A New Adaptive Density Estimator for Particle-Tracing Radiosity

Wong Kam Wah
Department of Computer Science and Information Systems
The University of Hong Kong
Pokfulam Road, Hong Kong
kwwong@csis.hku.hk

Abstract

In particle-tracing radiosity algorithms, energy-carrying particles are traced through an environment for simulating global illumination. Illumination on a surface is reconstructed from particle "hit points" on the surface, which is a density estimation problem [11]. Several methods can be used to solve this problem, such as the adaptive meshing method [14], the kernel method [15], and the orthogonal series estimator [3]. In this paper, a new orthogonal series estimator is proposed to tackle the problem. In the new method, the appropriate number of terms that should be used in the series is determined adaptively and automatically. Moreover, a surface subdivision scheme is combined with the estimator to increase the accuracy of estimation. The new method has several advantages over other existing methods: (1) it requires less memory than the adaptive meshing method; (2) it does not store all the particle-hit points as in the kernel method; (3) it determines automatically how many terms should be used in the orthogonal series; (4) it incorporates surface subdivision to further increase the accuracy of estimation.

1. Introduction

The global illumination problem can be solved by using the particle-tracing radiosity method [9]. Energy-carrying particles are emitted from each light source, traced through the environment, hit and reflected from surfaces until they are absorbed probabilistically. Illumination on a surface can be estimated from the density of all particle-hit points on the surface.

The particle-tracing radiosity method has several advantages over the classical hierarchical radiosity method [5]. Most complex optical effects, such as non-uniform luminaries, light scattered from a non-diffuse surface, and refraction, can easily be simulated [9]. The huge memory requirement for storing the "links" in the classical hierarchical radiosity method can also be avoided [12]. Moreover, since particles interact with the raw input geometry only, the particle tracing stage and surfaces' illumination estimation are independent of each other. As a result, the error in the surfaces' illumination estimation is not propagated [10, 15].

The illumination on a surface is proportional to the density of the particle-hit points on the surface. Regions with more particle-hits should have brighter illumination. Heckbert [7] first noted that reconstructing illumination from particle-hits is a density estimation problem [11]. Several techniques are given in the statistics literature for solving the density estimation problem [11]. Some of these techniques have already been applied by computer graphics researchers; for example, the histogram method [7, 14], the kernel method [1, 15], and the orthogonal series estimator [3].

The new method proposed in this paper belongs to the class of the orthogonal series estimator. The main contribution of the new method is that the appropriate number of terms that should be used in the orthogonal series is determined adaptively and automatically by the algorithm. Furthermore, in order to avoid using an overly large number of terms, an adaptive surface subdivision scheme is incorporated in the new method.

The remainder of this paper is organized as follows: in section 2 some typical methods for solving the density estimation problem are reviewed, and computer graphics algorithms that utilize these methods are described. In section 3 a new density estimation method is presented. Its implementation and experiment results are given in section 4. Finally the conclusion is drawn in section 5.

2. Density estimation

The density estimation problem can be stated as follow: derive a density function $\hat{f}(x)$ that approximates an unknown probability density function $f(x)$ from which independent samples $X_1, X_2, \ldots, X_n$ are drawn. Researchers in computer graphics community [7, 10] have realized that

0-7695-0868-5/00 $10.00 © 2000 IEEE

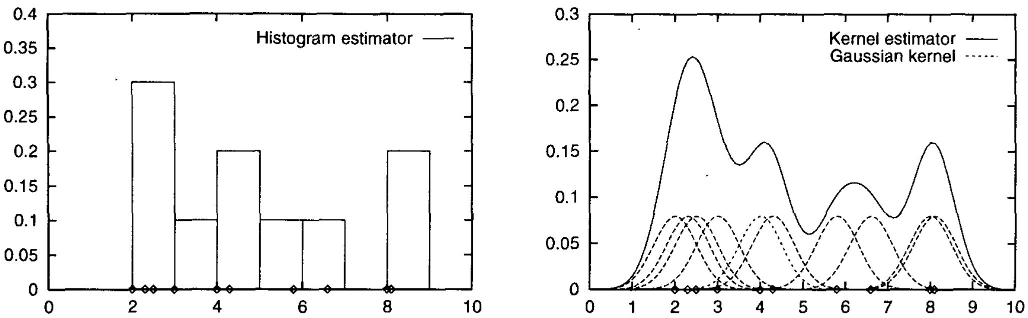


Figure 1. The histogram estimator and the kernel estimator (using Gaussian kernel) for the samples $\{2, 2.3, 2.5, 3, 4, 4.3, 5.8, 6.6, 8, 8.1\}$.

in a particle-tracing radiosity method the illumination reconstruction process is a density estimation problem. The particles-hit points are the sample points $\{X_i\}$, and the illumination function is a scaled probability density function $f(x)$. In this section, some of the commonly used techniques for solving the density estimation problem and implementations of these techniques in computer graphics applications are reviewed.

2.1. Histogram

Histogram is the oldest and most widely used density estimator. Suppose that the domain $[a, b]$ of a variable x is divided into sub-intervals $[a+mh, a+(m+1)h)$, each with width h. Let n denote the number of samples drawn from $[a, b]$. The histogram estimator is defined by

$$\hat{f}(x) = \sum_{i=1}^{n} \frac{1}{nh} * (\textit{no. of } X_i \textit{ in the same bin of } x)$$

Figure 1 shows an example of the histogram estimator. The "meshing" technique in computer graphics can be viewed as an implementation of the histogram method: the approximate illumination of a patch is proportional to the number of particle-hits on the patch. More sophisticated "adaptive meshing" approaches, such as the one in [14] can also be viewed as a variation of the histogram method. In that case it allows sub-intervals to have different widths, and the number of sub-intervals is increased adaptively.

2.2. Kernel estimator

A problem with the histogram approach is that the resulting estimation is a piecewise constant function. In order to obtain a smooth estimation, a kernel estimator can be used. Let $K(x)$ be a kernel function satisfying the condition $\int K(x)dx = 1$. The scaled kernel $K_h(x)$ with kernel width h is defined as $K_h(x) = \frac{1}{h} * K(\frac{x}{h})$ and the estimated function is defined as

$$\hat{f}(x) = \frac{1}{n} * \sum_{i=1}^{n} K_h(x - X_i)$$

We can think that, for each sample point X_i, a kernel is placed centered at the sample point. The estimated function is the sum of these n kernels. There are different choices for the kernel function, and figure 1 shows an example of the kernel estimator using a commonly used kernel, the Gaussian kernel, defined as $K(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$. Note that the choice of h affects the resulting estimation, and choosing a suitable h is not an easy task [10]. The kernel width should be narrow enough to capture the details in density distribution, and should be wide enough to avoid spurious fine structures. Walter et al [15] gives a particle-tracing radiosity algorithm that uses a kernel estimator to reconstruct surface illumination.

2.3. Orthogonal series estimator

The kernel method needs to store all the samples for evaluating $\hat{f}(x)$. When a large sample size is involved, the storage requirement will be very high. For this reason, Shirley et al [10] proposed to store the particle-hits on a disk. Another disadvantage of the kernel method is that an estimate of the illumination can be obtained only when all particle-hits are captured. If a current estimation is required during the solution process, the orthogonal series estimator is preferred [4].

Let $\{\phi_i, i \geq 0\}$ be a complete set of orthonormal basis functions on an interval I. Suppose that density function $f(x)$ is represented by:

$$f(x) = \sum_{i=0}^{\infty} a_i \phi_i(x) \qquad x \in I$$

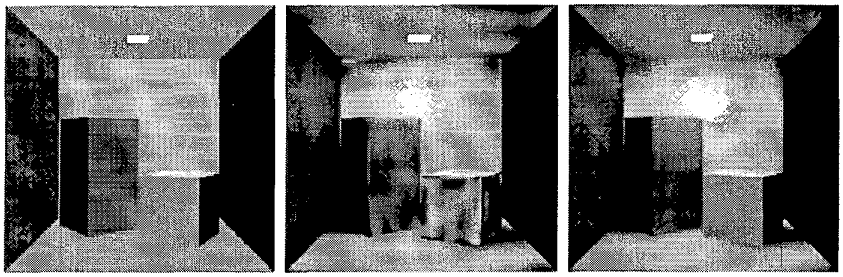


Figure 2. Inappropriate chosen of m. From left to right: all surfaces use one single term, 45 terms, and the number of terms determined adaptively. The same amount of particles (10^4) are traced in the three pictures.

where $a_i = \int_I f(x)\phi_i(x)dx$. An orthogonal estimator is then given by:

$$\hat{f}(x) = \sum_{i=0}^{m} \hat{a}_i \phi_i(x) \tag{1}$$

where $\hat{a}_i$ is an estimation of a_i. Since $\{X_i\}$ is drawn from the probability density function f,

$$a_i = \int_I f(x)\phi_i(x)dx = E[\phi_i(X)] \approx \frac{1}{n}\sum_{j=1}^{n} \phi_i(X_j) \equiv \hat{a}_i$$

Note that only a finite number of terms (m in equation (1)) are used in the estimation. An advantage of this estimator is that m is usually much smaller than the sample size n. Therefore, unlike the kernel estimator method, it only has to store on each surface a small number of coefficient $\{\hat{a}_i : i = 0, \ldots, m\}$, instead of storing all particle-hit points.

Feda [3] gives a particle-tracing radiosity algorithm based on this orthogonal series method. In his algorithm, the parameter m is specified by the user. Just like deciding how to subdivide a surface in the histogram method or choosing an appropriate width h in the kernel method, choosing an appropriate value of m for each surface is not an easy task. Moreover, when illumination discontinuity exists on a surfaces, the orthogonal series estimator cannot give a good result without surface subdivision. These are the two main problems to be addressed by the new method.

3. The new method

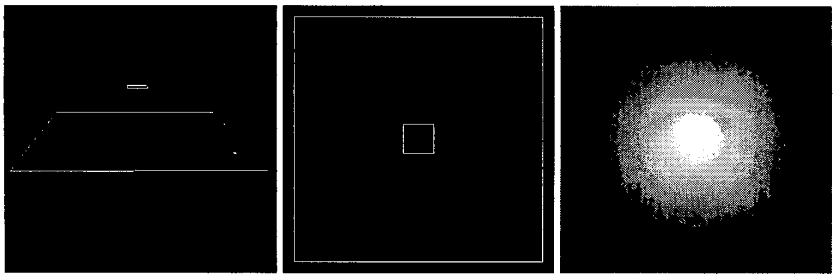
3.1. Overview

The new method is an extension of Feda's algorithm in [3]. Two modifications are added to the orthogonal series estimator. First, the new method determines automatically for each surface the appropriate number of terms that should be used in the orthogonal series. Second, when illumination discontinuity entails an unacceptably large number of terms, the new algorithm subdivides the surface adaptively to better capture the illumination discontinuity. Hence the new method combines in a natural manner the advantage of orthogonal series estimator for modeling smooth variation of illumination and that of adaptive subdivision for modeling illumination discontinuity.

Legendre polynomials are chosen as the orthogonal basis functions due to the following reason: Fourier series and Legendre series are two common choices in the orthogonal series estimator for a finite domain. However, Hall [4] points out that Fourier series has the problem of "edge effects", referring large bias towards the endpoints of the domain. Legendre polynomials do not suffer from this deficiency. His paper concludes that Legendre series is better than Fourier series in the density estimator method. Appendix A gives the equations of one-dimensional and two-dimensional Legendre basis functions. Note that the new method is independent of the choice of basis functions, and other basis functions can also be used. However, it is worthwhile to study the influence of different basis functions on the method, but that is out of the scope of this paper.

3.2. Automatic determination of m

For each surface, the number of terms m used in the series should be chosen carefully. Previous studies [8, 6, 2] show that for a small sample size, setting m too high will result in too many oscillations in the estimated function. For a large sample size, or for surfaces with high gradient changes, setting m too small will result in over-smoothing of the estimation. Figure 2 shows the result of inappropriately chosen m. In the left image all surfaces use the first term in series (i.e. the constant term) only, and the shad-



m	0	1	2	3	4	5	6	7	8	9
$H(m)$	13555.4	13555.0	13554.7	15004.6	15004.3	16452.3	16451.9	16451.6	16451.3	16450.9

Figure 3. A simple scene in which the Kronmal-Tarter "stopping rule" does not work. Top (from left to right): front view, top view, analytical solution of the receiver's illumination. Bottom: the function $H(m)$ of the receiver.

ing becomes too flat. In the middle image all surfaces use 45 terms. For those surfaces with fewer particle-hits, there are too many terms in the series and oscillations of higher degree polynomials in the basis functions cause visual artifacts.

Kronmal and Tarter [8] propose a method to choose the number m for Fourier series estimator automatically. They start with the integrated-mean-square error (IMSE) of the estimation

$$IMSE = J(m) = \int_I E[\hat{f}_m(x) - f(x)]^2 dx$$

and show that

$$J(m) = \sum_{i=0}^{m} Var(\hat{a}_i) + \sum_{i=m+1}^{\infty} a_i^2$$

The formula can be further expanded as

$$J(m) = \frac{1}{n-1} \sum_{i=0}^{m} [\, 2\hat{d}_i - (n+1)\hat{a}_i^2 \,] + \sum_{i=0}^{\infty} a_i^2 \quad (2)$$

where $d_i = \frac{1}{n}\sum_{j=1}^{n} {\phi_i}^2(X_j)$. Details of the expansion of (2) can be found in [8, 2]. The aim is to choose a number m for each surface which minimize $J(m)$, or equivalent, maximize the function

$$H(m) = \sum_{i=0}^{m} [\, (n+1)\hat{a}_i^2 - 2\hat{d}_i \,]$$

as the second term on right hand side of (2) is independent of m. The Kronmal-Tarter method starts at $m = 0$ and increases m one-by-one until $H(m+1) < H(m)$. However, this simple term-by-term "stopping rule" does not always give a good choice of m because the function $H(m)$ may have multiple peaks. The simple example shown in figure 3 illustrates the problem. A light source is placed right over the center of the receiver. Since the illumination is symmetric on the receiver, there is no linear variation over the entire surface. So including the linear terms ($m = 1, 2$) in the series will not reduce $J(m)$. However, including later terms in the series (e.g. the quadratic terms when $m = 5$) will reduce the error $J(m)$.

Hart [6] studied the problem of Kronmal-Tarter method and proposed an improvement of it. He considers all $m \leq M$ (where M is a global user-defined constant) and chooses the number m that maximizes $H(m)$. This method is practical only when M is small enough for computation. We have implemented his method, using $M = 45$, and found that it gives a good result in the radiosity application. It is because in a typical radiosity application, most surfaces need a fewer terms only, so M needs not to be large. It is found that the value of M may not be sufficient large only when illumination discontinuities exist on a surface. But this case is dealt with by the adaptive subdivision scheme in the new method.

3.3. Adaptive surface subdivision

When illumination discontinuity exists on a relatively large surface, the estimated function based on orthogonal series alone cannot fit the illumination function very well unless an impractically large number of terms are used. Figure 4 shows two simple scenes to illustrate this situation. The images show that even though 45 terms are used, the illumination still fails to produce a good approximation. Moreover, in the region with low gradient the visual artifact is even worse. That is because using too many number of terms in the series results in noticeable oscillations inherent to high order terms.

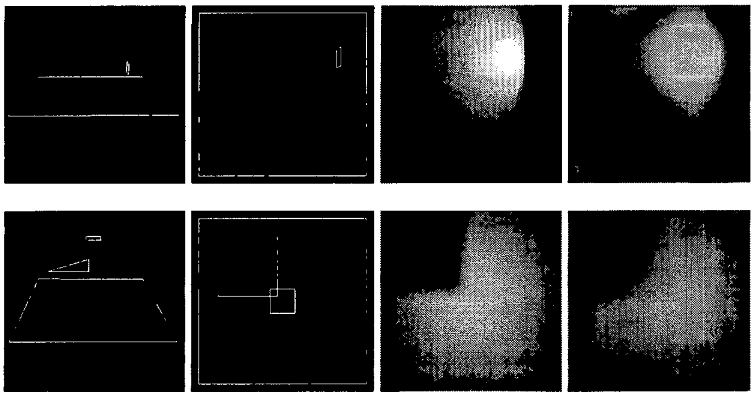


Figure 4. Results without surface subdivision. From left to right: the front view, top view, analytical solution, and adaptive orthogonal function approximation. In both cases $m = M$ is reached.

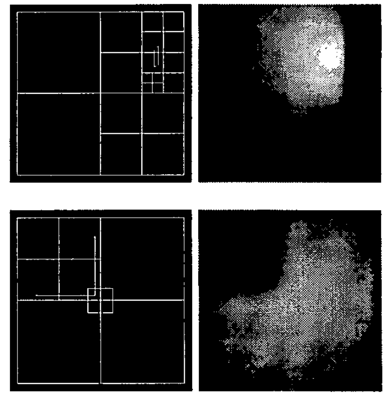


Figure 5. The result of the new estimator with surface subdivision.

To overcome this problem, a surface subdivision scheme is added to the estimator. When an appropriate number of terms determined by the algorithm reaches a pre-defined value M^*, the algorithm subdivides the surface. Note that M^* can be any value less than or equals to M mentioned in the previous subsection. After subdivision, each sub-surface will have its own density estimation function, and the late-coming particles hit on the surface will be captured by a sub-surface. We avoid to "push" the coarse estimate at the parent's level to the children's level, because any naive "push" will introduce error. Instead, the parent surface keeps the coarse estimate, and this estimated function will be superimposed to the children's illumination estimation at the rendering stage.

There is one problem if surface subdivision is added to the estimator: the estimated functions of the sub-surfaces may not be guaranteed to be consistent along subdivision boundaries. As a result, visual artifacts will appear in the final image. An interpolation technique is used to alleviate the problem: at a region close to the subdivision boundary, the estimated functions of the sub-surfaces will be interpolated by a blending function. Figure 6 illustrates the process in 1D. Note that interpolation is applied only at the image-rendering stage as a tools to remove visual artifact. It will not have any effect on the density estimator. Figure 5 shows the results of the new estimator with surface subdivision.

3.4. Comparison with Tobler et al [14] approach

Tobler et al [14] give an adaptive meshing algorithm for the particle-tracing radiosity method. Their algorithm allows the surface to have higher order illumination function terms rather than constant ones. Despite the resemblance of the new method to their method, these two methods have the following differences:

- In their method the number of terms used for each surface is specified by the user, whereas in the new method it is determined automatically. This makes the new method more practical for complex scenes.
- Their method focuses on adaptive meshing. At each level (the "current level"), the algorithm uses a "preview level" to keep track of particle-hits in a finer surface resolution. By comparing the illumination function of the "preview level" and the "current level", the algorithm decides whether the surface should be subdivided or not. In contrast, the new method emphasizes on determining the suitable number of terms used in the series for each surface. Thus no "preview level" is necessary, and a surface is subdivided only when the number of terms determined is larger than acceptable.
- In their method, after a surface is subdivided, the sub-surfaces use the same number of terms in the series as their parent surface. In the new method, the number of terms used by each of the sub-surfaces may differ, and is adaptively determined. This is a proper treatment, because when a surface is subdivided, the sub-surfaces will tend to have smoother illumination than their parent; consequently, they should use a smaller number of terms than their parent.

4. Implementation and results

Four different methods have been implemented for comparison purpose: (1) *AM-C*: Tobler et al [14] adaptive mesh-

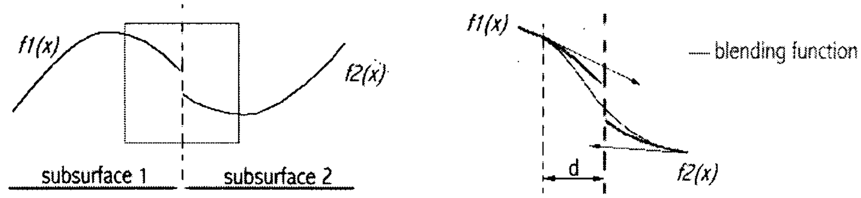


Figure 6. The interpolation of two functions near a subdivision boundary. Left: the two estimated functions are discontinuous across the subdivision boundary. Right: zoom view of the rectangular region in the left image. A blending function is used to interpolation the two functions in the region with distance $\leq d$ from the boundary (d is a pre-defined constant).

ing method with constant illumination; (2) *AM-H*: same as (1), but with higher order function on each surface (the number of terms m is defined by the user); (3) *FOSE*: Feda [3] orthogonal series estimator with a user defined m and without surface subdivision; (4) *NEW*: the new method. These methods are applied to the Cornell-box scene, and the running time, memory used, and L_2 $error$ (IMSE) of two surfaces against the number of particles traced are plotted. The surfaces chosen for error analysis are the floor, which has large variation in illumination, and the right red wall, which has smooth illumination variation. For the methods *AM-H* and *FOSE* 45 terms for the floor and 15 terms for the red wall are used. A reference solution is generated by shooting 10^8 particles to the scene, where each surface has a texture of size 200×200 to capture particle hits.

The results are shown in figures 7 and 8. A sequence of images are generated, as shown in figure 13 in the color page, by increasing the number of particles traced. The images and the statistics results show that:

- *error:* the *FOSE* method gives the largest error. The error will be bounded from below by some value and cannot be improved anymore if surface subdivision is not used, as the degree is limited. Adaptive meshing methods (*AM-C* and *AM-H*) produce smaller error due to surface subdivision, but *AM-C* uses lot of memory to store the meshes structure in order to achieve this small error level. The *AM-H* method avoids the storage problem, but that the user has to choose a suitable m for each surface makes this method impractical for complex scenes. The new method generates smaller error than the other three methods (if we ignore the case when there are too few particles traced, e.g. less than 10000), and it does not require the user to specify the number m for each surface.

- *running time:* as shown in figure 8, the *AM-H* method takes the longest time to trace the same number of particles. That is because for every particle-hit the method has to evaluate a high order illumination function for both the "preview level" and the "current level". The new method takes longer time than *AM-C* and *FOSE* because it has to evaluate the function $H(m)$ for all $m \leq M$, in return for a more accurate illumination estimation.

In summary, to achieve the same error level, the new method requires to shoot fewer particles and runs faster. Figure 9 plots the L_2 error against the running time used in the four methods. These graphs show the superiority of the new method.

Figure 11 shows a synthesized image of the graphics laboratory at the University of Hong Kong generated by the new method. The model contains 16000 triangles. In total 10^7 particles are shot from the light sources. Each particle is traced through the scene until it is absorbed probabilistically, or a maximum of 10 rebounds is reached. The computation takes 5.5 hours on a SGI MIPS R10000 195MHZ processor, with 512MB main memory. The new method is more suitable than the other three methods for complex scenes, because it uses less memory, produces a better rendering quality, and there is no need to specify the parameter m for each surface.

5. Conclusions

A new method is proposed for estimating the illumination of a surface in particle-tracing radiosity applications. The method adds two modifications to the standard orthogonal series estimator. This makes the orthogonal series estimator more practical for the radiosity problem. First, the method determines adaptively and automatically the appropriate number of terms should be used in the series for each surface. Second, when illumination discontinuity exists on a surface, this method subdivides the surface to improve accuracy. As a result, to achieve the same error level, the new

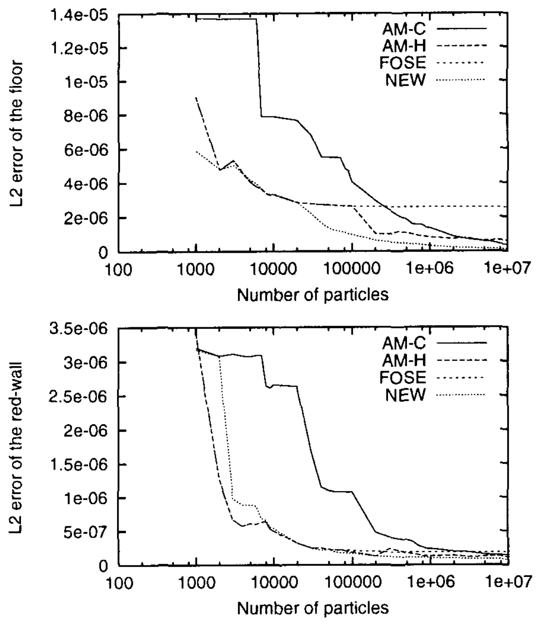


Figure 7. The L_2 error of the methods.

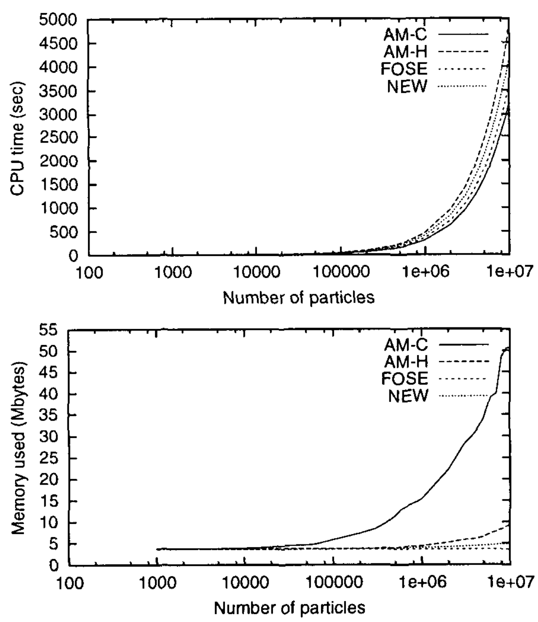


Figure 8. CPU time and memory used by different methods.

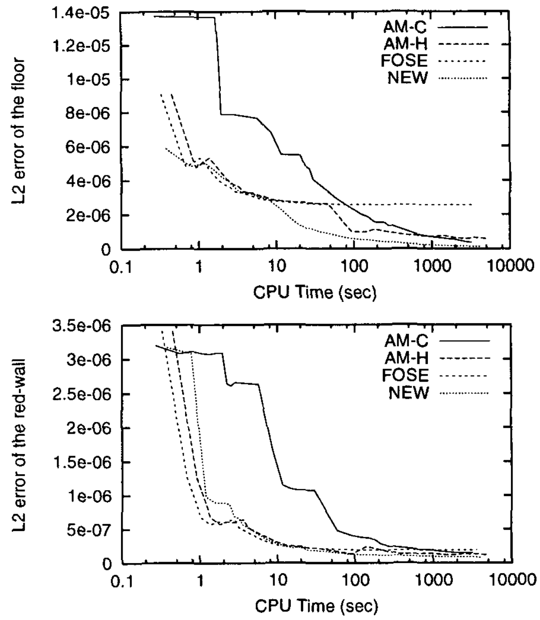


Figure 9. L_2 error against running time.

method uses less memory, requires fewer particles, and runs faster.

There is still lot of work to be done with this method. One major problem of the particle-tracing radiosity method is that tiny surfaces may not get enough particle-hits even though a large number of particles have been traced. The chairs in figure 11 shows the problem. One way to solve it is to trace more particles, as shown in figure 12, but this method increases the computational cost as well. A more sophisticated method has to be developed to solve this problem. Moreover, a better method has to be developed to rectify the inconsistent shading value across subdivision boundaries. Lastly, there are many kinds of orthogonal basis functions, either polynomial or non-polynomial. Different sets of basis functions have different properties. It is worthwhile to study other orthogonal series to see if any one is the most suitable for the particle-tracing radiosity applications.

A. Legendre polynomials

The one-dimensional Legendre polynomials are generated by the following recursive formula [13]:

$$P_0(x) = 1$$
$$P_1(x) = x$$
$$(n+1)P_{n+1}(x) = (2n+1)\, x\, P_n(x) - n\, P_{n-1}(x)$$

The normalized Legendre polynomials are

$$P_n^*(x) = \sqrt{n + \frac{1}{2}}\, P_n(x)$$

A two-dimensional basis set $\{Q_i(x, y)\}$ can be generated by multiplying two one-dimensional polynomials in different variables [13]. Therefore, the first 6 terms in the 2D Legendre basis set are:

$$\begin{aligned} Q_0(x,y) &= P_0^*(x)P_0^*(y) & Q_3(x,y) &= P_2^*(x)P_0^*(y) \\ Q_1(x,y) &= P_1^*(x)P_0^*(y) & Q_4(x,y) &= P_1^*(x)P_1^*(y) \\ Q_2(x,y) &= P_0^*(x)P_1^*(y) & Q_5(x,y) &= P_0^*(x)P_2^*(y) \end{aligned}$$

and so on. Figure 10 shows the 3D plotting of these 6 terms.

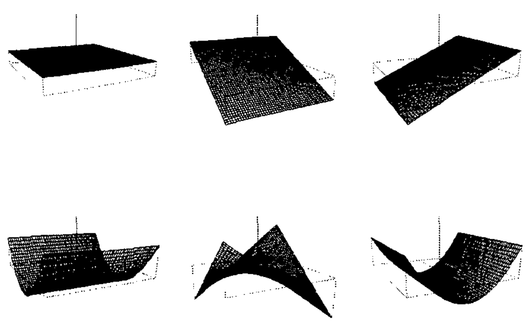


Figure 10. The first six terms of two-dimensional Legendre basis functions.

References

[1] S. E. Chen, H. E. Rushmeier, G. Miller, and D. Turner. A progressive multi-pass method for global illumination. In *Computer Graphics (SIGGRAPH '91 Proceedings)*, volume 25, pages 165–174, July 1991.

[2] P. J. Diggle and P. Hall. The selection of terms in an orthogonal series density estimator. *Journal of American Statistical Association*, 81:230–233, 1986.

[3] M. Feda. A Monte Carlo approach for Galerkin radiosity. *The Visual Computer*, 12(8):390–405, 1996.

[4] P. Hall. Compatison of two orthogonal series methods of estimating a density and its derivatives on an interval. *Journal of Multivariate Analysis*, 12:432–449, 1982.

[5] P. Hanrahan, D. Salzman, and L. Aupperle. A rapid hierarchical radiosity algorithm. In *Computer Graphics (SIGGRAPH '91 Proceedings)*, volume 25, pages 197–206, July 1991.

[6] J. D. Hart. On the choice of a truncation point in fourier series density estimation. *Journal of Statistical Computing and Simulation*, 21:95–116, 1985.

[7] P. S. Heckbert. Adaptive radiosity textures for bidirectional ray tracing. In *Computer Graphics (SIGGRAPH '90 Proceedings)*, volume 24, pages 145–154, Aug. 1990.

[8] R. A. Kronmal and M. E. Tarter. The estimation of probability densities and cumulatives by fourier series methods. *Journal of American Statistical Association*, 63:925–952, 1968.

[9] S. N. Pattanaik and S. P. Mudur. Computation of global illumination by monte carlo simulation of the particle model of light. *Third Eurographics Workshop on Rendering*, pages 71–83, May 1992.

[10] P. Shirley, B. Wade, P. Hubbard, D. Zareski, B. Walter, and D. P. Greenberg. Global illumination via density estimation. In *Eurographics Rendering Workshop 1995*, June 1995.

[11] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, 1986.

[12] M. Stamminger, H. Schirmacher, P. Slusallek, and H.-P. Seidel. Getting rid of links in hierarchical radiosity. *Computer Graphics Forum*, 17(3):165–174, 1998.

[13] P. K. Suetin. *Orthogonal polynomials in two variables*. Gordon & Breach, 1999.

[14] R. F. Tobler, A. Wilkie, M. Feda, and W. Purgathofer. A hierarchical subdivision algorithm for stochastic radiosity methods. In *Eurographics Rendering Workshop 1997*, pages 193–204, June 1997.

[15] B. Walter, P. M. Hubbard, P. Shirley, and D. F. Greenberg. Global illumination using local linear density estimation. *ACM Transactions on Graphics*, 16(3):217–259, July 1997.

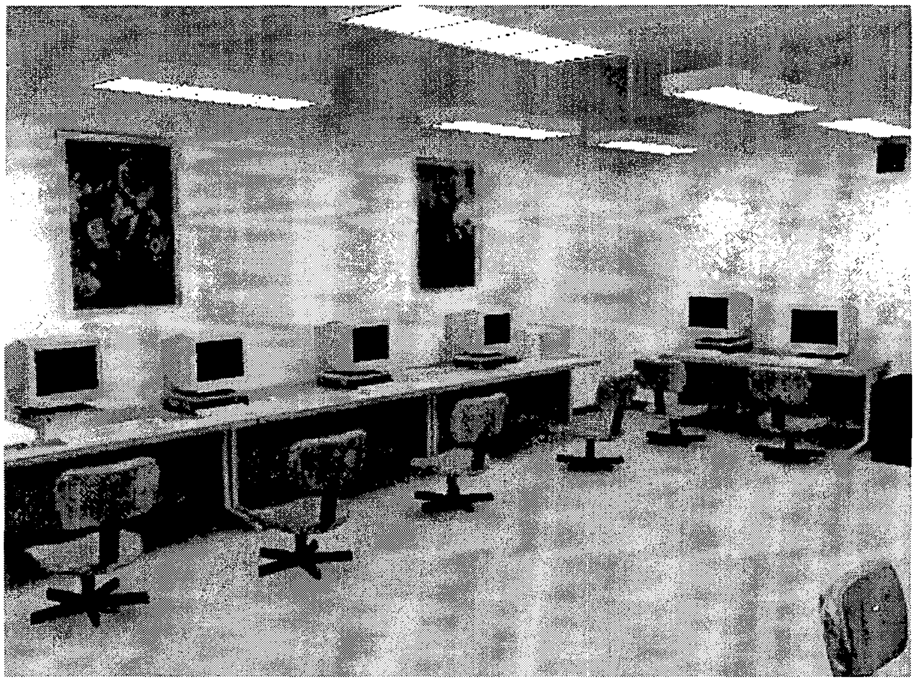


Figure 11. A complex scene rendered by the new method.

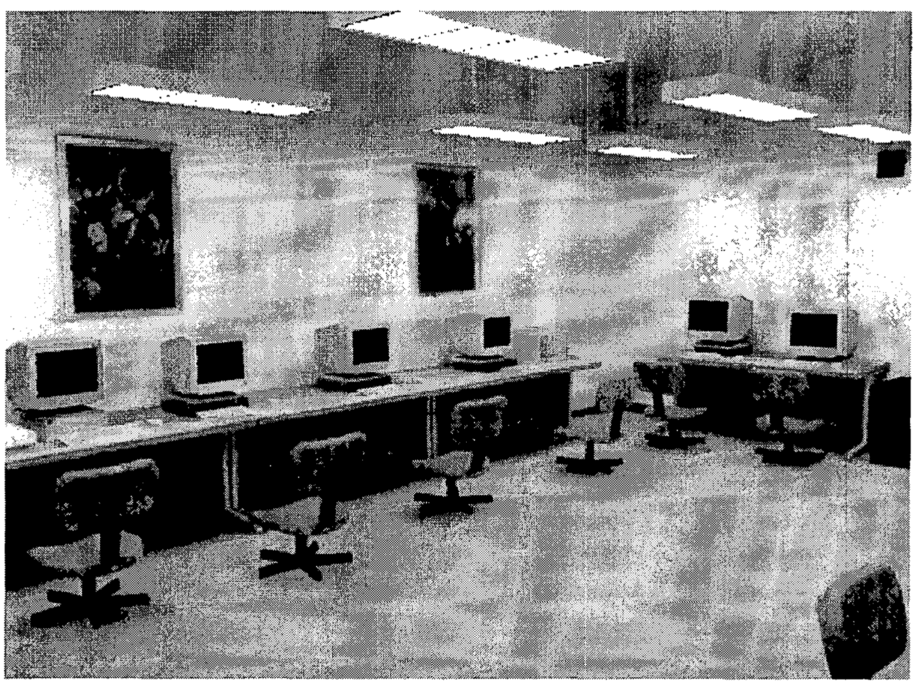


Figure 12. The same scene shown in figure 11, with five times number of particles has been traced.

Figure 13. Final rendered images. From left to right: *AM-C, AM-H, FOSE, NEW*. **From top to bottom:** 10^3, 10^4, 10^5, 10^6, **and** 10^7 **particles are traced.**