EUROGRAPHICS '97 / D. Fellner and L. Szirmay-Kalos
(Guest Editors)

Volume 16, (*1997*), *Number 3*

An Analysis of Quasi-Monte Carlo Integration Applied to the Transillumination Radiosity Method

László Szirmay-Kalos, Tibor Fóris, László Neumann and Balázs Csébfalvi

Department of Process Control, Technical University of Budapest,
Budapest, Műegyetem rkp. 11, H-1111, HUNGARY
szirmay@fsz.bme.hu

Abstract

This paper presents an enhanced transillumination radiosity method that can provide accurate solutions at relatively low computational cost. The proposed algorithm breaks down the double integral of the gathered power to an area integral that is computed analytically and to a directional integral that is evaluated by quasi-Monte Carlo or Monte-Carlo techniques. The paper also analyses the requirements of the convergence, presents theoretical error bounds and proposes error reduction techniques. The theoretical bounds are compared with simulation results.

Keywords: Radiosity method, error analysis, transillumination, quasi-Monte Carlo quadrature.

1. Introduction

Classical radiosity algorithms, including the hemisphere [22], the hemicube [6] or the cubic-tetrahedron methods [2], assume that the double integral of the form factor or the exchanged power can be approximated by the inner integral multiplied by the area of the patch. This approximation introduces errors proportional to the size of patches. In exact or semi-exact methods [1,20], on the other hand, the computational overhead is significant.

In order to make a radiosity method both accurate and fast, surface and directional integrals must be computed effectively. Recently Keller [9] proposed the application of quasi-Monte Carlo quadrature [17] for the radiosity problem as a promising alternative to Monte-Carlo or classical integration rules.

For the normalized, s-dimensional integration domain $[0,1]^s$, the quasi-Monte Carlo approximation is:

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

Sample points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M$ should be selected to minimize the error of the integral quadrature.

If the integrand f has finite variation in the sense of Hardy and Krause, then the error of the quasi-Monte Carlo quadrature can be bounded using the Koksma-Hlawka inequality:

$$\left| \int\limits_{[0,1]^s} f(\mathbf{x})\, d\mathbf{x} - \frac{1}{M}\sum_{i=1}^{M} f(\mathbf{x}_i) \right| \leq \mathcal{V}_{\mathrm{HK}}(f) \cdot D^*(\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_M), \tag{2}$$

where $\mathcal{V}_{\mathrm{HK}}$ is the variation in the sense of Hardy and Krause and $D^*(\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_M)$ is the star-discrepancy of the sample points [17]. The star-discrepancy is defined by

$$D^*(\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_M) = \sup_A \left| \frac{m(A)}{M} - V(A) \right| \tag{3}$$

where A is a s-dimensional "brick" parallel to the coordinate axes and originating at the center, and $m(A)$ is the number of sample points inside this "brick".

For carefully selected sample points, called low-discrepancy sequences, the discrepancy and consequently the error can be in the order of $O(\log^{d-1} M/M)$, which is much better than the $O(1/\sqrt{M})$ probabilistic error bound of Monte Carlo methods. Moreover, quasi-Monte Carlo methods quarantee this accuracy in a deterministic way, unlike Monte Carlo methods where the error bound is also probabilistic.

However, function f that needs to be integrated in rendering problems is usually discontinuous and thus has infinite variation (a function of finite variation can have discountinuities parallel to the coordinate axes only [4]). Although practical experiences show that quasi-Monte Carlo methods can still be better than Monte-Carlo methods [10], their advantages seem to be less than predicted by the theory for functions of finite variation.

In order to overcome this problem, this paper reformulates the radiosity equations to make the integrands have finite variation. The new formulation is based on the continuous version of the transillumination method originally proposed in [15]. In the new algorithm, the inner integral of the form factors is evaluated exactly by analytic techniques, while the outer integral is approximated by a quasi-Monte Carlo quadrature formula. Since the integrand of outer integral is continuous and its mixed derivatives are bounded and piece-wise continuous, its variation is finite [24], thus it is well suited for quasi-Monte Carlo techniques.

The paper also examines the convergence and error characteristics of the new method.

2. Transillumination method

Radiosity algorithms calculate the power exchange taking polygonal patches one-by-one and evaluating the gathered or shot energy by determining an integral of other surfaces visible in the directions of the hemisphere above the given patch. The transillumination method, on the other hand, takes the directions first, and evaluates the power exchange in this direction for all patches. In order to introduce the transillumination method, let us examine the power radiated onto patch A_i:

$$\Phi_i^{\text{in}} = \int_{\Omega_i} \int_{A_i} I^{\text{in}}(\vec{x}, \omega) \cos\theta_i(\omega) \, d\vec{x} \, d\omega \qquad (4)$$

where Ω_i is the set of directions from where patch A_i can be lit (it is actually a hemisphere above the plane of patch A_i), $I^{\text{in}}(\vec{x}, \omega)$ is the intensity of the radiation at point $\vec{x}$ from direction ω, $\vec{x}$ and ω are the running vectors on patch A_i and in Ω_i respectively and $\theta_i(\omega)$ is the angle between direction ω and the surface normal of patch A_i (see figure 1).

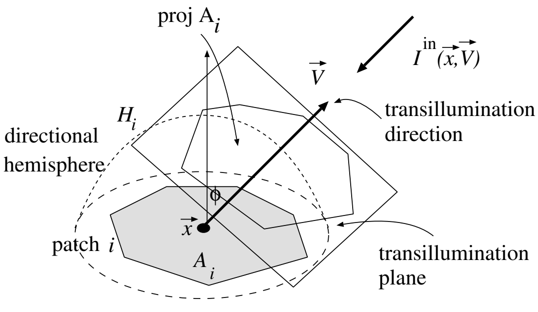

Figure 1: *The power radiated onto patch A_i*

The directional integral can be extended to the full sphere H if the function is multiplied by a characteristic function $h_i(\omega)$ that indicates whether ω is in the upper hemisphere of the given patch ($h_i(\omega) = 1$) or not ($h_i(\omega) = 0$). Assume that directions on the sphere are defined by two scalars u, v which are points in the domain of $[0, 1]^2$. Since $d\omega = d\theta \; d\psi \sin\theta$ where θ and ψ are the polar angles, we get:

$$d\omega = det \begin{bmatrix} \frac{\partial\psi}{\partial u} & \frac{\partial\psi}{\partial v} \\ \frac{\partial\theta}{\partial u} & \frac{\partial\theta}{\partial v} \end{bmatrix} \cdot \sin\theta(\omega(u,v)) \; du \; dv. \qquad (5)$$

Let us denote the determinant of the Jacobi matrix multiplied with $\sin\theta(\omega(u, v))$ by $J(u, v)$. Using this substitution, the incoming power is:

$$\Phi_i^{\text{in}} = \int_{u=0}^{1} \int_{v=0}^{1} \int_{A_i} I^{\text{in}}(\vec{x}, \omega) \cos\theta_i(\omega) \, d\vec{x} \; \cdot h_i \cdot J \; du \; dv. \qquad (6)$$

Integral $\int_{A_i} I^{\text{in}}(\vec{x}, \omega) \cos\theta_i(\omega) \; d\vec{x}$ expresses the power arriving at patch A_i from direction ω. This direction playes an important role in this method and is called the **transillumination direction**.

Let us place a plane perpendicular to the transillumination direction above patch A_i. This plane is called the **transillumination plane**. The same power reaching patch A_i can also be computed on this transillumination plane, just the range of integration should be changed from A_i to the projection of A_i on the transillumination plane, which is denoted by A_i^{proj}:

$$\int_{A_i} I^{\text{in}}(\vec{x}, \omega) \cos\theta_i(\omega) \; d\vec{x} = \int_{A_i^{\text{proj}}} I^{\text{in}}(\vec{x}\,', \omega) \; d\vec{x}\,' \qquad (7)$$

since $\cos\theta_i(\omega) \; dA_i = dA_i^{\text{proj}}$.

Let us turn our attention to the power emitted or reflected by patch A_i, denoted by Φ_i^{out}. If patch A_i is diffuse (which is a general restriction of radiosity methods), its reflection coefficient is ρ_i and the emission density is E_i, then the total power is:

$$\Phi_i^{\text{out}} = E_i \cdot A_i + \rho_i \cdot \Phi_i^{\text{in}}. \qquad (8)$$

By definition, the radiosity of patch i is the power of a unit surface area, that is:

$$B_i = E_i + \frac{\rho_i}{A_i} \cdot \int_{u=0}^{1} \int_{v=0}^{1} \int_{A_i^{\text{proj}}} I^{\text{in}}(\vec{x}\,', \vec{V}) \; d\vec{x}\,' \cdot h_i \cdot J \; du \; dv. \qquad (9)$$

The source of intensity $I^{\text{in}}(\vec{x}\,', \omega)$ is also a patch that is also assumed to be diffuse and to have radiosity $B(\vec{x}\,', \omega)$. For diffuse radiance $B = I\pi$ [26], thus equation (9) can also be written in the following form:

$$B_i = E_i + \frac{\rho_i}{\pi A_i} \cdot \int_{u=0}^{1} \int_{v=0}^{1} \int_{A_i^{\text{proj}}} B(\vec{x}\,', \omega)\; d\vec{x}\,' \cdot h_i \cdot J \; du\, dv. \quad (10)$$

This is an alternative form of the radiosity equation. As the classical radiosity equation, the alternative form can also be used as a basis of an iterational solution:

$$B_i^{(n+1)} = E_i + \frac{\rho_i}{\pi A_i} \cdot \int_{u=0}^{1} \int_{v=0}^{1} \int_{A_i^{\text{proj}}} B^{(n)}(\vec{x}\,', \omega_d)\; d\vec{x}\,' \cdot h_i \cdot J \; du\, dv. \quad (11)$$

Since the integral operator in equation (11) is a contractive mapping for physically correct models, this iteration scheme is convergent. The effective evaluation of this formula is the key to find an efficient algorithm. First the *directional integral* is approximated by a finite sum using some quadrature formula:

$$\int_{u=0}^{1} \int_{v=0}^{1} \int_{A_i^{\text{proj}}} B(\vec{x}\,', \omega)\; d\vec{x}\,' \cdot h_i \cdot J \; du\, dv \approx$$

$$\frac{1}{M} \cdot \sum_{d=1}^{M} \int_{A_i^{\text{proj}}} B(\vec{x}\,', \omega_d)\; d\vec{x}\,' \cdot h_{i,d} \cdot J_d \quad (12)$$

where ω_d, $h_{i,d}$ and J_d are calculated from $\omega(u,v)$, $h_i(u,v)$ and $J(u,v)$ taking samples in $[0,1]^2$. The samples of the unit square are mapped onto the directional sphere to find ω_d.

In order to generate a uniform distribution on the surface of the directional sphere, the following transformation can be used [23]:

$$\psi = 2\pi u, \qquad \theta = \arccos(1 - 2v). \quad (13)$$

In this case $d\omega = J(u,v) \cdot du\, dv = 4\pi \cdot du\, dv$, thus the distribution is uniform on the sphere if it was in the unit square.

Integrating over the whole sphere with the rejection of the lower hemisphere by function h and the application of uniform distribution may seem to be less efficient then the integration over just the upper hemisphere and the use of cosine or sine distributions incorporating importance sampling as applied in [12] or [11]. However, in the transillumination method the same set of directions are used parallely for all the patches, thus the distribution of directions cannot depend on the orientation of individual patches. It is obviously a disadvantage, which is compensated for by the possibility of handling all patches simultaneously.

According to the orientation relative to the transillumination direction, that is according to the value of $h_{i,d}$, patches can be classified as *front facing* ($h_{i,d} = 1$) and *back facing* ($h_{i,d} = 0$). Obviously, back facing patches cannot receive energy from a transillumination direction, thus only front facing patches must be examined. In order to numerically compute the integral $\int_{A_i^{\text{proj}}} B(\vec{x}\,', \omega_d)\; d\vec{x}\,'$ for a front facing patch, discrete or continuous methods can be used.

If discrete methods are applied, then the transillumination plane is discretized into finite rectangles of size δA, called "pixels". This discretization enables us to replace the integral of equation (12) by an approximation sum, which, in turn, can be evaluated by a generalised z-buffer [15] that holds lists of patches instead of depth values, or by using the painter's algorithm [25]. The latter has the advantage that the generated transillumination plane can be reused for subsequent patches, resulting in a radiosity algorithm of $O(N \log N)$ time-complexity (N is the number of patches in the scene). However, for complex and highly reflective scenes, the required resolution of the transillumination plane can be quite high. Thus, an objective of this paper is to present an approach that uses continuous computation on the transillumination plane. Two closely related versions, called the global and local visibility map methods, are presented. Then the algorithm of the discrete approximation of the directional integral is discussed.

3. Integration on the transillumination plane

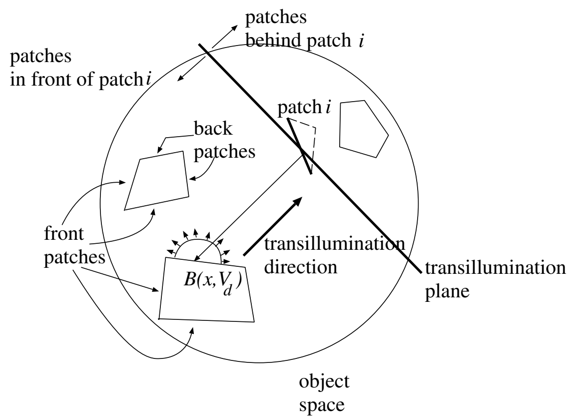

Figure 2: *Relation of the patches, the transillumination direction and the transillumination plane*

Let us first assume that the patches have constant radiosity, thus radiosity is approximated by a piece-wise constant function (piece-wise linear or polynomial approximation can also be handled easily). A piece-wise constant function can be integrated in closed form, thus we obtain:

$$\int_{A_i^{\text{proj}}} B^{(n)}(\vec{x}\,', \omega_d)\; d\vec{x}\,' = \sum_{j=1, i \neq j}^{N} A(i,j,\omega_d) \cdot B_j^{(n)} \cdot (1 - h_{j,d}) \quad (14)$$

where $A(i,j,\omega_d)$ is the area of patch j that is visible from inside of patch i at direction ω_d. Note that the term $(1 - h_{j,d})$

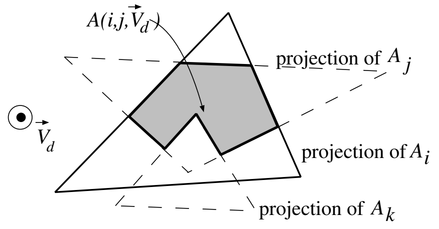

Figure 3: *Interpretation of* $A(i,j,\omega_d)$

is necessary since only back facing patches may contribute to the radiosity of a front facing patch.

In order to compute equation (14), a continuous visibility problem[26] must be solved assuming the transillumination plane to be the plane of the window. It requires the construction of the *visibility map* [3] of polygons in front of patch i on the transillumination plane. The visibility map is a planar graph that is generated by an orthogonal projection of patches onto the transillumination plane and by identifying those regions in which one patch is seen or no patch is seen. Having the visibility map of patches visible from A_i, the computation of $A(i,j,\omega_d)$ requires to determine which regions are inside the projection of A_i and to sum the areas weighted by the radiosity values of the associated patches.

The visibility map containing only the patches visible from A_i at the transillumination direction is called the local visibility map. The local visibility maps of different patches are obviously included in the global visibility map that contains all the patches. Thus generally we have two options to construct the visibility map seen from patch i:

1. A single, *global visibility map* is constructed for all patches.
2. The *local visibility maps* are constructed incrementally as the patches sorted in the transillumination direction are processed.

In the following subsections both the global and local visibility map methods are discussed. These methods can calculate the radiosity contribution of a single transillumination direction.

3.1. Global visibility map

This alternative constructs a single visibility map for all patches. Algorithms[5] are available that can do it in $O((N+i)\log N)$ [5] time where N is the number of patches (or edges) and i is the number of edge intersections, or even in $O(N^{1+\varepsilon}\sqrt{k})$ time [3] where k is the number of edges in the visibility map. The number of intersections i and the number of edges k are in $O(N^2)$ in the worst-case, but are in $O(N)$ in practical environments.

According to the properties of the visibility map, each region either corresponds to some patches that are completely visible in the region or corresponds to no patch at all. In each region there is a well-defined order of patches. Thus the evaluation of equation (14) requires the scanning of the list of patches in the different regions and finding where patch j is next to patch i. The algorithm of the global visibility map method is:

Compute the visibility map corresponding to direction ω_d
for each region R of the visibility map
  $A(R)$ = area of region R
  L = Sorted list of patches visible at region R
  for each front facing patch i in L **do**
    patch j = next of patch i in L
    if patch j is back facing **then**
      $B_i^{(n+1)} += \frac{A(R)}{A_i}\frac{\rho_i}{\pi M}\cdot J_d\cdot B_j^{(n)}$
  endfor
endfor

Note that in neighbouring regions the lists of patches may differ in a single patch. Thus the data-structure of the sorted lists can be updated efficiently as the visibility graph is traversed.

3.2. Local visibility map

The determination of the regions inside the projection of A_i can also be considered as a 2D clipping problem where the projections of patches should be clipped onto each-other. Let us assume that the patches are sorted in the transillumination direction. Such sorting is not always trivial because patches may overlap in this direction and cyclic overlapping might also occur, but the painter's algorithm [16] or the Binary Space Partitioning (BSP) trees [18] can handle this problem.

Suppose that the visibility map should be updated with patch i. Since the patches are sorted, all those patches must be clipped out from the visibility map, which fall into the interior of patch i since these parts will be hidden from subsequent patches. Then, if patch i can reflect energy onto the next patch (it is a back facing patch with respect to the transillumination direction), then patch i should be added to the visibility map, otherwise, the place of the projection of patch i will be empty. The local visibility map method is as follows:

list L = Sort patches in direction d
visibility map $V = \{\ \}$
for each patch i in L **do**
  Clip patches in V onto patch i
    and generate: O = outside list, I = inside list
  if patch i is front facing **then**
    // update with patches of the inside list
    $B_i^{(n+1)} += \sum_{j\in I}\frac{A(j)}{A_i}\frac{\rho_i}{\pi M}\cdot J_d\cdot B_j^{(n)}$
    $V = O$
  else
    $V = O$ + patch i
  endif

endfor

3.3. Integration on the directional hemisphere

In order to consider all transillumination directions, the $B_i^{(n+1)}(d)$ values of different directions must be summed. This is summarised in the following algorithm:

for each patch i **do** $B_i^{(0)} = E_i$;
for $n = 0$ **to** $n_{\max}$
  for each patch $B_i^{(n+1)} = E_i$
  for each direction d
    execute the global or local visibility map method
endfor

The presented algorithm consists of an iterational loop that, in turn, considers all directions. For a single direction, the visibility maps from all patches are determined. Thus the complexity of the algorithm will be proportional to the product of the number of iterational loops, directions and the complexity of the applied visibility map algorithm. Since the number of directions is constant, so is the number of iterational loops $n_{\max}$ needed to decrease the error below a given limit as it will be proven in the next section, the complexity of the algorithm is determined by the applied visibility map algorithm.

4. Convergence and number of iterational loops

The iteration scheme uses an equation that approximates the directional integral by a finite quadrature, which introduces some error in each step. Obviously, if this error is too big, not even the convergence of the iteration can be expected. In this section requirements are presented that can guarantee the convergence of the iteration of the approximated radiosity equation. It is also shown that if the iteration is convergent, then the number of iterational loops required to decrease the relative error between the actual and the limiting value of the radiosity vector below a given limit is independent of the number of patches.

Assume that the radiosity of patch i is $B_i^{(n)}$ in iteration step n. Thus, if the limiting radiosity value of the iteration at patch i were B_i^*, then the error between the actual radiosity and the solution of the approximate equation in step n would be $\Delta B_i^{(n)} = B_i^{(n)} - B_i^*$. Let $\Delta B^{(n)}$ be the maximum of absolute values of error in the radiosity estimates of all patches in step n. Substituting this into equation (11), the error in the next step can be computed as:

$$|\Delta B_i^{(n+1)}| \leq \frac{\rho_i}{MA_i\pi} \cdot \sum_{d=1}^{M} \int_{A_i^{\text{proj}}} \Delta B^{(n)}\, d\vec{x}\,' \cdot h_{i,d} \cdot J_d =$$

$$\Delta B^{(n)} \cdot \frac{\rho_i}{M\pi} \sum_{d=1}^{M} \cos\theta_{i,d} \cdot h_{i,d} \cdot J_d. \quad (15)$$

Note that the sum $1/M \sum_{d=1}^{M} \cos\theta_i(\omega_d) \cdot h_{i,d} \cdot J_d$ is an approximation of the following integral:

$$\int_{u=0}^{1} \int_{v=0}^{1} \cos\theta_i(\omega(u,v)) \cdot h_i(u,v) \cdot J(u,v)\, du\, dv = \pi. \quad (16)$$

The absolute error of quadrature depends on the number of sample points M, the distribution of the sample points and function $J(u,v)$ that describes how these sample points are mapped onto the surface of the directional hemisphere. Assume that the mapping keeps the uniform distribution, which makes $J(u,v)$ is constant and is equal to 4π, thus the following error formula can be obtained:

$$|\varepsilon(M)| = |\int_{u=0}^{1} \int_{v=0}^{1} \cos\theta_i \cdot h_i \cdot 4\pi\, du\, dv - \frac{1}{M} \sum_{d=1}^{M} \cos\theta_{i,d} \cdot h_{i,d} \cdot 4\pi|. \quad (17)$$

To bound this error, the Koksma-Hlawka inequality [17] is used. Krause is finite. For a 2-dimensional function $f(u,v)$ which is continuous and has piece-wise continuous and bounded mixed derivatives, the variation in the sense of Hardy-Krause $\mathcal{V}_{HK}(f(u,v))$ can also be expressed in the following form:

$$\int_0^1 \int_0^1 \left| \frac{\partial^2 f(u,v)}{\partial u \partial v} \right| du\, dv + \int_0^1 \left| \frac{\partial f(u,1)}{\partial u} \right| du + \int_0^1 \left| \frac{\partial f(1,v)}{\partial v} \right| dv. \quad (18)$$

Using the mapping defined by equation (13) and assuming that h_i selects $v \in [0, 0.5]$, function f is

$$f(u,v) = \cos\theta_i(u,v) \cdot h_i(u,v) \cdot 4\pi = \begin{cases} (1-2v) \cdot 4\pi \text{ if } v \leq 0.5, \\ 0 \text{ otherwise}, \end{cases} \quad (19)$$

thus it is continuously differentiable except for $v = 0.5$ where it is only continuous. Since $\partial^2 f(u,v)/\partial u \partial v$ and $\partial f(u,1)/\partial u$ are zero, and $|\partial f(1,v)/\partial v| = 8\pi$ if $v \leq 0.5$ and 0 otherwise, the Hardy-Krause variation is 4π.

The star-discrepancy $D^*([u_d, v_d])$ depends on the selection of the sample points in the unit square. For example, for the two-dimensional *Hammersley sequence* [17] applying 2 as base number, the discrepancy is bounded by:

$$D^*([u_d, v_d]) \leq \frac{1}{2M} \cdot (\log_2 M + 7). \quad (20)$$

Substituting this into the equation (17), the following inequality can be established:

$$\varepsilon(M) \leq \frac{2\pi}{M} \cdot (\log_2 M + 7). \quad (21)$$

Using this error term, inequality (15) can also be written in the following form:

$$|\Delta B_i^{(n+1)}| \leq \Delta B^{(n)} \cdot \rho_i \cdot (1 + \frac{|\varepsilon(M)|}{\pi}). \quad (22)$$

Let $\rho_{\max}$ be $\max_i \rho_i$. Since $\Delta B^{(n+1)} = \max_i |\Delta B_i^{(n+1)}|$, we get:

$$\Delta B^{(n+1)} \leq \Delta B^{(n)} \cdot \rho_{\max} \cdot (1 + \frac{|\varepsilon(M)|}{\pi}), \quad (23)$$

which means that the maximal error of patch radiosities is limited by a geometric series of coefficient $\rho_{\max} \cdot (1 + |\varepsilon(M)|/\pi)$. In order to make the bounding series convergent, the error term should satisfy:

$$\rho_{\max} \cdot (1 + \frac{|\varepsilon(M)|}{\pi}) < 1 \quad \Rightarrow \quad |\varepsilon(M)| < \pi(\frac{1}{\rho_{\max}} - 1). \quad (24)$$

This imposes a constraint on the number of samples used in the discrete approximation of the directional integration. Table (1) summarises the calculated minimum values of the number of transillumination directions M for environments of different maximum reflectivity.

$\rho_{\max}$	M
0.2	5
0.3	9
0.4	15
0.5	24
0.6	37
0.7	61
0.8	111
0.9	272

Table 1: *Required number of transillumination directions to converge for different* $\rho_{\max}$ *values*

If this error constraint is satisfied, then the coefficient of the geometric series is less than 1, providing convergence for any arrangements. It should be noted that this error bound is theoretical and it does not mean that if this constraint is not met, then the iteration is not convergent, and in practical cases significantly less severe conditions may also work.

Suppose that the termination criterion of the iteration is that the relative error between the actual value and the limiting value of the radiosity vector, which is defined by $||\Delta\mathbf{B}^{(n)}||/||\mathbf{B}^*||$, must be lower than a predefined limit. If $\mathbf{B}^{(0)} = 0$ and the infinite norm is used, then

$$\frac{||\Delta\mathbf{B}^{(n)}||}{||\mathbf{B}^*||} = \frac{\Delta B^{(n)}}{\max B_i^*} = \frac{\Delta B^{(n)}}{\Delta B^{(0)}} \leq (\rho_{\max} \cdot (1 + \frac{|\varepsilon(M)|}{\pi}))^n. \quad (25)$$

Note that the bound of the relative error is independent of the number of patches and their geometrical arrangements, thus the number of iterational steps required to generate estimates with some predefined precision is constant.

5. Error analysis

The presented approach follows an iteration scheme to determine the radiosities which satisfy an approximation of the original radiosity formula (equation (10)). In the presented method the approximation is due to the discretization of the directional hemisphere. Since not the real radiosity equation is iterated, some error is introduced in each step. At the end of the iteration the radiosity will be distorted by the total accumulated error. This error is estimated in this section.

The further analysis is based on the assumption that the iteration is convergent. Let us suppose that the following linear equation needs to be solved by iteration:

$$\mathbf{B} = \mathbf{E} + \mathbf{F} \cdot \mathbf{B}. \quad (26)$$

Let us denote the solution of this equation by $\tilde{\mathbf{B}}$. Suppose that matrix $\mathbf{F}$ is available only in an approximate form, thus the iterational formula used instead is:

$$\mathbf{B}^{(n+1)} = \mathbf{E} + \mathbf{F}^* \cdot \mathbf{B}^{(n)} \quad (27)$$

where the error of the matrix is $\delta\mathbf{F} = \mathbf{F}^* - \mathbf{F}$. For the time being, assume that the approximate matrix $\mathbf{F}^*$ is constant during the iteration. Using the terminology of the transillumination method, this means that the same set of transillumination directions is used in every iteration. Let us denote the norm of the approximate matrix by r:

$$r = |\mathbf{F}^*|. \quad (28)$$

Let us assume that both the original and the approximate iterations are convergent, thus the norms of $\mathbf{F}$ and $\mathbf{F}^*$ are less than 1. In this case, the iteration of the approximate form will converge to some $\mathbf{B}^*$ from any starting vector. The total accumulated error is $||\mathbf{B}^* - \tilde{\mathbf{B}}|| = \lim_{n\to\infty} ||\mathbf{B}^{(n)} - \tilde{\mathbf{B}}||$.

Since any starting vector can be selected, let us assume that the starting vector is the solution of the exact equation, thus $\mathbf{B}^{(0)} = \tilde{\mathbf{B}}$. Using the triangle inequality valid for any norm, we have:

$$||\mathbf{B}^{(n)} - \tilde{\mathbf{B}}|| = ||\mathbf{B}^{(n)} - \mathbf{B}^{(n-1)} + \mathbf{B}^{(n-1)} - \ldots + \mathbf{B}^{(1)} - \tilde{\mathbf{B}}|| \leq$$

$$\sum_{i=1}^{n} ||\mathbf{B}^{(i)} - \mathbf{B}^{(i-1)}||. \quad (29)$$

Taking into account that $\tilde{\mathbf{B}} = \mathbf{E} + \mathbf{F} \cdot \tilde{\mathbf{B}}$, the following equality can be established for $i = 1$:

$$||\mathbf{B}^{(1)} - \mathbf{B}^{(0)}|| = ||\mathbf{E} + \mathbf{F}^* \cdot \tilde{\mathbf{B}} - (\mathbf{E} + \mathbf{F} \cdot \tilde{\mathbf{B}})|| = ||\delta\mathbf{F} \cdot \tilde{\mathbf{B}}||. \quad (30)$$

For any $i > 1$, on the other hand:

$$||\mathbf{B}^{(i)} - \mathbf{B}^{(i-1)}|| = ||\mathbf{F}^* \cdot (\mathbf{B}^{(i-1)} - \mathbf{B}^{(i-2)})|| =$$

$$||(\mathbf{F}^*)^{i-1} \cdot (\mathbf{B}^{(1)} - \tilde{\mathbf{B}})|| \leq r^{i-1} \cdot ||\delta\mathbf{F} \cdot \tilde{\mathbf{B}}|| \quad (31)$$

if the matrix and vector norms are compatible. Substituting this into equation (29), we get:

$$||\mathbf{B}^* - \tilde{\mathbf{B}}|| = \lim_{n\to\infty} ||\mathbf{B}^{(n)} - \tilde{\mathbf{B}}|| \le ||\delta\mathbf{F}\cdot\tilde{\mathbf{B}}|| \cdot \sum_{i=1}^{\infty} r^i = \frac{||\delta\mathbf{F}\cdot\tilde{\mathbf{B}}||}{1-r}. \quad (32)$$

In the transillumination method, the absolute value of the ith element of $\delta\mathbf{F}\cdot\tilde{\mathbf{B}}$ is:

$$|\int_{u=0}^{1}\int_{v=0}^{1} f_i(u,v)\,du\,dv - \frac{1}{M}\sum_{d=1}^{M} f_i(u_d,v_d)|. \quad (33)$$

where

$$f_i(u,v) = \frac{4\rho_i}{A_i}\sum_{j=1}^{N} A(i,j,\omega(u,v))\cdot\tilde{B}_j\cdot(1-h_j(u,v))\cdot h_i(u,v). \quad (34)$$

Note that the integrand $f_i(u,v)$ has bounded and piecewise continuous mixed derivatives, which guarantees that the variation in the sense of Hardy and Krause is finite. This fact makes the integrand really appropriate for quasi-Monte Carlo integration, and also allows for quantitative error analysis.

The error of the integral quadrature depends on two independent factors according to the Koksma-Hlawka inequality: the discrepancy of the sample points describing how uniformly they are distributed, and the variation of the integrand. The variation of function $f_i(u,v)$ is also due to two different reasons. On the one hand, for different directions the projected area of A_i changes, thus $\sum_{j=1}^{N} A(i,j,\omega(u,v))$ decreases as the angle between the patch normal and the transillumination direction increases. On the other hand, in different directions different patches may be visible which may have significantly different radiosities.

Unfortunately, the integrand $f_i(u,v)$ is too complex to estimate its variation in the "general" case. Instead, two extreme cases are examined.

5.1. Error analysis in homogeneous-like environments

First, let us assume that in the variation of $f_i(u,v)$ the size of the projected areas is dominant, while the radiosity is approximately the same from every direction. This means that the environment consists of large, homogeneous areas (examples are a room, a terrain, the sky, etc). In this case the variation of f can be estimated by the variation of the following function:

$$f_i'(u,v) = \frac{4\rho_i}{A_i}\cdot A_i\cdot\cos\theta_i(\omega(u,v))\cdot\tilde{B}_{\max}\cdot h_i(u,v). \quad (35)$$

The error of the integral quadrature of f_i' is

$$4\rho_i\cdot\tilde{B}_{\max}\cdot|\int_{u=0}^{1}\int_{v=0}^{1}\cos\theta_i\cdot h_i\,du\,dv - \frac{1}{M}\sum_{d=1}^{M}\cos\theta_i(\omega_d)\cdot h_{i,d}| =$$

$$\rho_i\cdot\tilde{B}_{\max}\cdot\frac{\varepsilon(M)}{\pi} \quad (36)$$

where $\varepsilon(M)$ is the error of integral quadrature introduced in the section on convergence (equation (17)). Substituting this into equation (32), we get:

$$||\mathbf{B}^* - \tilde{\mathbf{B}}|| \le \frac{\rho_{\max}}{1-r}\cdot\frac{\varepsilon(M)}{\pi}\cdot||\tilde{\mathbf{B}}|| \le \frac{\rho_{\max}}{1-\rho_{\max}}\cdot\frac{\varepsilon(M)}{\pi}\cdot||\tilde{\mathbf{B}}|| \quad (37)$$

since $\rho_{\max} \ge r$. Using this inequality, the required number of directions M for different maximum reflectivity and accuracy can be calculated. Table (2) presents the results for $||\mathbf{B}^* - \tilde{\mathbf{B}}||/||\tilde{\mathbf{B}}|| \le 10^{-2}$.

$\rho_{\max}$	M
0.2	900
0.3	1600
0.4	2500
0.5	3800
0.6	5700
0.7	9000
0.8	15200
0.9	34000

Table 2: *Required number of transillumination directions in homogeneous-like environment*

5.2. Error analysis in strongly heterogeneous environments

In this section, the effect of seeing different patches in different directions is supposed to be dominant and the environment is assumed to be as "bad" as possible. Obviously, in the worst environment where the variation of f is maximal, $f(u,v)$ consists of separated peaks due to the contributions of other patches and $f(u,v)$ returns to 0 in between these peaks. Such an environment can be imagined as a set of stars - that is bright, small and separated objects.

Using equation 18, we obtain the following upper bound for the variation:

$$\mathcal{V}_{HK}(f_i) \le \max_{u,v}\left|\frac{\partial^2 f_i(u,v)}{\partial u\partial v}\right| + \max_{u}\left|\frac{\partial f_i(u,1)}{\partial u}\right| + \max_{v}\left|\frac{\partial f_i(1,v)}{\partial v}\right|. \quad (38)$$

Let the maximum distance between patches A_i and A_j be R, and let the maximum of the diameter of the circumcircle of patch be l. Using simple geometric argumens for $A(i,j,\omega(u,v))$ we obtain:

$$\max\left|\frac{\partial^2 f_i(u,v)}{\partial u\partial v}\right| \le \frac{4\rho_i}{A_i}\cdot\max_{j,u,v}\left(\tilde{B}_j\cdot\left|\frac{\partial^2 A(i,j,\omega(u,v))}{\partial u\partial v}\right|\right) \le$$

$$16\pi\rho_i \cdot \max_j \tilde{B}_j \cdot \frac{R^2}{A_i}, \quad (39)$$

$$\max_u \left|\frac{\partial f_i(u,1)}{\partial u}\right| \leq \frac{4\rho_i}{A_i} \cdot \max_{j,u} \left(\tilde{B}_j \cdot \left|\frac{\partial A(i,j,\omega(u,1))}{\partial u}\right|\right) \leq$$

$$8\pi\rho_i \cdot \max_j \tilde{B}_j \cdot \frac{lR}{A_i}, \quad (40)$$

$$\max \left|\frac{\partial f_i(1,v)}{\partial u}\right| \leq \frac{4\rho_i}{A_i} \cdot \max_{j,v} \left(\tilde{B}_j \cdot \left|\frac{\partial A(i,j,\omega(1,v))}{\partial v}\right|\right) \leq$$

$$8\rho_i \cdot \max_j \tilde{B}_j \cdot \frac{lR}{A_i}, \quad (41)$$

Since usually $R >> l$, the last two terms are negligible, thus we have:

$$Var(f_i(u,v)) \leq 16\pi \cdot \rho_i \cdot \frac{R^2}{A_i} \cdot \max_j \tilde{B}_j. \quad (42)$$

Since $\rho_{\max} \geq r$ and $\max_j \tilde{B}_j = ||\tilde{\mathbf{B}}||$, the norm of the radiosity error is

$$||\mathbf{B}^* - \tilde{\mathbf{B}}|| \leq \frac{16\pi\rho_{\max}}{(1-\rho_{\max})} \cdot \max_i \frac{R^2}{A_i} \cdot D^* \cdot ||\tilde{\mathbf{B}}||. \quad (43)$$

Note that $\max R^2/A_i$ expresses the ratio of the size of the scene and the size of the patches. From this inequality, in order to guarantee that $||\mathbf{B}^* - \tilde{\mathbf{B}}||/||\tilde{\mathbf{B}}|| \leq 10^{-2}$ for a scene where the ratio $\max R^2/A_i$ is 150, the required number of necessary transillumination directions is:

$\rho_{\max}$	M
0.2	$3 \cdot 10^6$
0.3	$5 \cdot 10^6$
0.4	$8 \cdot 10^6$
0.5	$12 \cdot 10^6$
0.6	$18 \cdot 10^6$
0.7	$29 \cdot 10^6$
0.8	$52 \cdot 10^6$
0.9	$120 \cdot 10^6$

Table 3: *Required number of transillumination directions in strongly heterogeneous environment*

These numbers are significantly higher than that of homogeneous-like environments. Thus we can conclude that the transillumination method is good for scenes consisting of large homogeneous patches but becomes inefficient for scenes of small, separated patches. However, if the scene is generally homogeneous, but has a few small objects with high emission power (like stars), then the transillumination method can still be used after preprocessing. In an appropriate preprocessing step, the power of the small sources should be shot onto the larger surfaces. Since these small objects receive just a small amount of power, they can be neglected in the further calculations when the transillumination method is applied.

6. Error reduction

So far, we have assumed that the approximate matrix is constant, that is, the same set of transillumination directions is used during the iteration. This means that if an accurate solution is required, then a large number of transillumination directions should be computed in each iteration step.

However, the computational burden can be reduced if different transillumination directions are used in different iteration steps. This means formally that the iteration formula is

$$\mathbf{B}^{(n+1)} = \mathbf{E} + \mathbf{F}^{(n)} \cdot \mathbf{B}^{(n)} \quad (44)$$

where $\mathbf{F}^{(n)}$ is the matrix used in step n. Note that $\mathbf{F}^{(n)}$ is computed from M number of transillumination directions. If $\lim_{n\to\infty} \mathbf{F}^{(n)} = \mathbf{F}$, then the iteration will converge to the real solution. It means practically that the number of directions must be increased in each step.

There is another alternative, however, that keeps the number of directions constant but uses a different set of transillumination directions in each iteration step. Let us note that $\mathbf{F}^{(n)}$ is not convergent in this case, thus neither can the iteration be convergent. However, applying a *semi-iterative scheme* [15], the iteration can be made convergent. The semi-iterative scheme at step K takes the average of the previous radiosity approximations:

$$\hat{\mathbf{B}}^{(K)} = \frac{1}{K}\sum_{n=1}^{K} \mathbf{B}^{(n+1)} =$$

$$\frac{1}{K}\Big(\sum_{n=1}^{K} \mathbf{F}^{(n)} \cdot \mathbf{E} + \sum_{n=2}^{K} \mathbf{F}^{(n)} \cdot \mathbf{F}^{(n-1)} \cdot \mathbf{E} + + \ldots\Big). \quad (45)$$

Obviously, if for any $i = 1,2,3\ldots$

$$\lim_{K\to\infty} \frac{1}{K}\sum_{n=i}^{K} \mathbf{F}^{(n)} \cdot \mathbf{F}^{(n-1)} \cdot \mathbf{F}^{(n-2)} \cdot \ldots \cdot \mathbf{F}^{(n-i+1)} \cdot \mathbf{E} = \mathbf{F}^i \cdot \mathbf{E}, \quad (46)$$

then $\hat{\mathbf{B}}^{(K)}$ will converge to the real solution ($\tilde{\mathbf{B}}$). Intuitively, this means that the $K \cdot M$ number of sample points can be distributed in K iteration steps if the directions used in different iterations steps are appropriate for $2K$-dimensional quadrature (factor 2 comes from the fact that the transillumination direction is a 2-variate function). Note that the proposed 2-dimensional Hammersly sequence does not meet this requirement since its coordinate is $u_n = n/M$, thus it can generate only those $\mathbf{F}^{(n)} \cdot \mathbf{F}^{(n-1)} \cdot \ldots \cdot \mathbf{F}^{(n-i+1)} \cdot \mathbf{E}$ products where the u-coordinate of transillumination direction increases for $\mathbf{F}^{(n-i+1)}, \ldots, \mathbf{F}^{(n-1)}, \mathbf{F}^{(n)}$. In order to avoid this effect, other sequences must be applied, or the Hammersley sequence must be scrambled.

Note that this problem is specific to quasi-Monte Carlo integration and does not occur when classical Monte Carlo method is used to select the sample points. A random matrix sequence of independent random variables $\mathbf{F}^{(n)}$ satisfying

$$E\{\mathbf{F}^{(n)}\mathbf{x}\} = \mathbf{F} \cdot \mathbf{x} \tag{47}$$

for any $\mathbf{x}$ also meets the requirement of equation (46).

7. Simulation results

The presented algorithm has been implemented and tested with different scenes and using different number of transillumination directions. In the actual implementation the simpler local visibility map method has been incorporated. In the simulations the error was approximated by the difference of the actual solution and a solution incorporating so high number of directions for which a further increase would cause only a negligible difference. It turned out that as long as the scene is not very inhomogeneous, the formula of homogeneous-like environments predicted the error quite accurately.

An image of a test scene and the simulation results are shown in figure 4, and in figure 5, respectively. The test scene consists of 145 diffuse spheres that are lit by an enclosing sphere which is divided into two hemispheres of different colours (lightsources cannot be seen in the image).

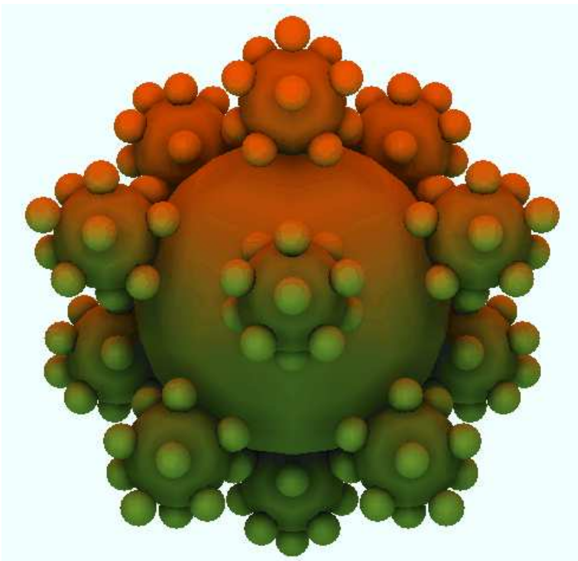

Figure 4: *A test scene: the "sphere-flake" (19976 patches, $\rho_{max} = 0.6$)*

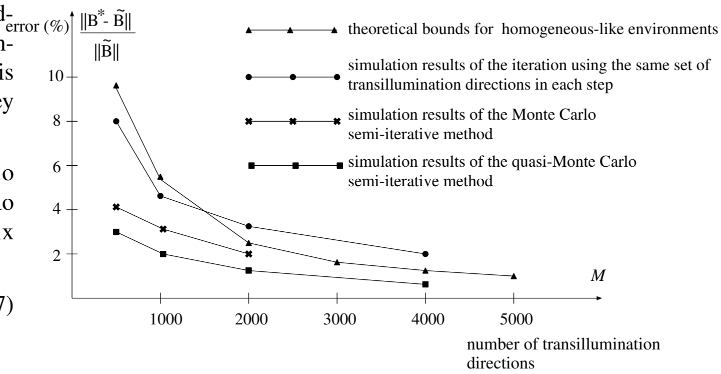

Figure 5: *Simulation results*

Acknowledgements

This work has been supported by the National Scientific Research Fund (OTKA), ref.No.: F 015884.

8. Conclusions and future work

This paper has presented a radiosity method where the double integral of the incoming power is calculated partly by analytical and partly by numerical, quasi-Monte Carlo techniques. The analytical integration prepared a function that is well suited for the quasi-Monte Carlo integration and error analysis. Theoretical bounds have been given for the required number of samples in the numerical integration to guarantee convergence and to find the solution with a given accuracy. In the error analysis two environment types were examined. We concluded that the presented method is efficient for environments that consists of large, homogeneous faces. For strongly heterogeneous environments of separated bright patches, the algorithm is recommended only if we can get rid of these separated bright patches by the proposed preprocessing.

The discussed semi-iteration step is promising, but more research needs to be done to find even more appropriate equidistribition sequences.

The actual implementation uses the local visibility map method that has $O(N^2)$ time and storage complexity. A future enhancement is the modification of this algorithm to avoid the majority of the cuttings of the patches. Instead of cutting, the radiosity of the patch can be corrected, resulting in a $O(N \log N)$ algorithm.

Another promising direction is the extension to non-diffuse environments. Since a transillumination step can be considered as tracing infinitely many parallel rays, the fundamental ideas of photon-tracing[?] or bi-directional ray-tracing[12] can be combined with this method. Unlike these stochastic ray tracing techniques, the non-diffuse transillumination method would require meshing, but could compute the contribution of infinitely many random walks.

References

1. D.R. Baum, H.E. Rushmeier, and J.M. Winget. Improving radiosity solutions through the use of analytically determined form-factors. *Computer Graphics (SIGGRAPH 89)*, pages 325–334, 1989.
2. J.C. Beran-Koehn and M.J. Pavicic. A cubic tetrahedra adaption of the hemicube algorithm. In James Arvo, editor, *Graphics Gems II*, pages 299–302. Academic Press, Boston, 1991.
3. Mark de Berg. *Efficient Algorithms for Ray Shooting and Hidden Surface Removal*. PhD thesis, Rijksuniversiteit te Utrecht, The Nederlands, 1992.
4. I. Deák. *Random Number Generators and Simulation*. Akadémia Kiadó, Budapest, 1989.
5. Ferenc Dévai. *Computational Geometry and Image Synthesis*. PhD thesis, Computer and Automation Institute, Hungarian Academy of Sciences, Budapest, Hungary, 1993.
6. E. Dutre. Bi-directional ray-tracing. In *Compugraphics '93*, Alvor, Portugal, 1993.
7. A. Glassner. *Principles of Digital Image Synthesis*. Morgan Kaufmann Publishers, Inc., San Francisco, 1995.
8. P. Hanrahan, D. Salzman, and L. Aupperle. Rapid hierachical radiosity algorithm. *Computer Graphics (SIGGRAPH proceedings '91)*, 1991.
9. A. Keller. A quasi-monte carlo algorithm for the global illumination in the radiosity setting. In H. Niederreiter and P. Shiue, editors, *Monte-Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, pages 239–251. Springer, 1995.
10. A. Keller. Quasi-Monte Carlo Radiosity. In X. Pueyo and P. Schröder, editors, *Rendering Techniques '96 (Proc. 7th Eurographics Workshop on Rendering)*, pages 101–110. Springer, 1996.
11. A. Keller. The fast Calculation of Form Factors using Low Discrepancy Sequences. In *Proc. Spring Conference on Computer Graphics (SCCG '96)*, pages 195–204, Bratislava, Slovakia, 1996. Comenius University Press.
12. E. Lafortune and E. Willems. Bi-directional ray-tracing. In *Compugraphics '93*, Alvor, Portugal, 1993.
13. D. Lischinski, B. Smits, and D.P. Greenberg. Bounds and error estimates for radiosity. In *Computer Graphics Proceedings, Annual Conference Series*, pages 67–75, 1994.
14. Cohen M.F. and D.B. Greenberg. The hemi-cube: A radiosity solution for complex environments. In *Proceedings of SIGGRAPH '85, Computer Graphics*, pages 31–40, 1985.
15. D. Mitchell. Ray Tracing and Irregularities of Distribution. In *Proc. 3rd Eurographics Workshop on Rendering*, pages 61–69, Bristol, UK, 1992.
16. L. Neumann. Monte carlo radiosity. *Computing*, 55:23–42, 1995.
17. M.E. Newell, R.G. Newell, and T.L. Sancha. A new approach to the shaded picture problem. In *Proceedings of the ACM National Conference*, page 443, 1972.
18. H. Niederreiter. *Random number generation and quasi-Monte Carlo methods*. SIAM, Pennsilvania, 1992.
19. M. S. Paterson and F. F. Yao. Efficient binary partitions for hidden-surface removal and solid modeling. *Discrete and Computational Geometry*, 5:485–503, 1990.
20. M. Sbert. *The Use of Global Directions to Compute Radiosity*. PhD thesis, Catalan Technical University, Barcelona, 1996.
21. G. Schaufler and W. Stürzlinger. Exact and error bounded approximation of local illumination. In *Compugraphics '95*, pages 327–333, Alvor, Portugal, 1995. GRASP.
22. P. Shirley. Discrepancy as a Quality Measure for Sampling Distributions. In *Eurographics '91*, pages 183–194, Amsterdam, North-Holland, 1991. Elsevier Science Publishers.
23. R. Siegel and J.R. Howell. *Thermal Radiation Heat Transfer*. Hemisphere Publishing Corp., Washington, D.C., 1981.
24. I. Sobol. *Die Monte-Carlo-Methode*. Deutscher Verlag der Wissenschaften, 1991. (in German).
25. J. Srejgyer. *Monte-Carlo Methods*. Fizmatgiz, Moscow, 1965. (in Russian).
26. L. Szirmay-Kalos and T. Fóris. Sub-quadratic radiosity algorithms. In *Winter School of Computer Graphics '97*, Plzen, Czech Republic, 1997. University of West Bohemia.
27. L. Szirmay-Kalos (editor). *Theory of Three Dimensional Computer Graphics*. Akadémia Kiadó, Budapest, 1995.