The Ambient Term as a Variance Reducing Technique for Monte Carlo Ray Tracing

Eric P. Lafortune and Yves D. Willems

Paper for the 5th Eurographics Workshop on Rendering

Please send any correspondence to:
Eric Lafortune
Department of Computer Science
Katholieke Universiteit Leuven
Celestijnenlaan 200A
B-3001 Heverlee
BELGIUM
Telephone: ++32 - 16 - 32 70 94
Fax: ++32 - 16 - 32 79 96
E-mail: Eric.Lafortune@cs.kuleuven.ac.be

The Ambient Term as a Variance Reducing Technique for Monte Carlo Ray Tracing

Eric P. Lafortune and Yves D. Willems

Department of Computer Science, Katholieke Universiteit Leuven
Celestijnenlaan 200A, 3001 Heverlee, Belgium
Eric.Lafortune@cs.kuleuven.ac.be

Abstract

Ray tracing algorithms often approximate indirect diffuse lighting by means of an ambient lighting term. In this paper we show how a similar term can be used as a variance reducing technique for stochastic ray tracing. In a theoretical derivation we prove that the technique is mathematically correct. Test results demonstrate its usefulness and effectiveness in practice.

1 Introduction

The original ray tracing technique as introduced by Whitted [1] was the first to be able to render some more complex lighting effects such as reflections and refractions. The resulting images often look spectacular albeit a bit artificial due to the limitations of the lighting simulation. Only perfectly specular reflections and direct diffuse lighting contributions are computed explicitly. Indirect diffuse contributions are approximated by a user-defined ambient lighting term.

Realistic rendering has come a long way since then. Cook et al. [2] presented distributed ray tracing which also renders glossy reflections by stochastically sampling the reflected directions. Kajiya later gave a theoretical foundation to the technique by introducing the rendering equation [3]. From this point of view stochastic ray tracing – or its special case path tracing – is simply a Monte Carlo method to solve an integral equation. Several other researchers have elaborated on this idea [4, 5]. Its advantages are its elegancy and generality. Monte Carlo ray tracing can render a host of lighting effects including depth of field and motion blur, all in a physically and mathematically correct way. Unfortunately the basic technique is slow to converge. Various optimised sampling techniques try to alleviate this problem [6, 7, 8, 9], but rendering typical global illumination effects such as indirect diffuse reflections still requires a lot of computational effort.

This paper presents a new optimisation which further reduces the variance and therefore improves on the required computation time. It elaborates on the idea of an ambient lighting term as it is used in the original ray tracing technique.

Rather than being used as a 'fudge factor' though, it will be shown that it lowers the variance without influencing the expected value of the result. Notably it improves the rendering of diffuse lighting effects.

First we will present the technique from a theoretical point of view. We will discuss the general principle as it is described in literature about Monte Carlo methods and subsequently apply it to the rendering equation. We will discuss what we can and what we cannot expect in terms of improvements. In the last section the theory will be verified by means of some practical test results.

2 Mathematical Principle

The basic idea of the technique is well-known in general Monte Carlo literature as *control variates* [10, 11] or *extraction of a regular part* [12]. Consider for instance the computation of a definite integral of a one-dimensional function $f(x)$ (Fig. 1):

$$I = \int_a^b f(x)dx$$

An estimate for the integral can be found by taking randomly distributed samples x_i over the domain (possibly stratified), evaluating their function values and taking a weighted sum:

$$\langle I \rangle = \frac{(b-a)}{N} \sum_{i=1}^{N} f(x_i)$$

Now suppose one knows a function $g(x)$ which more or less closely approximates $f(x)$ and which can be integrated analytically over the given domain:

$$J = \int_a^b g(x)dx$$

The original integral can then be rewritten as:

$$I = \int_a^b [f(x) - g(x)]dx + \int_a^b g(x)dx = \int_a^b [f(x) - g(x)]dx + J$$

Sampling the integral in the first term now yields another estimate:

$$\langle I \rangle = \frac{(b-a)}{N} \sum_{i=1}^{N} [f(x_i) - g(x_i)] + J$$

One can see intuitively that if $g(x)$ approaches $f(x)$ (by some metric) the variance of this estimate will go to 0. Finding an approximating function which can be integrated therefore may be worthwhile for Monte Carlo integration. Note that a constant control variate $g(x) = C$ does not change the value of the estimate at all and therefore cannot offer any improvement. This will have some consequences later for our application to the rendering equation.

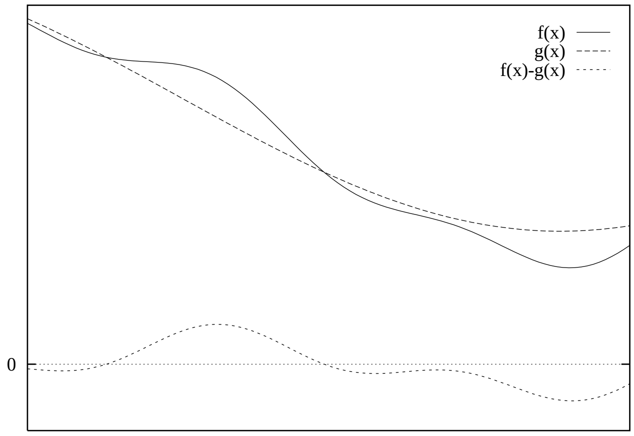

Fig. 1. Computing a definite integral of a function $f(x)$ using a Monte Carlo method. A function $g(x)$ which approximates $f(x)$ and which can be integrated analytically can serve as a control variate. It is more efficient to sample the difference between the functions rather than $f(x)$ itself to compute an estimate.

3 Application to the Rendering Equation

From a mathematical point of view Monte Carlo ray tracing samples the outgoing radiance values from surfaces that are seen through a given pixel. The radiance leaving a point x along direction Θ_x is determined by the rendering equation:

$$\begin{aligned} L(x, \Theta_x) &= L_e(x, \Theta_x) + L_r(x, \Theta_x) \\ &= L_e(x, \Theta_x) + \int_{\Omega_x^{-1}} f_r(x, \Theta_y, \Theta_x) L(y, \Theta_y) \, |\Theta_y \cdot N_x| \, d\omega_y \end{aligned}$$

where:

- $L(x, \Theta_x)$ = the emitted radiance at point x along direction Θ_x [W/m^{2}sr],
- $L_e(x, \Theta_x)$= the self-emitted radiance at point x along direction Θ_x [W/m^{2}sr],
- $L_r(x, \Theta_x)$ = the radiance reflected at point x along direction Θ_x [W/m^{2}sr],
- Ω_x^{-1} = the set of incoming directions around point x,
- $f_r(x, \Theta_y, \Theta_x)$ = the bi-directional reflection distribution function (*brdf*) at point x for light coming in from direction Θ_y and going out along direction Θ_x [1/sr],
- $|\Theta_y \cdot N_x|$ = the absolute value of the cosine of the angle between the direction Θ_y and the normal to the surface at point x,
- y = the point that 'sees' point x along direction Θ_y,
- $d\omega_y$ = a differential angle around Θ_y [sr].

In order to reduce the variance light sources are usually treated separately from other surfaces. For this purpose the terms in the rendering equation are rewritten. The reflected radiance $L_r(x, \Theta_x)$ is split up in a direct term $L_d(x, \Theta_x)$ and an indirect term $L_i(x, \Theta_x)$:

$$L_r(x, \Theta_x) = L_d(x, \Theta_x) + L_i(x, \Theta_x)$$

where:

– The direct term can be found by integrating over the light sources instead of over all incoming angles. It is usually estimated by sampling the surface area of the light sources.

$$\begin{aligned} L_d(x, \Theta_x) &= \int_{\Omega_x^{-1}} f_r(x, \Theta_y, \Theta_x) L_e(y, \Theta_y) \, |\Theta_y \cdot N_x| \, d\omega_y \\ &= \int_{A_l} f_r(x, \Theta_y, \Theta_x) L_e(y, \Theta_y) \frac{|\Theta_y \cdot N_y| \, |\Theta_y \cdot N_x|}{r^2} v(x, y) d\mu_y \end{aligned}$$

– The indirect term, which is of greater concern to us now, is:

$$L_i(x, \Theta_x) = \int_{\Omega_x^{-1}} f_r(x, \Theta_y, \Theta_x) L_r(y, \Theta_y) \, |\Theta_y \cdot N_x| \, d\omega_y$$

Monte Carlo ray tracing proceeds by recursively sampling the latter integral. It is in this sampling process that the variance reducing technique can be applied. As an approximation for the integrand – which is two-dimensional now – we select the *brdf* times a constant ambient radiance L_a. As in the one-dimensional example the indirect term can then be written as:

$$\begin{aligned} L_i(x, \Theta_x) &= \int_{\Omega_x^{-1}} f_r(x, \Theta_y, \Theta_x) [L_r(y, \Theta_y) - L_a] \, |\Theta_y \cdot N_x| \, d\omega_y \\ &\quad + \int_{\Omega_x^{-1}} f_r(x, \Theta_y, \Theta_x) L_a \, |\Theta_y \cdot N_x| \, d\omega_y \\ &= \int_{\Omega_x^{-1}} f_r(x, \Theta_y, \Theta_x) [L_r(y, \Theta_y) - L_a] \, |\Theta_y \cdot N_x| \, d\omega_y + \rho(x, \Theta_x) \times L_a \end{aligned}$$

where $\rho(x, \Theta_x)$ is the total reflectivity for light reaching point x along direction Θ_x. This factor only depends on the local reflective properties of the surface. It can be computed analytically for many models of *brdf*s. The general theory for control variates predicts that sampling the integrand in the new expression for $L_i(x, \Theta_x)$ will yield a smaller variance than with the original expression, on the condition that the approximation of the integrand is close enough.

4 Analysis for Specific Sampling Approaches

There are several alternatives for sampling the original and the new integral expressions. These influence the effectiveness of the variance reduction technique. The parameters that determine the sampling approach are the factors in which the integrand, $f_r(x, \Theta_y, \Theta_x) L_r(y, \Theta_y) |\Theta_y \cdot N_x|$, is split up:

- the probability P that the recursion is continued,
- the probability density function $pdf(\Theta_y)$ that is used to sample the reflected direction over the hemisphere (for the given (x, Θ_x)),
- the weight that forms the resulting estimate $\langle L_i(x, \Theta_x)\rangle$ for the sample if the recursion is continued.

Some of the most notable techniques use the following parameters:

- $P = 1$, $pdf(\Theta_y) = |\Theta_y \cdot N_x| / \pi$, $\langle L_i(x, \Theta_x)\rangle = \pi f_r(x, \Theta_y, \Theta_x) \langle L_r(y, \Theta_y)\rangle$ yields a straightforward technique without importance sampling. The recursion has to be cut off at a certain point, which theoretically introduces a bias. The ambient term may be expected to reduce the variance of the sampling process.
- $P = 1$, $pdf(\Theta_y) = f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| / \rho(x, \Theta_x)$, $\langle L_i(x, \Theta_x)\rangle = \rho(x, \Theta_x) \langle L_r(y, \Theta_y)\rangle$ yields an importance sampling technique. Again the recursion has to be cut off artificially. Importance sampling can be seen as a transformation of the integral. After transformation the *brdf* times the ambient term is reduced to a constant. As noted in the general discussion the ambient term therefore does not offer any improvement here. The technique on its own is not optimal though, since a lot of computational effort may be put in samples with small weights.
- $P = \rho(x, \Theta_x)$, $pdf(\Theta_y) = f_r(x, \Theta_y, \Theta_x) |\Theta_y \cdot N_x| / \rho(x, \Theta_x)$, $\langle L_i(x, \Theta_x)\rangle = \langle L_r(y, \Theta_y)\rangle$ yields an importance sampling technique with Russian roulette to end the recursion. This approach does not introduce a bias and samples are used more sparingly. Because of the Russian roulette the ambient term will affect the variance here.

If the ambient term is introduced the constant $\rho(x, \Theta_x) L_a$ is added to the estimate $\langle L_i(x, \Theta_x)\rangle$ and all occurrences of the estimate $\langle L_r(y, \Theta_y)\rangle$ are replaced by the delta term $\langle L_r(y, \Theta_y)\rangle - L_a$. This principle can be applied at each recursion level. The ambient radiance L_a may vary with the position in the scene.

For a path tracing algorithm which only takes a single sample at each recursion level the eventual estimate for the radiance $L(x, \Theta_x)$ with the last technique looks like:

$$\begin{aligned}\langle L(x, \Theta_x)\rangle = {} & L_e(x, \Theta_x) + \langle L_d(x_0, \Theta_{x_0})\rangle + \rho(x_0, \Theta_{x_0}) L_a(x_0) \\ & + \sum_{k=1}^{n} [\langle L_d(x_k, \Theta_{x_k})\rangle + \rho(x_k, \Theta_{x_k}) L_a(x_k) - L_a(x_{k-1})]\end{aligned}$$

where n is the depth of the recursion. Note that a value of 0 for L_a yields the original path tracing technique.

5 Results

We have verified our theoretical findings by means of an implementation of the path tracing algorithm. The basic implementation already contains some optimised sampling strategies such as multi-dimensional N-rooks sampling and importance sampling with Russian roulette.

The technique was tested on a simple scene of a room with coloured walls and a waxed floor. The room contains a white cube, a slightly specular green cylinder and a glass sphere (Fig. 2). Images were computed at a resolution of 100×100 pixels with 10 samples per pixel and compared against a reference image which was computed at 500 samples per pixel. The RMS error metric was used to measure the variance of the images:

$$RMS = \sqrt{\frac{\sum_p (L_p - L_{p,ref})^2}{N}}$$

Figure 3 shows how the ambient term affects the variance of the resulting images. Several images were computed using the same number of samples per pixel but with increasing estimates for the ambient term. For simplicity the same estimate was used for all objects in the scene.

An ambient term of 0 corresponds to the original path tracing algorithm. As expected from the theory the variance decreases with the introduction of a larger ambient term. It reaches a minimum at the optimal estimate and then increases again. For this example the RMS error is reduced with a maximum of almost 10% as compared to simple path tracing.

The variance can be further reduced by a few percent by estimating the ambient term more accurately for each of the individual objects or regions in the scene, either manually or in a preprocessing step. If the scene is illuminated indirectly by means of some spots pointing to the ceiling the reduction increases to 14%. This result may be expected since the ambient term specifically improves the accuracy for indirect lighting effects.

Another test with an outdoor scene under an overcast sky shows a more spectacular reduction of the RMS error by 87%. Note that the hemisphere is usually not sampled separately as a light source because of its size. Instead it is treated as a source of indirect lighting. The ambient term then proves to be an effective control variate for the skylight.

6 Conclusions

We have presented a variance reducing technique for Monte Carlo ray tracing which is based on an estimated ambient lighting term. In the recursive ray tracing process indirectly reflected radiance values are computed as a sum of two terms: reflected radiance resulting from this ambient term and a correction to this term which is computed in the remainder of the recursive process. In a theoretical discussion we have proven the mathematical correctness of the technique and looked at its potential for various stochastic ray tracing approaches.

Fig. 2. Test scene used for measuring the reduction of the variance.

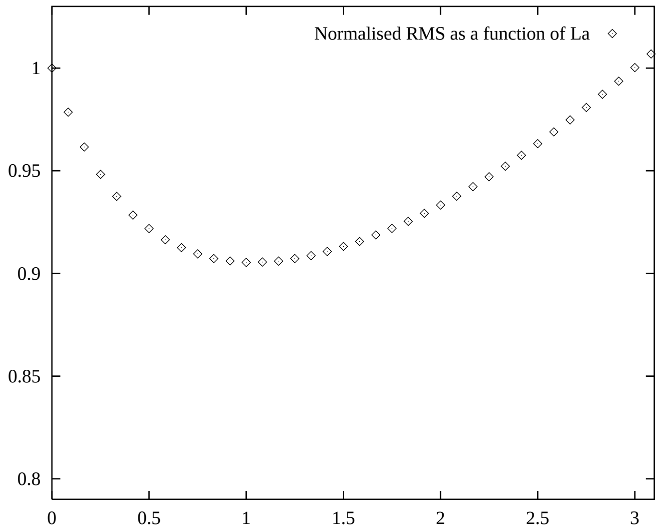

Fig. 3. The introduction of an ambient term reduces the variance for a fixed number of samples per pixel. The variance reaches a minimum at the optimal estimate.

Practical tests have shown that there really is a reduction of the variance. The improvement depends on the quality of the estimate of the ambient lighting and may be modest. Still, the technique has some distinct advantages:

- It mainly improves the rendering of diffuse interreflection, which has always been one of the weak points of Monte Carlo ray tracing. Because the convergence rate of Monte Carlo methods is generally limited to a slow $\sqrt{N}$ any reduction of the variance is more than welcome.
- The extra computational work involved is negligible since no extra rays have to be cast.
- Although an optimal ambient lighting term may be hard to find even a conservative estimate will improve the convergence. A simple preprocessing step may suffice for this purpose.
- Integration with other variance reducing techniques such as stratified sampling is straightforward. If importance sampling cannot be applied because of the complexity of the *brdfs* the technique may offer an even larger improvement. The only requirement is that the *brdf* can be integrated over the hemisphere.

The technique should also prove useful in related Monte Carlo algorithms. More in specific we are currently investigating its application in bi-directional path tracing [13, 14].

7 Acknowledgements

The first author would like to acknowledge the financial support by a grant from the Belgian "Instituut tot Aanmoediging van Wetenschappelijk Onderzoek in Nijverheid en Landbouw" (I.W.O.N.L. #910184). Implementation was done on equipment kindly made available by Hewlett Packard Belgium.

References

1. T. Whitted, "An improved illumination model for shaded display," *Communications of the ACM*, vol. 23, no. 6, 1980.
2. R. Cook, T. Porter, and L. Carpenter, "Distributed ray tracing," *Computer Graphics*, vol. 18, no. 3, pp. 137–145, 1984.
3. J. Kajiya, "The rendering equation," *Computer Graphics*, vol. 20, no. 4, pp. 143–150, 1986.
4. G. Ward, F. Rubinstein, and R. Clear, "A ray tracing solution for diffuse interreflection," *Computer Graphics*, vol. 22, no. 4, pp. 85–92, 1988.
5. P. Shirley and C. Wang, "Distribution ray tracing: Theory and practice," in *Proceedings of the Third Eurographics Workshop on Rendering*, (Bristol, UK), pp. 33–43, May 1992.
6. M. Lee, R. Redner, and S. Uselton, "Statistically optimized sampling for distributed ray tracing," *Computer Graphics*, vol. 19, no. 3, pp. 61–67, 1985.
7. P. Shirley, *Physically Based Lighting Calculations for Computer Graphics*. PhD thesis, University of Illinois, Nov. 1990.

8. P. Shirley and C. Wang, "Direct lighting by monte carlo integration," in *Proceedings of the Second Eurographics Workshop on Rendering*, (Barcelona, Spain), May 1991.
9. B. Lange, "The simulation of radiant light transfer with stochastic ray-tracing," in *Proceedings of the Second Eurographics Workshop on Rendering*, (Barcelona, Spain), May 1991.
10. M. Kalos and P. Whitlock, *Monte Carlo Methods*. Wiley & Sons, 1986.
11. J. Hammersly and D. Handscomb, *Monte Carlo Methods*. Chapman and Hall, 1964.
12. Y. Shreider, ed., *The Monte Carlo Method*. Pergamon Press, 1966.
13. E. Lafortune and Y. Willems, "Bi-directional path tracing," in *Proceedings of CompuGraphics*, (Alvor, Portugal), pp. 145–153, Dec. 1993.
14. E. Lafortune and Y. Willems, "A theoretical framework for physically based rendering," *Computer Graphics Forum*, vol. 13, pp. 97–107, June 1994.