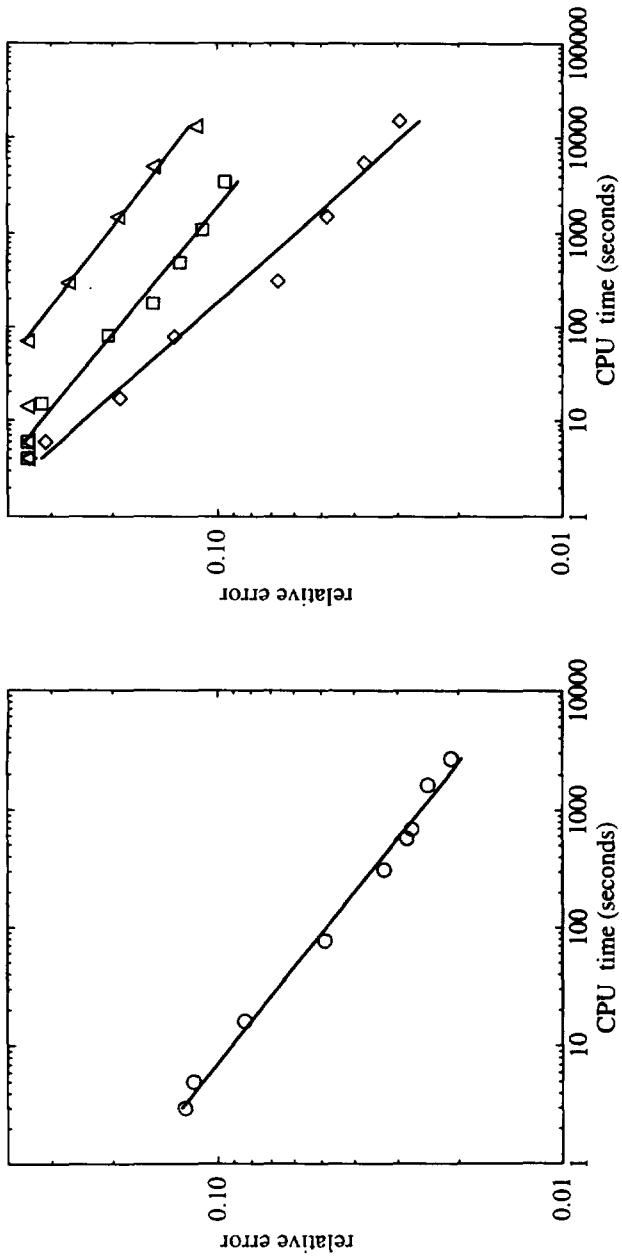

Fig. 13. (a) Convergence of radiance distribution on patch 2. (b) Convergence of radiance distribution on patch 3: top curve is the solution without importance, middle curve is the solution with importance, and bottom curve is the solution with final gather (without importance).

convergence of the distribution of radiance on patch 2. The CPU times were measured on a DEC 3000/400 "Alpha" computer.

The spatial variation of the radiance on patch 3, the large diffuse receiver, is shown in the rightmost image of Figure 15. 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 as 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 (because 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 third and fourth rows of Figure 15 show converging solutions and the corresponding difference images when the receiving patch emits importance. In this case, the interactions are refined more at the receiver than in the preceding test.

The fifth and sixth rows of Figure 15 show 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. Note that the final gather results in images that are still piecewise constant because there is only one radiance distribution visible to patch 3, namely, the piecewise-constant distribution of patch 2. The graph in Figure 13(b) illustrates the convergence of the radiance distribution on patch 3, with and without importance, and also with the final gathering step.

As these results show, a final gather improves the appearance of the solution, and gives a better image in the same amount of CPU time. However, the final gather is useful only for display, as the result is an image rather than a set of basis functions that could be used 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.

7.2 A More 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 16). 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 [Ward 1992], 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 16(a). The objective is to generate an image of

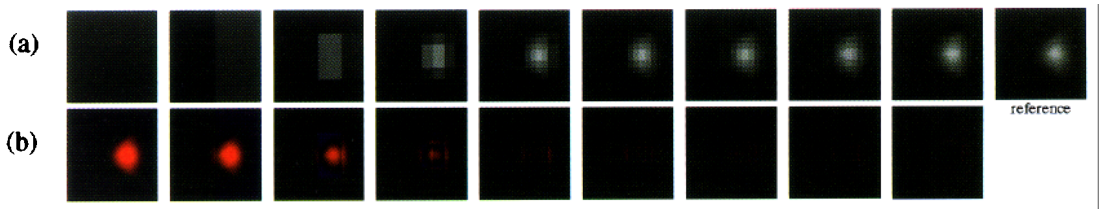

Fig. 14. (a) Refinement of radiance distribution on patch 2, displayed as a function of angular parameters for a fixed position. The rightmost image is the reference solution. (b) Difference between each image and the reference image (blue and red indicate positive and negative values, respectively).

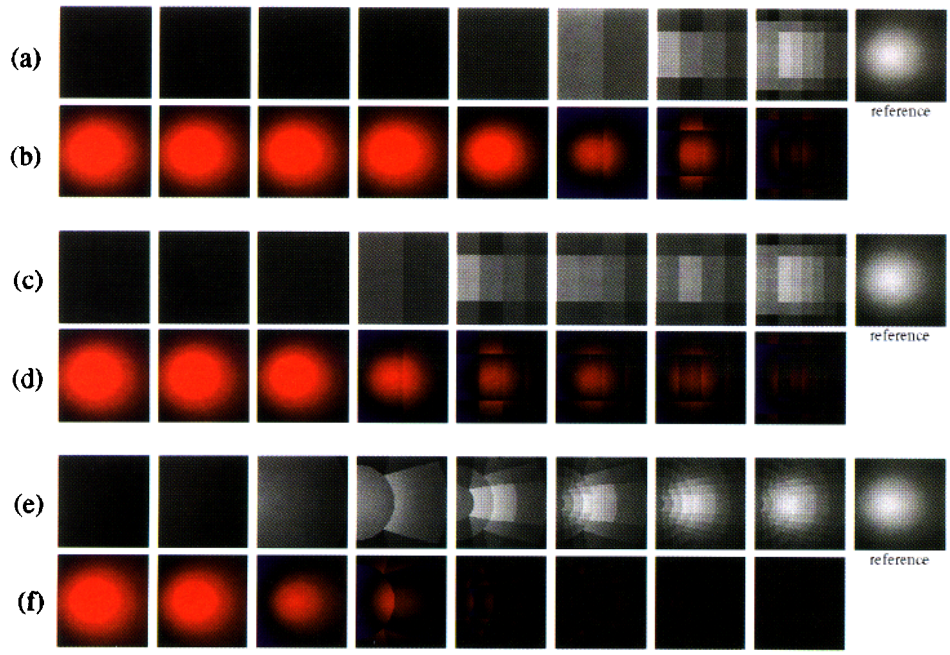

Fig. 15. (a) Refinement of radiance distribution on patch 3. (b) Difference between each image and the reference image. (c) Solutions when patch 3 emits importance. (d) Difference images with importance. (e) Solutions with final gather (without importance). (f) Difference images with final gather. The three rightmost images all show the reference solution for patch 3.

this 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 as gray.

Importance is emitted from the eye (just as a spotlight emits light) and reflected to the important parts of the scene, as shown in Figure 16(b). This figure demonstrates the small fraction of the overall model that significantly influences the visible scene. Figure 16(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

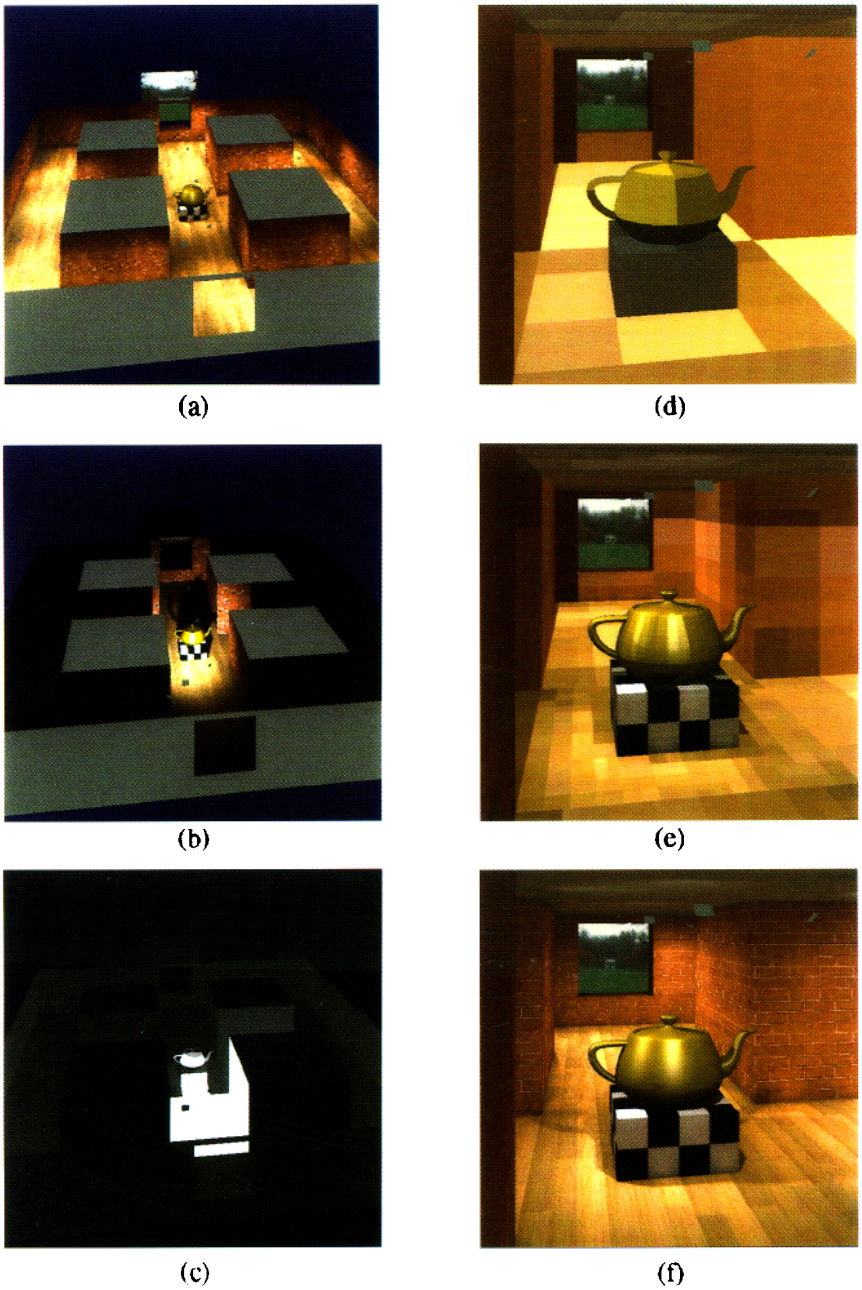

Fig. 16. Solutions for a complex scene: (a) radiance seen from above; (b) importance seen from above; (c) gray-scale representation of refinement; (d) initial radiance solution in space V^0; (e) refined radiance solution in space V^1; (f) refined radiance solution with final gather.

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 in which 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 16(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 16(e). In some refinements, new links are created only within existing spaces, so the solution space remains in 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. Once we obtain this solution, we need to create an image of it, either by evaluating the solution directly or by using a final gather step. Although it might be possible to take advantage of graphics hardware to render the solution directly, we use a ray-casting technique in order to preserve the quality of curved surfaces and correctly account for the directionality of radiance. The 600×600 image shown in Figure 16(e) was rendered in 15 minutes. The other alternative is to use a final gathering step for the rendering, which takes approximately two hours, making it comparable to the solution process itself. The result is shown in Figure 16(f). Note the significant color bleeding from the brick walls to the dim ceiling, as well as 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, as this representation has a lower initial cost than a representation using two-point transport intensities. For the finite elements, we used the Haar basis, the simplest wavelet basis. 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, and because we use a wavelet representation rather than scaling functions at all levels, our algorithm does not require "pushing" and "pulling" procedures. However, our algorithm requires that we update numerical integrations, and we described an adaptive integration method that reuses kernel samples to improve existing transport coefficients.

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

Light transport is formulated as a multidimensional Fredholm integral equation of the second kind, therefore 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. 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. Other wavelet bases, particularly those with more vanishing moments, could be tried. Wavelet bases, like all finite element bases, are not suited to the representation of ideal specular reflections; instead, a raytracing step for ideal specular reflection could be incorporated in the same fashion as in Sillion et al. [1991]. Finally, in a forthcoming article [Christensen et al. 1996], we describe how a clustering algorithm can be used to reduce the complexity of the initial linking phase of the simulation by grouping together nearby patches and representing each group's radiance with a single distribution. Such a method, combined with the work described in this article, can be used to compute glossy global illumination solutions for extremely complex scenes.

ACKNOWLEDGMENTS

We would like to thank Dani Lischinski and Steven Gortler for their careful reviews of an early version of this article. This work was supported by an NSF Young Investigator award (CCR-9357790), an NSF Presidential Faculty Fellowship (CCR-9553199), an ONR Young Investigator award (N00014-95-1-0728), an Alfred P. Sloan Research Fellowship (BR-3495), an NSF Graduate Research Fellowship, and a grant from the University of Washington Graduate Research Fund (75-1721).

REFERENCES

ALPERT, B. K. 1990. Sparse representations of smooth linear operators. Yale Univ., Ph.D. thesis.

AUPPERLE, L. AND HANRAHAN, P. 1993a. A hierarchical illumination algorithm for surfaces with glossy reflection. In *SIGGRAPH 93 Conference Proceedings* (Anaheim, CA, August), Computer Graphics Annual Conference Series, 155–162.

AUPPERLE, L. AND HANRAHAN, P. 1993b. Importance and discrete three point transport. In *Proceedings of the Fourth Eurographics Workshop on Rendering* (Paris, June), 85–94.

BEYLKIN, G., COIFMAN, R. R., AND ROKHLIN, V. 1992. Wavelets in numerical analysis. In *Wavelets and Their Applications*, Mary Beth Rushkai et al., Eds., Jones and Bartlett, Boston, MA, 181–210.

BEYLKIN, G., COIFMAN, R., AND ROKHLIN, V. 1991. Fast wavelet transforms and numerical algorithms I. *Commun. Pure Appl. Math. 44*, 141–183.

CHRISTENSEN, P. H. 1995. Hierarchical techniques for glossy global illumination. Univ. of Washington, Ph.D. thesis.

CHRISTENSEN, P. H., LISCHINSKI, D., STOLLNITZ, E. J. AND SALESIN, D. H. 1996. Clustering for glossy global illumination. *ACM Trans. Graph.* (to appear).

CHRISTENSEN, P. H., SALESIN, D. H., AND DEROSE, T. D. 1993. A continuous adjoint formulation for radiance transport. In *Proceedings of the Fourth Eurographics Workshop on Rendering* (Paris, June), 95–104.

CHRISTENSEN, P. H., STOLLNITZ, E. J., SALESIN, D. H., AND DEROSE, D. T. 1995. Wavelet radiance. In *Photorealistic Rendering Techniques*, G. Sakas, P. Shirley, and S. Müller, Eds., Springer-Verlag, Berlin, 295–309.

CHUI, C. K. 1992. *An Introduction to Wavelets*. Academic Press, Boston.

CHUI, C. K. AND QUAK, E. 1992. Wavelets on a bounded interval. In *Numerical Methods of Approximation Theory*, vol. 9, Birkhauser, Boston, MA, 53–75.

COHEN, M. F. AND WALLACE, J. R. 1993. *Radiosity and Realistic Image Synthesis*. Academic Press Professional, Cambridge, MA.

COOK, R. L., PORTER, T., AND CARPENTER, L. 1984. Distributed ray tracing. In *SIGGRAPH'84 Conference Proceedings* (Minneapolis, MN, July), Computer Graphics Annual Conference Series, 137–145.

DAUBECHIES, I. 1992. *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA.

GERSHBEIN, R., SCHRÖDER, P., AND HANRAHAN, P. 1994. Textures and radiosity: Controlling emission and reflection with texture maps. In *SIGGRAPH 94 Conference Proceedings* (Orlando, FL, July), Computer Graphics Annual Conference Series, 51–58.

GOLUB, G. H. AND VAN LOAN, C. F. 1989. *Matrix Computations*. Second edition. Johns Hopkins University Press, Baltimore, MD.

GORAL, C. M., TORRANCE, K. E., GREENBERG, D. P., AND BATTAILE, B. 1984. Modeling the interaction of light between diffuse surfaces. In *SIGGRAPH'84 Conference Proceedings* (Minneapolis, MN, July), Computer Graphics Annual Conference Series, 213–222.

GORTLER, S. J., SCHRÖDER, P., COHEN, M. F., AND HANRAHAN, P. 1993. Wavelet radiosity. In *SIGGRAPH 93 Conference Proceedings* (August), Computer Graphics Annual Conference Series, 221–230.

HANRAHAN, P., SALZMAN, D., AND AUPPERLE, L. 1993. A rapid hierarchical radiosity algorithm. In *SIGGRAPH'91 Conference Proceedings*, Computer Graphics Annual Conference Series (Las Vegas, NV, July), 197–206.

HECKBERT, P. S. 1991. Simulating global illumination using adaptive meshing. Univ. of California, Berkeley, Ph.D. thesis.

IMMEL, D. S., COHEN, M. F., AND GREENBERG, D. P. 1986. A radiosity method for non-diffuse environments. In *SIGGRAPH'86 Conference Proceedings* (Dallas, TX, August), Computer Graphics Annual Conference Series, 133–142.

JAFFARD, S., AND LAURENÇOT, P. 1992. Orthonormal wavelets, analysis of operators, and applications to numerical analysis. In *Wavelets: A Tutorial in Theory and Applications*, Charles K. Chui, Ed., Academic Press, 543–602.

LISCHINSKI, D., TAMPIERI, F., AND GREENBERG, D. P. 1993. Combining hierarchical radiosity and discontinuity meshing. In *SIGGRAPH 93 Conference Proceedings* (Anaheim, CA, August), Computer Graphics Annual Conference Series, 199–208.

MALLAT, S. 1989. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell. 11*, 7 (July), 674–693.

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

PATTANAIK, S. N. AND BOUATOUCH, K. 1995. Haar wavelet: A solution to global illumination with general surface properties. In *Photorealistic Rendering Techniques*, G. Sakas, P. Shirley, and S. Müller, Eds., Springer-Verlag, Berlin, 281–294.

PIESSENS, R., DE DONCKER-KAPENGA, E., ÜBERHUBER, C. W., AND KAHANER, D. 1983. *QUADPACK: A Subroutine Package for Automatic Integration*. Springer-Verlag, Berlin.

REICHERT, M. C. 1992. A two-pass radiosity method driven by lights and viewer position. Cornell University, Master's thesis.

SCHRÖDER, P., GORTLER, S. J., COHEN, M. F., AND HANRAHAN, P. 1994. Wavelet projections for radiosity. *Comput. Graph. Forum 13*, 2 (June), 141–151.

SCHRÖDER, P. AND HANRAHAN, P. 1995. Wavelet methods for radiance computations. In *Photorealistic Rendering Techniques*, G. Sakas, P. Shirley, and S. Müller, Eds., Springer-Verlag, Berlin, 310–326.

SCHRÖDER, P. AND SWELDENS, W. 1995. Spherical wavelets: Efficiently representing functions on the sphere. In *SIGGRAPH 95 Conference Proceedings* (Los Angeles, CA, August), Computer Graphics Annual Conference Series, 161–172.

SILLION, F. 1995. Clustering and volume scattering for hierarchical radiosity calculations. In *Photorealistic Rendering Techniques*, G. Sakas, P. Shirley, and S. Müller, Eds., Springer-Verlag, Berlin, 105–118.

SILLION, F., DRETTAKIS, G., AND SOLER, C. 1995. A clustering algorithm for radiance calculation in general environments. In *Rendering Techniques '95*, P. M. Hanrahan and W. Purgathofer, Eds., Springer-Verlag, Vienna, 196–205.

SILLION, F. X., ARVO, J. R., WESTIN, S. H., AND GREENBERG, D. P. 1991. A global illumination solution for general reflectance distributions. In *SIGGRAPH'91 Conference Proceedings* (July), Computer Graphics Annual Conference Series, 187–196.

SMITS, B. E. 1994. Efficient hierarchical radiosity in complex environments. Cornell Univ. Ph.D. thesis.

SMITS, B., ARVO, J., AND GREENBERG, D. 1994. A clustering algorithm for radiosity in complex environments. In *SIGGRAPH 94 Conference Proceedings* (Orlando, FL, August), Computer Graphics Annual Conference Series, 435–442.

SMITS, B. E., ARVO, J. R., AND SALESIN, D. H. 1992. An importance-driven radiosity algorithm. In *SIGGRAPH'92 Conference Proceedings* (Chicago, IL, July), Computer Graphics Annual Conference Series, 273–282.

STOLLNITZ, E. J., DEROSE, T. D., AND SALESIN, D. H. 1996. *Wavelets for Computer Graphics: Theory and Applications*. Morgan Kaufmann, San Francisco, CA (to appear).

STOLLNITZ, E. J., DEROSE, T. D., AND SALESIN, D. H. 1995. Wavelets for computer graphics: A primer. *IEEE Comput. Graph. Appl. 15*, 3 (May) (Part 1), 76–84; *15*, 4 (July) (Part 2), 75–85.

TELLER, S. AND HANRAHAN, P. 1993. Global visibility algorithms for illumination computations. In *SIGGRAPH 93 Conference Proceedings* (Anaheim, CA, August), Computer Graphics Annual Conference Series, 239–246.

TROUTMAN, R. AND MAX, N. L. 1993. Radiosity algorithms using higher order finite elements. In *SIGGRAPH 93 Conference Proceedings* (Anaheim, CA, August), Computer Graphics Annual Conference Series, 209–212.

WARD, G. J. 1992. Measuring and modeling anisotropic reflection. In *SIGGRAPH'92 Conference Proceedings* (Chicago, IL, July), Computer Graphics Annual Conference Series, 265–273.

ZATZ, H. R. 1993. Galerkin radiosity. In *SIGGRAPH 93 Conference Proceedings* (Chicago, IL, August), Computer Graphics Annual Conference Series, 213–220.

ZIENKIEWICZ, O. C. 1989. *The Finite Element Method*. Fourth edition, McGraw-Hill, London.

Received October 1994; revised October 1995; accepted November 1995