

Monte Carlo Approximation of Form Factors with Error Bounded *a Priori**

M. Pellegrini

Istituto di Matematica Computazionale del C.N.R.,
Via S. Maria 46, 56126 Pisa, Italy
pellegrini@imc.pi.cnr.it

Abstract. The exchange of radiant energy (e.g., visible light, infrared radiation) in simple macroscopic physical models is sometimes approximated by the solution of a system of linear equations (energy transport equations). A variable in such a system represents the total energy emitted by a discrete surface element. The coefficients of these equations depend on the *form factors* between pairs of surface elements. A form factor is the fraction of energy leaving a surface element which directly reaches another surface element. Form factors depend only on the geometry of the physical model. Determining good approximations of form factors is the most time-consuming step in these methods, when the geometry of the model is complex due to occlusions.

In this paper, we introduce a new characterization of form factors based on concepts from integral geometry. Using this characterization, we develop a new and asymptotically efficient Monte Carlo method for the simultaneous approximation of all form factors in an *occluded* polyhedral environment. The approximation error is bounded without recourse to special hypothesis. This algorithm is, for typical scenes, one order of magnitude faster than methods based on the hemisphere paradigm or on Monte Carlo ray-shooting.

Let A be any set of convex nonintersecting polygons in R^3 with a total of n edges and vertices. Let ε be the error parameter and let δ be the confidence parameter. We compute an approximation of each *nonzero* form factor such that with probability at least $1 - \delta$ the absolute approximation error is less than ε . The expected running time of the algorithm is $O((\varepsilon^{-2} \log \delta^{-1})(n \log^2 n + K \log n))$, where K is the expected number of *regular intersections* for a random projection of A . The number of regular intersections can range from 0 to quadratic in n , but for typical applications it is much smaller than quadratic. The expectation is with respect to the random choices of the algorithm and the result holds for any input.

* A preliminary version of this paper appeared in *Proceedings of the 11th ACM Symposium on Computational Geometry*, pp. 287–296, June 1995, Vancouver.

1. Introduction

1.1. *Radiant Energy Transport in Heat Engineering*

The determination of radiation interchange between surfaces is a central problem in heat transfer, illumination engineering, and applied optics [SH1]. Complex radiation interchange problems are simplified using a variety of assumptions and approximations. One of the simplest models in heat transfer is that in which (i) each surface is “black” (i.e., there is no reflected radiation and all emitted energy is diffuse) and (ii) each surface emits diffuse radiation uniformly in each direction and from each point of the surface (see [SH1] for an extended treatment). The balance of energy in the model is described by a system of linear equations (energy transport equations). For a surface i in a set of n surfaces, let E_i be the emissivity (i.e., the energy produced by the surface i), let ρ_i be the diffuse reflectance (i.e., the proportion of received energy that is re-emitted), and let B_i be the total energy emitted by surface i . Denoting by F_{ij} the form factor between surface i and surface j , the energy transport equation for surface i is

$$B_i = E_i + \rho_i \sum_j F_{ij} B_j,$$

where the summation is extended to all surfaces different from i . All the transport equations form a linear system in the variables B_1, \dots, B_n . The values of ρ_i and E_i are known for a given model. The form factor (also called *configuration factor* or *view factor*) is defined as *the fraction of the diffuse energy leaving one surface that directly reaches another surface*. Under the conditions of the model, form factors depend only on the geometry of the surfaces and are independent from all the other parameters of the model. The determination of the form factor is an essential preliminary step to the solution of the system.

1.2. *Radiosity Problems in Graphics*

In graphics, the goal of synthesizing realistic images has been pursued by computing the balance of light reaching each surface. Such an approach is known as *radiosity computation*. ACM SIGGRAPH and Eurographics annual conferences have sessions devoted to radiosity computations, and since 1990, Eurographics has organized annual workshops on rendering, featuring many papers on radiosity. The SIGGRAPH on-line bibliography lists more than 240 titles including the word “radiosity” in the period from 1983 to 1995. Some papers often cited in the literature are: [GTGB], [NN], [CG], [K], [WRC], [WEH], [SP], [HSA], [M1], and [Sh]. For more recent results we refer the reader to the proceedings of specialized conferences and workshops (e.g., [HP], [BJ], and [TT]).

When we consider only the effect of diffuse light we can use assumptions similar to those used in the study of heat transfer and we obtain the same model described above (in graphics known as the Lambertian model). Again, the determination of form factors is a central computational problem.

Exact analytical computation of form factors is difficult except for restricted special cases, none of which considers occlusions. References to analytic solutions for special

cases or to computer programs for general cases can be found in the book by Siegel and Howell [SH1]. Several approximation schemes have been devised to cope with realistic environments.

1.3. Current Approximation Algorithms

There are currently several methods available for computing approximations of form factors. *Monte Carlo ray tracing* is chronologically the first method proposed. This method is still much used and researched because of its flexibility and ease of implementation. The method is based on simulating the exchange of energy by performing ray-shooting operations. One of the most popular schemes for a direct estimation of form factors is the *hemicube algorithm* [CG], which is an elaboration of the hemisphere method of Nusselt [SH1]. Hanrahan *et al.* [HSA] have proposed methods for computing form factors based on defining each form factor as a matrix of subform factors and on an efficient hierarchical block decomposition of such matrices.

A detailed description of these and other algorithms is beyond the scope of this paper. We postpone some considerations to Sections 1.8 and 1.9. In the Appendix we give some more details while we attempt to compare these methods with the result presented in this paper.

1.4. New Results in this Paper

Within the discipline of design and analysis of algorithms, approximation methods are compared according to two main measures: one is the computational resources (time, storage) used, the second is the quality of the approximation, i.e., a bound on the error of the approximation. The two measures are not independent, in general. Until now the form factors problem has been unyielding to this type of analysis. In this paper we present the following results:

- (1) We give an integral geometric interpretation to the form factors using the style and notation of Santaló [Sa], and we show that the analytic and geometric formulations previously used are *specializations* of this formula.
- (2) We derive an efficient Monte Carlo algorithm to compute approximations of form factors for which we have *simultaneously* an asymptotic expected upper bound on the running time *and* an exact *a priori* upper bound, which holds with high probability, on the absolute approximation error of each form factor. The algorithm has better asymptotic running time, for typical scenes, than other methods for which a comparable analysis of the running time is possible (in particular, Nusselt's hemisphere method, and Monte Carlo ray tracing).

More precisely the algorithmic result in this paper is the following. Let A be a set of convex nonintersecting polygons in R^3 with a total of n edges and vertices.¹ We fix an

¹ We can think of these polygons as forming a covering of the facets of the input polyhedra.

error bound ε and a confidence δ . Let $K_r(A, v)$ be a certain subset of the intersection points among the orthogonal projections of the edges of A in direction v , which we call, following Mulmuley, *regular intersections*. We denote by K the expected size of $K_r(A, v)$ over a choice of v uniform at random in the set of all directions.

Theorem 1. *We can compute in expected time $O((\varepsilon^{-2} \log \delta^{-1})(n \log^2 n + K \log n))$ an approximation of each nonzero form factor such that with probability $1 - \delta$ the absolute error on each form factor is less than ε .*

The number of regular intersections can range from 0 to quadratic in n , but for typical applications it is much smaller than quadratic. A typical input is a set of polygons in which polygons are rather fat or, to put it negatively, not long and thin. This is a typical input because in these conditions the assumptions of the Lambertian model are more likely to be realistic.

1.5. What Is a Form Factor?

The first task we set out to accomplish is to give a simple definition of a form factor in the style and notation of Santaló [Sa]. The connection between the measure of sets of lines and form factors is the basis behind any derivation of the form factors integral. Here we apply this idea in a way fundamentally different from previous known derivations. The original definition of a form factor, as the fraction of diffuse energy leaving a surface that directly reaches another surface, belongs more to physics than to geometry. A further elaboration [SH1] equates the form factor to a double area integral in which the kernel function is diverging, thus creating problems of numerical instability. Nusselt gives a geometric interpretation of a form factor in terms of central projections on the surface of a sphere, which holds only when one of the two surfaces is infinitesimally small (a so-called differential area). Alternatively, researchers have used the connection between measure of lines and form factors to develop sampling strategies in spaces of rays or lines.

The author whose starting point is closest to the one used in this paper is Sbert [Sb] (see also [SPNP] and [SPP]). After noting that Lambert's cosine law is equal to one of the expressions for the differential element of lines, Sbert interprets the form factor as a geometric probability in a space of lines. Then a method is proposed based on choosing a set of lines randomly and counting the intersections between the lines and objects.

Starting from first principles, we show that form factors can be interpreted as the ratio between the measure of certain sets of lines in 3-space, *which is independent of any particular parametrization of lines*. We then show that the characterization of form factors previously used in the literature are equivalent to our "abstract" characterization for specific choices of line coordinates. The benefit of starting from first principles is that proving the correctness of the new integral expressions of form factors becomes almost trivial within the framework provided by Santaló's exposition of integral geometric theory.

1.6. *A New General Method*

Our definition of form factor via integral geometric theory is abstract in the sense that it is not immediately computable, and also in the sense that it is independent of any coordinates of lines in 3-space. When we decide on specific coordinates of lines we obtain an equivalent formulation that is, in principle, computable.

The next step in our research hinges on the choice of coordinates of the lines, so that the resulting concrete formulation for the form factor contains quantities that can be computed easily and efficiently using computation geometry techniques [PS]. In our case we use only orthogonal projections and the area of planar convex polygons, for this reason we refer to our method as the *orthogonal projection method*. This formulation involves an integral which is simpler than those previously known.

1.7. *An Efficient Algorithm*

We use Monte Carlo integration to approximate the value of the integral defining the form factor as derived by the previous analysis. The integration is based on sampling on the *unit sphere of directions*. This domain has only two parameters and has a bounded measure. These are important properties that we use to derive an exact bound on the variance of the integrand function and, consequently, a bound on the absolute error of the Monte Carlo integration.

For the case of two polygons, the algorithmic techniques needed are standard and can be found in [PS]. For the case of many polygons, we use the vertical cylindrical decomposition of Mulmuley [M2] to compute efficiently the terms of the summations in the Monte Carlo approximation of the integral. Such data structures are built via plane sweep and dynamic maintenance of polygonal planar maps.

1.8. *On the Difficulty of Comparing Algorithms*

Comparing the result presented in this paper with standard practice in graphics is difficult. The main contribution of our work on the theoretical side is a provable correlation between error and running time within the standard model used in computational geometry. An additional benefit is that the algorithm is almost linear in the number n of input vertices and in the average number K of regular intersections among edges.

Error analysis published in the graphics literature to my knowledge use additional assumptions, sometimes justifiable on pragmatic grounds, but usually avoided in computational geometry. For example, a widely referenced paper in the computer graphics community by Shirley [Sh] proves that a straightforward Monte Carlo radiosity computation requires a number of rays *linear* in the number of objects present in the scene to attain estimates of the energy emitted by each surface within a predefined variance. This result is obtained in a rather strong model for which, among other assumptions, the total surface of the objects remains constant as the number of objects increases [Sh, p. 462]. No such assumption is used in this paper.

In [Sb] the input objects are enclosed in a sphere S and a certain number of random

lines meeting the sphere is produced. The number of lines needed to obtain a predefined error bound with high probability on each form factor depends on the ratio between the area of the surface of the sphere and the smallest area of an object. A further case that is analyzed in [Sb] is one in which the number of patches increases as a result of a *subdivision* of the objects initially present at the scene. Our result does not depend on any hypothesis on the minimum area of the surfaces or on the diameter of the input; and it holds also when the number of objects increases, say, by the addition of new objects to the scene.

Most of the currently known algorithms are hard to analyze and therefore the correlation between running time and approximation error has been investigated by means of experiments, rather than algorithmic analysis. Another approach has been to produce error analysis *a posteriori*, that is, an error analysis based on values computed by the algorithm itself [LSG].

In the Appendix we attempt the following type of comparison. We translate some of the popular methods within the model used by our algorithm. Whenever possible we try to use exact computations instead of nonexact computations whose contribution to the error is not quantified.

1.9. On the Standard Monte Carlo Ray-Tracing Technique

In the standard Monte Carlo ray-tracing technique we send random rays from a surface and we count the number of such rays reaching a second surface. The ratio between the two numbers is an estimator of the form factor relative to the two surfaces.

It is easy to prove that for a predefined level of error and confidence the needed number of such rays is constant per surface (see, e.g., [Sb]), and roughly proportional to ε^{-2} . Thus to estimate all the form factors we need to trace $O(\varepsilon^{-2}n)$ rays among n surfaces. The complexity of ray tracing is still an open field of investigation. Currently the best asymptotic bounds in exact analytical models for ray tracing require roughly $O(m^{0.8}n^{0.8})$ operations to trace m rays among n triangles [P1], [AM]. These methods, besides being quite complex and mostly of theoretical interest, are not known to exhibit a dependence on other combinatorial parameters of the input, such as the number K of regular intersections. An interesting feature of our algorithm is that, still in an exact model, we are able to derive a simple algorithm whose running time is almost linear in n and K .

From a more practical point of view, it is well known that, as a rule of thumb, sampling in a lower-dimensional space reduces the variance of an estimate by a Monte Carlo method [BGS⁺]. In the case of a straightforward Monte Carlo ray tracing where the rays are sampled uniformly among those meeting the first object, the variance of the associated random variable is equal to $F_{12} - F_{12}^2$. The expression for the variance in our method, given in (11), is bounded by $2F_{12} - F_{12}^2$, and is much smaller than such bounds in certain cases. For example, when the area of the second surface is less than half of the area of the first surface the variance of our method is smaller than the variance of straightforward Monte Carlo ray-shooting. My conjecture is that for values of the form factor below a threshold the variance of our method is always better than the variance of the ray-shooting method.

1.10. *Organization of the Paper*

In Section 2 we give the two characterizations of form factors used most frequently in the literature. In Section 3 we introduce an “abstract” characterization via integral geometry. In Section 4 we derive a new “concrete” integral definition of the form factor. In Section 5 we derive the error bound on the Monte Carlo evaluation of such an integral. In Section 6 we give the algorithm to compute all form factors simultaneously. In Section 7 we speculate on possible extensions of our method to non-Lambertian models, where the total energy emitted by a surface is not uniform. In the Appendix we attempt a rough comparison between several methods for form factor computations.

2. **Currently Used Concrete Characterizations**

The algorithms known in the literature to compute form factors are mainly based on two formulations of form factors. In the first formulation (which is sometimes taken as the definition tout court) the form factor is a weighted double area integral [SH1]. Given a set of disjoint surfaces² \mathcal{S} in 3-space, the form factor between two surfaces S_i and S_j in \mathcal{S} is

$$F_{ij} = \frac{1}{A_i} \int_{p \in S_i} \int_{q \in S_j} f(p, q) dp dq, \tag{1}$$

where

$$f(p, q) = \frac{\cos \theta_i(p, q) \cos \theta_j(p, q)}{\pi |pq|^2} V(p, q, \mathcal{S});$$

A_i is the area of the surface S_i ; $\theta_i(p, q)$ is the angle between the normal vector to S_i at point p and the line through p and q ; $\theta_j(p, q)$ is the angle between the normal vector to S_j at point q and the line through p and q ; $|pq|$ is the distance between point p and point q ; and $V(p, q, \mathcal{S})$ is a predicate that has value 1 if the open segment pq does not meet any surface, and is 0 otherwise.

The second characterization due to Nusselt [SH1] defines the form factor between a differential surface of area dA_i and a finite surface A_j geometrically as follows. Consider the set of points of A_j visible from a point $p \in dA_i$ and the central projection from p of such a set onto the surface of a sphere centered at p . Orthogonally project such a set of points from the sphere onto the plane tangent to dA_i at p . The area of the set of points obtained from the second projection is the value of the (differential) form factor between dA_i and A_j . To obtain the value of the form factor between S_i and S_j we then need to integrate over S_i and divide by A_i . The analytic equivalent of Nusselt’s formulation is

$$F_{ij} = \frac{1}{A_i} \int_{p \in S_i} \int_{\omega \in \Omega} g(p, \omega) dp d\omega, \tag{2}$$

where

$$g(p, \omega) = \frac{\cos \theta_i(p, \omega)}{\pi} V(p, \omega, S_j, \mathcal{S});$$

² In this paper we consider mainly polygonal surfaces, but most observations hold for a larger class of surfaces.

Ω is the set of directions; $\theta_i(p, \omega)$ is the angle formed by the normal to S_i at p and the ray from p in direction ω ; and $V(p, \omega, S_j, \mathcal{S})$ is a predicate whose value is 1 if the ray from p in direction ω meets S_j before any other surface in \mathcal{S} . An analytic closed form of Nusselt's characterization of the differential form factor for polygonal objects in 3-space was found by Lambert [L] (see also [A]).

3. An Abstract Characterization

3.1. Unoccluded Case

We give an abstract characterization of form factors starting with the case in which we have only two polygons S_1 and S_2 and no occlusion between them. Let us go back to the original intuitive meaning of form factor as *the fraction of the diffuse energy leaving one surface that directly reaches another surface*. In radiation models with nonparticipating mediums energy is transferred along linear trajectories, i.e., along lines. Thus we can measure the total energy \mathcal{I}_1 leaving one surface S_1 by summing the energy carried by each single line. Since there are infinite lines meeting a given surface S_1 the measure is more precisely defined by an integral

$$\mathcal{I}_1 = \int_{L \cap S_1 \neq \emptyset} I(L) dL, \quad (3)$$

where $I(L)$ is the density of the energy on the line L . Under the assumptions of the Lambertian model, $I(L)$ is a constant \bar{I} . Thus we are left with a purely geometric integral to compute

$$\int_{L \cap S_1 \neq \emptyset} dL = m(\{L : L \cap S_1 \neq \emptyset\}). \quad (4)$$

We introduce the notation $m(\mathcal{L})$ to denote such an integral evaluated over the domain \mathcal{L} . It is convenient to look at formula (4) using the theory of integral geometry as revealed in the book by Santaló [Sa]. One of the concerns of integral geometry is measuring sets of a geometric object (i.e., associating a positive real number to a set) in such a way that the measure is invariant under rigid transformations of the space. In our case we want the measure of the set of lines meeting the surface S to be invariant under rigid transformation of the Euclidean three-dimensional space. The differential form dL denotes a differential for which this property of invariance holds. For any given parametrization of the lines in 3-space there is a corresponding formulation of dL that is unique up to constant multiplicative factors [Sa]. Thus formula (4) is nothing but the (invariant) measure of the set of lines meeting S_1 .

The amount of energy \mathcal{I}_{12} leaving a surface S_1 and reaching a surface S_2 , if there are no occlusions, is given by the following integral over the lines meeting S_1 and S_2 :

$$\mathcal{I}_{12} = \int_{L \cap S_1 \neq \emptyset \wedge L \cap S_2 \neq \emptyset} I(L) dL. \quad (5)$$

Since $I(L)$ is constant and equal to \bar{I} we are interested in the measure of the set of lines

meeting both surfaces given by the integral

$$m(\{L : L \cap S_1 \neq \emptyset \wedge L \cap S_2 \neq \emptyset\}). \quad (6)$$

The form factor F_{12} is given by the ratio of the two intensities (3) and (5)

$$F_{12} = \frac{\mathcal{I}_{12}}{\mathcal{I}_1} = \frac{\int_{L \cap S_1 \neq \emptyset \wedge L \cap S_2 \neq \emptyset} dL}{\int_{L \cap S_1 \neq \emptyset} dL}. \quad (7)$$

From now on we use (7) as our integral geometric definition of a form factor in the absence of occlusions. From a classical result of integral geometry adapted to polygons [Sa, p. 246] we have that

$$m(\{L \in \mathcal{L} | L \cap S_1 \neq \emptyset\}) = \pi A_1,$$

where A_1 is the area of S_1 . Since we can easily compute the area of a convex polygon, we concentrate our attention on the numerator of the fraction in (7).

Formula (7) is valid for any coordinatization of the lines. In order to obtain a more concrete formula (as opposed to the somewhat abstract formulation in (7)) we need to decide what coordinates we use for parametrizing the line L . As a consequence we have a unique expression for dL as a differential form in the chosen coordinates.

It is easy to prove that, when we use as coordinates of a line L the two intersection points of L with the surfaces S_1 and S_2 , formulation (7) becomes formula (1) [Sa, p. 230]. It is now clear that formula (1) is linked to a particular choice of line coordinates. The hemisphere characterization of Nusselt can be derived from (7) when we use as coordinates of a line the intercept of L on a fixed plane (the plane tangent to dA_1) and the direction of L [Sa, p. 211]. In Section 4 we use a different parametrization of lines in 3-space and the correspondent invariant differential form to derive a new concrete formulation of (7).

3.2. Occluded Case

Now we consider the general case in which we want to define the form factor F_{12} between S_1 and S_2 in the presence of other objects S_3, \dots, S_k . First of all, to simplify the situation, we notice that we need only consider the objects clipped within the convex hull of S_1 and S_2 , since (portions of) objects outside $CH(S_1 \cup S_2)$ cannot possibly occlude a line meeting S_1 and S_2 .

Therefore we want to compute the measure of the set of lines *meeting* S_1 and S_2 , and *missing* all the clipped polygons S_3^c, \dots, S_k^c . Let $\mathcal{L}_{12} = \{L | (L \cap S_1 \neq \emptyset) \wedge (L \cap S_2 \neq \emptyset) \wedge_{i=3}^k (L \cap S_i^c = \emptyset)\}$ be such a set of lines. The new abstract formula for the form factor is

$$F_{12} = \frac{m(\mathcal{L}_{12})}{m(\{L | L \cap S_1 \neq \emptyset\})}. \quad (8)$$

4. A New Concrete Characterization of Form Factors

4.1. Unoccluded Case

Let us consider two disjoint convex planar polygons in 3-space, namely, S_1 and S_2 . From the abstract definition in (7) we just need to evaluate the integral

$$m(\{L|L \cap S_1 \neq \emptyset \wedge L \cap S_2 \neq \emptyset\}).$$

From Santaló [Sa] we have that one of the possible parametrizations of the line L is by giving the plane $P(L)$ through the origin and orthogonal to L , and the intersection point $L \cap P(L)$. Using this coordinatization, the form of the invariant density for lines in 3-space is $dL = d\sigma \wedge du$, where $d\sigma$ is the invariant density of points on a plane through the origin and orthogonal to L , and du is the invariant density of unoriented planes through the origin (equivalent to the points on the unit two-dimensional sphere U centered at the origin where opposite points are identified).

We fix a point $u \in U$ and we integrate over the density $d\sigma$ of the plane through the origin and orthogonal to the vector \overline{Ou} , which we call $P(u)$. A line L orthogonal to $P(u)$ meets both S_1 and S_2 if and only if the intersection point $L \cap P(u)$ is in the intersection of the projection of S_1 onto $P(u)$ and the projection of S_2 onto $P(u)$. But the integral of the density of points in a given planar set is just the area of that set. So, denoting by $A_{12}(u)$ the area of the intersections of projections of S_1 and S_2 onto $P(u)$ we have that

$$\int_{\{L|L \cap S_1 \neq \emptyset \wedge L \cap S_2 \neq \emptyset\}} dL = \int_{u \in U} A_{12}(u) du. \quad (9)$$

Here we consider the case of two objects only and therefore there is no issue of occlusions. We can compute approximately the integral in formula (9) by using Monte Carlo integration [BGS⁺], [DR]. For a fixed direction u we can compute $A_{12}(u)$ in time $O(n)$, where n is the number of vertices of the polygons S_1 and S_2 , for example, by using a sweeping line approach. We choose uniformly at random, on the unit sphere U , a set of N points u_1, \dots, u_N , then the Monte Carlo approximation we obtain is

$$m(\{L|L \cap S_1 \neq \emptyset \wedge L \cap S_2 \neq \emptyset\}) \approx \frac{2\pi}{N} \sum_{i=1}^N A_{12}(u_i).$$

We can find an approximation to the form factor of two unoccluded polygons in 3-space in time $O(Nn)$. An upper bound on the error developed in Section 5 for the occluded case also holds in the unoccluded case. Here we show the high-level pseudocode of the algorithm to compute the approximate form factor between two polygons in 3-space without occlusions.

```

proc ff(A,B:3D-polygon; eps,de:real):real
begin
  N = floor(1/(eps * eps * de));
  ff = 0;
  for i = 1 to N do

```

```

    u = random-direction;
    A1 = projection of A along u;
    B2 = projection of B along u;
    ff = ff+ (area(intersection(A1, B2)));
end for;
return (2 * ff) / (N * area(A));
end

```

4.2. Importance Sampling

We can improve on the basic idea of algorithm `ff` by filtering out efficiently the directions for which the contribution to the summation is zero. We sketch here the method. We assume, without loss of generality, that the plane spanning S_1 does not meet S_2 and vice versa.³ Consider the convex hull $C = CH(S_1 \cup S_2)$. Assuming a general position, a facet f of C , except S_1 and S_2 , is incident to an edge e of S_1 and to a vertex v of S_2 , or vice versa. Moreover, the plane spanning f leaves S_1 and S_2 on the same side. We find, using a sort of binary search, the plane incident to e that instead leaves S_1 and S_2 on *opposite* sides (we call *positive* the side containing S_1). We find such planes for every facet f and we translate them to the origin. We take the intersection of all the translated positive half-spaces. We obtain a double cone with apex at the origin, which we call, for lack of a better name, the *anti-convex-hull* of S_1 and S_2 . It is easy to see that, for a projection direction u that falls outside the anti-convex-hull, the projections of the surfaces cannot meet. Instead, if the direction u falls within the anti-convex-hull, then the projections have a nonempty intersection. The convex hull and the anti-convex-hull can be built in time $O(n \log n)$. After preprocessing we can decide whether a sampled direction is inside the anti-convex-hull in time $O(\log n)$, since the problem is equivalent to locating a point in a planar convex polygon of $O(n)$ sides.

4.3. Occluded Case

We compute the measure of the set of lines in formula (8) by integrating the motion invariant unit element dL discussed in the previous subsection. We indicate by S_i^c the surfaces clipped within $CH(S_1 \cup S_2)$. We indicate by the superscript $''$ to a set, the set obtained by orthogonal projection in direction u . We denote by A'_{12} the area of the set $(S_1'' \cap S_2'') / \bigcup_{j=3}^k (S_j^c)''$. The numerator of the fraction in (8) is thus equal to

$$m(\mathcal{L}_{12}) = \int_{u \in U} A'_{12}(u) du. \quad (10)$$

In Section 6 we see how to compute the terms $A'_{ij}(u)$ simultaneously for all pairs ij for a fixed direction u . In the next section we give an upper bound to the error in a Monte Carlo approximation of the form factor using integral (10).

³ This assumption can be eliminated easily by splitting each surface in at most two parts.

5. Error Estimate

5.1. Bounding the Variance

In order to bound the error of our method we use an elementary analysis of Monte Carlo methods as described in [BGS⁺] and [DR]. This type of analysis relies on a bound on the variance of the integrand function over the domain of integration. This bound is hard to obtain for the integrand functions used in previous methods for form-factor computation. Luckily, in our case, a bound on the variance is obtained very easily. We recall that the measure of the set U is 2π . We introduce a function $\zeta(u)$

$$\zeta(u) = \frac{2A'_{12}(u)}{A_1}.$$

We have that the expected value E of $\zeta(u)$ over all directions U is the value of the form factor F_{12}

$$\frac{1}{2\pi} \int_{u \in U} \zeta(u) du = \frac{1}{2\pi} \int_{u \in U} \frac{2A'_{12}(u)}{A_1} du = F_{12}.$$

Now we estimate the variance D of the function ζ over the set of directions U

$$D(\zeta(u)) = E(\zeta(u)^2) - [E(\zeta(u))]^2 \quad (11)$$

$$= \left[\frac{1}{2\pi} \int_{u \in U} \frac{4(A'_{12}(u))^2}{A_1^2} du \right] - F_{12}^2 \quad (12)$$

$$= \frac{2}{2\pi} \int_{u \in U} \left(\frac{2A'_{12}(u)}{A_1} \right) \left(\frac{A'_{12}(u)}{A_1} \right) du - F_{12}^2 \quad (13)$$

$$\leq 2F_{12} - F_{12}^2 \leq 1, \quad (14)$$

where we exploit the fact that $A'_{12}(u) \leq A_{12}(u)$ is always smaller than or equal to A_1 .

5.2. A Bound Obtained Using Chebyshev's Inequality

For any random variable Y and $t > 0$, Chebyshev's inequality [F, p. 232] is

$$\text{Prob}(|Y| \geq t) \leq \frac{E(Y^2)}{t^2},$$

where $E(\cdot)$ indicate the expected value. Let $S_N = \sum_{i=1}^N \zeta(u_i)$, $\mu = E(\zeta)$, and let σ^2 be the variance of $\zeta(u_i)$. We apply Chebyshev's inequality, obtaining

$$\text{Prob} \left(\left| \frac{S_N - N\mu}{\sqrt{N}\sigma} \right| \geq t \right) \leq \frac{1}{t^2}.$$

Rearranging and dividing by N , we obtain

$$\text{Prob} \left(\left| \left(\frac{S_N}{N} - \mu \right) \right| \geq \frac{t\sigma}{\sqrt{N}} \right) \leq \frac{1}{t^2}.$$

Equivalently, if we call the absolute error ε , we have $\varepsilon = t\sigma/\sqrt{N}$ and $t = \varepsilon\sqrt{N}/\sigma$. The probability of exceeding the error bound is $\delta = 1/t^2 = \sigma^2/\varepsilon^2 N$. The probability of being within the error bound is conversely $1 - \delta$.

In the previous section we proved $\sigma^2 \leq 1$, thus we can derive the number of samples attaining the desired precision as a function of ε and δ . We have $N = \lceil 1/\varepsilon^2\delta \rceil$. We have proved the following lemma.

Lemma 1. *The Monte Carlo evaluation of formula (10) results in an approximation of the form factor within an absolute error ε , with probability at least $1 - \delta$, when we choose $N = \lceil 1/\varepsilon^2\delta \rceil$.*

5.3. Reducing the Number of Samples

From the previous section we have that the number of samples needed for an (ε, δ) approximation algorithm is $N = \lceil 1/\varepsilon^2\delta \rceil$. It is possible to reduce the number of samples to $N = O((1/\varepsilon)^2 \log(1/\delta))$ by using a trick reported in [KKK⁺], originally used in [JVV].

For completeness we describe this method. We use the algorithm described in Section 6 with a constant value for the confidence parameter, say $\delta = 1/4$. We run the algorithm Q times, and we take the *median* value as the outcome of the algorithm for F_{ij} . We obtain an (ε, δ) -approximation by choosing $Q = c \log(1/\delta)$ for an appropriate constant c . We give a sketch of the proof in the rest of this subsection.

We call a value X_i good if it has a distance ε from the expected value $E(X)$, and bad otherwise. In our experiment a sample will be bad with probability $1/4$ and good with probability $3/4$. If the median is bad, then there are at least $\lceil Q/2 \rceil$ bad samples. Therefore the probability of a bad median is less than the probability of having at least $\lceil Q/2 \rceil$ bad samples. To bound this second probability we use the Chernoff bound in a form reported in [Sp]. Let $Y_i = 1$ if X_i is bad and let $Y_i = 0$ if X_i is good. We have that

$$\text{Prob} \left(\sum_{i=1, Q} (Y_i - E(Y)) > a \right) < e^{-2a^2/Q}.$$

Using $E(Y) = 1/4$ and $a = Q/4$ we obtain

$$\text{Prob} \left(\sum_{i=1, Q} Y_i > \frac{Q}{2} \right) < e^{-Q/8}.$$

Thus, to reach a confidence level δ we just need to repeat the algorithm $Q = O(\log(1/\delta))$ times.

6. Measuring Vertical Visibility

The arguments in the previous sections reduce the problem of computing form factors to the problem of computing areas of intersections of planar polygons on a Euclidean plane. We compute such areas using an algorithm for vertical space partition as an algorithmic

skeleton. A good asymptotic performance is obtained using the cylindrical partition of Mulmuley [M2]. Let \mathcal{S} be a set of interior disjoint convex polygons in 3-space and let $\mathcal{E}(\mathcal{S})$ be the set of edges of polygons in \mathcal{S} . We choose u as the vertical direction. Let q be an intersection among the orthogonal vertical projections of $e_1 \in \mathcal{E}(\mathcal{S})$ and $e_2 \in \mathcal{E}(\mathcal{S})$. Such a point is called a *regular* intersection point if the vertical line through q does not intersect any polygon of \mathcal{S} between e_1 and e_2 . If we call k_r the number of regular intersections and k the number of intersections, we have that $0 \leq k_r \leq k \leq n^2$. In typical situations, though, k_r will be much smaller than n^2 . The vertical decomposition $H(\mathcal{S})$ for the set \mathcal{S} is a decomposition of R^3/\mathcal{S} into maximal vertical prisms⁴ such that: (i) each prism does not intersect \mathcal{S} ; and (ii) the two bases of the prism are subsets of two polygons in \mathcal{S} .

Clearly, for every pair of polygons S_i and S_j , the value of A'_{ij} is obtained by summing the area of the projection of cylinders with bases in S_i and S_j . The cylindrical partition $H(\mathcal{S})$ has size $O(n + k_r)$ and is built in time $O(n \log^2 n + k_r \log n)$ where k_r is the number of *regular intersection points* [M2].

6.1. Overall Algorithm

We summarize here the whole algorithm to compute form factors. C_{ij} is a counter for the pair of surfaces with indices $i < j$, which is initialized to 0. We organize such counters using an array indexed from 1 to n , holding pointers to dynamic binary search trees. The counters are at the leaves of the trees. Clearly we need to store only counters C_{ij} which, during the construction, are found to have a value different from 0. If a counter is not in the data structure its value is by default 0. In this way we can exploit the sparsity of the visibility structure and we avoid using quadratic storage if we do not need to. We use Q of such data structures.

The external loop is executed $Q = O(\log \delta^{-1})$ times. In the internal loop we select randomly and uniformly for $N = \lceil 4/\varepsilon^2 \rceil$ times a plane through the origin (i.e., a vertical direction). We compute the vertical cylindrical decomposition of Mulmuley for that choice of the vertical direction. For each prism whose bases are supported by S_i and S_j , we add the area of the projection of the prism to the counter C_{ij} .

At the exit of the external loop we save the median value of the Q values for C_{ij} . When we want to compute the approximation of F_{ij} we access C_{ij} and divide $2C_{ij}$ by NA_i . To approximate F_{ji} we divide by NA_j . We can easily find all nonzero approximate form factors by visiting the data structure. The size of the data structure is proportional to the number of nonzero approximations to form factors.

6.2. Observations

(1) By choosing a value of the confidence parameter $n^2\delta$ we obtain that, with probability $1 - \delta$, all the nonzero form factors have a simultaneous good approximation. The running time is increased only by a factor $O(\log n)$.

⁴ A prism here is a solid obtained by intersecting an infinite prism with two half-spaces.

(2) Our algorithm is particularly suitable in scenes with large emitting surfaces (e.g., a window). Sometimes scenes have small sources of large radiant emission. In this case our algorithm might estimate to 0 the form factor between this small source and small distant visible objects. On the other hand, the fact that the source is small causes no trouble in computing form factors between the source and large close visible surfaces.

(3) It is possible to customize the algorithm to compute only the form factors involving a specific polygon S_i by restricting the computation to the vertical cylinder based on S_i . The running time in this case would depend on the average number of regular intersections within the projection of S_i . This is desirable if we want to place more stringent error conditions on a form factor involving a bright surface S_i .

(4) Unlike methods based on formula (1) our method does not have any divergence problem due to surfaces too close to each other.

(5) Our algorithm does not spend time in computing *zero* form factors, thus it is particularly suited for scenes where the form-factor matrix is sparse. Moreover, if a form factor is zero, its approximation is by default also zero.

(6) Implementing the algorithm for the case of two surfaces without occlusions should be easy. For the general problem, the difficult part is the computation of the vertical cylindrical decomposition. The three-dimensional sweeping algorithm of Mulmuley, which gives us the best bound, is challenging to implement since it relies on the dynamic maintenance of planar maps [PT]. If we are willing to relax the time bound we can instead project all the surfaces onto a plane and use as an algorithmic skeleton the line sweep algorithm of Bentley and Ottman [BO].

7. Conclusions and Extensions

In this paper we have presented a new method, or paradigm, for computing the form factors among polygons in 3-space with occlusions (christened *the orthogonal projections method*). We derive a bound on the absolute approximation error under the assumption that the integrand function is computed using exact real arithmetic in the RAM model. We describe a simple algorithm for computing an approximation to the form factor between two disjoint nonoccluded polygons in 3-space. We describe an asymptotically efficient algorithm to approximate every nonzero form factor between disjoint polygons in an occluded three-dimensional scene.

The integrals considered in this paper can be seen as associated with a Galerkin method where the functional basis is piecewise constant. When one uses a Galerkin method over a more sophisticated functional basis (e.g., Legendre polynomials, Jacobi polynomials, and wavelets) more complex integrals are obtained [Z], [GSCH], [TM], [H]. Usually approximate quadrature methods are invoked in order to estimate such integrals. In the context of our method one has to redefine the kernels of such integrals as a function of lines meeting pairs of surfaces, as opposed to a function of points on such surfaces. Such an integral could then be split in an external integral on a domain of directions and an inner integral associated with an orthogonal projection. If our surfaces are polyhedral and the functional basis is polynomial, such inner integrals are easy to compute exactly (see [P2] for such a technique applied in a different context). One possible advantage

with respect to other methods is that dealing with occlusions without spoiling the quality of the approximation is easy within our framework.

Acknowledgments

My thanks to Nina Amenta and the Geometry Center at the University of Minnesota, where this research was initiated. Thanks also to the anonymous referees whose comments have helped in improving the presentation.

Appendix. Comparison of Algorithms for Approximating Form Factors

The algorithms developed for computing form factors come under many different families, some are randomized (Monte Carlo), some deterministic. A large variety of approximations in various aspects of the algorithm are used. Usually, neither the time bounds nor the error bounds are derived by analysis. For this reason we will not attempt to describe these methods as originally described in the literature. Still, it is instructive to compare our method of orthogonal projections with the other known methods. The comparison is based on the following criteria:

- (i) We consider the total cost of computing all form factors, that is, a form factor for each pair of surfaces, in the presence of occlusions, for an input set of disjoint convex polygons in 3-space with n vertices and edges.
- (ii) We classify algorithms into families according to the formulation of the form factor as an integral that is used as a guideline of the computation.
- (iii) We consider in each such family the Monte Carlo method in which the domain of integration is sampled uniformly at random and the integrand function is computed *exactly* in the real RAM model.
- (iv) For the exact computations of the integrand function we will refer to currently known method with known worst-case deterministic or expected time bounds.
- (v) When we are not able to bound the absolute error, we consider the less demanding requirement that the convergence rate of the error, as derived from simple analysis of the Monte Carlo method in [DR], is at least $O(\varepsilon)$ on each form factor. Here we rule out, for example, the usual assumption done in the hemicube algorithm that the center of a surface is sufficient to characterize form factors for the whole surface.

Many methods used in practice are obtained by those listed below by using approximations in determining also the integrand function and/or by using predefined lattice points instead of random points. A survey of methods used to compute diffuse radiation form factors is in [EJLA]. The hierarchical decomposition method of Hanrahan *et al.* [HSA] falls somewhat outside our Monte Carlo framework and the comparison of the running time of this algorithm with the running time of the method of orthogonal projections is more problematic.

1. **Analytical.** The double area integral is solved analytically. This method is used in the simple case where occlusion is not present. For example, the method of Schröder and Hanrahan [SH2] for two unoccluded convex polygons gives the exact value in time $O(n^2)$ where n is the number of edges of the polygons. Occlusions are not treated and the closed forms used contain transcendental functions (bilogarithmic functions).
2. **Monte Carlo Ray Tracing.** See Introduction (Section 1.9).
3. **Central Projection Methods (or Hemisphere Method).** This method is based on the Nusselt projection method [SH1] and is based on formula (2). An exact computation of the integrand function in formula (2) entails the computation of a spherical visibility map from a point. If we adopt an approach without preprocessing, we obtain a cost $O(n\epsilon^{-2} \times \text{Cost}(1\text{visibility map}))$. Computing a visibility map may take time between linear and quadratic in n depending on the input. Thus the hemisphere method ranges in complexity between $O(n^2\epsilon^{-2})$ and $O(n^3\epsilon^{-2})$. More complex implementations involving preprocessing and/or output sensitive algorithms can improve the running time, although not in the worst case.

The hemicube algorithm [CG] is based on an approximate (discrete) computation of the visibility map using a rectangular grid and on using tabulated values of the kernel function in (2).

4. **Double Area Integration.** This method corresponds to a direct Monte Carlo integration of formula (1). We choose randomly points on two surfaces S_i and S_j and, if the segment joining them does not intersect any other surface, we compute the integrand function in (1). The on-line approach to this method requires time $O(n^2\epsilon^{-2} \times \text{Cost}(1\text{ ray})) + \text{Preproc}(n)$. The considerations in Section 1.9 on ray shooting among polyhedra apply here as well.
5. **Hierarchical Radiosity Methods (HR).** Hanrahan *et al.* [HSA] describe a method for computing a matrix that approximates the form factor between two surfaces S_i and S_j . Surfaces S_i and S_j are subdivided recursively into subpatches until the form factor between two patches drops below a (small) threshold value. The form factor between patches is bounded using the value of the differential form factor between the center of one patch and a disk enclosing the other patch, for which a simple closed formula is available. The gist of the argument is that not every pair of subpatches need to be considered independently, instead, the resulting matrix of subform factors is structured into large rectangular blocks. Thus, even if we have expanded the computation of a single form factor into a (relatively) large matrix, the subsequent solution in terms of radiosity is speeded up by the block structure of the matrix. The correlation between the threshold values and the number of patches is studied experimentally. Similarly, the relation between the threshold value and the final error in estimating the form factor between the surfaces is explored by means of experiments.

Hierarchical radiosity methods of a second type deal with the problem of *clustering* the given surface into groups so to avoid computing all $O(n^2)$ form factors independently [SAG]. Here the error analysis is of the *a posteriori* type and is based on bounding the maximum value of the kernel function in (1) between well-separated clusters of surfaces.

6. **Integral Geometric Characterizations.** See the Introduction (Section 1.8).

- 7th ACM–SIAM Symposium on Discrete Algorithms, Atlanta, GA. SIAM, Philadelphia, PA, 1996, pp. 184–191.
- [PS] F. P. Preparata and M. I. Shamos. *Computational Geometry: An Introduction*. Springer-Verlag, New York, 1985.
- [PT] F. P. Preparata and R. Tamassia. Dynamic planar point location with optimal query time. *Theoret. Comput. Sci.* **74**:95–114, 1990.
- [Sa] L. A. Santaló. *Integral Geometry and Geometric Probability*. Addison-Wesley, Reading, MA, 1976.
- [SAG] B. Smits, J. Arvo, and D. Greenberg. A clustering algorithm for radiosity in complex environments. *Computer Graphics Proceedings, Annual Conference Series*, 1994, Orlando, FL. ACM SIGGRAPH, New York, 1994, pp. 435–442. *Proceedings of SIGGRAPH '94*.
- [Sb] M. Sbert. An integral geometry-based method for fast form-factor computation. In: *Eurographics 93*, R. J. Hubbard and R. Juan, eds., 1993, pp. 409–420. Also in *Computer Graphics Forum*, vol. 12, no. 3, 1993.
- [Sh] P. Shirley. Time complexity of Monte Carlo radiosity. In *Eurographics '91* (Werner Purgathofer, ed.). North-Holland, Amsterdam, 1991, pp. 459–465.
- [SH1] R. Siegel and J. R. Howell. *Thermal Radiation Heat Transfer*, 3rd edn. Hemisphere, Washington, 1992.
- [SH2] P. Schröder and P. Hanrahan. On the form factor between two polygons. *Computer Graphics Proceedings, Annual Conference Series*, 1993, Anaheim, CA. ACM SIGGRAPH, New York, 1993, pp. 163–164. *Proceedings of ACM SIGGRAPH '93*.
- [Sp] J. Spencer. *Ten Lectures on the Probabilistic Method*, vol. 52 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, PA, 1987.
- [SP] F. Sillion and C. Puech. A general two-pass method integrating specular and diffuse reflection. *Computer Graphics* **23**(3):335–344, 1989. *Proceedings of SIGGRAPH '89*.
- [SPNP] M. Sbert, X. Pueyo, L. Neumann, and W. Purgathofer. Global multi-path Monte Carlo algorithms for radiosity. Technical Report IMA-94-04. Departament d'Informàtica i Matemàtica Aplicada, Universitat de Girona, 1994.
- [SPP] M. Sbert, X. Pueyo, and W. Purgathofer. Adding progressivity to the global Monte Carlo radiosity method. Technical Report IMA-95-01. Departament d'Informàtica i Matemàtica Aplicada, Universitat de Girona, 1995.
- [TM] R. Troutman and N. L. Max. Radiosity algorithms using higher order finite element methods. *Computer Graphics Proceedings, Annual Conference Series*, 1993, Anaheim, CA. ACM SIGGRAPH, New York, 1993, pp. 209–212.
- [TT] N. M. Thalmann and D. Thalmann (editors). *Proceedings of Computer Graphics International '93*. Springer-Verlag, New York, 1993.
- [WEH] J. R. Wallace, K. A. Elmquist, and E. A. Haines. A ray tracing algorithm for progressive radiosity. *Computer Graphics*, **23**(3):3115–324, 1989. *Proceedings of SIGGRAPH '89*.
- [WRC] G. J. Ward, F. M. Rubinstein, and R. D. Clear. A ray tracing solution for diffuse interreflection. *Computer Graphics*, **22**(4):85–92, 1988. *Proceedings of SIGGRAPH '88*.
- [Z] H. R. Zatz. Galerkin radiosity: A higher-order solution method for global illumination. *Computer Graphics Proceedings, Annual Conference Series*, 1993, Anaheim, CA. ACM SIGGRAPH, New York, 1993, pp. 213–220.

Received March 17, 1995, and in revised form April 10, 1996.