Cost-Driven Multiple Importance Sampling for Monte-Carlo Rendering

Ferenc Csonka, László Szirmay-Kalos, György Antal

Department of Control Engineering and Information Technology,
Technical University of Budapest,
Budapest, Pázmány Péter s. 1/D, H-1117, HUNGARY
Email: szirmay@iit.bme.hu

Abstract

The global illumination or transport problems can also be considered as a sequence of integrals, while its Monte-Carlo solutions as different sampling techniques. Multiple importance sampling takes advantage of different sampling strategies and combines the results obtained with them. In this paper we propose the combination of very different global illumination algorithms in a way that their strengths can be preserved. To do this, we generalize the fundamental theory of multiple importance sampling for sequences of integrals and also take into account the computational cost associated with individual sampling techniques. The theoretical results are used to combine bi-directional path tracing and ray-bundles based stochastic iteration.

Keywords: Multiple importance sampling, stochastic iteration, random walk.

1. Introduction

Global illumination algorithms find the light paths connecting light sources to eye via reflections and compute the image as the integral of the contribution of these paths. Formally, they estimate integral: $\int_{\mathcal{P}} l(z)\, dz$ where $\mathcal{P}$ is the domain of the paths and $l(z)$ is the contribution of a path z. Within this context, different path building strategies have been published, and each of them is good for certain path types. Thus it is worth mixing existing algorithms of different strengths and weaknesses together in order to combine their merits [1, 2, 4].

2. Multiple importance sampling

In this section we recall the fundamental theory of multiple importance sampling [7]. Assume that integral $L = \int_{\mathcal{P}} l(z) dz$ needs to be evaluated. Monte-Carlo quadratures generate samples with certain probability density. Suppose that the ith sampling method uses density $p_i(z)$, and denote the primary estimator of the method i by $l(z)/p_i(z)$. The secondary estimator of a method i is obtained by taking N_i samples and averaging the results. The estimator combined from n sampling techniques is then calculated by weighting samples by appropriate functions $w_i(z)$, and summing the results:

$$\langle L \rangle_c = \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} w_i(z_{i,j}) \cdot \frac{l(z_{i,j})}{p_i(z_{i,j})}. \tag{1}$$

The combined estimator is unbiased if for all z values $\sum_{i=1}^{n} w_i(z) = 1$. In order to find an optimal weighting, the variance of the combined estimator $\langle L \rangle_c$ should be minimized by setting the weights appropriately and also taking into account the constraint of unbiasedness. Unfortunately, this optimization problem cannot be solved analytically, but different quasi-optimal solutions can be obtained. One such approximative solution is called the *balance heuristics*[6]:

$$w_i(z) = \frac{N_i \cdot p_i(z)}{\sum_{k=1}^{n} N_k \cdot p_k(z)}. \tag{2}$$

3. A cost driven approach

Multiple importance sampling assumes that the numbers of samples of the different sampling techniques are known a-priori. However, the definition of these numbers is not at all

presented at MCM 2001.

trivial. One of the objectives of this paper is to incorporate the sampling cost in multiple importance sampling.

Substituting the weights of balance heuristic into the variance formula we get the following variance for the multiple importance sampling method:

$$Var[\langle L\rangle_c] = \int_{\mathcal{P}} \sum_{i=1}^{n} \frac{w_i^2(z)\cdot l^2(z)}{N_i\cdot p_i(z)}\,dz - \sum_{i=1}^{n}\frac{1}{N_i}E^2[L_{i1}] =$$

$$\int_{\mathcal{P}} \frac{n\cdot l^2(z)}{\sum_k N_k\cdot p_k(z)}\,dz - \sum_{i=1}^{n}\frac{1}{N_i}\cdot\left(\int_{\mathcal{P}}\frac{N_i\cdot p_i(z)\cdot l(z)}{\sum_k N_k\cdot p_k(z)}\,dz\right)^2. \tag{3}$$

This variance is further minimized by finding the optimum N_i sample numbers under the constraint of the total computation time. This constraint be expressed as $C=\sum_{i=1}^{n} c_i\cdot N_i$, where c_i is the cost of a sample in method i. In order to carry out this optimization procedure, the integrals of equation (3) should be computed. However, we do not have the explicit form of the path contribution function $l(z)$ and can use only the point samples. Suppose that one method (say method 1) has been executed to obtain some initial N_1^* samples, and the variance is estimated empirically:

$$Var[\langle L\rangle_c] \approx \frac{1}{N_1^*}\cdot\sum_{j=1}^{N_1^*}\frac{n\cdot l^2(z_j)}{\sum_k N_k\cdot p_k(z_j)\cdot p_1(z_j)} -$$

$$\left(\frac{1}{N_1^*}\cdot\sum_{j=1}^{N_1^*}\frac{N_i\cdot p_i(z_j)\cdot l(z_j)}{\sum_k N_k\cdot p_k(z_j)\cdot p_1(z_j)}\right)^2.$$

Unfortunately, this function cannot be analytically minimized. However, its derivatives can be easily computed, thus a gradient search can provide quick approximations of the optimum. The optimization is an iteration where each step consists of a gradient step and a projection to the plane of the cost constraint.

4. Multiple importance sampling for the global illumination problem

Let us now examine how the general concepts can be applied to solve transport problems and particularly the global illumination problem. The solution of the global illumination problem can be obtained in the form of a Neumann series:

$$L = \sum_{i=0}^{\infty}\mathcal{T}^i L_e = L_e + \mathcal{T}(L_e+\mathcal{T}(L_e+\mathcal{T}(L_e+\ldots))).$$

where $\mathcal{T}$ is the integral operator of the light transport and L_e is the emission function. Note that here not a single integral, but a sequence of integrals should be computed. Due to the recursive formulation of the Neumann series, these integrals can be evaluated simultaneously and the samples of the first few variables of a higher dimensional integrand can be used for the integrands of the lower dimensional terms in the series. When the sampling process results in a sample and the integrand at this point is computed, the integrand value will represent the sum of the direct contribution, single reflection, double reflection etc. In many methods, the separation of the contributions of different reflections would be too cumbersome, thus we look for weighting techniques, which weight the total contribution of all terms.

The integrand is an infinite series which should be truncated to terminate the computation in finite time. When Russian roulette is used the computation of this series terminated randomly assuming that the sum of the rest of the terms is zero. This means that the probability of computing all terms is also zero. Looking at the formula of weights in multiple importance sampling, this means that those samples of light paths that have been truncated by Russian roulette always get zero weight, thus multiple importance sampling is not appropriate for integral series computed with Russian roulette.

In order to find a solution for this problem, let us consider a very simple case. Suppose that

$$I = \int_{[0,1)}\left[f(x)+\int_{[0,1)} g(x,y)\,dy\right]dx \tag{4}$$

is approximated. In the ith method we use N_i samples. First we decide whether or not we take any samples with probability $s_i^{(x)}$. If a sample is taken, then x is obtained with probability density $p_i^{(x)}(x)$ and $f(x)/s_i^{(x)}$ is computed. If no sample is generated, we assume that $x=1$ and $f(1)=0$ and $g(1,\cdot)=0$. Note that with this assumption the domains of f and g are extended to represent those cases when they are not sampled. This random termination can also be described by the following density:

$$\tilde{p}^{(x)}(x) = \begin{cases} s^{(x)}\cdot p^{(x)}(x) \text{ if } x\neq 1,\\ (1-s^{(x)})\cdot\delta(x-1) \text{ if } x=1.\end{cases}$$

where $\delta(x)$ is the Dirac-delta function. Having sample x computed, we decide with probability $s_i^{(y)}(x)$ whether or not a y sample is obtained. Sample y is generated with density $p_i^{(y)}(y|x)$ and $g(x,y)/s_i^{(y)}(x)$ is computed. If no sample is generated, we assume that $y=1$ and $g(x,y)=0$. The random termination and sampling of y can also be described by the following modified density:

$$\tilde{p}^{(y)}(y|x) = \begin{cases} s^{(y)}(x)\cdot p^{(y)}(y|x) \text{ if } y\neq 1,\\ (1-s^{(y)}(x))\cdot\delta(y-1) \text{ if } y=1.\end{cases}$$

These modified densities allow methods incorporating Russian roulette to be handled as the normal case. The probability of obtaining an x,y pair is the product of probabilities of sampling x, the continuation and then sampling y, that is $\tilde{p}^{(y)}(x,y)=\tilde{p}^{(x)}(x)\cdot\tilde{p}^{(y)}(y|x)$.

Using the modified density, the primary estimator of I in equation (4) is

$$I_{ij} = w_i(x_{ij}, y_{ij}) \cdot \left(\frac{f(x_{ij})}{\tilde{p}_i^{(x)}(x_{ij})} + \frac{g(x_{ij}, y_{ij})}{\tilde{p}_i^{(y)}(x_{ij}, y_{ij})} \right).$$

The combined estimate is $\langle I \rangle_c = \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} I_{ij}$. Ignoring the covariance between the samples of f and g, the variance of the estimate can be obtained in the following form:

$$Var[\langle I \rangle_c] = \sum_{i=1}^{n} \frac{1}{N_i^2} \sum_{j=1}^{N_i} Var[I_{ij}] = \sum_{i=1}^{n} \frac{1}{N_i} Var[I_{i1}] \approx$$

$$\int\limits_{[0,1]^2} \sum_{i=1}^{n} \frac{w_i^2(x,y)}{N_i} \cdot \left(\frac{f^2(x)}{\tilde{p}_i^{(x)}(x)} + \frac{g^2(x,y)}{\tilde{p}_i^{(y)}(x,y)} \right) dxdy -$$

$$\sum_{i=1}^{n} \frac{1}{N_i} E^2[I_{i1}].$$

This function should be minimized with the constraint $\sum_{i=1}^{n} w_i(z) = 1$. Unfortunately, this optimization problem cannot be solved analytically, thus quasi-optimal solutions are found. For example, we can aim at minimizing just the integrand of the above integral. Using the Lagrange multiplier method to incorporate the constraint, the minimum of

$$\sum_{i=1}^{n} \frac{w_i^2(x,y)}{N_i} \cdot \left(\frac{f^2(x)}{\tilde{p}_i^{(x)}(x)} + \frac{g^2(x,y)}{\tilde{p}^{(y)}(x,y)} \right) -$$

$$\lambda \cdot \left(\sum_{i=1}^{n} w_i(x,y) - 1\right)$$

is needed, which can be obtained by making the partial derivatives by w_i and λ equal to zero. The final result is

$$w_i(x,y) = \frac{v_i(x,y)}{\sum_{k=1}^{n} v_k(x,y)},$$

where

$$v_i(x,y) = \frac{N_i}{\frac{f^2(x)}{\tilde{p}_i^{(x)}(x)} + \frac{g^2(x,y)}{\tilde{p}_i^{(y)}(x,y)}}. \tag{5}$$

We shall call $v_i(x,y)$ as the *un-normalized weight* since weights w_i are obtained as their normalization. Let us interpret this formula. When the second integral of g is relevant, that is g is large, then f can be ignored, thus we have

$$w_i(x,y) \approx \frac{N_i \cdot \tilde{p}_i^{(y)}(x,y)}{\sum_{k=1}^{n} N_k \cdot \tilde{p}_k^{(y)}(x,y)}.$$

This is basically the original multiple importance sampling strategy applied for the integral of g.

However, when the second integral is negligible, i.e. g is small, then it can be omitted, thus

$$w_i(x,y) \approx \frac{N_i \cdot \tilde{p}_i^{(x)}(x)}{\sum_{k=1}^{n} N_k \cdot \tilde{p}_k^{(x)}(x)},$$

Note that we got back the original scheme as applied only to the first integral.

Let us try to apply this concept to the solution of the global illumination problem. The generalization of equation (5) for the Neumann series is straightforward, thus the un-normalized weights are as follows:

$$v_i(\omega_1, \omega_2, \ldots) = \frac{N_i}{\sum_j \frac{L_j^2(\omega_1,\ldots,\omega_j)}{\tilde{p}_i(\omega_1,\ldots,\omega_j)}}, \tag{6}$$

where $L_j(\omega_1, \ldots, \omega_j)$ is the contribution of a path which contains j reflection points and the light goes parallel to ω_k after the kth reflection. In the following sections, we combine bi-directional path tracing and ray-bundle based iteration using this formula.

5. Bi-directional path tracing

Random walk algorithms build up light paths connecting the light sources with the eye according to a random simulation of light scattering at the surfaces. At light-surface interaction points, the continuation direction is obtained with a probability density that is approximately proportional to probability density of photon reflection $w(\omega_{in}, \omega_{out})$. Russian roulette uses the integral of this function as the probability of termination. Bi-directional path tracing [3] simultaneously builds up an eye-subpath starting from the eye and a light-subpath initiated from the lightsources, and the two subpaths are connected deterministically. The strength of bi-directional path tracing is that it treats specular effects such as specular hightlights mirrors and caustics well and renders accurate shadows. The weakness of the algorithm is that it cannot exploit coherence thus it requires very many samples to get a noise free image.

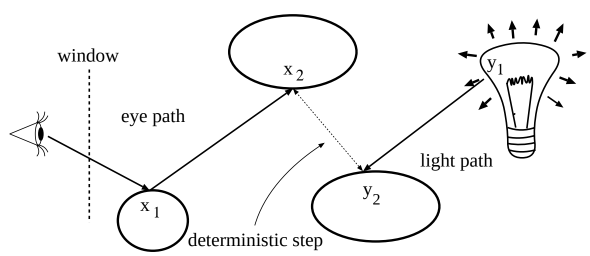


Figure 1: *Bi-directional path tracing*

6. Global ray-bundle iteration

Iteration algorithms are based on the fact that the solution of the rendering equation is the fixed point of the following iteration scheme $L_{i+1} = L^e + \mathcal{T} L_i$. Iteration requires the complete radiance function to be stored that results in astronomical storage requirements. To solve this problem, stochastic iteration replaces the determininistic operator $\mathcal{T}$ by a random operator $\mathcal{T}^*$, which behaves as the original light transport operator in the average case:

$$L_{i+1} = L^e + \mathcal{T}_i^* L_i, \qquad E[\mathcal{T}^* L] = \mathcal{T} L.$$

In this stochastic iteration scheme the radiance function and its functionals do not converge. However, if the estimates of subsequent iteration steps are averaged, this average will converge to the real solution [5].

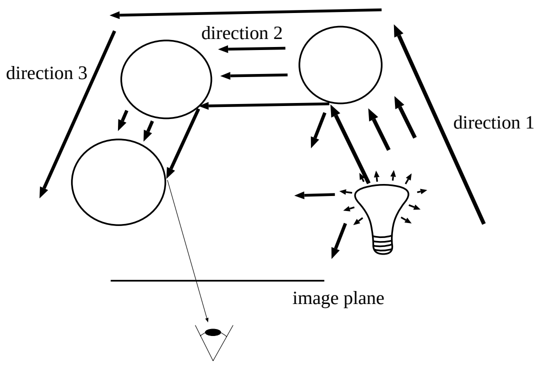


Figure 2: *A path of ray-bundles*

In ray-bundle iteration the randomization happens by choosing a random direction using a uniform distribution, and the radiance of all points is transferred parallel to this direction, then the transferred radiance is reflected towards the eye and to the next random direction. The strength of ray-bundle iteration is that the algorithm exploits coherence, therefore it is very fast. Combining with Gouraud or Phong shading the generated images are not noisy. The weakness of the algorithm is that by sampling global directions uniformly over the bounding sphere, it cannot take into account locally important directions. For example, it is very unlikely to sample near ideal mirror reflection directions.

7. The combined algorithm

According to equation (6), when computing the weight of a contribution of method i from the un-normalized weight, we should be able to determine the un-normalized weights for all methods to be combined. Let us denote bi-directional path tracing as method 1 and stochastic iteration as method 2. Consider path $z = (\omega_1, \omega_2, \ldots, \omega_J)$.

When a sample with random walk is obtained, we need the un-normalized weight of this sample which would be obtained with stochastic iteration. The probability of generating path $\omega_1, \ldots, \omega_J$ by random walk can be approximated as:

$$\tilde{p}_1(\omega_1, \ldots, \omega_J) = w(\omega_{eye}, \omega_1) \cdot w(\omega_1, \omega_2) \cdot \ldots \cdot w(\omega_{J-1}, \omega_J).$$

The radiance carried by the same path is:

$$L_j(\omega_1, \ldots, \omega_j) = L_e(\omega_1, \ldots, \omega_j) \cdot w(\omega_{eye}, \omega_1) \cdot w(\omega_1, \omega_2) \cdot \ldots \cdot w(\omega_{j-1}, \omega_j).$$

where $L_e(\omega_1, \ldots, \omega_j)$ is the emission at the end of the path in direction $-\omega_J$. Thus the un-normalized weight is:

$$v_1(\omega_1, \ldots, \omega_J) = \frac{N_1}{\sum_{j=1}^{J} L_e^2(\omega_1, \ldots, \omega_j) \cdot w(\omega_{eye}, \omega_1) \cdot \ldots \cdot w(\omega_{J-1}, \omega_J)}.$$

The probability of generating a path of length j by stochastic iteration is $\tilde{p}_2(\omega_1, \ldots, \omega_j) = (1/2\pi)^j$, and the carried power is obviously the same as before, thus the un-normalized weight of stochastic iteration is

$$v_2(\omega_1, \ldots, \omega_J) = \frac{N_2}{\sum_{j=1}^{J} (2\pi)^j \cdot (L_e(\omega_1, \ldots, \omega_j) \cdot w(\omega_{eye}, \omega_1) \cdot \ldots \cdot w(\omega_{j-1}, \omega_j))^2}.$$

Let us now consider the other case when the primary sample is obtained with ray-bundle iteration. The denominator of the un-normalized weight

$$v_i(\omega_1, \omega_2, \ldots, \omega_J) = \frac{N_i}{\sum_{j=1}^{J} \frac{L_e^2(\omega_1, \ldots, \omega_j) \cdot (w(\omega_1, \omega_2) \cdot \ldots \cdot w(\omega_{j-1}, \omega_j))^2}{\tilde{p}_i(\omega_1, \ldots, \omega_j)}}$$

is computed recursively backwards at each iteration step. In order to compute these values, the so called *incoming term*

$$I_i(\omega_1, \ldots, \omega_J) = \sum_{j=1}^{J} \frac{L_e^2(\omega_1, \ldots, \omega_j) \cdot (w(\omega_1, \omega_2) \cdot \ldots \cdot w(\omega_{j-1}, \omega_j))^2}{\tilde{p}_i(\omega_1, \ldots, \omega_j)}$$

is also transferred by the rays. When the stochastic iteration generates new direction ω_1', then these terms are multiplied by $w(\omega_1', \omega_1)/p_i(\omega_1')$ and those patches are identified which are in direction ω_1' from this patch. The identified patches will have this new incoming term.

Assuming the probability densities of random walk, the incoming term is:

$$I_1(\omega_1, \ldots, \omega_J) = \sum_{j=1}^{J} L_e^2(\omega_1, \ldots, \omega_j) \cdot w(\omega_1, \omega_2) \cdot \ldots \cdot w(\omega_{j-1}, \omega_j).$$

For stochastic iteration:

$$I_2(\omega_1, \ldots, \omega_J) = \sum_{j=1}^{J} L_e^2(\omega_1, \ldots, \omega_j) \cdot (2\pi)^j \cdot (w(\omega_1, \omega_2) \cdot \ldots \cdot w(\omega_{j-1}, \omega_j))^2.$$

From the incoming term, the un-normalized weights are computed by the following formulae. For random walk:

$$v_1(\omega_1, \omega_2, \ldots, \omega_J) = \frac{N_1}{w(\omega_{eye}, \omega_1) \cdot I_1(\omega_1, \ldots, \omega_J)}.$$

For stochastic iteration:

$$v_2(\omega_1, \omega_2, \ldots, \omega_J) = \frac{N_2}{2\pi \cdot (w(\omega_{eye}, \omega_1))^2 \cdot I_2(\omega_1, \ldots, \omega_J)}.$$

When combining ray-bundle iteration and bi-directional path tracing, we start with an initial ray-bundle iteration phase with N_1^* samples. At the end of this phase, the variances of the pixels are evaluated and we determine how many additional bi-directional samples per pixel and global ray-bundle steps are needed using the proposed cost-driven approach.

8. Implementation and Results

In order to demonstrate the proposed weighting technique, two scenes are used. The first is called the Cornell Chickens scene (25K patches), where the floor and the back wall has dominant specular characteristics (figure 3).

Ray-bundle iteration is particularly powerful at diffuse and glossy surfaces but get poorer at highly specular objects. On the other hand, bi-directional path tracing handles the specular objects well but results in high noise at diffuse and glossy surfaces. Thus the automatic weighting will downscale the results of ray-bundle iteration at highly specular surfaces and the results of bi-directional path tracing at diffuse or glossy surfaces. The final image is the sum of the weighted images.

The second example is an architectural scene modelled in ArchiCAD. The images in figure 4 are snapshots of an architectural walk-through.

9. Conclusions

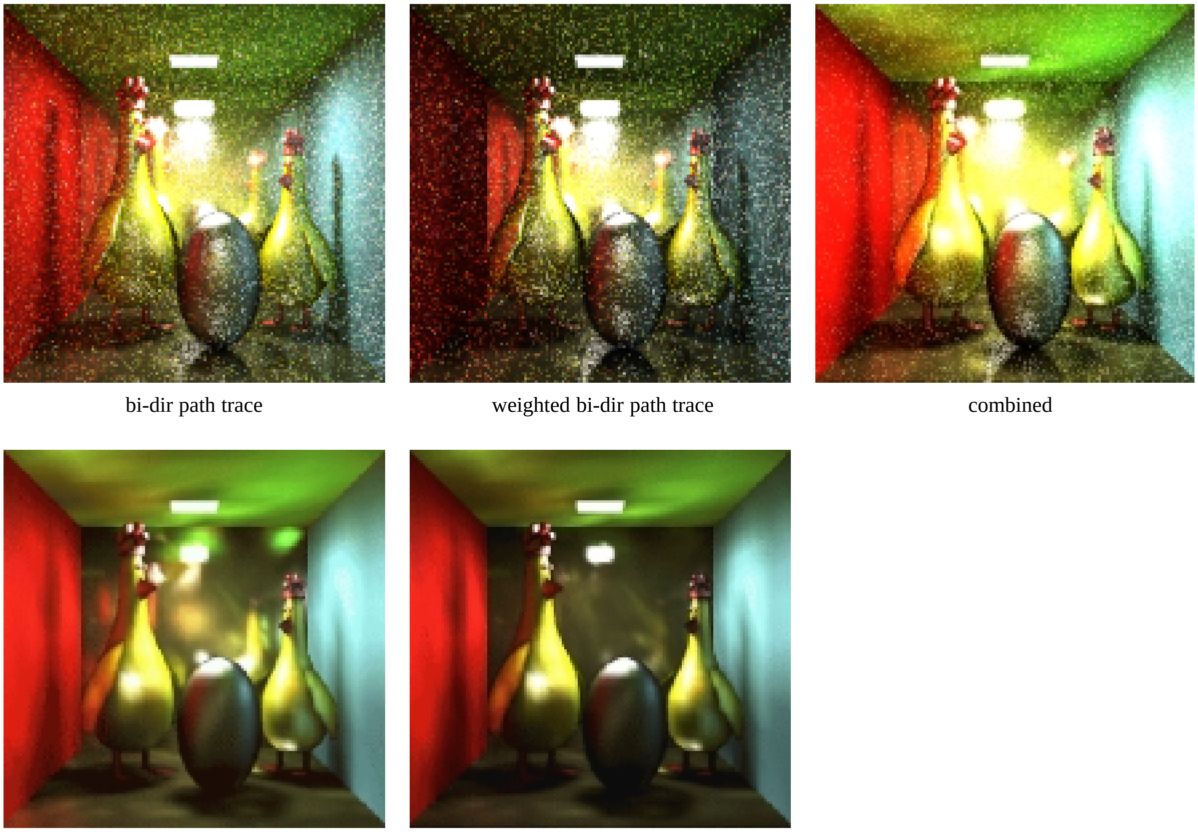
This paper proposed a combination of bi-directional path tracing and global ray-bundle iteration using an efficient strategy determining the weights of the different passes. The resulted algorithm exploits the advantages of both underlying algorithms, namely the fast image generation of ray-bundle iteration, and the precise specular artifact calculation of path tracing.

10. Acknowledgement

This work has been supported by National Scientific Research Fund (OTKA ref. No.: T029135), the Eötvös Foundation and the IKTA-00101/2000 project.

References

1. J. R. Wallace M. F. Cohen, D. P. Greenberg: A two-pass solution to the rendering equation: A synthesis of ray tracing and radiosity methods. Computer Graphics, 21(4), pp. 311-320, 1987.
2. Xavier Granier, George Drettakis and Bruce Walter: Fast Global Illumination Including Specular Effects. Eurographics Rendering Workshop'00, 2000.
3. Eric P. Lafortune, Yves D. Willems: Bi-directional Path-Tracing. Proceedings of Compugraphics '93, pp. 145-153, Alvor, Portugal 1993.
4. Frank Suykens, Yves D. Willems: Weighted Multipass Methods for Global Illumination. Computer Graphics Forum (Eurographics'99), 1999.
5. László Szirmay-Kalos: Stochastic Iteration for non-diffuse Global Ilumination. Computer Graphics Forum (Eurographics'99) 18(3):233-244, 1999.
6. Eric Veach, L. J. Guibas: Optimally Combining Sampling Techniques for Monte Carlo Rendering, SIGGRAPH '95 Proceedings, pp. 419-428, 1995.
7. Eric Veach: Robust Monte Carlo Methods for Light Transport Simulation. PhD thesis, Stanford University, 1997.



bi-dir path trace | weighted bi-dir path trace | combined

ray-bundle | weighted ray-bundle

Figure 3: *The evolution of images for the Cornell chickens scene*

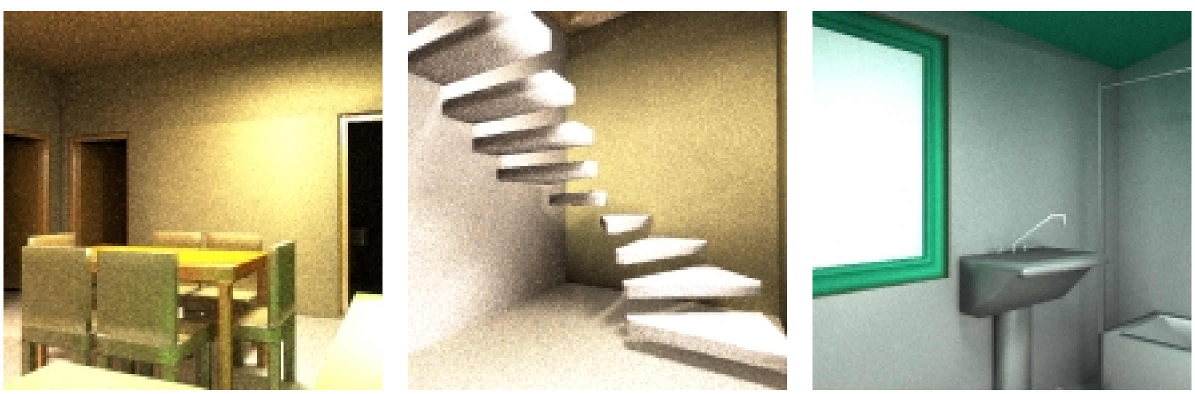


Figure 4: *An architectural scene rendered with the combined method*

presented at MCM 2001.