ADAPTIVE WAVELET DENSITIES FOR MONTE CARLO RAY TRACING

Georg Pietrek
Universität Dortmund
Informatik LS VII
Otto-Hahn-Str. 16
D-44221 Dortmund
Germany
pietrek@ls7.cs.uni-dortmund.de

Ingmar Peter
Universität Tübingen
WSI/GRIS
Auf der Morgenstelle 10, C9
D-72076 Tübingen
Germany
peter@gris.uni-tuebingen.de

ABSTRACT

Monte Carlo integration is a well established technique to solve the rendering equation. The efficiency of Monte Carlo integration strongly depends on the probability density functions (pdfs) used to control the stochastic process.
We will introduce a new method for representation and adaption of pdfs for Monte Carlo importance sampling based on a new mathematical approach for adaptive pdfs in basis representation. During the normal Monte Carlo integration process an approximation of the integrand is obtained that can be used to construct refined pdfs that tend to achieve better results. Based on this strategy we present a multi pass Monte Carlo algorithm using hierarchical function bases as known from wavelet applications. This approach is used to optimise the calculation of indirect illumination in a backward ray tracing application. The results show that the use of adaptive pdfs improves the image quality as well as the computational efficiency of the calculations.

CR Categories: G.1.2 [**Mathematics of Computing**]: Approximation - Least Squares Approximation G.3 [**Probability and statistics**]: Probabilistic Algorithms (including Monte Carlo) I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism - Ray Tracing
Keywords: Monte Carlo Ray Tracing, Importance Sampling, Adaptive PDFs, Wavelets.

INTRODUCTION

Monte Carlo integration is a well known technique to calculate integrals by stochastically sampling the integration domain [KW86]. Given a probability density function (pdf) p with the same support as the integrand f, the value of the integral can be estimated using

$$\int f(x)\,\mathrm{d}x \approx \frac{1}{m}\sum_{i=1}^{m}\frac{f(x_i)}{p(x_i)} \qquad (1)$$

where x_i, $i = 1, \ldots, m$ are random numbers independently drawn according to p.

One possible application of Monte-Carlo integration is the physical correct simulation of light transport, which is basically the solution of the rendering equation [Kaj86]. Calculating this equation using Monte Carlo sampling means to draw random directions over the hemisphere above a surface point, then determine the illumination arriving from these directions and sum up the sample values using (1).

Since Monte Carlo sampling is a random process, the variance σ^2 of the estimator depends strongly on the pdf used to select the sampling directions. From Monte Carlo theory it is known that the optimal pdf has to be proportional to the integrand. For the rendering equation this would require

$$p \propto f_r\, L_i\, \cos(\theta) \qquad (2)$$

where f_r is the brdf, L_i the incoming radiance and θ the angle between incoming direction and normal at the surface point. It is not possible to choose the pdf in exactly this way because this would require to know L_i which is recursively defined by the same equation. So efforts have to be made to find pdfs which closely match condition (2) to minimize the variance in the solution.

Kajiya [Kaj86] proposed a simple path tracing algorithm using a pdf $p \propto f_r$ to choose a direction when extending the path. This is only a partial good solution, because the incoming illumination is not taken into account in the pdf.

Various algorithms [Jen95, Jen96, UT97] were proposed to construct pdfs for Monte Carlo importance sampling, which take the incoming illumination into account. Most of them consist of two steps. In the first step an approximate solution for the global illumination in the scene is calculated. These results are used to obtain appropriate pdfs and calculate the final result in a second step. But if the calculation of the global illumination during the first step of these algorithms is too coarse (e.g. too few particles were used or the radiosity grid was too coarse), the constructed pdfs will cause a high variance.

To solve this problem Dutré and Lafortune proposed multi pass algorithms with adaptive pdfs. They present particle tracing approaches based on the VEGAS method [Lep78] or piecewise constant pdf representations [DW94, DW95] and backward ray tracing using an adaptive 5D tree [LW95]. During the sampling process these representations are adapted to approximate visual importance [DW94, DW95] or incident illumination [LW95].

Instead of piecewise constant approximations Bustillo used a neural gas structure, to store information about the illumination in the scene [Bus97]. In the first step rays are shot from the light sources and stored in the neural gas structure. In the second step of the algorithm an evolutionary importance sampling is carried out using the information in the gas structure.

Veach and Guibas present a method based on Metropolis sampling that converges to the optimal probability density without storing any pdf representation [VG97]. This is done by local exploration of the path space starting from a few seed light transport paths and applying random mutations to them.

In this paper we will introduce a new method for representation, storage, and adaption of pdfs for Monte Carlo importance sampling. Our approach is a multi pass Monte Carlo algorithm using adaptive pdfs. In each pass we get as a byproduct of the normal Monte Carlo integration process an approximation of the integrand that is used to construct a refined pdf for the following pass. The pdfs are representated in hierarchical function bases as known from wavelet applications. Furthermore our approach allows the use of higher order functions bases (eg. B-Splines). This approach is used to optimise the calculation of indirect illumination in a backward ray tracing application.

The remaining part of this paper is organized as follows: The next paragraph describes a multi-pass Monte Carlo integration method that works with successively refined probability density functions. After that a short outline of our rendering algorithm is given. The following section presents some results showing the advantages of our approach.

MATHEMATICAL APPROACH

In the following we will consider functions from the space $L^2(0,1)$ consisting of functions $f : [0,1] \to \mathbb{R}$ with

$$\int_0^1 |f(x)|^2 \, \mathrm{d}x < \infty$$

ADAPTIVE PDFS

In this section we show that it is possible to construct an approximation $\hat{f}$ of the integrand f from the samples obtained during a Monte Carlo integration.[1] From this approximation a new, adapted pdf p^+ is constructed that can be used in further Monte Carlo passes.

Let $N = \{N_j(x) \in L^2(0,1), j = 1, \ldots, n\}$ be a function basis where every basis function can be written as a scaled probability density function:

$$N_j(x) = w_j \, p_j(x) \tag{3}$$

This is true for non-negative basis functions, eg. for the B-spline basis functions. The scaling factors are then given by $w_j = \int_0^1 N_j(x) \, \mathrm{d}x$.

Let p be a probability density function represented in the given basis N:

$$p(x) = \sum_{j=1}^{n} c_j \, N_j(x)$$

[1] To review basic concepts of function approximation please look at the appendix.

We presume that we have methods for generating random numbers according to the pdfs p_j. Then a random number $x_i \sim p$ can be generated by first choosing an index j of a basis function and then generating a random number according to the chosen pdf p_j [KW86]. Index j has to be chosen according to

$$Prob(\text{choosing } j) = \frac{c_j\, w_j}{\sum_{k=1}^{n} c_k\, w_k} \tag{4}$$

In fact this defines a discrete pdf.

For each index j we get a set of random numbers $S^j = \{x_k^j, k = 1, \ldots, n_j\}$ with $x_k^j \sim p_j$. The union $S = \cup S^j = \{x_i\}$ of these sets gives a set of random numbers $x_i \sim p$ that can be used to get an estimate of the integral as described in equation (1).

Furthermore the sets S^j and the corresponding sample values $f(x_k^j)$ can be used to construct an approximation

$$\hat{f} = \sum_{j=1}^{n} c_j^\star\, N_j(x)$$

of the integrand f in the basis N. The coefficients $c_j^\star$ can be calculated with

$$\begin{aligned} c_j^\star &= \langle f, N_j \rangle \\ &\overset{(5)}{=} \int_0^1 f(x)\, N_j(x)\, \mathrm{d}x \\ &\overset{(1)}{\approx} \frac{1}{n_j} \sum_{k=1}^{n_j} \frac{f(x_k^j)\, N_j(x_k^j)}{p_j(x_k^j)} \\ &\overset{(3)}{=} \frac{1}{n_j} \sum_{k=1}^{n_j} w_j\, f(x_k^j) \end{aligned}$$

If N is an orthonormal basis, we know from (6) that $c_j^\star$ is the coefficient of the approximation $\hat{f}$ in the basis N.

For a non-orthonormal basis, (7) says that $c_j^\star$ is the coefficient of the approximation $\hat{f}$ represented in the dual basis $\tilde{N}$. In this case, changing back to the primal basis N is a simple operation that consists of multiplying the vector $[c_j^\star]$ with a matrix $M_{\tilde{N} \to N}$.

The coefficients $c_j^\star$ are stochastically calculated by Monte Carlo integration. So the approximation $\hat{f}$ is not the exact proximum. But from the law of large numbers we know that for a sufficient number of samples our approximation converges to the proximum.

Now we have an approximation $\hat{f}$ of f in the basis N. Our goal is to get a refined pdf p^+ from this approximation. In order to define proper probabilities in equation (4) the coefficients c^+ of p^+ have to be non-negative. This can be achieved by setting negative coefficients of $\hat{f}$ to zero. It is even better to set all coefficients below some threshold τ to the value τ, because then it is guaranteed that the pdf does not drop to zero in the integration domain.

After this correction of small coeffcients $\hat{f}$ has to be normalized to get the pdf p^+. The integral of $\hat{f}$ that is needed for this normalization is simply calculated as

$$\begin{aligned} \int_0^1 \hat{f}(x) &= \int_0^1 \sum_{j=1}^{n} c_j^\star\, N_j(x)\, \mathrm{d}x \\ &= \sum_{j=1}^{n} c_j^\star \int_0^1 N_j(x)\, \mathrm{d}x \\ &\overset{(3)}{=} \sum_{j=1}^{n} c_j^\star\, w_j \end{aligned}$$

So the coefficients of the new pdf p^+ are given by

$$c_j^+ = \frac{c_j^\star}{\sum_{k=1}^{n} c_k^\star\, w_k}$$

The approximating pdf p^+ is used for the next Monte Carlo pass.

HIERARCHICAL PDFS

If we choose the bases for the pdf representation to be the scaling bases of a wavelet hierarchy, we can refine the pdf on successively finer scales. This is simply done by changing the basis of p^+ to a higher level in the wavelet hierarchy before beginning the next Monte Carlo pass, a process known as refinement in wavelet applications [SDS96].

INTEGRATION IN HIGHER DIMENSIONS

The generalization of this approach to higher dimensions is straightforward. Let p be a pdf given in a two-dimensional tensor product representation:[2]

$$p(x, y) = \sum_{j=1}^{n} \sum_{k=1}^{n} c_{j,k} N_j(x) N_k(y)$$

[2] For simplicity of notation we assume the same basis N is taken for x and y direction, of course it is possible to use different bases N^x and N^y.

Instead of choosing an index j we have to choose a pair (j,k) according to

$$Prob\,(\text{choosing } (j,k)) = \frac{c_{j,k}\, w_j\, w_k}{\sum_{l=1}^{n}\sum_{m=1}^{n} c_{l,m}\, w_l\, w_m}$$

Then we have to generate sets $S^{j,k}$ of random number pairs $(x_l^{j,k}, y_l^{j,k}) \sim p_{j,k}(x,y)$. Because of the tensor product approach $p_{j,k}(x,y) = p_j(x)\, p_k(y)$ holds and the pair $(x_l^{j,k}, y_l^{j,k})$ can be generated independently according to $x_l^{j,k} \sim p_j(x)$ and $y_l^{j,k} \sim p_k(y)$. Again, $S = \cup S^{j,k}$ is a set of random pairs $(x_i, y_i) \sim p$ that is used to calculate an estimate of the two-dimensional integration problem

$$\int_0^1\!\!\int_0^1 f(x,y)\,\mathrm{d}y\,\mathrm{d}x \approx \frac{1}{m}\sum_{i=1}^{m}\frac{f(x_i,y_i)}{p(x_i,y_i)}$$

In the same way as for the one-dimensional case, the sets $S^{j,k}$ can be used to get coefficients for a two-dimensional approximation $\hat{f}$ of the integrand f in the tensor product basis. After correcting too small coefficients and normalizing we get a refined pdf p^+ that can be used in the next Monte Carlo pass.

ALGORITHM

Our approach for Monte Carlo integration using adaptive pdfs is used in a backward ray tracing algorithm. Rays are traced from the eye into the scene. Direct and indirect illumination are handled separately.

The direct contribution of light sources is calculated by sending shadow rays towards light sources. Since we are using area light sources shadow rays are sent to jittered points on the light sources. For our examples we used 4 shadow rays per light source.

The indirect illumination is calculated by a Monte Carlo path tracing approach. A new ray is sent into a stochastically chosen direction of the hemisphere. The incoming radiance from this ray direction is calculated by recursive application of the lighting calculation at the intersection point of this ray. The ray paths are terminated by russian roulette [AK90].

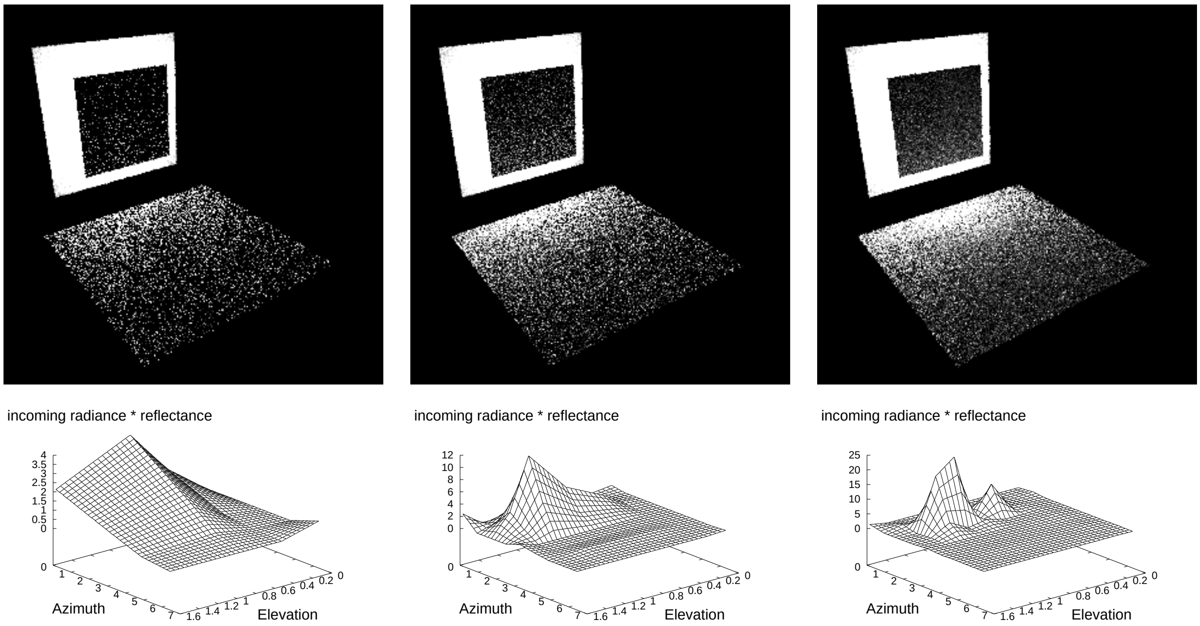
The probabilistic generation of ray directions for the indirect lighting contributions is controlled by adaptive pdfs. In a general approach this would require six-dimensional function representations for the pdfs because they depend on the location on a surface (local coordinates u, v), and the incoming (θ_{in}, ϕ_{in}) and outgoing (θ_{out}, ϕ_{out}) ray directions. Using a six-dimensional tensor product approximation leads to a large number of coefficients, eg. 262144 coeffients for a basis consisting of 8 basis functions. As stated in [DW95] using pure diffuse reflection eliminates the dependency on the outgoing direction.[3] The dependency on the location can be handled by discretizing the surfaces into patches and using a single pdf for each patch. So this leaves us with pdfs depending only on incoming ray directions. For these pdfs we used a two-dimensional tensor product representation as described above.

The image is rendered in different passes. The process starts with uniform pdfs. After each pass the pdfs are adapted and refined as described in the previous section. The pdfs adapt to the integrand of the rendering equation that means to $L_i\, f_r\, \cos(\theta)$. Since the incoming radiance is wavelength dependent but the pdf is a single valued function we adapt to the intensity of the incoming radiance. Another approach would be to establish different pdfs for different wavelengths of light but that would further require to trace independent rays for each wavelength.

In each pass a different number of samples per pixel can be used. In contrast to other approaches [LW95] the results are independent from the order of pixel evaluations because the pdfs do not alter during a single pass. As a rule of thumb more samples should be used in later passes because the better quality of the adapted pdfs should lead to higher efficiency of the Monte Carlo integration. The only requirement on the number of samples in early passes is that the number should be large enough to result in reliable function approximations. Using a too small number of samples leads to bad function approximations and perhaps higher variance. But even then the algorithm is still unbiased because these approximations are only used as pdfs for the Monte Carlo process.

The resulting image is generated as a weighted sum of the images resulting from the different passes, where each image is weighted according to the number of samples per pixel used in the corresponding pass.

[3] Please note that in a backward ray tracing application the direction of outgoing light is the direction from where the ray reached the intersection point.



first pass second pass third pass

Figure 1: Adaptive pdfs, linear B-spline basis

RESULTS

VERIFICATION

The first test was to verify that the adaptive pdfs have the desired properties. For this purpose a small scene consisting of a ground patch that is exclusively illuminated by indirect illumination reflected off from another patch is used (see Fig. 2).

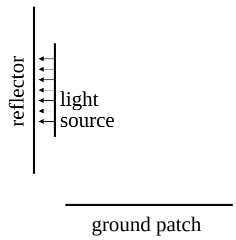


Figure 2: Geometry of scene

The upper row of Fig. 1 shows the images generated in the first three passes of our ray tracing process. The bottom row of Fig. 1 shows the adapted pdf positioned in the center of the ground patch resulting from these passes.

As the calculations progress, the pdf adapts better to the incoming light. Light reaches the ground patch only from visible portions of the reflector. The adapted pdfs show how the sampling process concentrates on those parts of the hemisphere where the light comes from.

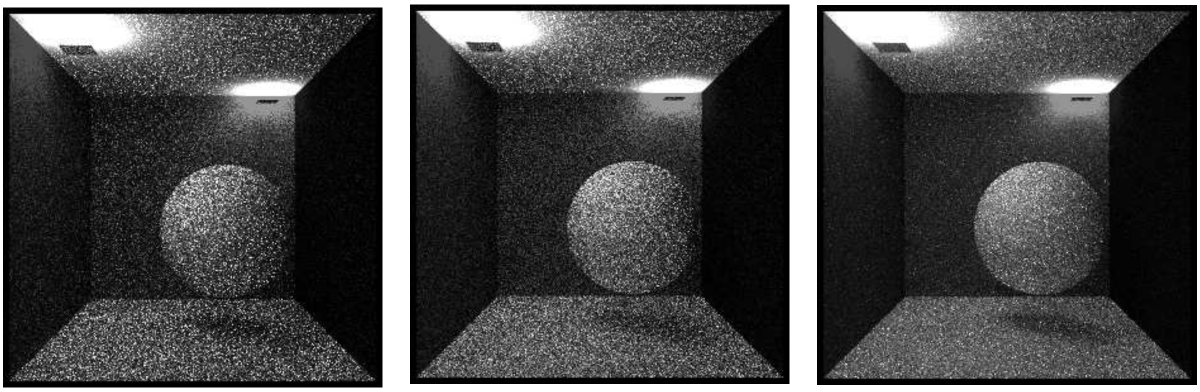
QUALITY OF ADAPTED PDFS

To evaluate the quality of the adapted pdfs we rendered a scene using different pdfs. To get a comparison with non-adaptive pdfs we generated images using uniform sampling (Fig. 3, left) and a pdf according to $f_r \cos(\theta)$ (Fig. 3, middle). Then our adaptive approach was used with the piecewise constant Haar basis (Fig. 3, right) and the linear B-spline basis (image omitted, only small visual difference to Haar basis). The adapted pdfs were generated by performing 3 passes with 10 samples per pixel before rendering the final image using 70 samples per pixel. The images show a visible decrease of noise when using adaptive pdfs.

pdf	L^1-error
uniform	0.0982
$f_r \cos(\theta)$	0.0812
adaptive, linear B-spline	0.0696
adaptive, Haar	0.0607

Table 1: Error measures for different pdfs

A radiosity solution for the scene was used as a reference to get an error measure of the images. The quality of a Monte Carlo integration is measured with the variance of the estimates. Variance is the squared expected error. Therefore we used the L^1-error norm, that is calculated as the mean value of the absolute pixel differences. The



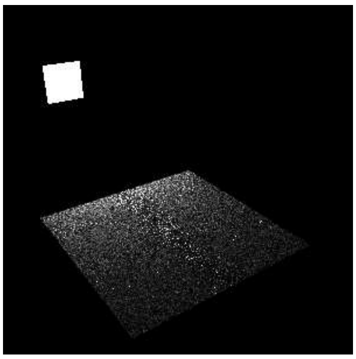
uniform $\qquad$ $f_r\ \cos(\theta)$ $\qquad$ adaptive, Haar basis

Figure 3: Scene rendered with different pdfs

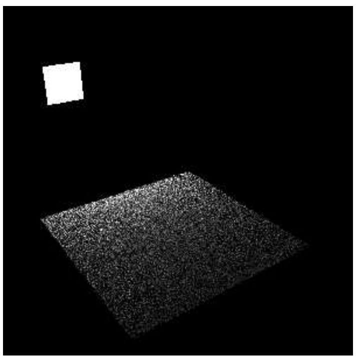
errors for the images generated with different pdfs are listened in Table 1. As we can see from these results the error decreases when using adaptive pdfs.

FUNCTION BASES

The results from the previous subsections suggest that there is no benefit from using higher order basis functions like B-splines instead of the piecewise constant Haar basis.



Haar basis



linear B-spline basis

Figure 4: Adaption to direct illumination

A small example can be used to prove that for some applications higher order basis functions can clearly increase the quality. In contrast to the previously described rendering algorithm the images shown in Fig. 4 were generated with pdfs adapted to direct illumination. When using the Haar basis noisy bands become visible on the ground patch that are eliminated through the use of the linear B-spline basis. We think these artifacts result from the discontinuities of the piecewise constant pdfs. To eliminate these artifacts when using the Haar basis a significantly higher number of samples is needed.

IMAGE QUALITY

Fig. 5 shows the error of images generated with different numbers of samples per pixel. The graph shows that the rendering algorithm using adaptive pdfs results in up to 17% smaller errors than the non-adaptive image generation. Best results are obtained using the Haar basis.

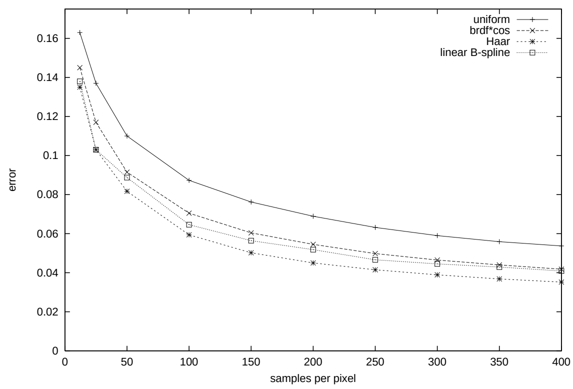


Figure 5: Error as a function of the number of samples per pixel

EFFICIENCY

Our previous results showed that the error decreases with adaptive pdfs for images generated using the same number of samples per pixel. But to evaluate the computational efficiency of the rendering algorithm we have to take the calculation time into account. One rendering method is better than the other if it produces better results in the same time or the same result in shorter time. Using adaptive importance sampling requires some overhead that only pays off if the increase in image quality is higher than the increase of calculation time.

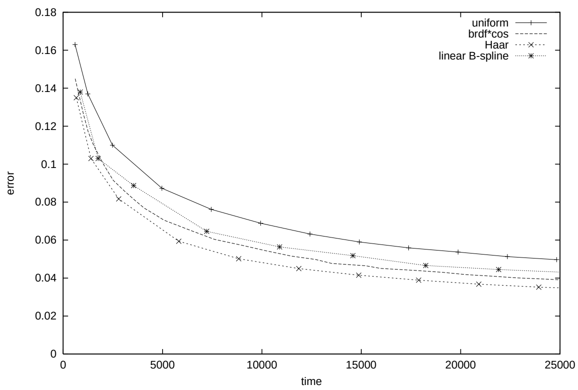


Figure 6: Error as a function of calculation time

Fig. 6 shows the error of images as a function of calculation time for different sampling strategies. As one can see the adaptive Monte Carlo sampling using the Haar basis yields for all calculation times better results than the non-adaptive techniques. Using pdfs according to $f_r \cos(\theta)$ (which is common for non-adaptive Monte Carlo ray tracing) needs up to 41 % more time than adaptive sampling using Haar wavelets.

CONCLUSION

We presented a new approach for Monte Carlo integration using adaptive probability density functions. Based on approximations of the integrand which are constructed from the samples of the Monte Carlo integration process, refined pdfs are obtained that lead to lower variance and thus higher efficiency of the integration calculations. We applied this approach to the calculation of indirect illumination in a backward ray tracing algorithm. We have shown that the same number of samples per pixel results in images revealing less error than images generated without adaptive pdfs. The use of adaptive pdfs adds some overhead due to the additional operations performed. It was possible to show that this overhead is more than compensated by the increase in image quality. That means our rendering algorithm is more efficient in the sense that it produces images with better quality in the same time as algorithms without adaptive pdfs.

Since our approach works with different function bases it was possible to investigate the question if using linear function bases instead of piecewise constant functions results in better quality. For a small example it could be shown that using the linear B-spline basis avoids typical artifacts that become visible when using the Haar basis. But the images of a larger scene showed no advantage of the higher order function basis.

A part of our ongoing research is to examine how the choice of parameters influences the efficiency of the rendering calculations. This includes investigating different function bases like higher order B-splines or tree wavelets, comparison of hierarchical and non-hierarchical function representations, and finding the optimal number of samples for the different passes.

An important task in the future is to test our new ray tracing approach with more complex scenes because as already stated in [LW95] the overhead of using adaptive pdfs is constant whereas the sample costs increase with scene complexity, so using adaptive pdfs becomes more worthwhile for complex scenes.

References

[AK90] James Arvo and David Kirk. Particle transport and image synthesis. *Computer Graphics*, 24(4):63–66, August 1990.

[Bus97] Eduardo Bustillo. A neuro-evolutionary unbiased global illumination algorithm. In Julie Dorsey and Philipp Slusallek, editors, *Eurographics Rendering Workshop 1997*, pages 263–274, New York City, NY, June 1997. Eurographics, Springer Wien.

[DW94] Philip Dutré and Yves D. Willems. Importance-driven monte carlo light tracing. In *Fifth Eurographics Workshop on Rendering*, pages 185–194, Darmstadt, Germany, June 1994.

[DW95] Philip Dutré and Yves D. Willems. Potential-driven monte carlo particle tracing for diffuse environments with adaptive probability density functions. In P. M. Hanrahan and W. Purgathofer, editors, *Rendering Techniques '95 (Proceedings of the Sixth Eurographics Workshop on Rendering)*, pages 306–315. Springer-Verlag, 1995.

[Jen95] Henrik Wann Jensen. Importance driven path tracing using the photon map. In P. M. Hanrahan and W. Purgathofer, editors, *Rendering Techniques '95 (Proceedings of the Sixth Eurographics Workshop on Rendering)*, pages 326–335. Springer-Verlag, New York, NY, 1995.

[Jen96] Henrik Wann Jensen. Global illumination using photon maps. In *Proceedings of the Seventh Eurographics Workshop on Rendering*, Porto, Portugal, June 1996.

[Kaj86] J. T. Kajiya. The rendering equation. *Computer Graphics*, 20(4):143–150, August 1986.

[KW86] M. Kalos and P. Whitlock. *Monte Carlo Methods.* Wiley & Sons, New York, 1986.

[Lep78] G. Lepage. A new algorithm for adaptive multidimensional integration. *Journal of Computational Physics*, 27:192–203, 1978.

[LW95] Eric P. Lafortune and Yves D. Willems. A 5D tree to reduce the variance of monte carlo ray tracing. In P. M. Hanrahan and W. Purgathofer, editors, *Rendering Techniques '95 (Proceedings of the Sixth Eurographics Workshop on Rendering)*, pages 11–20. Springer-Verlag, New York, NY, 1995.

[SDS96] Eric J. Stollnitz, Tony D. DeRose, and David H. Salesin. *Wavelets for Computer Graphics.* Morgan Kaufmann, 1996.

[UT97] Carlos Ureña and Juan C. Torres. Improved irradiance computation by importance sampling. In Julie Dorsey and Philipp Slusallek, editors, *Eurographics Rendering Workshop 1997*, pages 275–284, New York City, NY, June 1997. Eurographics, Springer Wien.

[VG97] Eric Veach and Leonidas J. Guibas. Metropolis light transport. *Computer Graphics*, 31(3A):65–76, August 1997.

APPENDIX FUNCTION APPROXIMATION

Let $N = \{N_j(x) \in L^2(0,1), j = 1,\ldots,n\}$ be a function basis and V_N be the space of all functions spanned by N. The proximum of a function $f \in L^2(0,1)$ in V_N is the function $\hat{f} \in V_N$ with

$$\hat{f} := \inf_{v \in V_N} \|f - v\|$$

This means that $\hat{f}$ is the function in V_N with minimal distance to f where the distance is measured by the L^2-norm

$$\|f\| := \sqrt{\langle f, f\rangle}$$

with the inner product

$$\langle f, g\rangle := \int_0^1 f(x)\, g(x)\, \mathrm{d}x \tag{5}$$

If N is an orthonormal basis (that means $\langle N_j, N_k\rangle = \delta_{jk}$ for $j,k = 1,\ldots,n$) the basis representation of $\hat{f}$ is calculated as

$$\hat{f}(x) = \sum_{j=1}^{n} \langle f, N_j\rangle N_j(x) \tag{6}$$

In the case of a non-orthonormal basis the set of dual basis functions $\tilde{N}_k$ (for which $\langle N_j, \tilde{N}_k\rangle = \delta_{jk}$, $j,k = 1,\ldots,n$ holds) has to be known. Then $\hat{f}$ is given by

$$\begin{aligned}\hat{f}(x) &= \sum_{j=1}^{n} \langle f, \tilde{N}_j\rangle N_j(x) \\ &= \sum_{j=1}^{n} \langle f, N_j\rangle \tilde{N}_j(x) \end{aligned} \tag{7}$$