Improving Sampling and Reconstruction Techniques for Radiosity

Dani Lischinski*
Filippo Tampieri*
Donald P. Greenberg*

TR 91-1202
August 1991

Program of Computer Graphics
Department of Computer Science
Cornell University
Ithaca, NY 14853-7501

*This research was supported by the National Science Foundation under a grant entitled "Interactive Input and Display Techniques" (CCR-8617880), and was performed using equipment generously donated by the Hewlett-Packard Corporation and the Digital Equipment Corporation.

Improving Sampling and Reconstruction Techniques for Radiosity*

Dani Lischinski *Filippo Tampieri* *Donald P. Greenberg*

Program of Computer Graphics
Cornell University
Ithaca, New York 14853

January 1991

Abstract

The view-independent global illumination problem is rephrased as one of determining a radiance function across each surface in the environment. A new methodology for diffuse environments, based on the sampling and reconstruction of these functions is introduced. Within this context, the following problems are investigated: (i) where the radiance functions should be sampled; (ii) how to evaluate a radiance function at each sample; and (iii) how to reconstruct a radiance function from the set of samples. The new methodology relaxes some of the assumptions built into current radiosity algorithms. Results are presented which show that the new methodology yields significantly higher accuracy than existing radiosity methods.

CR Categories and Subject Descriptors: I.3.7—[**Computer Graphics**]: Three-Dimensional Graphics and Realism; I.3.3—[**Computer Graphics**]: Picture/Image Generation.

Additional Key Words and Phrases: diffuse reflector, global illumination, radiance function, radiosity, reconstruction, rendering equation, sampling, shadows.

*This research was supported by the National Science Foundation under a grant entitled "Interactive Input and Display Techniques" (CCR-8617880), and was performed using equipment generously donated by the Hewlett-Packard Corporation and the Digital Equipment Corporation.

1 Introduction

Realistic image synthesis is one of the most popular branches of computer graphics. The development of lighting models that simulate the interaction of light with surfaces and of shading algorithms that apply these models to produce realistic looking images has resulted in sophisticated rendering methods. These methods are used in numerous applications that require realism and demand ever increasing image quality, fidelity, complexity, and computation speed.

Global illumination algorithms which play a major role in realistic image synthesis can be classified into two main groups: *view-dependent* methods such as ray-tracing which compute an image for a particular point of view, and *view-independent* methods such as radiosity. This latter approach solves the global illumination problem for all the surfaces of an environment, providing a solution that can be used to produce dynamic sequences.

Radiosity is a physically-based method that attempts to solve the global illumination problem according to the principles of transfer and conservation of energy; it was derived by techniques used in thermal engineering for the determination of radiative heat exchange in enclosures [13, 8, 5, 4]. First formulated [8] as a finite element method, it required that the environment be discretized into a set of elements, each having constant radiosity across its entire area. Current radiosity implementations start with an initial discretization based on local geometric considerations; then, as the solution process progresses and more information about the radiosity distribution is gained, adaptive subdivision schemes are used to refine regions of high radiosity gradients [6]. These schemes are entirely dependent on the size and shape of the elements of the initial mesh and frequently result in "floating" objects, "leaking" light, and the inability to capture fine illumination details such as, for example, the thin shadows cast from sunlight going through a venetian blind.

Based on the constant radiosity assumption, the contribution of an element to another point in the environment can be computed as the product of the average element radiosity and a purely geometrical term known as the form-factor. For efficiency it was proposed that elements be grouped into larger patches and the overall contribution of each patch be computed as if radiosity was constant across its area [6]. Considerable effort has subsequently been devoted to the accurate computation of form-factors [1, 15], but little attention has been given to the fact that patches composed of several elements do not, in general, satisfy the constant radiosity assumption.

One can think of radiosity as a method which computes a *radiance function* for each surface in the environment. The radiance function (which will be formally defined later) gives the *radiance* (commonly referred to as *intensity* in computer graphics) at each location on the surface, and thus can be used for shaded display.

To yield a display, the computation process involves sampling and reconstruction stages. Each radiance function is sampled at a set of locations on the corresponding surface. These locations can either be centers of elements [6], or mesh vertices [15]. The computation of the radiance value at a given location involves another sampling process: light sources are sampled in order to numerically approximate their contributions at the location. This has been done using the hemi-cube [5], or via ray-tracing [15]. Displaying the results involves reconstruction of the radiance functions from the samples. This is typically done by means of bilinear interpolation.

In this context, the following observations can be made with respect to the weaknesses of current radiosity implementations:

1. Dependence on the initial discretization, which takes into account only the local geometry of each surface, frequently leads to undersampling producing erroneous results.

2. Sampling the light sources is driven by the geometry of each light source, and not by its radiance function, computing contributions incorrectly.

3. Bilinear interpolation schemes for rendering cause perceptual anomalies (Mach banding) in regions where the actual radiance function varies smoothly.

Because of these problems, high quality results often require fine meshes, containing topological as well as geometric information, which cause excessive space and time requirements. In practice, for achieving balance between quality and space/time requirements, user intervention is frequently needed to help create a satisfactory mesh. With the exception of a limited solution to the meshing problem [3], none of these issues have been sufficiently addressed in the published literature.

Methods striving for high quality results should:

1. Employ sampling strategies guided by the properties and the behavior of the radiance function across illuminated surfaces as well as light sources.

2. Automatically determine initial sampling locations from which convergence to correct results is guaranteed.

3. Use reconstruction techniques to produce results which comply with the characteristics of actual radiance functions.

To achieve these goals, a novel approach to the global illumination problem has been investigated by the authors. The new approach, which is currently restricted to ideal diffuse reflectors, parts from the current constant finite elements formulation and returns to a continuous domain to represent the radiance across each surface in the environment as a function of location on the surface. A methodology based on sampling and reconstruction techniques and led by the insight gained from a simple two-dimensional analysis of the properties and behavior of such functions is then developed to compute a solution to the global illumination problem. The ideas presented imply possible solutions to common problems that afflict current radiosity algorithms. Preliminary results indicate that the advocated approach is considerably more accurate than standard radiosity algorithms.

2 Problem Formulation

Assume that the environment to be rendered is defined by a collection of surfaces S. Let L_i be the *radiance function* over surface $s_i \in S$. $L_i(x, x'', \lambda)$ is the light energy flux per unit solid angle per unit projected area leaving surface s_i from point x towards point x'', at wavelength λ. Radiance (often named *intensity* in the computer graphics community) is the quantity that is used for assigning pixel values in image synthesis. Therefore, it is convenient to formulate the rendering equation [9] in terms of radiance [2, 12]:

$$L_i(x, x'', \lambda) = L_i^e(x, x'', \lambda) + \sum_{s_j \in S} \int_{x' \in s_j} \rho_i(x', x, x'', \lambda) L_j(x', x, \lambda) v(x, x') \frac{\cos\theta_i \cos\theta_j}{r^2} dA(x') \qquad (1)$$

where

$L_i^e(x, x'', \lambda)$ is the emitted radiance

$\rho_i(x', x, x'', \lambda)$ is the bidirectional reflectance distribution function

$v(x, x')$ is 1 if x and x' are visible to each other and 0 otherwise

$dA(x')$ is a differential area element centered at x'

r is the distance between x and x'

θ_i and θ_j are the angles between the surface normals at x and x' and the line connecting the two points.

Note that θ_i, θ_j, and r are functions of x and x', rather than constants (see Figure 1).

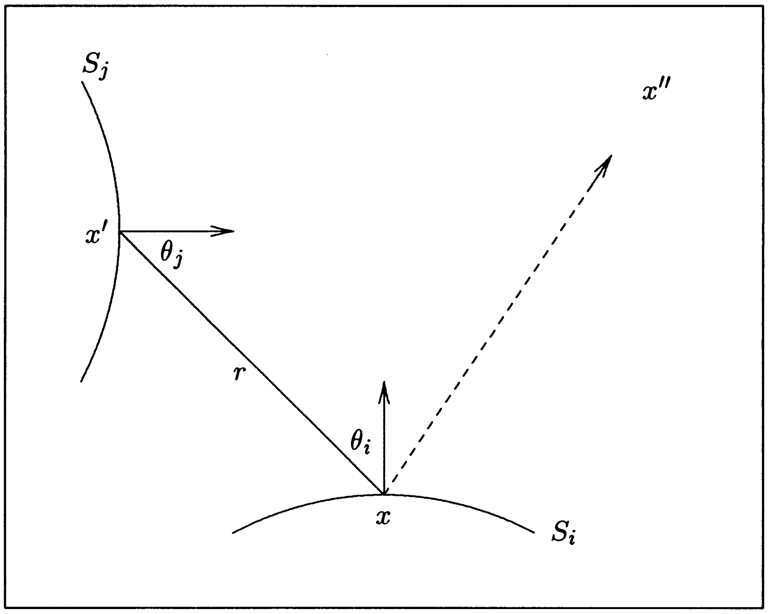


Figure 1: Geometry for Equation (1).

If all the surfaces are ideal diffuse (Lambertian) reflectors, with constant reflectivity across their areas, then for a single wavelength λ the following simplified equation is obtained:

$$L_i(x) = L_i^e(x) + \rho_i \sum_{s_j \in S} \int_{x' \in s_j} L_j(x') v(x, x') \frac{\cos\theta_i \cos\theta_j}{r^2} dA(x') \tag{2}$$

This equation can also be expressed in terms of radiosity. Recall that the radiosity, $B_i(x)$, is the total (hemispherical) light energy flux per unit area leaving surface s_i from point x. It is easy to show by integrating the radiance over the hemisphere [8] that $B_i(x) = \pi L_i(x)$. Thus, the area averaged radiosity B_i for a finite area surface i is given by

$$B_i = \frac{1}{A_i} \int_{x \in s_i} \pi L_i(x) dA(x)$$

The usual discrete radiosity equation is derived by assuming constant radiosity (and hence constant radiance) across each surface and by integrating each side of Equation (2) over the area of s_i:

$$\begin{aligned}
B_i A_i &= \int_{x \in s_i} \pi L_i(x) dA(x) \\
&= \int_{x \in s_i} \pi L_i^e(x) dA(x) + (\pi \rho_i) \sum_{s_j \in S} \int_{x \in s_i} \int_{x' \in s_j} L_j(x') \frac{\cos\theta_i \cos\theta_j}{r^2} v(x, x') dA(x') dA(x) \\
&= B_i^e A_i + (\pi \rho_i) \sum_{s_j \in S} B_j \int_{x \in s_i} \int_{x' \in s_j} \frac{\cos\theta_i \cos\theta_j}{\pi r^2} v(x, x') dA(x') dA(x)
\end{aligned}$$

$$= B_i^e A_i + (\pi \rho_i) \sum_{s_j \in S} B_j A_i F_{ij} \quad (3)$$

where B_i^e is the emitted radiosity and F_{ij} is the form-factor from surface i to surface j. Dividing each side of Equation (3) by A_i yields

$$B_i = B_i^e + (\pi \rho_i) \sum_{s_j \in S} B_j F_{ij} \quad (4)$$

Note that $\pi \rho_i$ is the *hemispherical to hemispherical* (as opposed to *bidirectional*) reflectivity of surface i.

3 Methodology

The solution to Equation (2) can be expressed as an infinite series, where the n-th term of the series sums up the light reflected by the surfaces in the environment after $(n-1)$ bounces [9, 14]. Such a summation, in effect, simulates the propagation of light through the environment starting from the light sources. The solution to Equation (2) can then be approximated to any accuracy by reordering the terms of the summation according to the importance of their contribution and limiting the computation to a finite number of terms, yielding an algorithm whose structure is very similar to that of progressive refinement radiosity [4]:

```
Initialization:
     For all surfaces i {
          /* Set total and unshot radiance to the emitted radiance */
          L_i := L_i^e;
          ΔL_i := L_i^e;
     }
Until convergence {
     Choose a surface j with highest unshot radiance over its area ∫_{x∈s_j} ΔL_j(x)dA(x);
     For each surface i {
          Compute the contributed radiance L_ij;
          /* Update the total and unshot radiance functions */
          L_i := L_i + L_ij;
          ΔL_i := ΔL_i + L_ij;
     }
     ΔL_j := 0;
}
```

where the contributed radiance function L_{ij} is defined as follows:

$$L_{ij}(x) = \rho_i \int_{x' \in s_j} \Delta L_j(x') v(x, x') \frac{\cos \theta_i \cos \theta_j}{r^2} dA(x') \quad (5)$$

Note that in contrast to radiosity, radiance is *not* assumed constant since ΔL_j is left inside the integral. While this might seem a minor difference, it is actually very important, since it allows working with the entire surfaces, without discretizing them into patches and elements first.

By examining the above procedure it can be seen that the computed radiance function L_i across surface i is actually a sum of functions L_{ij}. Each function L_{ij} represents the contribution of the unshot radiance, ΔL_j across surface j (a shooting surface) to the radiance across surface i (a receiving surface) at a particular iteration. Except for the simplest cases, the functions L_{ij} cannot be computed analytically. An obvious alternative is to sample each function at a finite number of locations and then to reconstruct it from these sample values.

A natural question to ask at this point is whether it is possible to reconstruct L_{ij} exactly from a finite number of sample points. The answer depends on the Fourier transform of the function. The sampling theorem [11] states that if the function is band-limited in its frequency domain, and its highest frequency is ν, it can be reconstructed exactly, if sampled at a frequency higher than 2ν. Unfortunately, even in the simplest cases, radiance functions can have infinite bandwidths in the frequency domain. A radiance function can exhibit first derivative discontinuities (cusps) at penumbra and shadow boundaries (i.e., it is only C^0 continuous). Each term in a Fourier series is a smooth function, and therefore no finite number of terms suffices for the exact representation of these cusps.

In spite of this fact, as we will see in the following sections, good approximations to radiance functions can be obtained by carefully choosing the sample locations, and the application of appropriate reconstruction techniques. In the next section we will discuss the issue of placing the sample points on the receiving surfaces, i.e., finding the locations at which radiance functions should be sampled. Section 5 presents two methods for evaluating the contributed radiance function at each sample location. Techniques for reconstructing these functions from the sampled values are investigated in Section 6. In each of the following three sections, we demonstrate the approach by means of simple two-dimensional test-cases and contrast the proposed solutions with the corresponding techniques employed by current radiosity algorithms.

4 Placement of Samples on a Receiver

The methodology presented in the previous section computes each radiance function L_i as the summation of contributed radiance functions L_{ij}. Since the contributed functions are induced by the unshot radiance across various light sources they can considerably differ from each other. Thus, it is difficult to find *a-priori* a single set of sample points that would correctly capture the contributions from each illuminating source. The approach that we propose in this section, is to find a good set of samples for each L_{ij}. $L_i(x)$ can then be computed at any point x by reconstructing the L_{ij}'s and summing their values at x.

Consider the problem of choosing a set of sample locations, sufficient for an accurate reconstruction of L_{ij}, given a receiver i and a shooting surface j. The general idea is to identify disjoint regions on the receiver, delimited by the edges of the surface and the penumbra and umbra boundaries. Initial sample points are placed on the boundaries of these regions, and each region is then adaptively supersampled. The adaptive supersampling stops when the radiance function across a subregion is approximately planar.

In contrast to the approach outlined above, in the traditional radiosity approach each surface in the environment is meshed into patches and elements before the solution process starts. The meshing process typically takes into account only the local geometry of the surface being meshed. The vertices of the resulting mesh are the locations at which radiosity is sampled.

If the initial mesh is too coarse, undersampling occurs and features such as shadow boundaries are not captured well by the solution. Radiosity algorithms typically use *adaptive subdivision* [6] to alleviate this problem: if comparison of adjacent vertices indicates a radiosity gradient above a user specified threshold, the corresponding element is subdivided into smaller elements. However, even when adaptive subdivision is used, features such as small shadows and bright spots on the surfaces might be missed entirely by the solution. If a feature is detected by at least one vertex of the mesh, then adaptive subdivision is triggered, more vertices are introduced in the areas covered by and adjacent to the feature, and it will appear in the solution. However, if the feature falls entirely between vertices no radiosity gradient is ever detected, and no subdivision takes place.

Thus, to enable the detection of small features, the initial mesh must be fine. In fact, for good results, the size of a mesh element should be smaller than the size of the feature to be detected. When the initial mesh is fine over all the surfaces in a complex environment, because of the need to store and repeatedly update many vertices, solutions result in excessive space and time requirements. In practice, in order to produce a good radiosity image, user intervention is needed.

The user must guess in which areas high radiosity gradients occur, and specify to the program that fine meshing is needed in these areas. This guessing task becomes increasingly difficult and tedious as the complexity of the environment grows.

Any automated solution to the initial mesh problem must consider the light sources illuminating a given surface and the objects that occlude these sources, casting shadows on the surface. Such information cannot be precomputed before the solution process, because each surface in the environment can become a light source. The only published solution to the problem was suggested by Campbell and Fussel [3]. In their method, each finite area light source is approximated by a set of point light sources on its surface. The mesh is created and refined by shadow boundaries that are computed with respect to these point light sources across all the surfaces in the environment. In order to get accurate results on surfaces close to the source, the number of points approximating it must be large. This creates a mesh which is unnecessarily fine on surfaces far away from the source. As new surfaces become light sources and the mesh is refined accordingly, the memory requirements of this algorithm may become prohibitive.

In the rest of this section we describe and analyze the performance of the new sampling approach by means of several two-dimensional test environments. The results obtained are compared to those produced by standard radiosity algorithms.

We begin with a simple example without occlusion. Let the receiver be the x-axis, and the light source be a line segment parallel to it. Figure 2 shows two solutions computed using the new approach and by standard radiosity procedures for such an environment. The new method starts with two initial sample points (the endpoints of the receiving segment). Adaptive supersampling then proceeds by inserting additional sample points, terminating whenever insertion of a new point in the middle of an interval implies linearity of the radiance function there. The standard radiosity approach starts with uniformly spaced initial samples (three) and proceeds by refining each interval whose endpoints values differ by more than a given threshold. The threshold was chosen so that the total number of sample points would be approximately the same used by the new method. Comparison of the results reveals that while the new method concentrates more samples in areas of higher curvature, radiosity undersamples the curved part of the radiance functions while placing unnecessarily many samples in parts which are almost linear. Thus, samples are wasted, and the results are less accurate (note the cusp at the top of the bump).

Figure 3 shows a simple two-dimensional configuration, consisting of a light source, a single occluding object and a receiver. Superimposed is the corresponding radiance function (computed

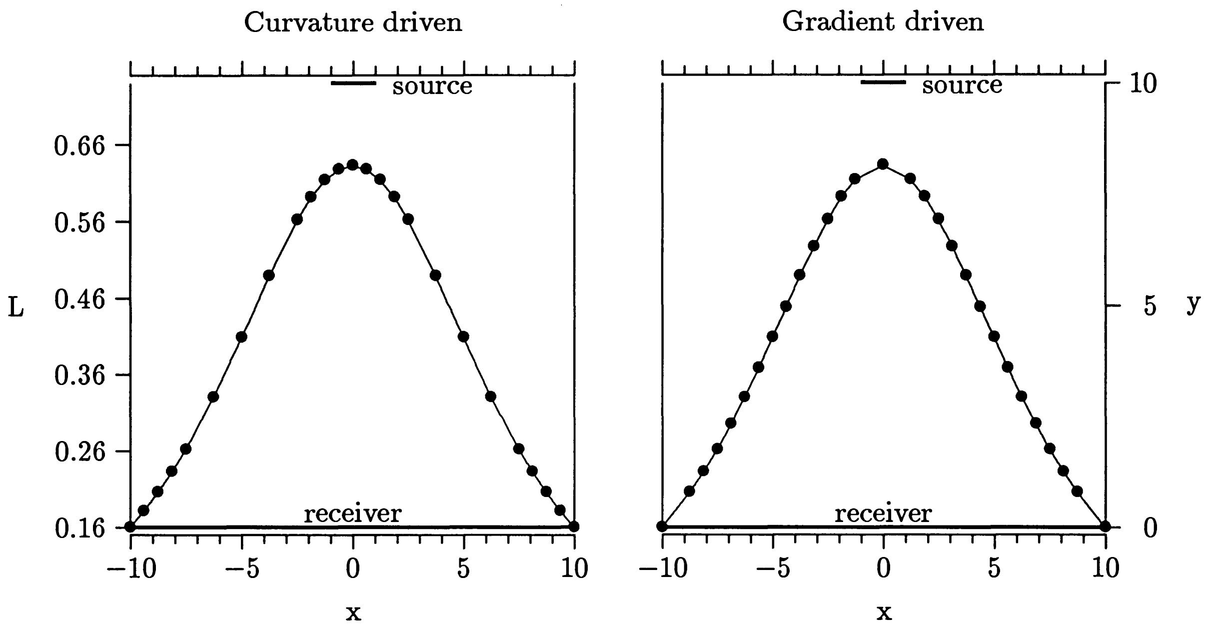


Figure 2: Solutions produced using the new approach (left) and standard radiosity (right) for a simple unoccluded 2D configuration. The samples taken by each method are marked by solid circles on each curve.

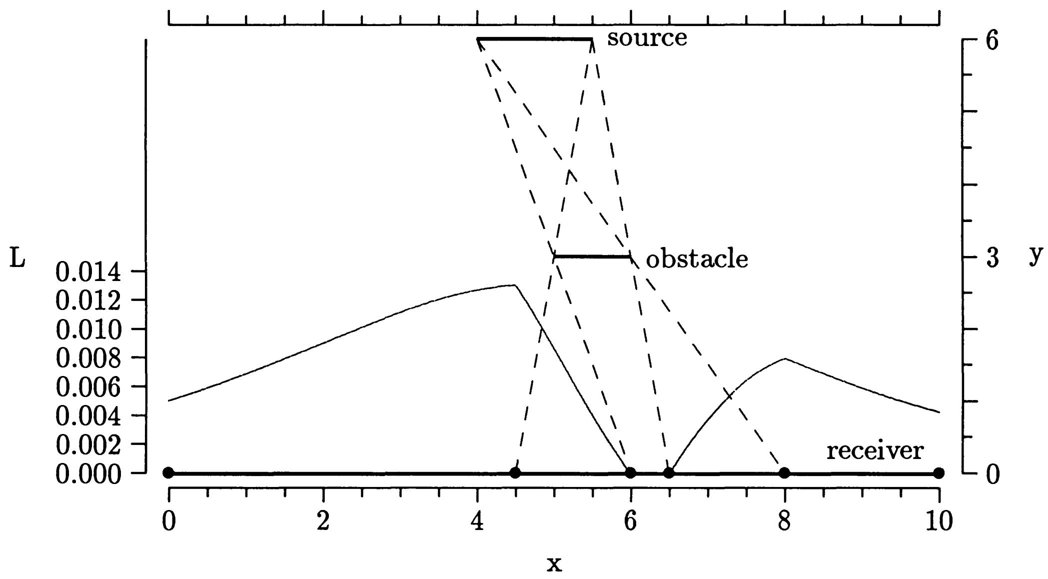


Figure 3: A simple 2D environment and the corresponding radiance function.

using brute-force numerical integration). It can be seen that the radiance function exhibits cusps (discontinuities in its first derivative) at penumbra boundaries, but is smooth elsewhere. This property suggests the following approach for placement of samples in the presence of occlusions. First, penumbra boundaries must be identified on the receiver. These can be obtained by projecting each of the obstacles on the receiver from each of the two end-points of the light source (see Figure 3). Penumbra boundaries divide the receiver into disjoint areas: regions completely unoccluded from the source, completely shadowed regions, and various penumbra regions. Placing initial sample points on the boundaries of these regions ensures that the cusps in the corresponding radiance function are captured. In the simple case of Figure 3 the initial sample points would be in the locations marked by the solid circles. The rest of the sample points are placed inside each region using the adaptive supersampling scheme described earlier. The adaptive supersampling scheme alone, as well as the adaptive subdivision schemes used by traditional radiosity systems, are liable to stop early and miss some feature of the radiance function; when applied separately on the regions enclosed between penumbra boundaries, though, its effectiveness is much greater because the radiance function over each area is smoother overall.

To illustrate the effectiveness of the suggested sampling scheme and to contrast it with standard radiosity, a two-dimensional environment consisting of a light source and seven small occluding objects was used. The geometry of the environment and the corresponding radiance function appear in Figure 4. Figure 5 shows the solution computed using the new approach. Two sample points were placed at the ends of the receiver, and additional points were located at penumbra boundaries as described above. Additional points were added by adaptive supersampling. Several results produced by standard radiosity on the same environment are shown in Figure 6. These results were obtained using various numbers of equally spaced initial samples. Because the points are spaced evenly, rather than being concentrated in important regions, undersampling occurs and the results exhibit severe errors missing both shadows and peaks of brightness on the receiver. Ironically, in this particular example, increasing the number of initial sample points only makes the result worse.

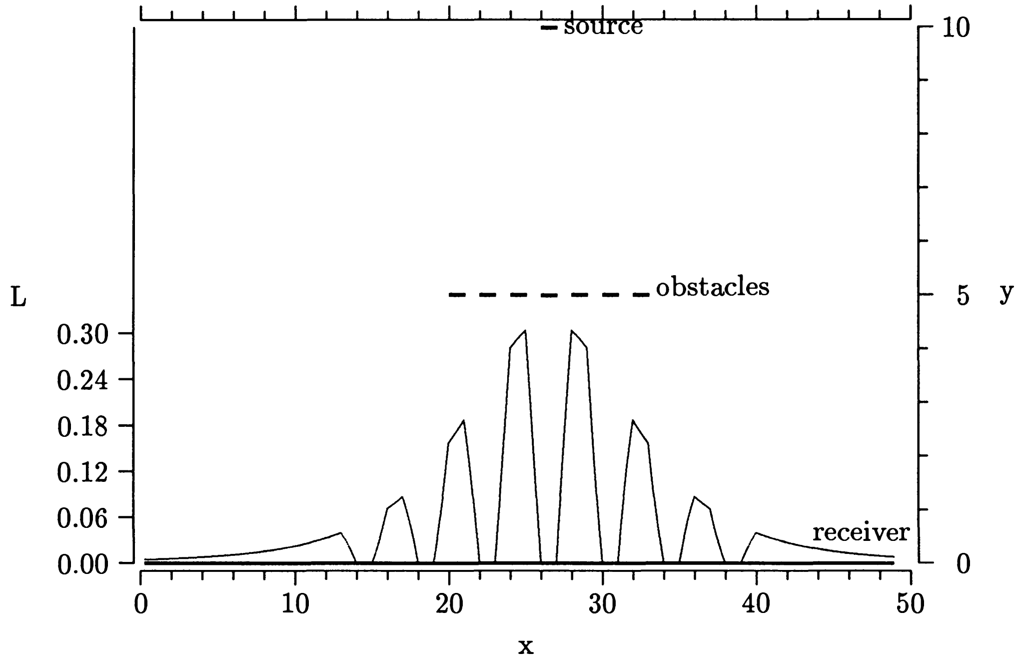


Figure 4: A more complicated 2D environment and the corresponding radiance function (obtained by numerical integration).

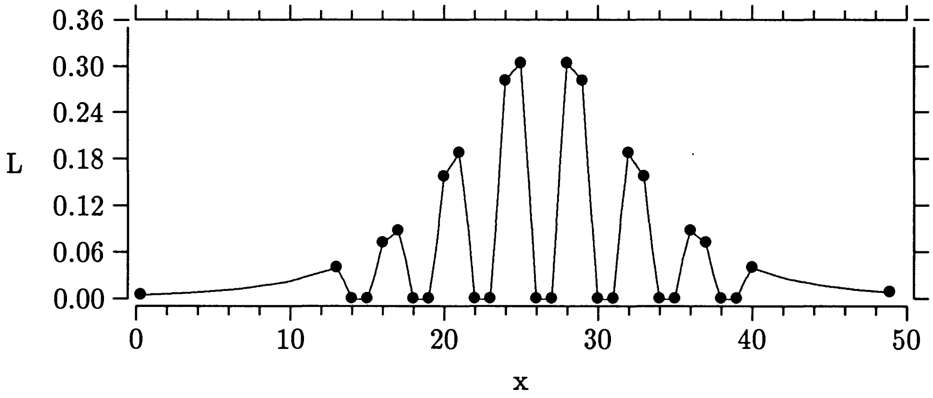


Figure 5: Location of initial sample points, and the radiance function computed by the new method.

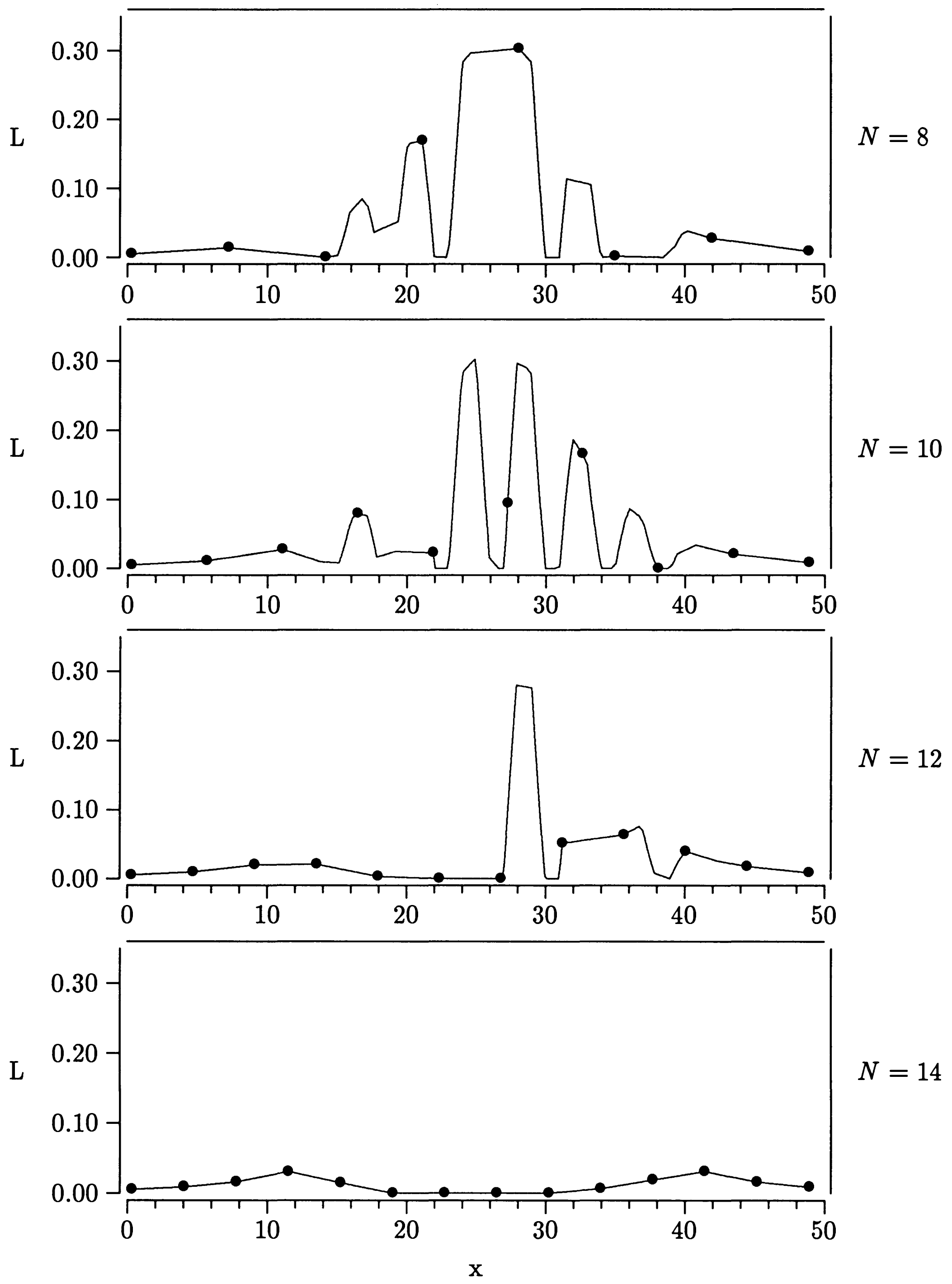


Figure 6: The radiance functions computed by standard radiosity using 8,10,12, and 14 initial samples.

5 Evaluating the Contributed Radiance Function

Given a location x on a receiving surface i and a shooting surface j, we need a method for accurately computing $L_{ij}(x)$. In standard radiosity this corresponds to computing the contribution of a shooting patch j to the radiosity of a vertex on a receiving patch i, given by $\rho_i dF_{j,di} B_j A_j$, where $dF_{j,di}$ is the form-factor from (finite area) patch j to a (differential area) vertex on patch i. However, when the distribution of energy on the shooting surface is non-uniform, the traditional radiosity formulation which assumes that the radiosity B_j is constant across patch j, is inherently inaccurate.

Consider the simple example illustrated in Figure 7. Assume that the right half of patch j is brightly illuminated, while the left half is shadowed. Let vertex l face the dark half, and vertex k face the bright half. Clearly, from symmetry, the form factors $dF_{j,dl}$ and $dF_{j,dk}$ are equal. Thus the same contribution will be computed for both vertices, when, in fact, the contribution to vertex l should be much smaller than to vertex k. Naturally, the inaccuracies diminish when patches are small, distances between patches are large, or the radiosity on the shooting patch is relatively constant.

When considering a shooting surface with unshot radiance ΔL_j, an accurate estimate of the contribution at a given location x can be computed by using a straightforward extension of the method used by Wallace *et al.* [15]:

$$L_{ij}(x) = \rho_i \sum_{k=1}^{n} \Delta L_j(x_k) A_{jk} \pi \frac{\cos\theta_{ik}\cos\theta_{jk}}{\pi r_k^2 + A_{jk}} v(x, x_k) \tag{6}$$

This formulation assumes adaptive subdivision of the source for each location x into possibly unequal areas A_{jk}. The subdivision is performed essentially in the same way as described in [15]. Note that each area A_{jk} is multiplied by the unshot radiance in its center x_k, rather than

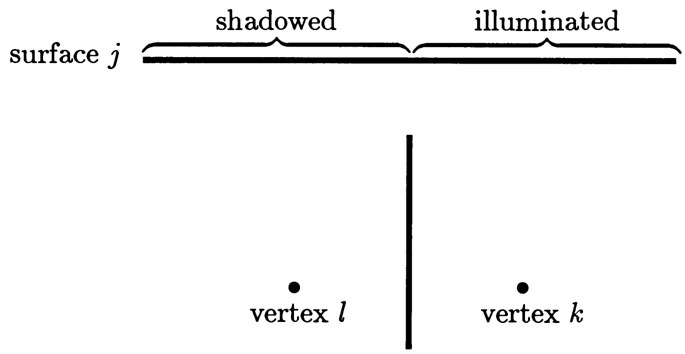


Figure 7: A simple example where the constant radiosity assumption leads to inaccurate results.

by an average value of radiance across the entire surface. The term $\Delta L_j(x_k)A_{jk}$ is in fact an approximation to the integral $\int_{x \in A_{jk}} \Delta L_j(x)dA(x)$. More accurate results can be obtained if better estimates of these integrals are available. Figure 8 shows a comparison between the new method and standard radiosity procedures. Note how the new method places more samples in the brighter part of the source, while the standard one samples the source uniformly because of the constant radiosity assumption.

Equation (6) produces very good results with a small number of samples on the source when the source is unoccluded (or totally occluded) with respect to the point x on the receiver. However, as with traditional radiosity, in the case of partial occlusion, results are prone to error due to the fact that the placement of the samples x_k on the source takes no account of the occluding geometry. For instance, small occluding objects can be missed by all the samples from one location on the receiver, but detected from a point next to it, thus resulting in aliasing. Inaccuracies in evaluating L_{ij} have another undesirable side-effect: they slow down the convergence of the adaptive supersampling of the receiver's radiance function.

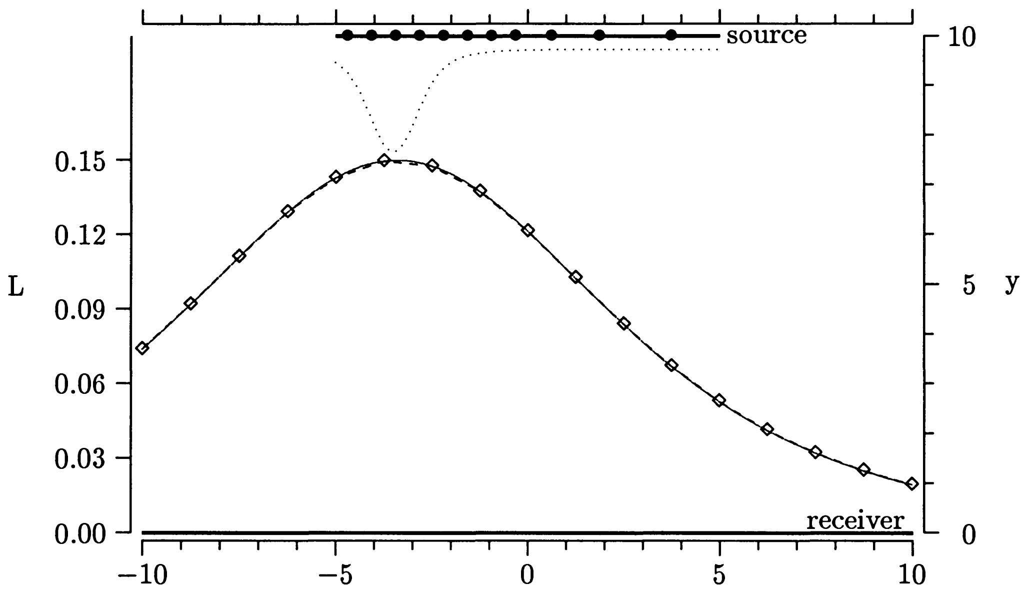
More accurate computation of $L_{ij}(x)$ can be performed by identifying the regions on the shooting surface visible from x. Then for each of these regions Equation (6) can be used, omitting the $v(x, x_k)$ term, since visibility for each region has already been resolved. Although this approach requires solving a restricted version of a hidden surface problem for each sample point on the receiver, it was found to be more efficient than expected. This fact is explained by the following considerations:

- For each location x, only the objects in the frustum defined by x and the shooting surface j need be considered.

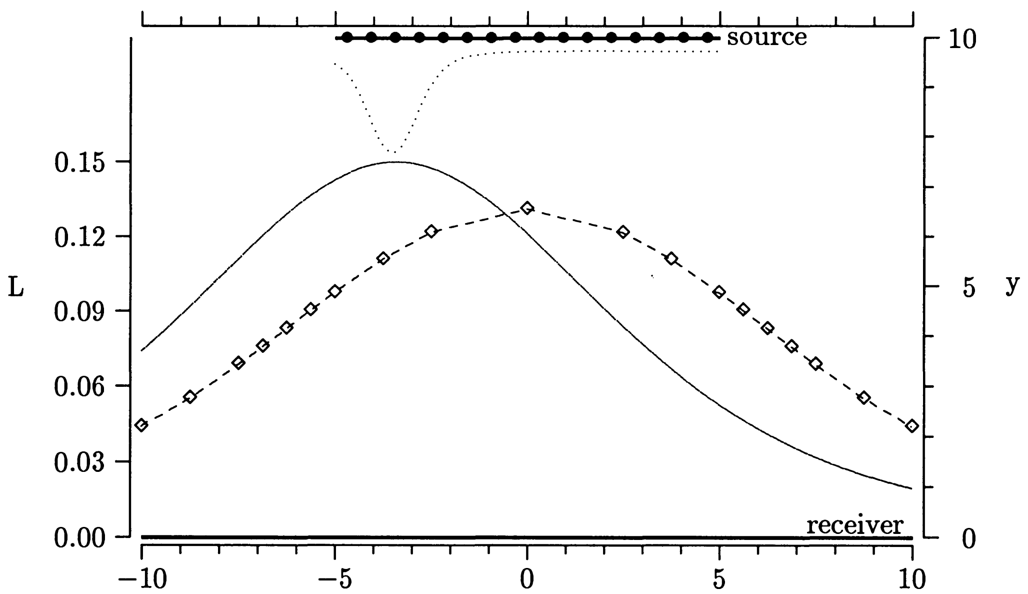
- Once the regions visible from x have been identified, no rays need to be shot (no need for ray-object intersections).

- Since the radiance at points on the receiver is computed accurately, adaptive supersampling converges faster and thus the computation needs to be performed less frequently.

To summarize, for good results, the latter, more accurate method should be used when the shooting surface is one of the primary light sources, e.g. emitters and bright, large surfaces. The former, cheaper method can be used with secondary light sources whose contribution is less significant.



(a)



(b)

Figure 8: A source with nonuniform radiance (shown dotted) results in an non-symmetric, shifted, bump-shaped radiance function (solid) over the receiver. The new method (a) produces virtually the same result (dashed). Standard radiosity (b) produces a symmetric bump (dashed) significantly different from the actual radiance function. Marked by solid circles are the locations at which the light source was sampled from point $x = 0$ on the receiver.

6 Function Reconstruction

In the previous two sections, attention was given to methods of accurately capturing the behavior of a radiance function across a surface by a finite set of opportune samples. For display the function needs to be reconstructed and then resampled at locations determined by the particular display algorithm used, e.g. z-buffer, ray tracing, or other.

Both filtering techniques used in image processing [16] and surface fitting and interpolation methods [7] can be used to reconstruct a continuous function from a set of samples. However, to date, little attention has been devoted to the problem of reconstructing a radiance function over the surfaces of a three-dimensional environment.

Filtering is usually expressed as the convolution of a set of regularly spaced samples with a continuous interpolation kernel. Interpolation can also be formulated as a convolution operation, but it is simpler to apply it in its original form when reconstructing a function from scattered sample data. Since the sampling methods examined in Section 4 produce a set of irregularly spaced samples, interpolation techniques are suggested for the reconstruction process.

A good reconstruction technique should yield a function that not only closely approximates the actual radiance function, but also exhibits the properties characteristic of the actual function; (1) non-negativity, (2) continuity, and (3) piecewise-smoothness, i.e. at least C^1 continuity over the regions delimited by shadow and penumbra boundaries. While the first characteristic simply constrains the interpolated values to be physically meaningful, the other two relate to the ability of our visual system to detect rapid changes of the illumination gradient; first derivative discontinuities introduced by the reconstruction process result in disturbing visual artifacts, e.g. Mach banding, that severely affect the realism of synthesized images. The fact that the penumbra boundaries correspond to first derivative discontinuities of the radiance function allows piecewise smooth interpolation schemes to be used.

The shadow and penumbra boundaries found by the sampling strategy presented in Section 4 define a set of disjoint subregions over which the radiance function is smooth. An interpolation method that provides smoothness can be applied independently on each of such subregions and care can be taken to maintain continuity at subregion boundaries, thus resulting in an overall interpolation scheme that possesses the characteristic properties of an actual radiance function.

Interpolating cubic splines, cubic B-splines using the sampled radiance values as control points, and C^1 piecewise cubic Hermite interpolation [7] were tested for the interpolation over a single subregion.

Again, for simplicity and clarity of presentation, the analysis and comparison of the interpolation schemes was carried out in two dimensions. Cubic splines were found to exhibit undesirable oscillations. Furthermore, the resulting curves are not guaranteed to be non-negative. The curves generated using B-splines are non-negative, but could present a significant error in the proximity of local maxima or minima unless additional control points are added for a tighter fit. The best results were obtained using C^1 piecewise cubic Hermite interpolation. An estimate of the first derivative at each sample point was computed by fitting a parabola through the point and its two neighbors and computing the value of its derivative at the point of interest. The resulting curve can be negative, but the offending curve segment can be easily detected and replaced preserving C^1 continuity. Figure 9 shows the radiance across a receiving surface contributed by a single source in the presence of an obstacle. The interpolating curve produced by piecewise-smooth Hermite interpolation is shown superimposed over the reference solution.

7 Conclusions

In this paper, a new formulation for the view independent diffuse global illumination problem was presented. This new formulation does not require the discretization of the surfaces in the environment into patches and elements of constant radiosity. Rather, an attempt is made to compute a continuous radiance function for each surface.

A methodology was developed that suggests solving the problem by means of sampling and reconstruction techniques. Such an approach requires solving three major subproblems:

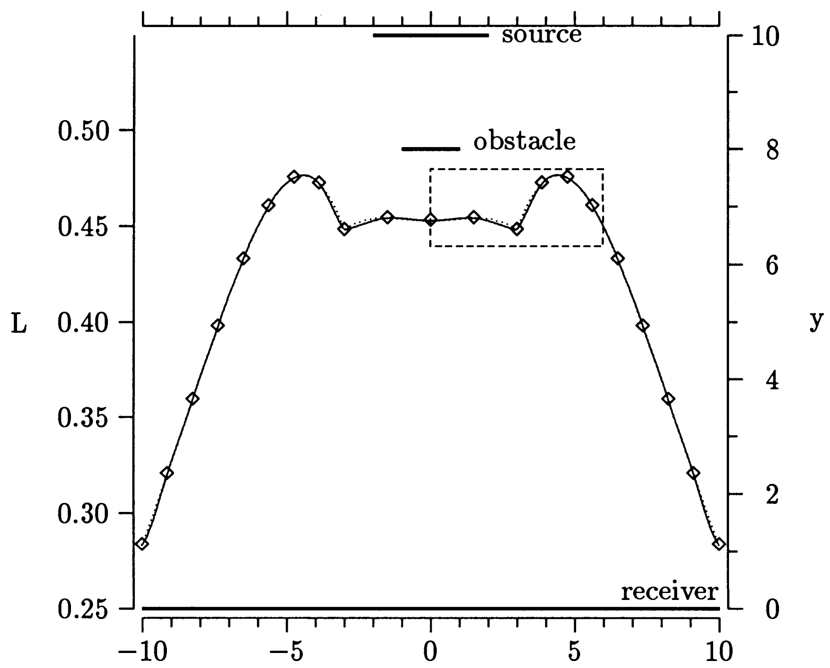
1. Decide where each radiance function should be sampled.

2. Evaluate the function at each sample point.

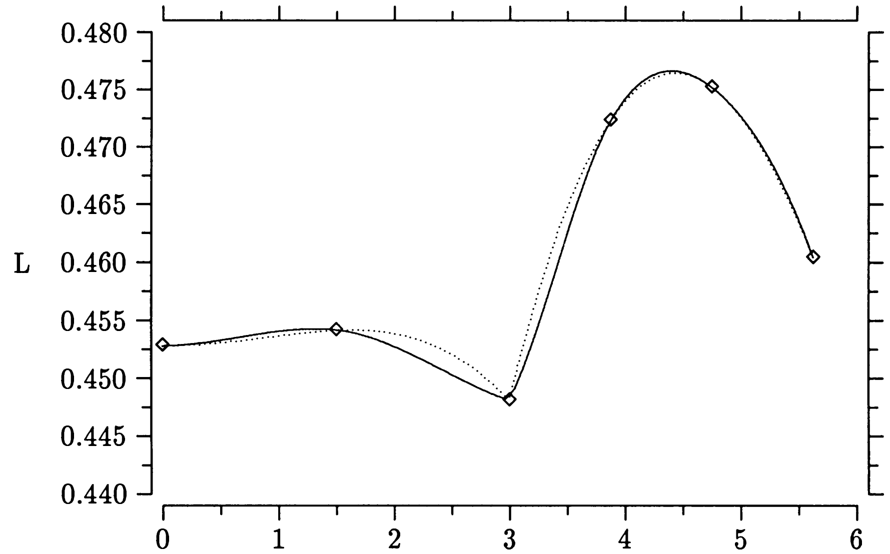
3. Reconstruct the function from its sampled values.

Solutions for these subproblems were suggested. Experiments were performed on simple two-dimensional test-cases which compared the solutions provided by the advocated techniques and by their respective counterparts in existing radiosity methods with the actual radiance functions. The results of these comparisons clearly indicate that the new approach produces a much tighter match than that produced by standard radiosity algorithms.

The development of a full three-dimensional implementation is currently underway. Although the ideas and principles described in the paper are restricted neither to two dimensions nor to



(a)



(b)

Figure 9: (a) Superposition of an actual radiance function (dotted) and a piecewise-smooth Hermite interpolated curve (solid) for a simple test configuration. The interpolated points are marked by small diamonds. (b) Enlargement of the area marked by the dashed rectangle.

planar geometries, extending the implementation to three dimensions is not straightforward. For environments containing only planar polygons, the shadow and penumbra boundaries can be found in a way similar to that described by Nishita and Nakamae [10], but more general geometries will require more sophisticated techniques.

The methods presented in the paper yield a radiance function representing the contribution from a given source to a receiving surface. To compute a complete solution, many such functions must be summed, one for each contributing source. Since each function is represented by samples taken at a possibly different set of locations on the receiver, their addition is not straightforward and efficient solutions are under investigation.

Finally, the approach is currently limited to ideally diffuse reflectors and future research should be directed at extending it to more general bidirectional reflectance functions.

Acknowledgements

The authors acknowledge the support of the National Science Foundation under a grant entitled "Interactive Input and Display Techniques" (CCR8617880) and the Hewlett-Packard Corporation and the Digital Equipment Corporation for generous donations of equipment.

References

[1] Baum, Daniel R., Holly E. Rushmeier, and James M. Winget. "Improving Radiosity Solutions Through the Use of Analytically Determined Form-Factors," Proceedings of SIGGRAPH'89 (Boston, Massachusetts, July 31–August 4, 1989), in *Computer Graphics*, 23(3), July 1989, pages 325–334.

[2] Bouville, Christian, Kadi Bouatouch, Pierre Tellier, and Xavier Puyeo. "Theoretical Analysis of Global Illumination Models," in Proceedings of Eurographics Workshop on Photosimulation, Realism and Physics in Computer Graphics (Rennes, France, June 11–12, 1990), June 1990, pages 53–66.

[3] Campbell, A. T., III and Donald S. Fussell. "Adaptive Mesh Generation for Global Diffuse Illumination," Proceedings of SIGGRAPH'90 (Dallas, Texas, August 6–10, 1990), in *Computer Graphics*, 24(4), August 1990, pages 155–164.

[4] Cohen, Michael F., Shenchang Eric Chen, John R. Wallace, and Donald P. Greenberg. "A Progressive Refinement Approach to Fast Radiosity Image Generation," Proceedings of SIGGRAPH'88 (Atlanta, Georgia, August 1–5, 1988), in *Computer Graphics*, 22(4), August 1988, pages 75–84.

[5] Cohen, Michael F. and Donald P. Greenberg. "The Hemi-Cube: A Radiosity Solution for Complex Environments," Proceedings of SIGGRAPH'85 (San Francisco, California, July 22–26, 1985), in *Computer Graphics*, 19(3), July 1985, pages 31–40.

[6] Cohen, Michael F., Donald P. Greenberg, and David S. Immel. "An Efficient Radiosity Approach for Realistic Image Synthesis," *IEEE Computer Graphics and Applications*, 6(2), March 1986, pages 26–35.

[7] Farin, Gerald. *Curves and Surfaces for Computer Aided Geometric Design*, *Computer Science and Scientific Computing*, Academic Press, Inc., San Diego, California, second edition, 1990.

[8] Goral, Cindy M., Kenneth E. Torrance, Donald P. Greenberg, and Bennet Battaile. "Modeling the Interaction of Light Between Diffuse Surfaces," Proceedings of SIGGRAPH'84 (Minneapolis, Minnesota, July 23–27, 1984), in *Computer Graphics*, 18(3), July 1984, pages 213–222.

[9] Kajiya, James T. "The Rendering Equation," Proceedings of SIGGRAPH'86 (Dallas, Texas, August 18–22, 1986), in *Computer Graphics*, 20(4), August 1986, pages 143–150.

[10] Nishita, Tomoyuki and Eihachiro Nakamae. "Continuous Tone Representation of Three-Dimensional Objects Taking Account of Shadows and Interreflections," Proceedings of SIGGRAPH'85 (San Francisco, California, July 22–26, 1985), in *Computer Graphics*, 19(3), July 1985, pages 23–30.

[11] Shannon, Claude E. "A Mathematical Theory of Communication," *Bell System Tech. Journal*, Vol. 127, 1948, pages 379–423, 623–656.

[12] Shirley, Peter. "Physically Based Lighting Calculations for Computer Graphics: a Modern Perspective," in Proceedings of Eurographics Workshop on Photosimulation, Realism and Physics in Computer Graphics (Rennes, France, June 11–12, 1990), June 1990, pages 67–82.

[13] Siegel, Robert and John R. Howell. *Thermal Radiation Heat Transfer*, Hemisphere Publishing Corp., Washington D.C., 1981.

[14] Sillion, François and Claude Puech. "A General Two-Pass Method Integrating Specular and Diffuse Reflection," Proceedings of SIGGRAPH'89 (Boston, Massachusetts, July 31–August 4, 1989), in *Computer Graphics*, 23(3), July 1989, pages 335–344.

[15] Wallace, John R., Kells A. Elmquist, and Eric A. Haines. "A Ray Tracing Algorithm for Progressive Radiosity," Proceedings of SIGGRAPH'89 (Boston, Massachusetts, July 31–August 4, 1989), in *Computer Graphics*, 23(3), July 1989, pages 315–324.

[16] Wolberg, George. *Digital Image Warping*, IEEE Computer Society Press, Los Alamitos, California, 1990.