A Survey on Recent Developments in Global Illumination Techniques

Tomoyuki NISHITA† (Fellow), Hisanari OTSU††

† Hiroshima Shudo University / UEI Research, †† The University of Tokyo / UEI Research

Summary: The paper presents recent developments in global illumination techniques for producing photo-realistic images. In pursuit of the increasing requirement for photo-realism in various industries, research and development into global illumination techniques has received considerable attention within the graphics community from the early days. In this survey, we review various global illumination techniques, from the early attempts to recent advances in the two major categories of global illumination techniques: those based on the finite element method (FEM) and those based on the Monte Carlo method. Recently, in the field of the Monte Carlo based global illumination techniques, some advanced techniques incorporating photon density estimation or Markov chain Monte Carlo method is actively researched in the graphics community. We attempted to pave the way for the reader to understand such advanced topics by describing historical background and motivation behind these techniques. We hope this survey to help the readers to comprehend the overview of the field and the characteristics of these techniques, and to indicate the next direction of the research in the field of global illumination computation.

Keywords: survey, rendering, global illumination, light transport, finite element method, Monte Carlo method

1. Introduction

From the early days of computer graphics, rendering photo-realistic images has been one of the most important topics in the field. In order to achieve photo-realism, the rendering techniques categorized as global illumination techniques, which simulates the propagation of light energy taking into account inter-reflections of light between surfaces, has been researched and developed. Research into global illumination techniques started from the research on the Radiosity method, which is known to originate from Nishita and Nakamae’s technique†† (Fig. 1). In some literature, global illumination techniques are also known as the light transport techniques, since these techniques actually simulate the transportation of light emitted from light sources to the sensor. Global illumination techniques can drastically improve the appearance of the rendered images, so nowadays their application covers various industries including the movie production industry, the game production industry, and the medical industry. In order to pursue all of the possible demands from these industries and because of their increasing importance in the field of photo-realistic rendering, development and research into global illumination techniques will still be an important topic in future research of the field of computer graphics.

However it is a challenging problem to develop efficient global illumination techniques due to the underlying complexities of materials or the geometries in the rendered scenes. Various criteria can be assumed in order to define the efficiencies of global illumination techniques. For example, Robustness is one such criterion, and it defines the ability to render a scene properly irrespective of the scene configuration. For instance, a scene with inter-reflections between glossy surfaces or a scene containing complex occlusions are known to be difficult cases and even the latest techniques often show poor performance. Rendering speed is another important criterion, which indicates how fast the rendering process is completed. Accuracy and precision are also important criteria for numerical techniques. We note that some of these criteria are not always independent and global illumination techniques often compromise conflicting criterion in some way.
The purpose of this survey is to review the various global illumination techniques from early attempts to recent developments, and to help the readers to comprehend an overview of the field and the characteristics of these techniques. In this survey, we mainly focus on the global illumination effects occurred on surfaces. Although the theory and techniques with participating media has been actively researched, we will not review these techniques in details. Moreover unfortunately due to the lack of space and time, we omit some types of global illumination techniques including pre-computation based techniques, virtual point light (VPL) based techniques, real-time global illumination techniques. The types of global illumination techniques handled in the survey is summarized in Fig. 2.

The organization of the paper is as follows. First in section 2, we will describe the theory of global illumination computation. In section 3, we review the finite element method (FEM) based techniques. In section 4, we review the Monte Carlo method based techniques. Finally, in section 5, we conclude the survey.

![Fig. 2](image)

Classification of global illumination techniques in the survey

<table>
<thead>
<tr>
<th>Global illumination techniques</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEM based techniques</td>
<td>4</td>
</tr>
<tr>
<td>Discretization</td>
<td>4.2</td>
</tr>
<tr>
<td>Form factor computation</td>
<td>4.3</td>
</tr>
<tr>
<td>Linear system computation</td>
<td>4.4</td>
</tr>
<tr>
<td>Monte Carlo method based techniques</td>
<td>5</td>
</tr>
<tr>
<td>Path sampling</td>
<td>5.2</td>
</tr>
<tr>
<td>Irradiance / radiance caching</td>
<td>5.3</td>
</tr>
<tr>
<td>Photon density estimation</td>
<td>5.4</td>
</tr>
<tr>
<td>Markov chain Monte Carlo (MCMC)</td>
<td>5.5</td>
</tr>
</tbody>
</table>

### 2. Theory of Global Illumination

In this section we will briefly describe the theoretical aspects of global illumination computations. Global illumination computation was initially formulated as the rendering equation by Kajiya\(^2\), which is now known as the hemispherical formulation of global illumination. The concept of the rendering equation was derived from the literature of the radiative heat transfer. With the development of global illumination techniques, the theory of global illumination has become more sophisticated. Arvo\(^3\) precisely formulated radiometry and rendering equation using measure theory, and proposed the linear operator formulation of global illumination computation, which is useful for studying the properties of global illumination computation using techniques of functional analysis. Veach\(^4\) extended Arvo’s linear operator formulation, named the self-adjoint linear operator formulation, and unifies bidirectional light transports by self-adjoint operators to achieve unifying support for non-symmetric BSFDs. Veach also proposed the path integral formulation for global illumination computation, which simplifies the hemispherical form of the rendering equation such that measurements can be obtained from the direct solution rather than the solution of an integral equation. Some research has focused on the theory of 2D global illumination, which is useful for the detailed analysis of the global illumination techniques\(^5\).

From among the above formulations, we review two major formulations: the hemispherical formulation and the path integral formulation. For simplicity, we define the useful terms beforehand in Table 1.

![Table 1](image)

<table>
<thead>
<tr>
<th>Notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M)</td>
<td>Number of pixels in the image</td>
</tr>
<tr>
<td>(\mathcal{M})</td>
<td>Union of all surfaces in the scene</td>
</tr>
<tr>
<td>(\mathcal{S}^2)</td>
<td>Set of outward directions in the unit sphere</td>
</tr>
<tr>
<td>(N(x))</td>
<td>Geometry normal at surface point (x)</td>
</tr>
<tr>
<td>(x, x', x'')</td>
<td>Points on (\mathcal{M})</td>
</tr>
<tr>
<td>(x_0, x_1, ..., x_n)</td>
<td>Light path with vertices (x_0, x_1, ..., x_n)</td>
</tr>
<tr>
<td>(x)</td>
<td>Light path in (\mathcal{P})</td>
</tr>
<tr>
<td>(x_n)</td>
<td>Light path in (\mathcal{P}_n)</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>Solid angle measure</td>
</tr>
<tr>
<td>(\sigma^\perp)</td>
<td>Projected solid angle measure</td>
</tr>
<tr>
<td>(W_e^{(j)}(x, \omega)), (W_e^{(j)}(x \leftrightarrow x'))</td>
<td>Responsively function for (j)-th pixel at (x)</td>
</tr>
<tr>
<td>(L_i(x, \omega))</td>
<td>Incident radiance function at (x) and direction (\omega)</td>
</tr>
<tr>
<td>(L_o(x, \omega))</td>
<td>Outgoing radiance function at (x) and direction (\omega)</td>
</tr>
<tr>
<td>(x_{m}(x, \omega_o))</td>
<td>First point in (\mathcal{M}) from a point (x) to a direction (\omega_o) (ray-casting function)</td>
</tr>
<tr>
<td>(f_s(x, \omega_i \rightarrow \omega_o))</td>
<td>BSDF (bidirectional scattering distribution function)</td>
</tr>
<tr>
<td>(f_s(x \rightarrow x' \rightarrow x''))</td>
<td>Three-point form BSDF</td>
</tr>
<tr>
<td>(B(x))</td>
<td>Radiosity at (x)</td>
</tr>
<tr>
<td>(E(x))</td>
<td>Radiant exitance at (x)</td>
</tr>
<tr>
<td>(\rho(x))</td>
<td>Diffuse reflectance at (x)</td>
</tr>
</tbody>
</table>

#### 2.1 Hemispherical formulation

Among the available formulations, the so-called hemispherical formulation is one of the most common, and various literature have adopted this formulation as an introductory theory
for global illumination computation\(^5\), \(^6\), \(^7\). Due to a lack of the space, we will only describe a brief overview of the formulation, please refer to above-mentioned literature for details.

The purpose of global illumination computation is to compute measurements of the light energy that contribute to each pixel. We let the measurements for \(j\)-th pixels be \(I_j\) \((j = 0, ..., M)\). Then the measurement equation is defined as

\[
I_j = \int_{M \times S^2} W_e^{(j)}(x, o) L_o(x, o) dA(x) d\sigma^+_g(o).
\] (1)

Then we define the light transport equation. This equation is defined as an integral equation with respect to the outgoing radiance function \(L_o(x, o)\), which is obtained from energy conservation between the incident and emitted and scattered radiance at a point, assuming that we observe the equilibrium radiance:

\[
L_o(x, o_a) = L_e(x, o_a) + \int_{S^2} L_o(x, o_a) W_e(x, o_1, o_2) d\sigma^+_g(o_1) d\sigma^+_g(o_2).
\] (2)

Converting the projected solid angle measure \(\sigma^+_g\) to the area measure \(A\), we obtain the three-point form of the light transport equation using three surface points \(x, x', x''\) instead of angles:

\[
L(x' \rightarrow x'') = L_e(x, o_a) + \int_{S^2} L(x \rightarrow x') f_g(x, x' \rightarrow x'') dA(x),
\] (3)

where \(G(x' \rightarrow x'')\) is the geometry term obtained as a result of the measure conversion (Jacobian), which is defined as

\[
G(x' \rightarrow x'') = V(x' \rightarrow x'') \frac{\cos \theta_0 \cos \theta_1}{||x - x'||^2}
\] (4)

where \(\theta_0\) and \(\theta_1\) are the angles between the surface normal at \(x'\) and \(x''\) and the outgoing vectors from \(x'\) to \(x\) and \(x''\), respectively. \(V(x' \rightarrow x'')\) is the visibility term defined as \(V(x' \rightarrow x'') = 1\) if \(x\) and \(x'\) are mutually visible, otherwise \(V(x' \rightarrow x'') = 0\).

### 2.2 Path integral formulation

The path integral formulation of global illumination computation was proposed by Veach\(^8\). Unlike the hemispherical formulation, the path integral formulation simplifies the measurement equation and the light transport equation to use direct integration rather than an integral equation. Some light path sampling based global illumination techniques, such as bidirectional path tracing\(^9\), can be well described by using this formulation.

The measurement equation in the path integral formulation is obtained by recursively expanding Equation (3), which is defined as the integral on the light path space \(P\) with the light path measure \(\mu\). The light path space is defined as \(P = \bigcup_{n=1}^\infty P_n\) where \(P_n = \{x_0 x_1 \cdots x_n | x_0, x_1, ..., x_n \in \mathcal{M}\}\) is the set of light paths with \(n + 1\) edges (or length). The light path measure is defined as \(\mu(D) = \sum_{n=1}^\infty \mu_k(D \cap P_n)\) where \(D \subset P\) and \(\mu_k\) is the area-product measure\(^9\) on \(P\) defined as

\[
\mu_k(D) = \int_D dA(x_0) \cdots dA(x_n).
\] (5)

The function \(f_j\) is the measurement contribution function and is defined for a light path with length \(n\) as

\[
f_j(x_n) = L_e(x_0 \rightarrow x_1) \cdots \int_{x_{k-1}}^{x_k} G(x_{k-1} \rightarrow x_k) f_j(x_{k-1} \rightarrow x_k) \rightarrow x_{k+1})
\] (6)

\[
\cdot G(x_{n-1} \rightarrow x_n) W_e^{(j)}(x_{n-1} \rightarrow x_n).
\]

### 3. FEM Based Techniques

In this section we review the major global illumination techniques according to our classifications. The first section in this classification considers finite element method (FEM) based techniques, which achieves global illumination computation by using FEM. These techniques were originally derived from the Radiosity method\(^{1, 10, 11}\). Through the development of the Radiosity method, various techniques have been proposed. These techniques can be separated into groups according to the focus of the execution process used in the method: discretization, form factor computation, and linear system computation. In this survey, we adopted these focuses as a way of classifying the sub-techniques in the Radiosity method. Although other focuses such as rendering are still important, they are omitted due to a lack of space.

#### 3.1 Theory of FEM based technique

In the ordinary Radiosity method, all surfaces are assumed to be Lambertian diffuse surfaces. From this assumption, we obtain the radiosity equation from Equation (3):

\[
B(x) = E(x) + \rho(x) \int_{\mathcal{M}} B(x') G'(x \rightarrow x') dA(x'),
\] (7)

where \(G'(x \rightarrow x') = G(x \rightarrow x')/\pi\).

In order to approximate \(B(x)\) following the standard FEM setup, the domain of \(B(x)\), that is, the surface mesh \(\mathcal{M}\) is subdivided into small elements and some nodes are assigned on those elements. Then \(B(x)\) is approximated by a linear sum of the basis functions:
\[
B(x) \approx \bar{B}(x) = \sum_{i=1}^{n_\mathcal{B}} B_i N_i(x),
\]

where \(n_\mathcal{B}\) is the number of basis functions and \(N_i(x)\) is the \(i\)-th basis function. Various selections for the basis function are possible, and this has influence on the final result. The simplest of these is the constant basis function and is defined as \(N_i(x) = 1\) if \(x\) is within the element and otherwise \(N_i(x) = 0\). Although many techniques adopt the constant basis function, some research has investigated the possibility of using non-constant basis functions.\(^{26}\)

By applying a weighted residual method such as the point collocation or the Galerkin method, we obtain a way to fit \(\bar{B}(x)\) to \(B(x)\) by solving a linear equation. Specifically, when the constant basis function and the Galerkin method are used, we obtain \(^3\) a form of the radiosity equation that can be seen various literatures:

\[
B_i = E_i + \rho_i \sum_{j=1}^{n_\mathcal{B}} B_j F_{ij},
\]

where \(n_\mathcal{E}\) is the number of elements and \(\rho_i\) is the diffuse reflectance associated with the \(i\)-th element, and \(F_{ij}\) is the form factor defined as

\[
F_{ij} = \frac{1}{A_i} \int_M \int_M G(x \leftrightarrow x') dA(x') dA(x),
\]

where \(A_i\) is the area of the \(i\)-th element.

Some techniques uses the hemispherical form of the Equation (10), which converts the area measure of the inner integral into solid angle measure:

\[
F_{ij} = \frac{1}{A_i} \int_M \int_S G''(\mathbf{x}, \omega) d\sigma(\omega) dA(x),
\]

where \(G''(\mathbf{x}, \omega) = \pi^{-1} \cdot \cos \theta_{\mathbf{x} \omega} \cdot V(\mathbf{x}, \mathbf{x}_M(\mathbf{x}, \omega))\).

### 3.2 Discretization

The first classification involves in discretization of the radiosity equation with FEM. We now review some techniques related to this classification.

First we focus on the discretization of the mesh. It is known that various conditions that are introduced in the discretization step on a mesh, such as mesh density or continuity, have a bad influence on the accuracy of the rendered images.\(^5\) Various techniques have been investigated in an attempt to alleviate this inaccuracy. Lischinski et al.\(^{15}\) and Heckbert\(^{7}\) proposed a mesh subdivision technique for discontinuities introduces by change of illumination. Campbell and Fussell\(^{16}\) proposed an adaptive mesh generation scheme utilizing a binary space partitioning (BSP) tree. Smits et al.\(^{21}\) utilize the idea of importance transport, which is a dual concept of the radiance concept for adaptive mesh subdivision. More recently, Dobashi et al.\(^{12}\) proposed a Radiosity method for point-sampled geometries which does not need explicit mesh subdivision (Fig. 3).

Another important concept related to discretization is to employ a hierarchical structure. Cohen et al.\(^{18}\) introduced a two-level hierarchical mesh structure and proposed a modified version of the form factor computation. Hanrahan et al.\(^{19}\) extended and generalized Cohen’s two-level hierarchical radiosity technique to multi-level. By introducing multi-level hierarchy, the computational complexity is drastically reduced to linear time.

Smits et al.\(^{22}\) utilize clustering to accelerate the hierarchical radiosity. Some methods focus on the hierarchical structure of the basis functions used in the FEM formulation. Gortler et al\(^{20}\). Propose wavelet radiosity which utilizes hierarchical structure for wavelet basis functions converted from the normal finite element basis functions, and which also allows us to construct a linear time technique.

### 3.3 Form factor computation

One way to compute the form factor is to use an analytical solution. Goral et al.\(^{16}\) utilized the contour integral form of Equation (10) to compute the form factor between rectangular elements without occlusions. Nishita and Nakane\(^1\) resolved the restriction on occlusions by adding an extra visibility handling scheme. Baum et al.\(^{23}\) partially used an analytical approach to solve the inner part of Equation (10). Another way is to use numerical solution techniques. This approach was initially attempted by Cohen et al.\(^{11}\) and is known as the hemi-cube technique. This technique utilizes the hemispherical form of the form factor (Equation (11)) under the assumption that each element has a constant inner integral across the element, which means the form factor for an element can be evaluated as a single integral. The inner integral is computed numerically by projecting scene surfaces into hemi-cube, which is defined as a

![Fig. 3](image-url)
half-cube structure whose faces are separated into different elements. Projection onto the hemi-cube can be achieved using
standard rasterization techniques such as the Z-buffer algorithm, or
even with graphics hardware. Various techniques that stem from
this approach have been proposed. Max\textsuperscript{20} proposed an optimal
sampling scheme for the hemi-cube based on some observations on
the arrangement of the sample directions. Some methods utilize
other shapes that are different from the hemi-cube. Silvon and
Puech\textsuperscript{23} proposed an alternative technique to the hemi-cube
 technique that utilizes a plane parallel to the surface of the element
with the observation that the angular contribution of the geometry
term is relatively small in the area that is nearly-tangential to the
surface.

Some numerical solution techniques for the form factor
computation utilize the Monte Carlo method. Sherr\textsuperscript{22, 28} makes use
of knowledge of integral geometry (or geometric probability) to
calculate the form factor by using the Monte Carlo method.

### 3.4 Linear system computation

In the Radiosity method, iterative methods known as
relaxation methods, such as the Jacobi, Gauss-Seidel or Southwell
methods are often used for linear system computation. However in
order to facilitate the potential of these techniques, the execution
steps in these methods are often rearranged and interleaved. For
instance, the progressive refinement radiosity method proposed by
Cohen et al.\textsuperscript{26} relaxes the execution flow of classical linear system
solution methods to achieve progressive rendering. In this
technique, the form factor computation and the linear system
computation are partially interleaved.

The stochastic relaxation radiosity\textsuperscript{30-32} technique makes use
of the Monte Carlo method to solve the linear system of the
radiosity equation. This technique alleviates the intermediate step
of the Jacobi iteration to converge in probability by replacing the
matrix-to-vector multiplication with an unbiased Monte Carlo
estimator, which significantly reduces the evaluation cost of
handling large form factor matrix directly.

The discrete random walk radiosity\textsuperscript{33-35} technique is another
way of employing the Monte Carlo method to a linear system
solution. This technique utilizes the discrete random walk process
to estimate the solution of the linear system. In this technique,
several counting strategies are considered to obtain an estimated
value of the solution with low variance.

### 4. Monte Carlo Method Based Techniques

Monte Carlo method based techniques is another major way
to achieve global illumination computation, which utilizes the
Monte Carlo method to numerically solve the rendering equation.

Although various techniques have been developed and researched,
and classification criteria are vast, we adopted the classification
with 4 categories: path sampling based techniques, irradiance
caching based techniques, photon density estimation based
techniques, and MCMC (Markov chain Monte Carlo) based
techniques. Unfortunately, due to a lack of space and time we
omitted some important categories of techniques: Pre-computation
based techniques (e.g. see Fig. 4) or VPL based techniques. Also,
we will not delve into the topics like participating media (e.g. see
Fig. 5) or adaptive sampling and reconstruction (a.k.a. filtering),
although they are very fascinating field and have been actively
researched in the latest research. For techniques in these categories,
please refer to other literature.

#### 4.1 Theory of Monte Carlo method

In this section we briefly introduce some selected theoretical
concepts related to the Monte Carlo method.

1. **Monte Carlo integration**

In global illumination computation with the Monte Carlo
method, the central tool is the Monte Carlo integration, which is a
technique for numerically integrating arbitrary functions. We let
the target integral we want to estimate be

\[ I = \int_{\Omega} f(x) \, dx, \]

where \( f \) is a real-valued function defined on \( \Omega \), and \( \mu \) is some
measure on \( \Omega \). We let a probability density function (PDF) be
\( p = dP/du \) on \( \Omega \) with some probability space \((\Omega, \mathcal{F}, P)\) where
\( \mu \ll P \) (\( \mu \) is absolutely continuous to \( P \)) and \( \text{supp}(f) \subseteq \text{supp}(\mu) \)
(non-zero region of \( \mu \) contains non-zero region of \( f \)). We
consider to estimate the integral \( I \) with \( N \) number of
independent and identically distributed (i.i.d.) random variables
\( X_i (i = 1, \ldots, N) \) generated from the distribution \( p \). Then an
estimator \( \langle I \rangle_N \) for the integral \( I \) can be defined as

\[ \langle I \rangle_N = \frac{1}{N} \sum_{i=1}^{N} f(X_i) \]

We note that \( \langle I \rangle_N \) converges to \( \mathbb{E}_p[I] = I \) almost surely
as \( N \rightarrow \infty \) by the strong law of large number (SLLN), where
\( \mathbb{E}_p[Y] \) denotes the expected value of \( Y = f(X) \) where \( X \sim p \),
which is defined as

\[ \mathbb{E}_p[Y] = \int_{\Omega} f(x) p(x) \, dx. \]

The variance of \( Y \) is another important quantity which
indicates how \( Y \) is dispersed around \( \mathbb{E}_p[Y] \).
\[ \text{Var}_p[Y] = \mathbb{E}_p \left[ (Y - \mathbb{E}_p[Y])^2 \right] \]
\[ = \int_\Omega (f(x) - \mathbb{E}_p[X])^2 p(x) d\mu(x). \]  

In Monte Carlo based techniques, the variance is one of the most important quantities because the error convergence speed is closely related to the variance. The asymptotic result of this relation can be easily confirmed by the central limit theorem (CLT). According to the theorem, \( N^{-1/2} \cdot (\langle l \rangle_N - 1) \) is converged in distribution to a normal distribution \( N(0, \text{Var}_p(\langle l \rangle_N)) \).

2) Error and bias

We let \( l \) be a target quantity to be estimated and its estimator be \( \langle l \rangle_N \). Then the error of the estimator (\( l \)) is defined by \( \langle l \rangle_N = l \). The bias of the estimator \( \langle l \rangle_N \) is defined as the expected value of the error:

\[ B(\langle l \rangle_N) = \mathbb{E}_p(\langle l \rangle_N - l). \] 

(16)

The estimator \( \langle l \rangle_N \) is called unbiased if \( B(\langle l \rangle_N) = 0 \) for all \( N \), i.e., \( \mathbb{E}_p(\langle l \rangle_N) = l \) for all \( N \). The estimator \( \langle l \rangle_N \) is called consistent if \( \langle l \rangle_N \) is converged to \( l \) in probability as \( N \to \infty \), that is, for all \( \varepsilon > 0 \)

\[ \lim_{N \to \infty} P(|\langle l \rangle_N - l| > \varepsilon) = 0. \] 

(17)

We note that unbiasedness and consistency are orthogonal concepts, although it is often confused, so we can assume a unbiased and inconsistent estimator, for instance. In global illumination computation, a techniques is called unbiased or consistent according to unbiasedness or consistency of the underlying estimator in the technique.

3) Variance reduction

As we described above, in order to increase the efficiency of the Monte Carlo estimator, it is important to reduce variance of the estimator. In this part of the section, we briefly introduce some variance reduction techniques which is often used in the global illumination computation. Stratified sampling is one of variance reduction techniques. In this technique the sample space \( \Omega \) is subdivided into \( m \) disjoint sub-spaces \( \Omega_i (i = 1, ..., m) \) called stratum, and for each \( \Omega_i \) random sample \( X_{i,j} \) is generated. However, stratified sampling is not effective on a high dimensional space because the number of stratum and the number of samples required for each stratum are increased exponentially. Latin hypercube sampling addresses the problem by generating samples so that only one sample is assigned for each row and each column. Quasi-Monte Carlo (QMC) utilizes a low discrepancy sequence instead of a sequence of random numbers for estimating an integral. A low discrepancy sequence is deterministically chosen so that the samples are uniformly distributed over the sample space. An advantage of QMC over standard Monte Carlo is faster convergence ratio. Although there is little application in global illumination computation, variance reduction techniques such as control variate or Rao-Blackwellization is also known and widely used in the other fields.

4) Importance sampling

Selection of the distribution \( p \) in Equation (13) has considerable influence on the variance of the estimator. A techniques of improving the estimator by choosing good distribution \( p \) is called the importance sampling. Although it is one of the variance reduction techniques, due its importance and effectiveness in global illumination computation we separate the technique in an independent part. A known result for the importance sampling improve efficiency if the distribution \( p \) is chosen so that \( p \) is similar to the integrant \( f \). More specifically, the optimal choice of the distribution \( p^* \) that minimizes the variance is known to be

\[ p^*(x) = \frac{|f(x)|}{\int_\Omega |f(x)| d\mu(x)}. \]  

(18)

We note that in global illumination computation the integrant is always positive, so the optimal \( p^* \) simply becomes the normalized version of \( f \), and in this case the variance is always zero.

Multiple importance sampling proposed by Veach\(^{(4)}\) is an importance sampling technique which combines multiple samples from different distributions (or techniques). These samples are combined with weights associated with distributions. The technique is useful for an integrant that is difficult to sample with a single importance distribution. We let the distributions be \( p_i (i = 1, ..., N) \) and the weighing function associated with the distribution \( p_i \) be \( w_i \). Then the estimator for multiple importance sampling is defined as

\[ \langle l \rangle_{\text{MIS}} = \sum_{i=1}^{N} \frac{1}{N_i} \sum_{j=1}^{N_i} w_i(X_{i,j}) \frac{f(X_{i,j})}{p_i(X_{i,j})}. \]  

(19)

where \( N_i \) is a number of samples taken from the distribution \( p_i \) and \( X_{i,j} \sim p_i \). The estimator is unbiased and consistent if following conditions on weights are satisfied: 1) for all \( x \in \text{supp}(f) \), \( \sum_{i=1}^{N} w_i(x) = 1 \), 2) \( \sum_{i=1}^{N} \text{supp}(w_i) \subseteq \text{supp}(p_i) \). We note that these conditions imply that we do not need to sample whole domain of \( \Omega \) for all techniques. More specifically, for all \( x \) in non-zero part of \( f \), at least one distribution generates \( x \), that is,
\[ \text{supp}(f) \subseteq \bigcup_{i=1}^{N} \text{supp}(p_i). \] (20)

Another focus is the selection of the weights. Veach proposed some simple but effective techniques based on heuristics. One technique is balance heuristics:

\[ w_i(x) = \frac{N_i p_i(x)}{\sum_{k=1}^{N} N_k p_k(x)} \] (21)

which is proved to have good variance bound (Theorem 9.2 of Veach\textsuperscript{4}).

5) Markov chain Monte Carlo (MCMC)

Although importance sampling is effective for many configuration, sampling from the general path space is still challenging. In order to resolve the problem, some recent global illumination techniques begin to use Markov chain Monte Carlo (MCMC) method. In this part of the section, we briefly review the theoretical aspect of MCMC that is useful for understanding recent MCMC based global illumination techniques.

The basic idea of MCMC is to use a correlated sequence of random variables called a Markov chain, which is a sequence of random variable whose samples depend only on samples one before. Specifically, a Markov chain is defined as a sequence of random variables \( X_1, X_2, \ldots \) such that \( X_{i+1} \) depends only on \( X_i \) for \( i = 1, 2, \ldots \), that is, \( X_{i+1} \sim K(X_i \rightarrow X_{i+1}) \) where \( K(x \rightarrow y) \) is a conditional probability density function called the transition kernel. The algorithm which generates stationary Markov chain according to some distribution is called Markov chain Monte Carlo (MCMC). A Markov chain is called stationary if there exists a distribution \( \pi \) such that

\[ \pi(x) = \int_{\Omega} K(y \rightarrow x) \pi(y) d\mu(y). \] (22)

In order to obtain a stationary Markov chain, many MCMC techniques utilize a stronger condition called detailed balance condition. A transition kernel \( K \) satisfies the detailed balance condition if there exists a distribution \( \pi \) for all \( x, y \in \Omega \),

\[ K(x \rightarrow y) \pi(x) = K(y \rightarrow x) \pi(y). \] (23)

A Markov chain which satisfies this condition is called a reversible Markov chain, and it can be shown that in this case the distribution \( \pi \) is stationary.

In order to numerically compute Equation (12) by using Monte Carlo integration with MCMC, we want to generate samples according to some distribution proportional to \( f \). In order to achieve the goal, now we review one of the most famous techniques: the Metropolis-Hastings (MH) algorithm. The transition kernel for the algorithm, a.k.a., Metropolis-Hastings update, is defined as follows. Given the current sample \( X_i \), we choose a tentative sample \( X'_i \sim T(X_i \rightarrow X'_i) \), where \( T(x \rightarrow y) \) is transition function. Then the next sample \( X_{i+1} \) is defined as

\[ X_{i+1} = \begin{cases} X'_i & \text{with probability } a(X_i \rightarrow X'_i), \\ X_i & \text{otherwise,} \end{cases} \] (24)

where \( a(X \rightarrow X') \) is acceptance probability defined as

\[ a(X \rightarrow X') = \min \left( 1, \frac{f(X')T(X' \rightarrow X)}{f(X)T(X \rightarrow X')} \right). \] (25)

We note that Equation (25) implies \( f \) need not to be normalized, and we can observe the transition kernel associated with the algorithm:

\[ K(x \rightarrow y) = a(x \rightarrow y)T(x \rightarrow y) + \left( 1 - \int_{\Omega} a(x \rightarrow y)T(x \rightarrow y)d\mu(y) \right) \delta_x(y) \] (26)

preserves Equation (23) with \( f \) as the stationary distribution.

As for the convergence property of a Markov chain, we can show that if a Markov chain satisfied the following two conditions: 1) \( f \)-irreducible which means that any part of the state space can be reached from the selection of the initial state, and 2) aperiodic which means the Markov chain does not contain cycles, a sequence of samples is converged to follow the stationary distribution. And similar to SLLN for i.i.d. random number case (see Equation (13)), the expected value of the estimator \( \langle I \rangle_N \) converges to \( I \) even when a Markov chain satisfying above conditions is used as a sequence of samples, which is known as Ergodic Theorem. We also note that MH update satisfies these conditions with a loose conditions with \( f \) and the transition function\textsuperscript{75}. Convergence property of Markov chain is important topic but we will not handle more it in the survey, for details please refer to Meyn and Tweedie\textsuperscript{76}, Robert and Casella\textsuperscript{77} etc.

4.2 Local sampling based techniques

In this section, we introduce local sampling based techniques, which focus on solving the rendering equation by Monte Carlo method with sampling light paths with a sequence of local sampling on scene surfaces. Here we call the techniques in this section are local because all sampling processes are based on the distributions that is defined on a surface point. It implies a light path is generated by sequentially sampling next surface points according to some distribution associated with the points. Specifically, we focus on the global illumination techniques with Monte Carlo method in which the techniques sample the light paths with i.i.d. random variables. This type of the techniques does not facilitate the correlation of samples like techniques described in
section 4.5, and many of them are statistically unbiased, so analysis and implementation of the techniques is relatively simple.

An idea of sampling light path is originated from the Whitted’s ray tracing\(^ {30} \) which handles global illumination restricted to perfect specular surfaces. While the techniques itself is deterministic, he mentioned the possibility of using random sampling to simulate specular or glossy surfaces. Initial attempt to utilize the Monte Carlo method for global illumination computation is the distributed ray tracing proposed by Cook et al.\(^ {37} \) which can handle graphics effects such as depth-of-filed or motion-blur in the framework of the ray tracing. As we describes in section 2, first comprehensive framework for global illumination computation is developed by Kajiya\(^ {2} \), whose technique is now known as the path tracing. Similar to the ray tracing, the path tracing samples a light path by tracing a ray from a camera, and when the ray intersects with the scene surface, next ray direction is sampled from the distribution associated with the BSDF on the intersected point, and the next ray is propagated to the direction until the ray hits a light source. Along with the propagation, the radiance carried along the sampled path is computed according to the equations describes in section 2.1. Following Kajiya’s path tracing, various techniques have been proposed. As an reverse approach of the path tracing, Arvo and Kirk\(^ {39} \) proposed the inverse path tracing (a.k.a. the light tracing, or the particle tracing). The bidirectional path tracing\(^ {41, 42} \) combines both light paths traced from a camera and a light source, which can efficiently generate a light path which is difficult to sample by path tracing or light tracing, e.g., sampling from the scene that a camera and a light source are placed in different rooms connected by a small window. According to the combination of the vertex connection between sampled light paths, some light paths with same lengths sampled from different denses are obtained. These light paths are combined by the multiple importance sampling (MIS) described in section 4.1.

4.3 Irradiance / radiance caching based techniques

From this section to section 5.4, we will introduce the techniques based on caching, which stores intermediate data into some data structure and utilize them for rendering. In this section we introduce the techniques based of the caching of irradiance or radiance, which is the radiometric quantity handled in the light transport.

1) Basics

The irradiance caching proposed by Ward et al.\(^ {43} \) is the first technique of this kind, and the techniques is applicable to scenes with diffuse surfaces like classical radiosity method. A key observation is that on diffuse surfaces the indirect illumination due to the diffuse inter-reflection has relatively low variation or frequency. For those scene with diffuse surfaces, as we described in section 3.1, the rendering equation can be rewritten with irradiance, which means the transported quantity can be represented as irradiance. Irradiance values are associated with positions on surface. Although the technique is biased due to the interpolation of irradiance values, using the technique, a rendered image without high frequency noise can be obtained in relatively small running time. Execution steps of this techniques is as follows. If a ray emitted from a camera is intersected with scene surface, the technique collects cached irradiance values from the octree storage. Only the values which can be well interpolated at the position using a metric with the error estimation based on the split-sphere heuristics are collected. Reciprocals of estimated error values are also used as weights for interpolation. If at least a cached value is collected, the interpolation using the samples are dispatched. Otherwise if no cached value is found, a new irradiance value is computed with ordinary techniques like path tracing, and computed irradiance value is stored into the octree storage.

2) Radiance caching

Some recent techniques attempt to alleviate the restriction that the irradiance caching is only applicable to scenes with diffuse surfaces, and to support the material like glossy surfaces. Krivanek et al.\(^ {44} \) proposed a technique called the radiance caching which extends the irradiance caching to support glossy surfaces by precomputing all scene BSDFs as spherical harmonics (SH) representation. In the octree storage, radiance is recorded instead of irradiance. Extending their technique, Scherzer et al.\(^ {46} \) utilizes improved the efficiency of the radiance caching by alleviating the evaluation cost of SHs by introducing pre-filtered MIP maps. Some techniques attempt to support glossy surfaces without SHs. The radiance-cache splatting proposed by Gautron et al.\(^ {45} \) improved the rendering part of the irradiance / radiance caching by eliminating nearest-neighbor queries to the data structure. In this technique, all stored values are splatted to the image plane and interpolation is computed with the information. The technique can be implemented in GPUs except for irradiance / radiance cache recording parts. The anisotropic radiance-cache splatting proposed...
by Herzog et al.\textsuperscript{46} extended the idea to utilize the orientation of splatted cache items in order to reduce errors along illumination gradient.

3) Improving error estimation

An important focus on the irradiance / radiance caching techniques is how to design the error estimation scheme. Ward et al.\textsuperscript{60} proposed a new interpolation scheme utilizing the irradiance gradient, which improves the accuracy of error estimation utilizing first order gradient of irradiance. In this technique, additional information for gradient computation is stored in the octree storage as well as irradiance values, and later in the interpolation step these information is used for gradient computation. Tabellion and Lamorlette\textsuperscript{49} proposed an interpolation technique utilizing the coherency of the caching which is effective for the surfaces with highly geometric details such as bump-mapped surfaces. More recently, Schwarzhaupt et al.\textsuperscript{56} proposed sophisticated error metric utilizing Hessians (a.k.a. second order gradients). This technique utilizes extended version of an analysis of irradiance Hessian in 2D global illumination\textsuperscript{56}. They also utilizes occlusion information in order to improve the accuracy of the Hessian computation.

4.4 Photon density estimation based techniques

Next, we review second caching based techniques: photon density estimation based techniques. This type of techniques is originated from the photon mapping proposed by Jensen\textsuperscript{51}.

1) Basics

The photon mapping is executed by two steps: 1) photon tracing step and 2) photon density estimation step. In the first step, light particles a.k.a. photons are traced from the light sources. In the propagation of a photon, intermediate radiance value is computed. When a photon is intersected with scene surface, the photon is stored into the kd-tree storage called the photon map. In the second step, the image is reconstructed by ray tracing with photon density estimation. When a ray is intersected with diffuse surface through propagation, the photons near the intersected point is gathered by nearest-neighbor query to the photon map and used for the density estimation. Various techniques originated from the photon mapping is proposed. Havran et al.\textsuperscript{52} proposed the reverse photon mapping which utilizes the ray tracing for the initial step, and the photon tracing in the second step. A new kd-tree structure called reverse photon map is introduced for preserving information of the ray tracing step, and in the second step nearest-neighbor query is issued on the reverse photon map.

2) Improving structure

Some techniques introduces a new structure as a replacement of the photon map to improve the efficiency. Havran et al.\textsuperscript{53} proposed a technique with a new structure that stores photon ray paths instead of photons, which is effective for eliminating bias on density estimation due to topology of the geometry. Herzog et al.\textsuperscript{54} extended the approach to utilize splatting, projection onto the image plane. Chen et al.\textsuperscript{55} utilizes clustering of photons according to the intersection history, which is used for generating the polygonal boundaries for expressing complex caustics. Spencer and Jones\textsuperscript{56} improve the final gathering by introducing hierarchical structure to the photon map.

3) Improving photon distribution

Another focus on improving the technique is improvement of photon distribution in the photon map. Some techniques attempt to achieve this goal by refinement of the photon map with redistribution of photons after photon tracing step. Spenser et al.\textsuperscript{60} proposed a technique called the photon relaxation that iteratively redistributes photons according to the blue noise distribution. They later extended\textsuperscript{57} the technique to support complex illumination structure containing discontinuity by storing photons with their trajectory information into high-dimensional kd-tree. Another approach to improve photon distribution is to improve the photon tracing step itself. Peter and Pietrek\textsuperscript{61} improved photon distribution in photon tracing step by introducing the importance map. Suykens and Willems\textsuperscript{62} controlled photon density by a criterion based on importance. Hachisuka and Jensen\textsuperscript{59} proposed a photon tracing technique utilizing Markov chain Monte Carlo (MCMC) sampling according to the visibility of photon paths.

4) Improving density estimation kernels

Another way to improve the photon mapping is to select appropriate density estimation kernels. An obvious idea for the kernel selection problem is to modify the bandwidth of kernels. The idea of bandwidth selection has been employed in the context of the photon density estimation. Jensen and Christensen\textsuperscript{59} initially attempt to control the bandwidth by the technique called differential checking, which limits the bandwidth by checking the difference in irradiance. Schregle\textsuperscript{63} proposed a bandwidth selection technique based on the bias estimation of the reconstructed illumination with binary search. Schijstj et al.\textsuperscript{64, 65} proposed a technique based on diffusion filtering, which is an density estimation technique with anisotropic kernel. Hey and Purgatho\textsuperscript{65} proposed a technique utilizing surface geometry information, which is effective for edge or corners. More recently, Kaplanyan and Dachsbacher\textsuperscript{66} proposed an technique with adaptive bandwidth selection for progressive photon mapping with plug-in bandwidth selection.

5) Progressive estimation

Although Jensen’s photon mapping is consistent in the sense that if we can take infinite number of samples in the photon map,
However it is virtually impossible due to memory restriction. In order to resolve the problem and to develop a consistent technique with realistic configuration, Hachisuka et al.\(^{60}\) proposed a technique called *progressive photon mapping*, which progressively diminishes kernel radius while increasing number of photons to be used for the estimation. The technique is extended\(^{69}\) to support some rendering effects such as depth-of-field or motion blur. Later Hachisuka et al.\(^{70}\) developed an error estimation framework for the technique which can be utilized as a criterion for termination of rendering. Knaus and Zwicker\(^{70}\) proposed a variant of the progressive photon mapping which is not dependent on local statistics and can be utilize arbitrary kernel with a probabilistic analysis of the technique. More recently, Georgiev et al.\(^{72}\) and Hachisuka et al.\(^{73}\) independently proposed a generalized framework for the progressive photon mapping, which combines photon density estimation and multiple importance sampling in bidirectional path tracing.

### 4.5 MCMC based techniques

In this section, we review the techniques based on Markov chain Monte Carlo (MCMC) method described in section 4.1.

1) **Basics**

MCMC is initially brought to the global illumination computation by *Metropolis light transport (MLT)* proposed by Veach and Guibas\(^{77}\). MLT utilizes the Metropolis-Hastings algorithm for sample generation and introduces some light path specific mutation strategies such as the *bidirectional mutation or perturbations*, which correspond to *transition function* in section 4.1. In these strategies, light paths are modified directly in the light path space, for example by slightly changing the next ray directions or light path lengths, which enables the local exploration in the light path space.

2) **Changing sample space**

Some variants of MLT have been developed for changing the sampling space and the target distribution. *Primary sample space MLT* proposed by Kelemen et al.\(^{78}\) introduces the *primary sample space*, which is a space of high-dimensional unit cube of uniform random numbers used for sampling a light path, and Metropolis-Hastings algorithm is dispatched in the space. This technique simplifies the mutation strategy and reduces high energy region of the light path space thank to the importance sampling. More recently, Hachisuka et al.\(^{84}\) proposed a technique called the *multiplexed MLT* on Kelemen’s framework. In this technique, a primary sample space is extended to multiple spaces, and in each space corresponding light paths are sampled to follow the distributions associated with the sampling techniques used in the bidirectional path tracing. A sample is assigned one of the space and can move beyond the spaces, similarly to *serial tempering* which is one of MCMC technique with tempered distributions. *Manifold exploration* proposed by Jakob and Marschner\(^{78, 82}\) introduces the specular manifold, which is the space defined in the local coordinates of the tangent spaces for each intersection points with the constraints of reflection or refraction vectors for specular surfaces. In the technique, a new mutation strategy is introduced to explore on the space with a iterative equation solving technique like Newton’s method. Very recently, the technique is extended and improved by Kaplanyan et al.\(^{89}\) using a space of the projected half vector representation of light paths with constraints. *Gradient-domain MLT* proposed by Lehtinen et al.\(^{79}\) introduced image space gradient distribution of the energy contribution function of the light paths. Instead of directly sampling from energy distribution, the technique samples a light path in the gradient space. Metropolis-Hastings sampler driven by distribution on the gradient space enables for the samples to concentrate the on the portion of the light path space that energy is changed drastically, *e.g.*, edges on the mesh. The final rendered image is reconstructed by Poisson solver from the gradient image.

3) **Utilizing advanced MCMC techniques**

Some technique attempts to use advanced MCMC techniques other than MH algorithm. *Energy preserved path tracing (ERPT)* proposed by Cline et al.\(^{80}\) alleviates the detailed balance and ergodicity condition to achieve the techniques which depends only on Veach’s perturbations. Base on the path sampled from the
ordinary light path sample techniques, some short burns of the perturbations are dispatched followed by the scatter of the energy over the image plane without introducing any bias. Replica exchange light transport proposed by Kitaoaka et al.\(^5\) applies replica exchange method to the global illumination computation. In the technique, a set of paths divided by the feature of the paths such as random paths or caustic paths, which preserves coherency of the samples in the same type of the paths. Lai et al.\(^11\) proposed a technique using the population Monte Carlo (PMC) method, which is a variant of MCMC technique. Unlike ordinary MCMC techniques like MH algorithm PMC utilizes a number of samples in a batch (a.k.a. population), which enables to exploit information collected from the samples. Their technique utilize D-kernel PMC, which adapts kernel function with populations per iteration. They apply D-kernel PMC to ERPT and iteratively adapts energy-redistributed region.

5. Conclusion

In this survey, we review various types of global illumination techniques as well as underlying theoretical concepts. As we review in the previous sections, there are various technique and topic on the global illumination computation and further researches are required to solve remaining difficulties. In that means, the research on the global illumination computation will certainly remain to be one of the central topics of the computer graphics. We hope this survey to help the readers to comprehend the overview of the field and the characteristics of these techniques, and to indicate the next direction of the research in the field of global illumination computation.

References

7) M. Pharr and G. Humphreys, Physically Based Rendering, Morgan Kaufmann (2010).


Tomoyuki NISHITA (Fellow)
He is a Professor in Department of Economic Information of Hiroshima Shudo University and a Director of UEI Research since 2013. He was a Professor in the Department of Complexity Science and Engineering (also in the Department of Information Science) at University of Tokyo, Japan since 1998, and now is a Professor Emeritus of The University of Tokyo. He taught at Fukuymama University from 1979 to 1998. He was an associate researcher in the Engineering Computer Graphics Laboratory at Brigham Young University from 1988 to 1989. His research interests center in computer graphics including lighting models, hidden-surface removal, and anti-aliasing. He received his B.E., M.E. and Ph.D. in Engineering in 1971, 1973, and 1985, respectively, from Hiroshima University. He received Research Award on Computer Graphics from Information Processing Society of Japan in 1987, and also received Steven A. Coons Award from SIGGRAPH in 2005. He was a president of The Institute of Image Electronics Engineers of Japan in 2009.

Hisanari OTSU
He received the B.E. degree in Science from Department of Information Science, Faculty of Science, The University Tokyo in 2013. He is currently working toward the M.S. degree at The University of Tokyo. His research interests include physically based rendering and light transport simulation.