Adaptive Progressive Photon Mapping

ANTON S. KAPLANYAN and CARSTEN DACHSBACHER
Karlsruhe Institute of Technology

This article introduces a novel *locally adaptive* progressive photon mapping technique which optimally balances noise and bias in rendered images to minimize the overall error. It is the result of an analysis of the radiance estimation in progressive photon mapping. As a first step, we establish a connection to the field of recursive estimation and regression in statistics and derive the optimal estimation parameters for the asymptotic convergence of existing approaches. Next, we show how to reformulate photon mapping as a spatial regression in the measurement equation of light transport. This reformulation allows us to derive a novel data-driven bandwidth selection technique for estimating a pixel's measurement. The proposed technique possesses attractive convergence properties with finite numbers of samples, which is important for progressive rendering, and it also provides better results for quasi-converged images. Our results show the practical benefits of using our adaptive method.

Categories and Subject Descriptors: I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism—*Raytracing*

General Terms: Algorithms, Experimentation, Performance, Theory

Additional Key Words and Phrases: Global illumination, photon mapping

ACM Reference Format:

Kaplanyan, A. S. and Dachsbacher, C. 2013. Adaptive progressive photon mapping. ACM Trans. Graph. 32, 2, Article 16 (April 2013), 13 pages.
DOI: http://dx.doi.org/10.1145/2451236.2451242

1. INTRODUCTION

Global illumination algorithms have experienced significant advances in recent years. We have now reached a point where predictive renderings can be achieved within reasonable time of minutes to hours, which allows for the use of global illumination in many fields including movie production and industrial and architectural design. Monte-Carlo techniques such as (bidirectional) path tracing [Lafortune and Willems 1993; Veach and Guibas 1994] or metropolis light transport [Veach 1998] compute unbiased solutions, however, they converge only slowly to noise-free images.

Photon mapping [Jensen 1996] traces and stores photons, and uses an on-surface estimation to obtain radiance values. By this estimation, it essentially trades the variance (visible as noise in the images) for bias (usually visible as overblurring of the illumination), which is a well-known "artifact" in the fields of estimation and regression in statistics [Silverman 1986]. However, this property makes photon mapping robust with complex illumination, such as specular-diffuse-specular paths, which are difficult (or even impossible) to sample with pure Monte-Carlo-based approaches. And in theory, the bias vanishes in the limit after tracing an infinite number of photons. The noise of an image generated with photon mapping usually vanishes faster compared to unbiased methods [Hachisuka et al. 2012], which is achieved by trading the variance for bias. Although in theory this bias could be removed by any consistent estimator in the limit, the problem of storing an infinite number of samples makes the bias elimination practically impossible.

Recently, Hachisuka et al. [2008] introduced *Progressive Photon Mapping* (PPM), which uses an iterative radiance estimation to merge the estimates from a series of photon maps, and thus avoids storing all traced photons at the same time. Similar to the original approach, it has low variance due to the reuse of light subpaths. In this article, we focus on the important question: what is the optimal reuse strategy for these subpaths? To this end we make the following contributions.

—We establish a connection between PPM and the field of recursive estimation in statistics. This enables us to employ techniques from a huge body of previous work from this field.

—Based on an asymptotic analysis of the class of PPM methods, we derive the optimal asymptotic convergence rate and the values for previously user-specified parameters in PPM.

—We further show that PPM can be expressed as a local regression over the pixel measurement equation.

—This generalized measurement estimation serves as a basis for an adaptive data-driven bandwidth (gather radius) selection, which automatically and locally balances bias and noise in order to minimize the total error of a rendered image.

—The results of our method show that it converges faster and has less noise and bias than existing PPM methods.

The article is structured as follows: the next section describes the most relevant previous work in both graphics and statistics. Section 3 introduces the background theory of progressive photon mapping. In Section 4 we establish a relation to existing recursive estimators in statistics and present an asymptotic analysis of PPM, leading to the optimal asymptotic convergence rate. In order to approach this convergence rate as closely as possible with a limited number of samples, an optimal bandwidth selection is necessary. To this end, we reformulate PPM as a spatial regression in path space in Section 5. Using this reformulation, we derive our novel adaptive bandwidth selection method in Section 6. Section 7 provides implementation details of our method. We present results in Section 8 which demonstrate the benefits of our adaptive method.

The research was supported by the Intel Visual Computing Institute, Saarbrücken, Germany.

Authors' addresses: A. S. Kaplanyan (corresponding author) and C. Dachsbacher, Karlsruhe Institute of Technology, Germany; email: anton.kaplanyan@kit.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.
© 2013 ACM 0730-0301/2013/04-ART16 $15.00
DOI: http://dx.doi.org/10.1145/2451236.2451242

2. PREVIOUS WORK

2.1 Photon Mapping, Bias and Variance Reduction

Shirley et al. [1995] were among the first who proposed to use density estimation for global illumination. Based on the original photon mapping technique [Jensen 1996], numerous publications focused on reducing the bias inherent to the method. Myszkowski [1997] proposed to use multiple radii during radiance estimation to estimate bias, and then selected the estimate that minimized an error heuristic. Walter [1998] reduced the boundary bias further using a locally weighted linear least squares regression together with a perceptual metric. Suykens and Willems [2000] control the query radii by reducing the photon density in unimportant and over-dense regions. Schregle [2003] proposed a bandwidth selection for radiance estimation based on error estimates extracted from the reconstructed irradiance. Ray maps [Havran et al. 2005] reduce bias by taking into account spatial and directional distribution of photons, while Wong and Wang [2005] use an orthogonal series estimator. Schjoth [2009] reduces the topological bias and bias at radiance discontinuities using a diffusion-based anisotropic estimator. Similarly, Spencer and Jones [2009] improve caustics rendering with photon relaxation. For participating media, Jakob et al. [2011] approximate a photon map with a hierarchical Gaussian mixture for volumetric radiance estimation. Recently, Vorba [2011] showed that the radiance estimation for different types of surfaces can be combined using multiple importance sampling in a bidirectional manner.

2.2 Progressive Photon Mapping

Hachisuka et al. [2008] introduced progressive photon mapping, which does not require storing all photons for radiance estimation at the same time. It has been generalized to handle distributed ray-tracing effects [Hachisuka and Jensen 2009], and the photon tracing has been improved by using adaptive mutation of visible photon paths [Hachisuka and Jensen 2011]. Belcour and Soler [2011] proposed to enrich emitted photons with the spectrum of the carried radiance to adaptively choose the radiance estimation radius. Knaus and Zwicker [2011] developed a probabilistic framework, studied the asymptotic behavior of bias and variance, proposed a memoryless estimator, and generalized PPM to participating media. Jarosz et al. [2011] combined PPM with photon beams for improved rendering of participating media. Hachisuka et al. [2010] describe an error estimation framework for bias and variance estimation and early rendering termination. However, their approach was demonstrated for a single-point error estimation and has not been used for an adaptive bandwidth selection. From our analysis we derive additional conditions that complement their theory (Section 6.3). In contrast to previous work, we analyze the bias and variance as well as convergence of PPM methods in general. And most importantly, we propose a method for a better bandwidth selection which speeds up the convergence for progressive rendering. Georgiev et al. [2011] investigate the combination of PPM and unbiased Monte-Carlo estimators. Independently and parallel with our work, Hachisuka et al. [2012] also propose to combine PPM with unbiased techniques and derive optimal values of some PPM parameters for the case of multiple importance sampling.

2.3 Adaptive Estimation

In kernel estimation methods, the smoothing radius of a kernel is called *bandwidth*, and the optimal choice thereof is a challenging problem. The development of *adaptive estimators* started in the 1960s, when Parzen [1962] showed the optimal rate of convergence and asymptotically optimal bandwidth reduction rate for the kernel estimator. Rudemo [1982] suggested the first cross-validation method, however, these methods usually undersmooth the data and are very sensitive to outliers. Silverman [1986] suggested a *rule of thumb* bandwidth selection, which assumes that the underlying function can be well approximated by a normal distribution. Quickly evolving fields, such as computational finance, artificial intelligence, robotics, machine learning, and computer vision, created much activity in this field in the 1990s. Novel methods for estimating the precision of sample statistics (medians, variances, percentiles) by using subsets of available data (*jackknifing*), or randomly drawing from a set of data points (*bootstrapping*) were developed. Popular methods include *wavelet shrinkage* [Donoho et al. 1993], an estimation technique based on the idea of thresholding the wavelet coefficients; or the *Intersection of Confidence Intervals* (ICI) [Katkovnik and Shmulevich 2002], which is based on the calculation of estimates of the function and their variances. Neural networks and support vector machines [Cortes and Vapnik 1995] can also be used to learn an unknown function. The ICI method has been recently used in graphics by Rousselle et al. [2011] to determine the optimal smoothing radius in screen space for BiDirectional Path Tracing (BDPT). However, due to using BDPT they inherit difficulties with complex specular paths.

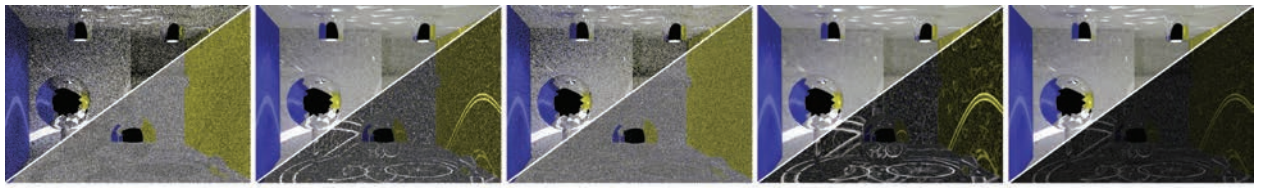
The *plug-in bandwidth selection* [Park and Marron 1990; Sheather and Jones 1991] for kernel estimation selects the optimal bandwidth by "plugging in" the estimate for higher-order derivatives of a function into the formula for the asymptotically optimal bandwidth. This approach inspired us to develop the adaptive progressive photon mapping method which we propose in this article.

Our work carries on the asymptotic analysis started by Knaus and Zwicker [2011] and the error estimation of Hachisuka et al. [2010]. We make use of adaptive recursive estimation methods [Jones and Sheather 1991; Hall and Patil 1994; Wand and Jones 1994] to develop a novel data-driven bandwidth selection for PPM.

3. PRELIMINARIES

Our work is inspired by progressive photon mapping [Hachisuka et al. 2008], its extension stochastic PPM [Hachisuka and Jensen 2009], and the probabilistic analysis of PPM [Knaus and Zwicker 2011]. These techniques are able to robustly sample difficult paths in light transport, for example, reflected caustics, without suffering the high memory requirements of the original photon mapping method. In this section, we will briefly recapitulate the most relevant aspects for our work.

The main issue in conventional photon mapping is that there is a trade-off between the variance, or noise, and the expected error, or bias, in the radiance estimate. A photon map is used to obtain an approximation of the true reflected radiance and one can either achieve a low variance or a low expected error, but not both. The main insight of progressive photon mapping is that we can obtain a solution with vanishing variance and expected error by averaging the results over many photon maps [Hachisuka et al. 2008], that is, progressive photon mapping is a multipass method that solves the rendering equation by accumulating statistics of photons. The initial eye pass traces rays from the camera and stores the nonspecular hit points. The following photon passes trace photons from light sources and update statistics on the hit points using these photons. The photon statistics include position of a hit point $\mathbf{x}$, the incident radiance times the BRDF $f(\mathbf{x};\omega)$, and the search radius r_N. Here, N is the number of the photon tracing passes and the number of eye samples. At the N-th pass, the radiance value at the position $\mathbf{x}$



(a) $\alpha = 0.5, k = 10$ (b) $\alpha = 0.5, k = 50$ (c) $\alpha = 0.8, k = 10$ (d) $\alpha = 0.8, k = 50$ (e) Adaptive PPM

Fig. 1. The Box scene rendered with standard PPM with different values of the user-specified parameters α and k (for the k-NN search to determine the initial bandwidth r_1). Result images (top left part) and differences to the converged images (bottom right part) are shown after emitting 10 million photons. Here our adaptive method (e) shows the benefits of minimizing the overall estimation error.

Table I. Notation Used Throughout This Article

Symbol	Description
N	Number of estimation iterations (eye subpaths)
J	Photon map size (emitted photons per iteration)
$M \equiv NJ$	Total number of emitted photons
$\mathbf{x}, \omega$	Surface point and outgoing direction
$\mathrm{Var}[X]$	Variance of expression X
$\mathrm{E}[X]$	Expected value of expression X
$\epsilon_N, \bar{\epsilon}_N$	One-step and average radiance estimation error at step N
$\mathbf{e}, \mathbf{l}$	Vertex of an eye subpath and light subpath
$I, \hat{I}_N$	Exact and estimated measurement of a pixel value
W	Importance (contribution of eye subpaths)
L	Outgoing radiance
γ_j	Contribution of the light subpath (i.e., photon) j
i	Eye subpath index
j	Light subpath index, photon index
r_N	Kernel radius (bandwidth) at step N
k, k_r	Canonical and normalized kernel with bandwidth r
Ψ	Contribution of all full paths through the estimation points
$\psi_{i,j}$	Contribution of eye subpath i with light subpath j

towards the direction ω is estimated as

$$\hat{L}(\mathbf{x}, \omega) = \frac{1}{J} \sum_{j=1}^{J} k_r(\mathbf{x}_j - \mathbf{x}) \gamma_j, \tag{3.1}$$

where the notation of variables is listed in Table I. This estimation is based on the contributions γ_j of all light subpaths around the point $\mathbf{x}$ applying a smoothing filter with kernel k_r.

Knaus and Zwicker [2011] rewrote this approximation as $\hat{L}(\mathbf{x}, \omega) = L(\mathbf{x}, \omega) + \epsilon$, where ϵ represents the error introduced by the radiance estimation. Then they obtain a complete pixel measurement estimate $\hat{I}_N$ as

$$\hat{I}_N = \frac{1}{N} \sum_{i=1}^{N} \frac{W_i}{p_{e,i}} \hat{L}(\mathbf{x}_i, \omega_i) \tag{3.2}$$

for which the expected value $\mathrm{E}[\hat{I}_N]$ converges to the true pixel value I and the variance $\mathrm{Var}[\hat{I}_N]$ approaches zero as the number of eye samples N increases. Here, $W_i \equiv W(\mathbf{x}_i, \omega_i)$ is the importance of the i-th eye subpath ending at the point $\mathbf{x}_i$ and $p_{e,i}$ is the probability of sampling this eye subpath. A crucial property of progressive photon mapping is that each sample is evaluated using a new photon map.

Knaus and Zwicker rederived PPM based on a more general probabilistic framework which allowed them to analyze the variance and the expected error of the radiance estimation in photon mapping from a probabilistic perspective.

They showed that if the variance and the expected value of the average error $\bar{\epsilon}_N$ both approach zero as $N \to \infty$ then the estimator is consistent.

$$\mathrm{Var}[\bar{\epsilon}_N] \to 0 \implies \mathrm{Var}[\hat{I}_N] \to 0$$
$$\mathrm{E}[\bar{\epsilon}_N] \to 0 \implies \mathrm{E}[\hat{I}_N] \to I$$

As a consequence, with each new sample the variance of the error ϵ_N at every iteration is allowed to increase by a factor

$$\frac{\mathrm{Var}[\epsilon_{N+1}]}{\mathrm{Var}[\epsilon_N]} = \frac{N+1}{N+\alpha}$$

for some constant α with $0 < \alpha < 1$[1]. This parameter α controls how quickly the variance is allowed to increase in each iteration. Knaus and Zwicker showed that α effectively determines a trade-off between vanishing variance and expected value of the average error. In summary, the variance and the expected value of the average error vanish as desired if $0 < \alpha < 1$, where α controls the relative rate of decrease. In addition, their asymptotic analysis showed that: (1) the variance vanishes to the order of $O(1/N^{\alpha})$, and (2) the expected value of the average error is proportional to the square radius and vanishes to the order of $O(1/N^{1-\alpha})$. That is, for $0 < \alpha < 1$ the squared radius decreases slower than the increase in number of photons. As a consequence, more and more photons are collected with increasing N, which guarantees $L(\mathbf{x}, \omega) = \lim_{N \to \infty} \hat{L}_N(\mathbf{x}, \omega)$. Note that this equation neither converges for $\alpha = 0$, because the initial radius will not be decreased, nor for $\alpha = 1$ as the noise level increases.

As we will see, the choice of the parameter α is crucial and directly affects the convergence rate of the whole rendering with PPM. Moreover, we will also show that the choice of the initial radius r_1 affects progressive rendering as well as the absolute error of an image.

4. PROGRESSIVE PHOTON MAPPING AS A RECURSIVE ESTIMATOR

In this section we analyze the radiance estimation in PPM and its asymptotic behavior. PPM methods comprise two user-specified parameters: (1) the constant $\alpha \in (0; 1)$ controls the bandwidth reduction, effectively balancing between variance and bias reduction, and (2) the initial per-pixel bandwidth r_1 which defines the initial "smoothing" of the radiance estimate. Figure 1 shows the influence of differently chosen values on the rendering convergence.

The choice of the proper value for the parameter α seemed to be nonintuitive, which can be observed from varying values given in previous works, for example, see Hachisuka et al. [2008] and Knaus and Zwicker [2011]. In this section, we analyze the progressive

[1]This is asymptotically equivalent to the initial radius reduction scheme by Hachisuka et al. [2008], but easier to analyze.

radiance estimate and show that a unique asymptotically optimal α value exists. This is the first step towards our new method, as we will see that for progressive rendering, that is, when displaying results prior to convergence, the selection of the second parameter, the bandwidth r_1, is of utmost importance (Section 6).

We begin by rewriting the radiance estimator used by Knaus and Zwicker [2011, Eq. (3.1) and Eq. (3.2)] as a recursive equation using the notation that we will use throughout this article. We have

$$\hat{I}_N = \frac{1}{N}\sum_{i=1}^{N}\frac{W_i}{p_{e,i}}\hat{L}(\mathbf{e}_i,\omega_i) = \frac{1}{N}\sum_{i=1}^{N}\frac{W_i}{p_{e,i}}\left(\frac{1}{J}\sum_{j=1}^{J}k_{r_i}\left(\mathbf{e}_i-\mathbf{l}_j^i\right)\gamma_j^i\right)$$
$$= \frac{N-1}{N}\hat{I}_{N-1} + \frac{1}{N}\frac{W_N}{p_{e,N}}\frac{1}{J}\sum_{j=1}^{J}k_{r_N}\left(\mathbf{e}_N-\mathbf{l}_j^N\right)\gamma_j^N, \tag{4.1}$$

where $\hat{L}(\mathbf{e}_i,\omega_i)$ is the reflected radiance at the point $\mathbf{e}_i$ and is approximated using kernel estimation (second equality); γ_j^i is the contribution of the j-th light subpath from the i-th photon map, landing at the point $\mathbf{l}_j^i$; all other variables are listed in Table I. The second line is obtained by regrouping and rearranging the terms, where the last summand is the contribution of the new photon map obtained from the step N.

The term $k_{r_N}(t) = \mu(r_N)^{-1}k(t/r_N)$ is the scaled and normalized version of the canonical kernel $k(t)$, and $\mu(r_N) = \pi r_N^2$ is the normalization factor for an isotropic, circular on-surface estimation. If the argument of $k(\cdot)$ is a vector, we assume its Euclidean length is used to evaluate the kernel. For the sake of consistency with the existing work in statistics, we call the radius r_N *bandwidth*. In PPM it shrinks with every iteration and converges to zero as $N\to\infty$. The kernel k should be continuous, bounded, and normalized [Parzen 1962]. Similarly to Hachisuka et al. [2010], we use the Perlin kernel $k(t) = 1+t^3(-6t^2+15t-10)$ [Perlin 2002] in our work. Note that the choice of the bandwidth r_N, but not the particular shape of the kernel, is important for the efficiency of the estimator [Silverman 1986].

PPM combines the results of photon maps using a shrinking bandwidth, controlled by the parameter α. Intuitively, the bandwidth trades both sources of error: bias and variance. In a first step, we will derive the asymptotically optimal value for α. However, as we will see, it only balances bias and variance in the limit. Thus our goal is to select the bandwidth for every pixel measurement and every iteration more precisely such that we can minimize the error on finite samples as well.

4.1 Recursive Estimation

Recursive estimators, such as the radiance estimator in Eq. (4.1), have been proposed by Wolverton and Wagner [1969] and Yamato [1971]. Such an estimator is *consistent*, that is, it yields unbiased results in the limit, if, for some $\lambda \geq 0$ and a kernel with at least two existing nonzero derivatives, the sequence of bandwidths $\{r_N\}$ obeys the following conditions in d dimensions [Yamato 1971; Hall and Patil 1994]:

$$\lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^{N}r_i^{-(d+\lambda)}\to 0, \qquad \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}r_i^2\to 0. \tag{4.2}$$

These two conditions guarantee that variance and bias, respectively, vanish in the limit. The parameter λ should be zero for estimating the function value [Hall and Patil 1994]. Later in Section 6.3 we will use the values $\lambda > 0$ for the estimation of the function derivatives. An equivalent result to Eq. (4.2) for $\lambda = 0$, that is, only for the estimation of the function value but not its derivatives, can be obtained from Eqs. (3), (5), (9), and (10) in Knaus and Zwicker's work [2011].

When updating r_N as in Knaus and Zwicker [2011] with

$$r_{N+1}^2 = r_N^2\frac{N+\alpha}{N+1}, \tag{4.3}$$

the conditions in Eq. (4.2) are met only for $\alpha \in (0;1)$, which is in accordance with previous PPM methods. In the following, we replace this sequence with a simpler, yet asymptotically equivalent one, $r_N = r_1 N^{(\alpha-1)/2}$ (see Appendix C). It simplifies the asymptotic analysis as it can be directly computed for any iteration N.

4.2 Asymptotic Convergence Rate of PPM

The preceding observations allow us to derive the asymptotic convergence rate for PPM and the asymptotically optimal value of the parameter α. Throughout the article we base our analysis on the minimization of the *Asymptotic Mean Squared Error* (AMSE). The *Mean Squared Error* (MSE) is the most widely used loss function and well-suited for analyzing the performance of estimators as it enables partitioning of variance and bias.

We show in Appendix B that, in terms of the MSE, the asymptotically optimal α for shrinking the bandwidth is

$$\alpha_{\text{opt}} = 2/3. \tag{4.4}$$

From Eqs. (C.1) and (4.4) we can obtain the asymptotically optimal shrinking rate of the bandwidth.

$$\boxed{r_N = r_1 N^{-1/6} \propto O(N^{-1/6})} \tag{4.5}$$

This result is important, as it shows that using α_{opt} for the bandwidth update yields the same asymptotically optimal rate that we find in statistics literature [Wand and Jones 1994].

Note that Knaus and Zwicker [2011] thoroughly studied the dependence of bias and variance on the parameter α *independently* of each other. They show that different α-values trade the vanishing rate between variance and bias. Yet α_{opt} balances these rates such that the MSE is asymptotically minimized, that is, variance and (squared) bias vanish with the same rate. Note that any nonoptimal α unbalances asymptotic bias and variance and slows the convergence rate.

It is also noteworthy that the optimal value of α and the asymptotic behavior of the sequence $\{r_N\}$ remain the same with sparse updates of the bandwidth, that is, when the bandwidth is updated once per J samples, as the conditions in Eq. (4.2) are still met. Consequently, the asymptotic convergence of PPM does not depend on the size of the photon map. However, in order to obtain consistent results independent of J, one can employ the compensation procedure proposed by Jarosz et al. [2011].

Convergence rate of progressive photon mapping. Substituting α_{opt} into the AMSE leads to the optimal convergence rate for PPM (see Appendix B for details).

$$\boxed{\text{AMSE}[\hat{I}_N] \propto O(N^{-2/3})} \tag{4.6}$$

Again, this result aligns with convergence rates found in existing literature for 2D estimators [Wand and Jones 1994; Silverman 1986]. Note that this is the *optimal theoretical asymptotic convergence rate* for photon mapping techniques using a density estimation on surfaces. An experimental measurement of the convergence rate with different values of α is shown in Figure 2.

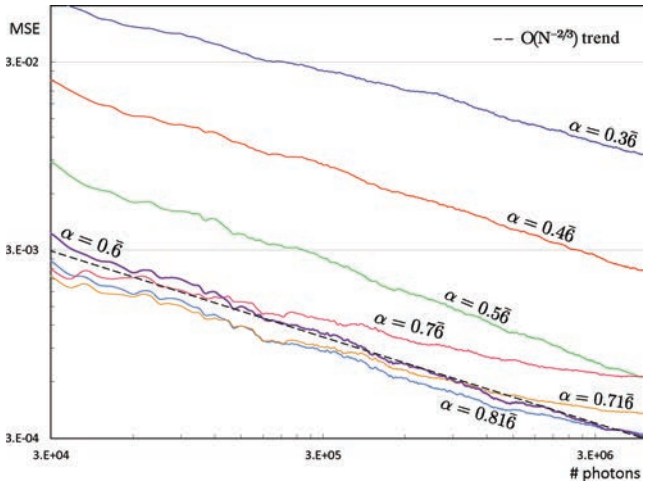


Fig. 2. Mean Squared Error (plotted in log-log scale) for the image shown in Figure 5(a) rendered with PPM with different values of α. The initial bandwidth is selected with a k-NN search with $k = 10$ and different values for α: $0.3\bar{6}/0.4\bar{6}/0.5\bar{6}/0.\bar{6}$ (optimal)$/0.71\bar{6}/0.7\bar{6}/0.81\bar{6}$. Note that the optimal convergence rate can be outperformed by suboptimal α-values depending on the choice of the initial bandwidth.

It is interesting to compare this result to the asymptotic convergence of unbiased Monte-Carlo techniques such as path tracing. The AMSE of these techniques (consisting of variance only) is $\mathrm{AMSE}[MC] = \mathrm{Var}[MC] = O(M^{-1})$, where M is the number of sampled full paths for a pixel. The theoretical asymptotic convergence rate of PPM is $O(N^{-2/3})$ and thus slower than that of unbiased Monte-Carlo techniques. This is because the probability of constructing a full path in PPM is also proportional to the kernel footprint πr_N^2, which shrinks as $O(N^{-1/3})$, thus reducing the probability of constructing a full path.

However, this does not mean that unbiased Monte-Carlo estimators should always be preferred due to their superior rate of convergence: there exist *undecidable* paths in path space, such that no algorithm can construct all of them. One example is pure SDS paths, for example, a reflected caustic seen from a pinhole camera where the light is coming from a point light source. In all unbiased Monte-Carlo techniques such paths require finding the path from one given point to another one. It is possible to find such paths numerically in simple cases [Mitchell and Hanrahan 1992], but this introduces bias and cannot construct these paths exactly. Note that generally there exist configurations of perfect specular surfaces where this was proven undecidable [Reif et al. 1994]. And this is not just a theoretical issue: almost all example scenes in this article contain such paths (for example those shown in Figure 7).

Finite number of samples. Obviously we are interested in an optimal result for a finite number of iterations and photons when rendering images. Most importantly, α_{opt} only defines the asymptotic shrinkage rate of the bandwidth, but not the optimal value for a particular (finite) number of iterations, as the initial bandwidth r_1 is an unknown in the bandwidth update (Eq. (4.5)). With finite numbers of samples it is very important to directly select the optimal bandwidth in every iteration and not rely on the asymptotic shrinkage rate. This can be observed from the results shown in Figure 2: due to a suboptimal initial bandwidth r_1 the optimal α-parameter outperforms others only when many samples are taken. In order to select the bandwidth more precisely, we reformulate PPM in the next section. In Section 6 we derive our new adaptive method which is asymptotically optimal as well, but also yields near-optimal results significantly faster. As we will see, selecting the bandwidth is of utmost importance.

5. PHOTON MAPPING AS A PATH-SPACE REGRESSION

In this section we express photon mapping as a local regression over the measurement equation in path space. Unlike the previous formulations of PPM, we express the kernel smoothing as a unified operator on top of the pixel measurement equation. The new formulation allows us to apply methods from statistics to derive our new adaptive estimator in Section 6. A reformulation is required, for example, to eliminate the weighted sum in Eq. (4.1), which prevents us from analyzing the same kernel bandwidth for different passes.

The measurement equation introduced by Veach [1998], generalized for paths consisting of eye and light subpaths of arbitrary lengths, can be written as

$$I = \int_{\mathcal{M}\times\mathcal{S}^2} W(\mathbf{x}, \omega) L_o(\mathbf{x}_{\mathcal{M}}(\mathbf{x}, \omega), -\omega)\, \mathrm{d}\mathbf{x}\mathrm{d}\omega, \tag{5.1}$$

where $\mathcal{S}^2$ is the unit sphere and $\mathcal{M}$ is the manifold of all scene surfaces in $\mathbb{R}^3$; $W(\mathbf{x}, \omega)$ is the importance of the eye subpaths of all lengths at $(\mathbf{x}, \omega)$; $L_o(\mathbf{x}_{\mathcal{M}}(\mathbf{x}, \omega), -\omega)$ is the outgoing radiance (caused by the light subpaths of all lengths) at the point $\mathbf{x}_o = \mathbf{x}_{\mathcal{M}}(\mathbf{x}, \omega)$, and $\mathbf{x}_{\mathcal{M}}(\mathbf{x}, \omega)$ is the ray-casting function. We denote incident values (in the direction of light flow) with the subindex i (e.g., $\mathbf{x}_i$, L_i), while the subscript o denotes outgoing values (e.g., ω_o, $\mathbf{x}_o$, L_o). The outgoing radiance at a surface point $\mathbf{x}_o$ in direction ω_o can be estimated [Pharr and Humphreys 2010] with

$$\begin{aligned} L_o(\mathbf{x}_o, \omega_o) &= \int_{\mathcal{S}^2} f(\mathbf{x}_o, \omega_o, \omega_i) \int_{\mathcal{M}} \delta(\mathbf{x}_o - \mathbf{x}_i)\, L_i(\mathbf{x}_i, \omega_i)\, \mathrm{d}\mathbf{x}_i \mathrm{d}\omega_i \\ &\approx \int_{\mathcal{S}^2} f(\mathbf{x}_o, \omega_o, \omega_i) \int_{\mathcal{M}} k(\mathbf{x}_o - \mathbf{x}_i)\, L_i(\mathbf{x}_i, \omega_i)\, \mathrm{d}\mathbf{x}_i \mathrm{d}\omega_i, \end{aligned} \tag{5.2}$$

where δ is the Dirac delta function, which we replace by the smoothing kernel $k(t)$. After inserting Eq. (5.2) into Eq. (5.1), we apply Fubini's theorem multiple times and obtain

$$\begin{aligned} I &\approx \int_{\mathcal{M}\times\mathcal{S}^2} W(\mathbf{x}_{\mathcal{M}}(\mathbf{x}_o, -\omega), \omega) \int_{\mathcal{S}^2} f(\mathbf{x}_o, -\omega, \omega_i) \\ &\qquad \int_{\mathcal{M}} k(\mathbf{x}_o - \mathbf{x}_i) L_i(\mathbf{x}_i, \omega_i)\, \mathrm{d}\mathbf{x}_i \mathrm{d}\omega_i \mathrm{d}\mathbf{x}_o \mathrm{d}\omega \\ &= \int_{\mathcal{M}\times\mathcal{M}} k(\mathbf{x}_o - \mathbf{x}_i) \int_{\mathcal{S}^2} W(\mathbf{x}_{\mathcal{M}}(\mathbf{x}_o, -\omega), \omega) \\ &\qquad \int_{\mathcal{S}^2} f(\mathbf{x}_o, -\omega, \omega_i) L_i(\mathbf{x}_i, \omega_i)\, \mathrm{d}\omega_i \mathrm{d}\omega \mathrm{d}\mathbf{x}_i \mathrm{d}\mathbf{x}_o \\ &= \int_{\mathcal{M}\times\mathcal{M}} k(\mathbf{x}_o - \mathbf{x}_i) \Psi(\mathbf{x}_o, \mathbf{x}_i)\, \mathrm{d}\mathbf{x}_i \mathrm{d}\mathbf{x}_o, \end{aligned} \tag{5.3}$$

where $\Psi(\mathbf{x}_o, \mathbf{x}_i)$ is the contribution of all possible full paths constructed by connecting the endpoint $\mathbf{x}_o$ of eye subpaths to the endpoint $\mathbf{x}_i$ of light subpaths. Such full paths are weighted by $k(\mathbf{x}_i - \mathbf{x}_o)$. Consider the example shown in Figure 3: the path $\mathbf{e}_0$(eye)–$\mathbf{e}_1$–$\mathbf{e}_2$–$\mathbf{l}_2$–$\mathbf{l}_1$–$\mathbf{l}_0$(light) is one possible path constructed in $\Psi(\mathbf{e}_2, \mathbf{l}_2)$ and weighted by $k(\mathbf{e}_2 - \mathbf{l}_2)$.

Let us now assume that we draw N eye subpaths from the importance distribution and J random light subpaths for each eye

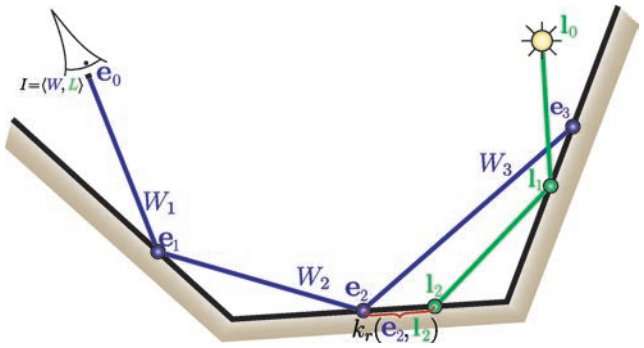


Fig. 3. Path-space estimation of the incident radiance in the measurement equation: and eye subpath, for example, eye $-\mathbf{e}_1-\mathbf{e}_2$ is connected to a light subpath light $-\mathbf{l}_1-\mathbf{l}_2$ using the kernel k_r.

subpath from the radiance distribution. Then we can estimate the measurement with

$$\hat{I}_N = \frac{1}{NJ}\sum_{i=1}^{N}\sum_{j=1}^{J} k_{r_i}\left(\mathbf{e}_i - \mathbf{l}_j^i\right)\psi_{i,j}, \tag{5.4}$$

which can be written in a recursive manner similarly to Eq. (4.1). Here, $\psi_{i,j}$ is the contribution of the full path constructed using a single eye subpath ending at $\mathbf{e}_i$ and a single light subpath ending at $\mathbf{l}_j^i$ (see Figure 3). The superscript i denotes that a new sample set (a photon map) $\{\mathbf{l}^i\}_J$ is used for every new eye subpath i.

This new formulation enables us to express the optimal bandwidth selection using the generalized kernel smoothing operator on top of pixel measurement as shown in the next section.

6. ADAPTIVE MEASUREMENT ESTIMATION

The kernel estimation in photon mapping provides lower variance at the price of introducing a systematic error (bias). The main goal of our work and the preceding derivations is to analyze this trade-off and to obtain the optimal strategy for balancing variance and bias at every step of the recursive estimation. In statistics, this problem is known as *bandwidth selection* [Silverman 1986; Jones and Sheather 1991], where the optimal bandwidth r is selected in order to minimize the estimation error.

Knaus and Zwicker [2011] studied several methods for determining the *initial* bandwidth r_1 (starting condition for Eq. (4.3)): (1) globally constant bandwidth, (2) local bandwidth based on k-Nearest-Neighbors (k-NN) search, and (3) local bandwidth based on ray differentials—and none of them leads to the optimal choice (see Knaus and Zwicker [2011] for discussion). Figure 7 shows an example where the result is still oversmoothed even after emitting 10^{11} photons. The reason is that with an improper initial bandwidth, the recursive estimator accumulates suboptimal results, while subsequent iterations (with smaller bandwidth) are weighted less and less; thus the error remains large even for large N.

It is also intuitive that the bandwidth selection should be locally adaptive to smooth the signal in sparse regions as well as preserve crisp radiance discontinuities. In the following, we derive the optimal bandwidth to improve the efficiency of the measurement estimation with a finite number of samples.

6.1 Asymptotic MSE of the Measurement

We derive our adaptive method by restarting from the AMSE for the pixel measurement estimator $\hat{I}$ (Eq. (5.4)). Our goal is a bandwidth selection procedure which minimizes the AMSE. The AMSE for $\hat{I}$ can be found in Appendix D where we show that it consists of the variance of the Monte-Carlo estimation inside the measurement equation and the bias and variance of the kernel estimator. However, we show that the latter variance is negligible and thus simplify the AMSE for our derivation by omitting it (see Appendix D for details):

$$\mathrm{AMSE}[\hat{I}_N] \approx \frac{1}{\pi p_l J}\,\mathrm{Var}[\psi]\,\frac{1}{N^2}\sum_{i=1}^{N} r_i^{-2} + \mathscr{B}[\hat{I}_N]^2, \tag{6.1}$$

where $\mathscr{B}[\hat{I}] = \mathrm{E}[\hat{I}] - I$ is the bias of the estimator; p_l is the average density of light subpaths in the proximity of all estimation points of the pixel measurement. The goal of our work is to minimize the AMSE by balancing between the measurement variance and the bias of the kernel estimation. Note that this is possible as the bandwidth also influences the Monte-Carlo variance of the path integral: the larger the bandwidth, the more full paths are constructed.

Using existing approximations (e.g., Hall and Patil [1994], Hachisuka et al. [2010], and Knaus and Zwicker [2011]) the asymptotic approximation of the bias is expressed as

$$\mathscr{B}\left[\hat{I}_N\right] \approx \frac{1}{2}k_2\Delta I\,\frac{1}{N}\sum_{i=1}^{N} r_i^2, \tag{6.2}$$

where ΔI is the Laplacian of the measurement with respect to the kernel estimation which we discuss in detail in Section 6.3; and $k_2 = \int_{\mathbb{R}} t^2 k(t)\mathrm{d}t$ is the kernel-dependent constant [Silverman 1986]. The exact value for the Perlin kernel can be found in Hachisuka et al. [2010]. Using Eqs. (6.1) and (6.2) we obtain

$$\mathrm{AMSE}[\hat{I}_N] \approx \frac{1}{\pi p_l J}\mathrm{Var}[\psi]\frac{1}{N^2}\sum_{i=1}^{N} r_i^{-2} + \frac{1}{4}k_2^2(\Delta I)^2\left(\frac{1}{N}\sum_{i=1}^{N} r_i^2\right)^2.$$

One can minimize this equation with respect to r_N, when assuming all other r_i to be constant (please see the online appendix in the ACM Digital Library.). However, the solution is computationally expensive and we will make further simplifying assumptions.

Simplified bandwidth selection. We make the assumption that the bandwidth r_N approaches the optimal value $\bar{r}_N$ quickly enough such that we can assume $\bar{r}_N \approx r_i \approx 1/N\sum_{i=1}^{N} r_i$. In other words, we assume that r_i quickly stabilizes and does not change significantly during later iterations (and thus the difference in error introduced by using $\bar{r}_N$ instead of r_i is negligible). Thus we replace all r_i in the previous equation by $\bar{r}_N$ and obtain the new AMSE′:

$$\mathrm{AMSE}'[\hat{I}_N] \approx \frac{1}{\pi p_l J N \bar{r}_N^2}\,\mathrm{Var}\,[\psi] + \frac{1}{4}k_2^2(\Delta I)^2\bar{r}_N^4. \tag{6.3}$$

Minimizing AMSE′ means that we always minimize the current error while using the previously estimated parameters for computing it. We provide an experimental validation of this approximation in the online appendix in the ACM Digital Library. After minimizing this equation with respect to $\bar{r}_N$, we obtain the asymptotically optimal bandwidth

$$\boxed{\bar{r}_N = \left(\frac{2\,\mathrm{Var}[\psi]}{\pi J p_l k_2^2(\Delta I)^2}\right)^{1/6} N^{-1/6},} \tag{6.4}$$

where the density p_l is estimated using a recursive estimator identical to Eq. (4.1), but without the actual photon values γ_j^i.

As the measurement estimator in Eq. (5.4) is a reformulation of PPM, it is not surprising that it has the same asymptotic convergence rate as the one we obtained earlier in Section 4.2. However, Eq. (6.4)

provides an explicit expression for bandwidth selection which will allow us to obtain faster convergence and better results for quasi-converged images.

The Laplacian ΔI in Eq. (6.4) is unknown, yet important for the bandwidth selection. To this end, we briefly discuss different bandwidth selection approaches and show how to estimate ΔI in Section 6.3.

6.2 The Plug-In Method for Bandwidth Selection

Many adaptive estimation techniques are known in literature (see Jones et al. [1996] for a survey), however, most of them are not suitable for our purposes as they either require to store a huge amount of data (e.g., Donoho et al. [1993] and Katkovnik and Shmulevich [2002]), or require access to all previous samples, for example, (biased) cross-validation and expectation-maximization methods. We decided to use the recursive modification [Hall and Patil 1994] of the *plug-in method* [Sheather and Jones 1991; Wand and Jones 1994]. It can be computed fully recursively and efficiently, has low variance, and good behavior for finite samples.

The main idea of the plug-in method is to plug an estimate $\Delta\hat{I}_N$ of the unknown Laplacian ΔI into the original estimator Eq. (6.4). This requires an estimation of the radiance derivatives on the surface. However, as noted by Jones and Sheather [1991], this requires selecting yet another bandwidth r' for the estimation of the derivatives, which is different from the optimal bandwidth r used for estimating the measurement itself. The necessity of selecting another optimal bandwidth seemingly leads to a nested hierarchy of such bandwidths for higher derivatives. However, as pointed out in previous work [Wand and Jones 1994; Jones et al. 1996], the second bandwidth r' is less critical. In fact, we can estimate it using a normal distribution fit to the estimated derivatives (the so-called rule of thumb), without causing a significant loss in convergence performance. In the nomenclature of Wand and Jones [1994], we use a simplified $\mathcal{S}-\mathcal{N}$ scheme: in the $\mathcal{N}$-step, we use a normal distribution fit to select the bandwidth r' for the estimation of the Laplacian. And the $\mathcal{S}$-step uses the plug-in method to select a single bandwidth r based on the estimated Laplacian.

6.3 Recursive Estimation of the Laplacian

In order to estimate the bias introduced by the kernel smoothing, it is necessary to estimate the curvature of the underlying function. In our case, the estimation is done by constructing full paths using a spatial kernel smoothing to connect eye subpaths to light subpaths. This means that smoothing happens around every vertex of every possible path of the measurement equation I. To this end, we use a two-step approach: first, we estimate the Laplacian at the surface at every path vertex of the measurement (denoted as *vertex-Laplacian*, $L_{\Omega(\mathbf{e}_i)}$, Figure 4). Next, we compute a sum of all estimated vertex-Laplacians weighted by the respective importance function. Throughout the article, we denote this as the *Laplacian of the pixel measurement* with respect to the on-surface incident radiance (ΔI in Figure 4). It determines the average curvature of the incident radiance at all path vertices of paths going through the pixel weighted by the respective importance. Thus, intuitively, it determines the flatness of the incident radiance for the measured pixel.

We compute the vertex-Laplacian $\Delta\hat{L}_{\mathbf{e}_N}$ by approximating the second derivatives along two orthogonal directions $\Delta\hat{L}_{\mathbf{e}_N} = \hat{L}^{(2)}_{\mathbf{e}_N,\mathbf{u}} + \hat{L}^{(2)}_{\mathbf{e}_N,\mathbf{v}}$, where $\mathbf{e}_N$ is the estimation point at the step N; $\mathbf{u}$ and $\mathbf{v}$ are two orthonormal basis vectors defined on the surface at the estimation point; and $\hat{L}^{(2)}_{\mathbf{e}_N,\mathbf{u}}$, $\hat{L}^{(2)}_{\mathbf{e}_N,\mathbf{v}}$ are the estimators for the second

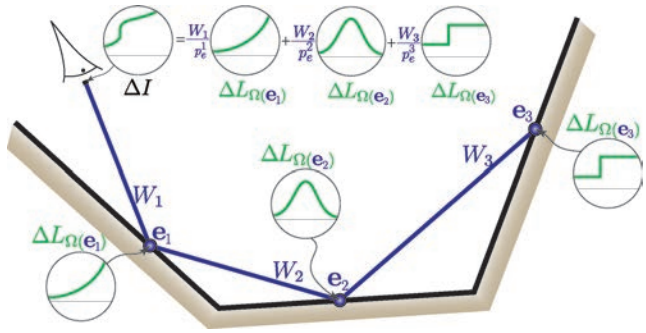


Fig. 4. The Laplacian of a pixel measurement is computed as the sum of the Laplacians of the on-surface incident radiance at the path vertices, $L_{\Omega(\mathbf{e}_i)}$, weighted by the respective importance W_i.

derivative along $\mathbf{u}$ and $\mathbf{v}$, respectively. Note that the vertex-Laplacian is invariant to the rotation of the basis $(\mathbf{u},\mathbf{v})$. Each derivative is estimated recursively using shrinking finite differences as proposed by Ngerng [2011]. Furthermore, we replace $\hat{L}_N$ with $\hat{I}_N$ and $\hat{L}^{(2)}_{\mathbf{e}_N}$ with $\hat{I}^{(2)}_{\mathbf{e}_N}$ after weighting by the importance, and denote the value of the measurement with radiance estimated at the point $\mathbf{e}_i + \mathbf{u}r'_i$ as $\hat{I}_{+\mathbf{u}r'_i}$. The equation is provided for $\hat{I}^{(2)}_{\mathbf{u},N}$ (and is analogous for $\hat{I}^{(2)}_{\mathbf{v},N}$).

$$\begin{aligned}\hat{I}^{(2)}_{\mathbf{u},N} &\approx \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{I}_{+\mathbf{u}r'_i} + \hat{I}_{-\mathbf{u}r'_i} - 2\hat{I}}{r'^2_i}\\ &= \frac{1}{NJ}\sum_{i=1}^{N}\sum_{j=1}^{J}\psi_{i,j}\frac{1}{r'^2_i}\Big(k_{r'_i}(\mathbf{e}_i-\mathbf{l}^i_j+\mathbf{u}r'_i)\\ &\quad + k_{r'_i}(\mathbf{e}_i-\mathbf{l}^i_j-\mathbf{u}r'_i) - 2k_{r'_i}(\mathbf{e}_i-\mathbf{l}^i_j)\Big)\end{aligned} \tag{6.5}$$

Consequently, we compute the Laplacian of the measurement as

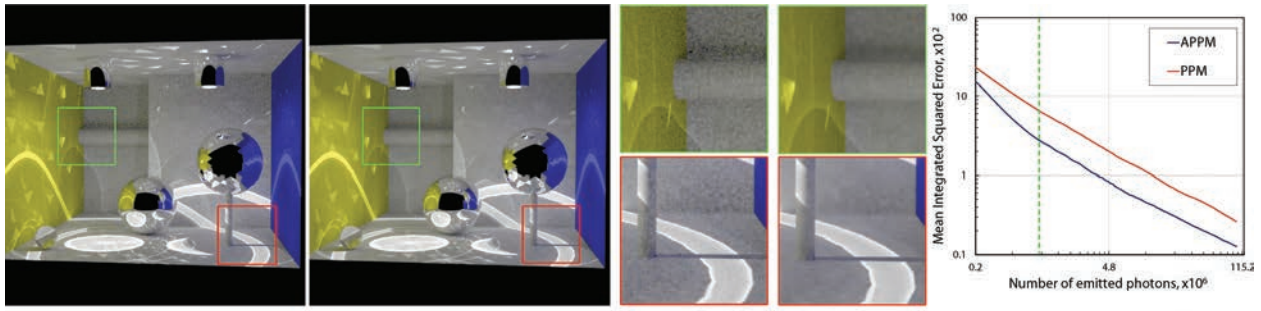
$$\Delta\hat{I}_N = \hat{I}^{(2)}_{N,\mathbf{u}} + \hat{I}^{(2)}_{N,\mathbf{v}}. \tag{6.6}$$

Radiance steps (discontinuities). The asymptotic behavior of the AMSE at discontinuities has a higher theoretical rate of convergence. Van Eeden [1985] showed that for a point on the discontinuity, the optimal bandwidth shrinkage rate is $O(N^{-1/2d})$, where d is the dimensionality of the estimator. Due to the robustness of the Laplacian estimator the discontinuities are handled automatically: the finite difference in the numerator of Eq. (6.5) becomes constant and the optimal shrinkage rate of r'_N leads to the growth rate of the Laplacian $O(N^{1/4})$. After substituting this rate into Eq. (6.4), one can validate that the asymptotic shrinkage rate of r_N becomes indeed $O(N^{-1/4})$.

Consistency of the derivative estimation. In general, for a consistent recursive estimation of the s-th-order derivative, the asymptotic bandwidth reduction rate of r' should obey Eq. (4.2) with $\lambda = 2s$ [Hall and Patil 1994]. This restricts the valid range for the power of the shrinkage rate of r' in $d=2$ dimensions to

$$O(N^{-1/(2s+2)}) < r'_N < O(N^0).$$

Consequently, for a consistent estimation of the second derivative the rate must be slower than $O(N^{-\frac{1}{6}})$, which is by coincidence equal to the optimal shrinking rate for the estimation bandwidth r. This means that we cannot use the same bandwidth for the estimation of



(a) progressive photon mapping (b) adaptive PPM (our method) (c) PPM (d) our method (e) mean squared error

Fig. 5. The Box scene is challenging for global illumination algorithms because of its specular-diffuse-specular paths. These images show renderings with depth-of-field using progressive photon mapping and adaptive progressive photon mapping (our method) using 20 million photons and 300 iterations. From left to right: (a) PPM with $\alpha = 2/3$ and $k_{\mathrm{NN}} = 10$; (b) adaptive PPM; (c) and (d) show close-ups; (e) plots of the L_2 loss (mean squared error) of (a) and (b). The dashed green line roughly separates two phases: (1) where the bandwidth adapts to the lighting in the scene, and (2) the equalization of the convergence rate.

both the function and its second derivative, otherwise the estimation of the latter diverges.[2]

Analogous to Section 4.2, we obtain the asymptotically optimal values by minimizing the AMSE of the second derivative estimator as in Wand and Jones [1994] (where r'_1 is some constant).

$$\boxed{r'_{N,\mathrm{opt}} = r'_1 N^{-1/8}}$$

Selecting the optimal bandwidth for the derivative estimation. In the $\mathcal{N}$-step of the $\mathcal{S}-\mathcal{N}$ scheme we estimate the bandwidth r'_N by applying Silverman's [1986] rule of thumb to the second derivative. That is, we assume that the actual function of the Laplacian ΔI can be approximated by a normal distribution and select the optimal bandwidth r' based on this assumption. For the 2D Perlin kernel the bandwidth is

$$r'_N \approx 1.9635\, \hat{\sigma}_N N^{-1/8}, \tag{6.7}$$

where $\hat{\sigma}_N = \sqrt{\mathrm{Var}[\Delta \hat{I}_N]}$ is the standard deviation of the Laplacian estimate; the variance can be estimated recursively by updating the first two moments of the derivative estimator.

6.4 Convergence Rate and Consistency

Our adaptive scheme does not change the asymptotic convergence rate of PPM (Eq. (4.6)). However, it provides better results with finite samples (Figure 6) compared to initial bandwidth selection used in previous work. Moreover, in contrast to our method, the bias in the previous PPM methods is still noticeable even in quasi-converged images after emitting 10^{11} photons (Figure 7). The reason is inherent to the recursive estimator: the contribution of the N-th iteration contributes to the solution with a weight of $1/N$, that is, an initially bad choice of the bandwidth can hardly be corrected. In contrast, our method corrects the bandwidth based on the input data at every iteration.

In order to ensure a robust behavior of our adaptive method, we bound the optimal bandwidth r_N (Eq. (6.4)) within the interval $[r_{\min} N^{-1/6}; r_{\max} N^{-1/6}]$, where $r_{\min}$ and $r_{\max}$ are (user-defined) minimum and maximum initial bandwidths. This guarantees convergence in regions where no reliable statistics are available, for example, in completely black areas of the image. Seemingly, this introduces two new parameters, however, these bounds are not crucial for the method and are reached very seldomly, usually in some dark places during initial iterations when the estimates are noisy. We found that setting $r_{\min}$ and $r_{\max}$ to $10^{-6} R_{\mathrm{scene}}$ and $10^{-1} R_{\mathrm{scene}}$, respectively, where R_{scene} is the bounding radius of the scene, works well for all scenes we have tested. In principle, if the bounds are too narrow, the benefits of APPM might be smaller, as there is less freedom to adjust the bandwidth; if the upper bound is too high, the performance might drop during the first iterations due to extensive queries to the photon map.

[2]This also shows that when using Eq. (4.3) for updating r' for second derivative estimation (as in Hachisuka et al. [2010]), the valid range for the α parameter is $(2/3; 1)$, otherwise the estimation diverges. The optimal value for the second derivative is $\alpha_{\mathrm{opt}} = 3/4$.

7. IMPLEMENTATION DETAILS

In this section we detail the differences of an APPM implementation compared to standard PPM. APPM requires 6 additional single precision floating point values per pixel to store $\mathrm{E}[\psi]$, $\mathrm{E}[\psi^2]$, $\Delta \hat{I}_N$, $\mathrm{E}[\Delta \hat{I}_N]$, $\mathrm{E}[\Delta \hat{I}_N^2]$, and $\hat{p}_{l,N}$. At the first iteration we initialize the values of r_1 and r'_1 with the maximum bandwidth $r_{\max}$ (Section 6.4). In the subsequent iterations, we update the estimates for these values and recompute the bandwidth r_N using Eq. (6.4). We use Eq. (6.5) for the estimation of the Laplacian ΔI, where the bandwidth r'_N is selected according to Silverman's rule of thumb (Eq. (6.7)).

The pseudocode shown in Algorithm 1 summarizes our bandwidth selection process. All variances are computed as a difference of the corresponding moments, that is, as $\mathrm{Var}[X] = \mathrm{E}[X^2] - E[X]^2$.

ALGORITHM 1: Adaptive bandwidth selection for a single pixel

```
Initialization stage (before the first frame)
  {r'_1, r_1} ← r_max        ▷ Both bandwidths start with maximum value
  {E[ψ], E[ψ²], E[ΔÎ_0], E[ΔÎ_0²], p̂_{l,0}} ← 0    ▷ Clear other values
Estimation pass N
  Generate photon map
  for each pixel do
    Trace eye subpath
    {E[ΔÎ_N], E[ΔÎ_N²]} ← Eq. 6.6 ▷ Update Laplacian's moments
    r'_N ← Eq. 6.7               ▷ Update Laplacian's bandwidth r'_N
    E[ψ], E[ψ²] ← Eq. 5.4 ▷ Update moments of contribution ψ
    p̂_{l,N} ← 1/(NJ) Σ Σ k_{r_i}(e_i − I^i_j) ▷ Update density p_l (Eq. 5.4 w/o ψ)
    r_N ← Eq. 6.4                ▷ Update PPM bandwidth r_N
    Obtain radiance estimate using bandwidth r_N
    Update pixel value
  end for
```

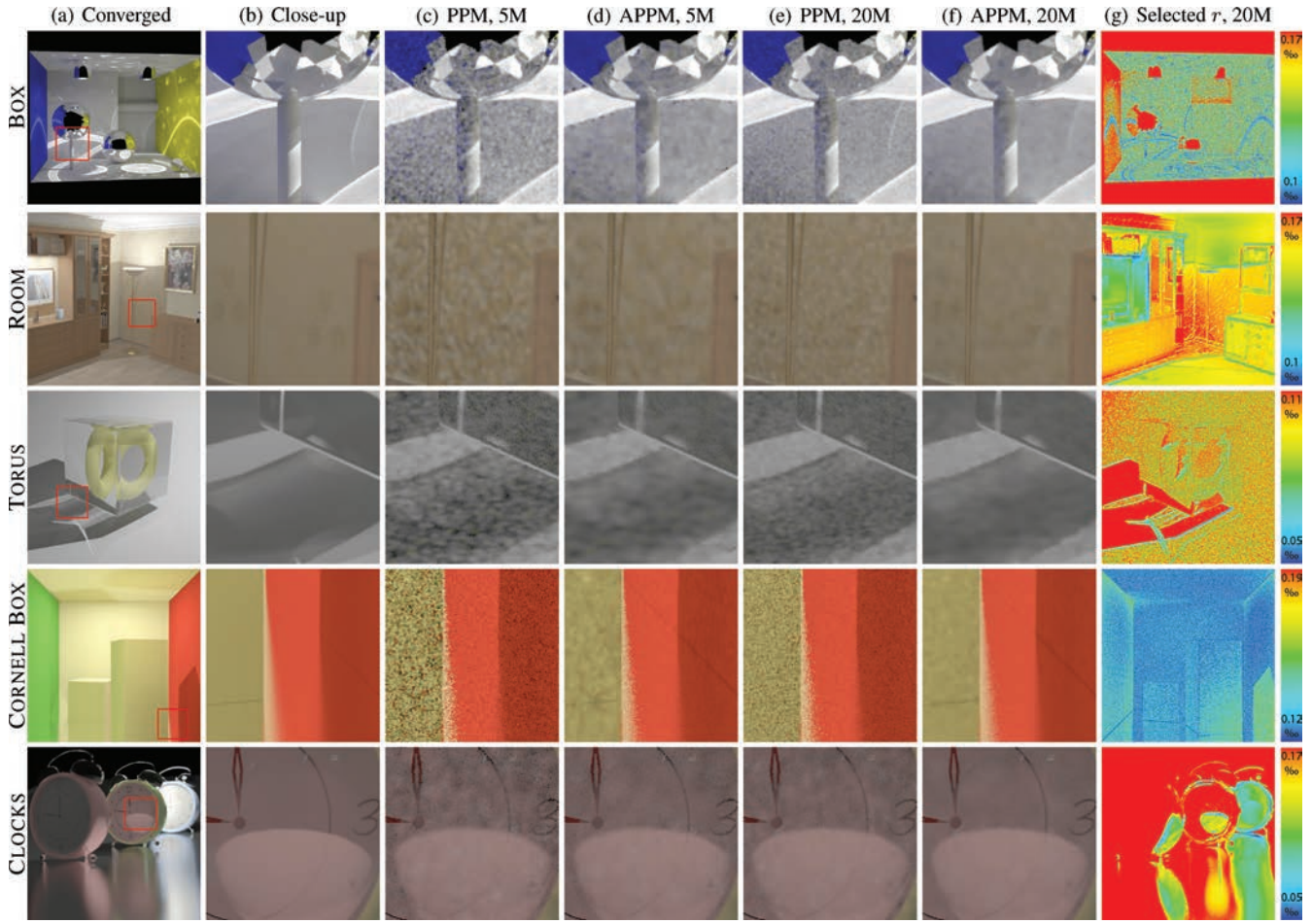


Fig. 6. We show a comparison of the progressive convergence for different scenes using the original PPM and our adaptive PPM method 5 million and 20 million emitted photons, respectively. The initial bandwidth for PPM has been selected using a k-NN search with $k = 10$ and the parameter $\alpha = \alpha_{\text{opt}} = 2/3$. The last column shows the selected bandwidth using the APPM method. All scenes are illuminated by point light sources and have perfectly reflective or refractive materials which makes it impossible to render them with existing unbiased techniques. The bandwidth range is specified in per mille with respect to the bounding radius of the corresponding scene.

In our implementation we perform the estimations for luminance only, however, this procedure can be trivially extended to the estimation of an (r, g, b)-triplet or spectral distribution. In our experiments the additional computation for our method in all presented scenes added an overhead of 2% to 7% (see Table II). Knaus and Zwicker [2011] also pointed out that local per-pixel statistics have only minimal impact on performance. The measured performance loss and memory cost of storing the statistics per pixel proved to be dominated by ray tracing costs even with simple scenes.

8. RESULTS AND DISCUSSION

In this section, we discuss the results of APPM and provide a comparison to standard PPM. In our implementation we use NVIDIA OptiX [Parker et al. 2010] as a GPU-accelerated ray tracing platform and all measurements were taken on an Nvidia GTX 570 GPU. All images in the article were rendered at a resolution of 768^2. In all tests we used a photon map of size $J = 256^2$ (the total number of photons shot per iteration). Table II shows performance measurements and MSE for our test scenes.

We can observe that the selected bandwidth is quite noisy in some scenes (e.g., in the BOX and TORUS scenes in Figure 6), however, this does not lead to a noisier image. The reason is that this noise mostly appears in uniformly lit regions and is caused by noisy statistical data. Thus the bandwidth might fluctuate around the optimal value, but is increased to capture more statistics, that is, more photons which also result in smooth estimates.

Figure 7 shows a comparison of the bias in quasi-converged images. We can observe that the initial radius selection results in visible bias in PPM even after emitting 100 billion photons in total. This is due to the weakness of PPM that the initial bias vanishes very slowly. This is a well-known disadvantage of the recursive estimator (Eq. (4.1)) where the contribution of later iterations is weighted by the quickly vanishing term $1/N$. Thus the convergence slows down significantly when drawing a large number of samples. Although theoretically the bias is eliminated in the limit, practically this does not happen due to the finite number of rendering iterations and finite precision of the rendering pipeline.

The close-ups in Figure 6 clearly show more noise for PPM than for APPM. However, in some regions smaller kernel bandwidths are

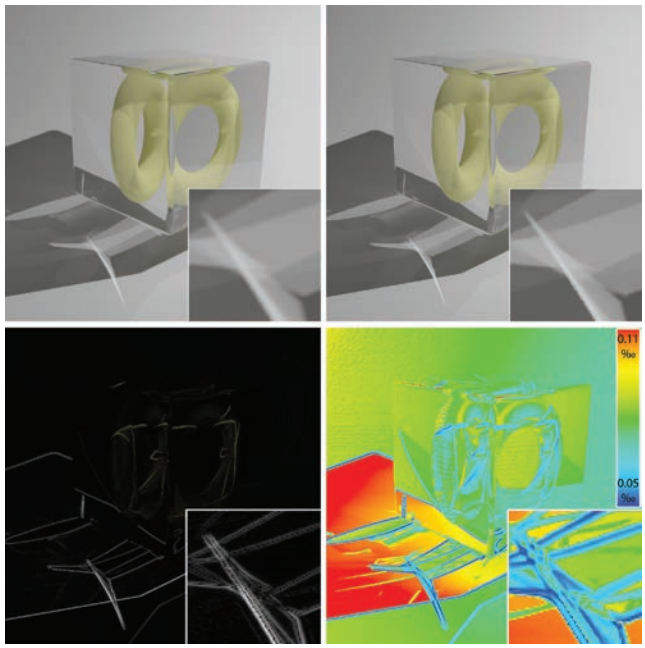


Fig. 7. The TORUS scene illuminated with a single point light source, rendered with 100 billion emitted photons. For standard PPM (top-left) the initial bandwidth was chosen using the k-NN search with $k = 10$, and $\alpha = 2/3$. Top-right: adaptive PPM; bottom-left: the absolute difference to PPM, scaled by a factor of 20—note that the bias is still clearly visible for standard PPM; bottom-right: the selected bandwidth using APPM.

selected (e.g., close to the caustic edges in the BOX scene) and hence APPM exhibits more high-frequency noise along radiance discontinuities as fewer photons are considered for the estimation. This is a consequence of selecting the bandwidth such that it minimizes the AMSE. As this is obviously not a perceptual metric, two images with the same error computed with PPM and APPM cannot be considered as equally visually good (please see the additional material for a comparison using a perceptual visible differences predictor). We believe that adaptive photon shooting [Hachisuka and Jensen 2011] could augment APPM here to make an image converge more uniformly.

The benefit of APPM for progressive results can also be observed in Figure 6. APPM achieves results with significantly less noise and reduced blur compared to standard PPM. This is because the bandwidth is chosen per pixel such that bias and variance are balanced and thus the MSE is minimized in every iteration. This leads to a faster convergence in regions with flat or very smooth lighting where the Laplacian of the pixel measurement is low.

9. LIMITATIONS AND FUTURE WORK

One limitation of our method is the assumption that there is a single optimal bandwidth for a pixel. As shown in Eq. (5.3), the measurement can be represented as a weighted sum of different modes (weighted by the importance of the eye subpaths). As a potentially difficult case, imagine some dim caustics on top of some brightly and uniformly lit diffuse surface: the optimal bandwidth would mostly be selected to ensure a good convergence of the diffuse illumination, thus making the caustics blurrier. However, according to the experience of many practitioners of the plug-in method, it can gracefully handle complex and even discontinuous multimodal distributions. Our experience with the adaptive bandwidth selector (based on the plug-in method) is similar and can also be observed from the results, for example, the illumination on the alarm clock glass in Figure 6.

Table II. Performance and Error of PPM and Our Method after 5 Million Emitted Photons

	Time, sec		MSE, $\times 10^2$	
Scene	PPM	Our method	PPM	Our method
BOX	10.14	10.18	1.91	0.79
CLOCKS	12.13	12.76	6.17	1.76
CORNELL	13.49	13.57	0.032	0.004
ROOM	20.73	21.34	4.61	1.12
TORUS	6.32	6.76	0.073	0.016

The MSE is computed against the reference solution, which is obtained from the original PPM with large number of photon passes as in Hachisuka and Jensen [2009] and the global bandwidth smaller than the pixel footprint.

Another limitation is that regions with discontinuities might appear noisier than with the original PPM method (see Figure 6, Box-scene for APPM with 5 million photons). This is because a small bandwidth is selected due to a large bias. Even though such regions converge quickly even with moderate numbers of photons (Figure 6, APPM with 20 million photons), we believe that photons should ideally be shot adaptively, for example, in spirit of Hachisuka et al. [2011] in order to provide better convergence; combining both algorithms would be an interesting future work.

Although the provided APPM derivation is demonstrated only for on-surface estimation, it is trivial to show (using a procedure similar to Appendix B) that the optimal bandwidth for volumetric radiance estimation [Knaus and Zwicker 2011] is $r_{N,\text{vol}} \propto O(N^{-1/7})$ and $AMSE_{\text{vol}} \propto O(N^{-4/7})$. For progressive photon beams [Jarosz et al. 2011] the optimal shrinkage rate can be similarly obtained as $r_{N,\text{ppb}} \propto O(N^{-1/5})$ and the convergence rate as $AMSE_{\text{ppb}} \propto O(N^{-4/5})$. Interestingly, the convergence rate for progressive photon beams is faster than the one of progressive photon mapping, which means that the participating media rendering with such transport has faster convergence (given the same number of samples) than the surface rendering with PPM.

Even though the adaptive method can be similarly derived for volumetric phenomena, the *curse of dimensionality* comes into play and influences the adaptivity more than for on-surface estimation. We believe that a thorough analysis of every such method is required, which we leave for the future work.

Similar to other PPM methods, we currently assume to have only one eye subpath for a single photon map. An interesting question is to what degree we can reuse a single photon map with different eye subpaths, as dropping the photon map after performing an estimation for a single eye subpath is not always the most efficient strategy. This is especially interesting for scenes with glossy surfaces or out-of-focus regions. Both PPM and APPM can be trivially modified to reuse a photon map for several eye subpaths, however, this ratio could be adaptively estimated by means of statistical methods.

We believe that our work opens a valley for new research in the field of path-space kernel estimation. As future work, it would be interesting to use more elaborate bandwidth selection procedures, for example, the $\mathcal{D}-\mathcal{S}-\mathcal{N}$ scheme [Wand and Jones 1994], which would lead to an anisotropic bandwidth selection with a matrix of bandwidths. Such an improvement can significantly help in the regions with highly anisotropic signal in image space, such as heavily motion-blurred objects.

Another improvement would be to combine different estimators of the measurement integral using multiple importance sampling

as proposed by Vorba [2011], while using our method to select the bandwidth either for the weighted sum of these estimators or for each estimator separately.

Alternative estimation techniques, such as memory-optimized wavelet estimation or neural networks, are also an interesting direction. The use of perceptual metrics instead of the MSE would be another important and promising step towards more efficient bandwidth selection criterion. Lastly, our estimation technique can also be used for early rendering termination in spirit of Hachisuka et al. [2010], as well as for adaptive sampling of eye and light subpaths.

10. CONCLUSIONS

In this article we deduced the optimal convergence rate and consistency conditions for progressive photon mapping techniques. We presented a data-driven adaptive measurement estimation technique, which requires no crucial user-specified parameters. Moreover, we showed how, compared to previously existing techniques, the convergence can be sped up on finite numbers of photons, and how quasi-converged images of better quality can be obtained.

APPENDIX

In this appendix we use the same notation as Knaus and Zwicker [2011]; the initial equations used for the asymptotic analysis are taken from this article. Note that all approximations in this appendix are done in the scope of an *asymptotic analysis*.

A. ASYMPTOTIC VARIANCE AND EXPECTATION OF THE ESTIMATION ERROR

We can simplify the asymptotic variance of the estimation error (based on two last equations in Appendix E of Knaus and Zwicker [2011]). We have

$$\begin{aligned}\mathrm{Var}\,[\bar{\epsilon}_N] &\approx \frac{\mathrm{Var}[\epsilon_1]}{N^2}\left(1+\sum_{i=2}^{N} i\alpha B(\alpha,i)\right)\\ &= \frac{\mathrm{Var}[\epsilon_1]}{N^2}\left(1+\frac{2\Gamma(N+\alpha)-\alpha\Gamma(\alpha)\Gamma(N+2)}{(\alpha-2)\Gamma(N+\alpha)}\right)\\ &\approx \frac{\mathrm{Var}[\epsilon_1]}{N^2}\frac{\alpha\Gamma(\alpha)\Gamma(N)(N+1)(N+2)}{(2-\alpha)\Gamma(N+\alpha)}\\ &\approx \frac{\mathrm{Var}[\epsilon_1]}{N^2}\frac{\alpha B(\alpha,N)}{2-\alpha} \approx \mathrm{Var}[\epsilon_1]\frac{\alpha\Gamma(\alpha)N^{-\alpha}}{2-\alpha}\\ &\approx \frac{\mathrm{Var}[\epsilon_1]}{(2-\alpha)N^{\alpha}},\end{aligned}$$

where $B(\cdot)$ and $\Gamma(\cdot)$ are the Beta and the Gamma functions, respectively. The expected estimation error can also be simplified (based on last equation, second line in Appendix F of Knaus and Zwicker [2011]).

$$\begin{aligned}\mathrm{E}[\bar{\epsilon}_N] &= \frac{\mathrm{E}[\epsilon_1]}{N}\left(\frac{\alpha+N}{B(\alpha,N)\alpha^2 N}-\frac{1}{\alpha}\right)\\ &\approx \frac{\mathrm{E}[\epsilon_1]}{N}\frac{\alpha+N}{\Gamma(\alpha)N^{-\alpha}\alpha^2 N} \approx \frac{\mathrm{E}[\epsilon_1]}{\alpha^2\Gamma(\alpha)N^{1-\alpha}} \approx \frac{\mathrm{E}[\epsilon_1]}{\alpha N^{1-\alpha}}\end{aligned}$$

B. MINIMIZING THE MEAN SQUARED ERROR OF THE MEASUREMENT

The *asymptotic* mean squared error can be approximated using the previous observations and results for variance and expectation of the measurement from Knaus and Zwicker [2011, Appendices C and D].

$$\begin{aligned}\mathrm{AMSE}\,[\hat{I}_N] = \mathrm{Var}[\hat{I}_N] + \mathscr{B}[\hat{I}_N]^2 &\approx \frac{1}{N}\,\mathrm{Var}\left[\frac{W}{p_e}L\right]\\ &+ \mathrm{E}\left[\frac{W}{p_e}\right]^2\left(\frac{\mathrm{Var}\,[\epsilon_1]}{(2-\alpha)N^{\alpha}}+\frac{E\,[\epsilon_1]^2}{\alpha^2 N^{2-2\alpha}}\right)\end{aligned} \tag{B.1}$$

After equating the first derivative to zero and performing asymptotic simplifications we obtain

$$N^{3\alpha-2} = \frac{\alpha^2\,\mathrm{Var}\,[\epsilon_1]}{2(2-\alpha)\,\mathrm{E}[\epsilon_1]^2}.$$

The outcome of this result is twofold: first, as the right part is finite, the only local minimum with respect to α is

$$\alpha_{\mathrm{opt}} = 2/3.$$

Second, if we insert α_{opt} back into the AMSE we obtain

$$\begin{aligned}\mathrm{AMSE}\,[\hat{I}_N] &\approx \frac{1}{N}\,\mathrm{Var}\left[\frac{W}{p_e}L\right]+\frac{3}{4N^{\frac{2}{3}}}\,\mathrm{E}\left[\frac{W}{p_e}\right]^2\left(3\,\mathrm{E}\,[\epsilon_1]^2+\mathrm{Var}\,[\epsilon_1]\right)\\ &\propto O\left(N^{-\frac{2}{3}}\right).\end{aligned} \tag{B.2}$$

C. ASYMPTOTIC BEHAVIOR OF THE PPM BANDWIDTH SEQUENCE

Both original PPM methods [Hachisuka et al. 2008; Knaus and Zwicker 2011] use an asymptotically equivalent formula for bandwidth shrinkage.

$$r_{N+1}^2 = r_N^2\frac{N+\alpha}{N+1} = r_N^2\left(1+\frac{\alpha-1}{N+1}\right)$$

It can be replaced with a simplified yet asymptotically equivalent explicit formula which allows for an easier analysis.

To this end, we set $t_N \equiv r_N^2$ and solve the asymptotic equation as $N\to\infty$.

$$t_{N+1}-t_N = t_N\frac{\alpha-1}{N+1} \Leftrightarrow t'(N) = t(N)\frac{\alpha-1}{N+1} \Rightarrow t(N) = c(N+1)^{\alpha-1}$$

for some constant c. Thus the new sequence $\{r_N\}$ ($r_N = \sqrt{t_N}$) is

$$r_N = r_1 N^{\frac{\alpha-1}{2}} \propto O\left(N^{\frac{\alpha-1}{2}}\right). \tag{C.1}$$

D. VARIANCE OF THE AMSE

We begin with the AMSE as in Appendix B.

$$\mathrm{AMSE}\,[\hat{I}_N] \approx \mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{W}{p_e}L\right]+\mathrm{E}\left[\frac{W}{p_e}\right]^2\left(\mathrm{E}\,[\bar{\epsilon}_N]^2+\mathrm{Var}\,[\bar{\epsilon}_N]\right)$$

The term L in the first summand is the radiance at a single iteration i. If we denote as S_i the average number of full paths for the Monte-Carlo integration of the pixel measurement constructed at the i-th iteration, then the first summand of the AMSE becomes

$$\mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{W}{p_e}L\right] = \mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{W}{p_e}\frac{1}{S_i}\sum_{j=1}^{S_i}\gamma_j^i\right] = \mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{S_i}\sum_{j=1}^{S_i}\psi_{i,j}\right].$$

The samples $\psi_{i,j}$ are independent and identically distributed and thus we apply the Bienaymé formula. Also the number of samples S_i at the i-th iteration can be approximated as $S_i \approx \pi r_i^2 J p_l$ where p_l

is the mean density of light subpaths (photons) around all estimation points. By this we obtain

$$\mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{S_i}\sum_{j=1}^{S_i}\psi_{i,j}\right]\approx\frac{1}{\pi p_l J}\,\mathrm{Var}\left[\psi\right]\frac{1}{N^2}\sum_{i=1}^{N}r_i^{-2}.$$

The variance $\mathrm{Var}[\bar{\epsilon}_N]$ can be expressed using Appendix B from Knaus and Zwicker [2011], and by using Hall and Patil [1994] we obtain

$$\mathrm{Var}\left[\bar{\epsilon}_N\right]\approx\left(\mathrm{Var}\left[\gamma\right]+\mathrm{E}[\gamma]^2\right)\frac{p_l k_3}{J}\frac{1}{N^2}\sum_{i=1}^{N}r_i^{-2},$$

where $k_3=\int_{\mathbb{R}}k(t)^2\mathrm{d}t$. Thus the variance of the AMSE becomes

$$\left(\aleph p_l k_3+\frac{1}{\pi p_l}\,\mathrm{Var}\left[\psi\right]\right)\frac{1}{JN^2}\sum_{i=1}^{N}r_i^{-2}\approx\frac{1}{\pi p_l J}\mathrm{Var}\left[\psi\right]\frac{1}{N^2}\sum_{i=1}^{N}r_i^{-2},$$

where $\aleph=\mathrm{E}[W/p_e](\mathrm{Var}\left[\gamma\right]+\mathrm{E}[\gamma]^2)$. The last step is based on the assumption (in spirit of Knaus and Zwicker [2011]) that $1/p_l\gg p_l$. Using the equality $\mathrm{E}[\frac{W}{p_e}]\,\mathrm{E}[\bar{\epsilon}_N]=\mathrm{E}[\frac{W}{p_e}\bar{\epsilon}_N]=\mathscr{B}[\hat{I}_N]$, we get

$$\mathrm{AMSE}\left[\hat{I}_N\right]\approx\frac{1}{\pi p_l J}\,\mathrm{Var}\left[\psi\right]\frac{1}{N^2}\sum_{i=1}^{N}r_i^{-2}+\mathscr{B}[\hat{I}_N]^2.$$

From this we can observe that the variance of the AMSE is dominated by the Monte-Carlo integration of the measurement.

ELECTRONIC APPENDIX

The electronic appendix to this article is available in the ACM Digital Library.

ACKNOWLEDGMENTS

We would like to thank László Györfi and Harro Walk for discussions on the state-of-the-art recursive estimators and bandwidth selectors, Peter-Pike Sloan and Peter Shirley for proof-reading. We are grateful to Toshiya Hachisuka for sharing the Box, the Torus, and the Clocks scenes (Figure 6) with the community. Finally, we would like to thank the anonymous reviewers for their insightful comments and suggestions.

REFERENCES

Belcour, L. and Soler, C. 2011. Frequency-Based kernel estimation for progressive photon mapping. In A*CM SIGGRAPH Asia Poster Program*.

Cortes, C. and Vapnik, V. 1995. Support-Vector networks. *Mach. Learn. 20*, 3, 273–297.

Donoho, D., Johnstone, I., and Johnstone, I. M. 1993. Ideal spatial adaptation by wavelet shrinkage. *Biometrika 81*, 425–455.

Georgiev, I., Krivanek, J., and Slusallek, P. 2011. Boidirectional light transport with vertex merging. In *SIGGRAPH Asia Sketches.* Vol. 27, 27:1–27:2.

Hachisuka, T., Jarosz, W., and Jensen, H. W. 2010. A progressive error estimation framework for photon density estimation. *ACM Trans. Graph. 29*, 6, 144:1–144:12.

Hachisuka, T. and Jensen, H. W. 2009. Stochastic progressive photon mapping. *ACM Trans. Graph. 28*, 5, 141:1–141:8.

Hachisuka, T. and Jensen, H. W. 2011. Robust adaptive photon tracing using photon path visibility. *ACM Trans. Graph. 30*, 5, 114:1–114:11.

Hachisuka, T. Ogaki, S., and Jensen, H. W. 2008. Progressive photon mapping. *ACM Trans. Graph. 27*, 5, 130:1–130:8.

Hackisuka, T., Pantaleoni, J., and Jensen, H. W. 2012. A path space extension for robust light transport simulation. *ACM Trans. Graph. 31*.

Hall, P. and Patil, P. 1994. On the efficiency of on-line density estimators. *IEEE Trans. Inf. Theory 40*, 5, 1504–1512.

Havran, V., Bittner, J., Herzog, R., and Seidel, H.-P. 2005. Ray maps for global illumination. In *Proceedings of the Eurographics Symposium on Rendering Techniques*. 43–54, 311.

Jakob, W., Regg, C., and Jarosz, W. 2011. Progressive expectation maximization for hierarchical volumetric photon mapping. *Comput. Graph. Forum 30*, 4.

Jarosz, W., Nowrouzezahrai, D., Thomas, R., Sloan, P.-P., and Zwicker, M. 2011. Progressive photon beams. *ACM Trans. Graph. 30*, 6.

Jensen, H. W. 1996. Global illumination using photon maps. In *Proceedings of the Eurographics Workshop on Rendering Techniques*. 21–30.

Jones, M. C., Marron, J. S., and Sheather, S. J. 1996. A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc. 91*, 433, 401–407.

Jones, M. C. and Sheather, S. J. 1991. Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. *Statist. Probab. Lett. 11*, 6, 511–514.

Katkovnik, V. and Shmulevich, I. 2002. Kernel density estimation with adaptive varying window size. *Pattern Recogn. Lett. 23*, 14, 1641–1648.

Knaus, C. and Zwicker, M. 2011. Progressive photon mapping: A probabilistic approach. *ACM Trans. Graph. 30*, 3, 25:1–25:13.

Lafortune, E. P. and Willems, Y. D. 1993. Bi-Directional path tracing. In *Proceedings of the 3rd International Conference on Computational Graphics and Visualization Techniques (COMPUGRAPHICS).* 145–153.

Mitchell, D. and Hanrahan, P. 1992. Illumination from curved reflectors. *ACM Trans. Graph. 26*, 2, 283–291.

Myszkowski, K. 1997. Lighting reconstruction using fast and adaptive density estimation techniques. In *Proceedings of the Eurographics Workshop on Rendering Techniques*. 251–262.

Ngerng, M. H. 2011. Recursive nonparametric estimation of local first derivative under dependence conditions. *Comm. Statist. Theory Methods 40*, 7, 1159–1168.

Park, B. U. and Marron, J. S. 1990. Comparison of data-driven bandwidth selectors. *J. Amer. Statist. Assoc. 85*, 409, 66–72.

Parker, S. G., Bigler, J., Dietrich, A., Friedrich, H., Hoberock, J., Luebke, D., McAllister, D., McGuire, M., Morley, K., Robison, A., and Stich, M. 2010. Optix: A general purpose ray tracing engine. *ACM Trans. Graph. 29*, 4, 66:1–66:13.

Parzen, E. 1962. On estimation of a probability density function and mode. *Ann. Math. Statist. 33*, 3, 1065–1076.

Perlin, K. 2002. Improving noise. *ACM Trans. Graph. 21*, 3, 681–682.

Pharr, M. and Humphreys, G. 2010. *Physically Based Rendering: From Theory to Implementation* 2nd Ed. Morgan Kaufmann, San Fransisco, CA.

Reif, J. H., Tygar, J. D., and Yoshida, A. 1994. Computability and complexity of ray tracing. *Discr. Comput. Geom. 11*, 1, 265–288.

Rousselle, F., Knaus, C., and Zwicker, M. 2011. Adaptive sampling and reconstruction using greedy error minimization. *ACM Trans. Graph. 30*, 6, 159:1–159:12.

Rudemo, M. 1982. Empirical choice of histograms and kernel density estimators. *Scand. J. Statist. 9*, 2, 65–78.

Schjoth, L. 2009. Anisotropic density estimation in global illumination. Ph.D. thesis, University of Copenhagen.

Schregle, R. 2003. Bias compensation for photon maps. *Comput. Graph. Forum 22*, 4, 729–742.

KATKOVNIK, V. AND SHMULEVICH, I. 2002. Kernel density estimation with adaptive varying window size. *Patt. Recog. Lett. 23*, 14, 1641–1648.

KNAUS, C. AND ZWICKER, M. 2011. Progressive photon mapping: A probabilistic approach. *ACM Trans. Graph. 30*, 3, 25:1–25:13.

LAFORTUNE, E. P. AND WILLEMS, Y. D. 1993. Bi-Directional path tracing. In *Proceedings of the 3rd International Conference on Computational Graphics and Visualization Techniques (COMPUGRAPHICS'93).* 145–153.

MITCHELL, D. AND HANRAHAN, P. 1992. Illumination from curved reflectors. *SIGGRAPH Comput. Graph. 26*, 2, 283–291.

MYSZKOWSKI, K. 1997. Lighting reconstruction using flat and adaptive density estimation techniques. In *Proceedings of the Eurographics Workshop on Rendering Techniques*. 251–262.

NGERNG, M. H. 2011. Recursive nonparametric estimation of local first derivative under dependence conditions. *Comm. Statist. Theory Methods 40*, 7, 1159–1168.

PARK, B. U. AND MARRON, J. S. 1990. Comparison of data-driven bandwidth selectors. *J. Amer. Statist. Assoc. 85*, 409, 66–72.

PARKER, S. G., BIGLER, J., DIETRICH, A., FRIEDRICH, H., HOBEROCK, J., LUEBKE, D., MCALLISTER, D., MCGUIRE, M., MORLEY, K., ROBISON, A., AND STICH, M. 2010. Optix: A general purpose ray tracing engine. *ACM Trans. Graph. 29*, 4, 66:1–66:13.

PARZEN, E. 1962. On estimation of a probability density function and mode. *Ann. Math. Statist. 33,* 3, 1065–1076.

PERLIN, K. 2002. Improving noise. *ACM Trans. Graph. 21*, 3, 681–682.

PHARR, M. AND HUMPHREYS, G. 2010. *Physically Based Rendering: From Theory to Implementation* 2nd Ed. Morgan Kaufmann, San Fransisco, CA.

REIF, J. H., TYGAR, J. D., AND YOSHIDA, A. 1994. Computability and complexity of ray tracing. *Discr. Comput. Geom. 11*, 1, 265–288.

ROUSSELLE, F., KNAUS, C., AND ZWICKER, M. 2011. Adaptive sampling and reconstruction using greedy error minimization. *ACM Trans. Graph. 30*, 6, 159:1–159:12.

RUDEMO, M. 1982. Empirical choice of histograms and kernel density estimators. *Scand. J. Statist. 9*, 2, 65–78.

SCHJOTH, L. 2009. Anisotropic density estimation in global illumination. Ph.D. thesis, University of Copenhagen.

SCHREGLE, R. 2003. Bias compensation for photon maps. *Comput. Graph. Forum 22*, 4, 729–742.

SHEATHER, S. J. AND JONES, M. C. 1991. A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. B53*, 3, 683–690.

SHIRLEY, P., WADE, B., HUBBARD, P. M., ZARESKI, D., WALTER, B., AND GREENBERG, D. P. 1995. Global illumination via density-estimation. In *Proceedings of the 6th Workshop on Rendering*. 219–230.

SILVERMAN, B. 1986. *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability. Chapman and Hall.

SPENCER, B. AND JONES, M. W. 2009. Into the blue: Better caustics through photon relaxation. *Comput. Graph. Forum 28,* 2, 319–328.

SUYKENS, F. AND WILLEMS, Y. D. 2000. Density control for photon maps. In *Proceedings of the Eurographics Workshop on Rendering Techniques*. 23–34.

VAN EEDEN, C. 1985. Mean integrated squared error of kernel estimators when the density and its derivative are not necessarily continuous. *Ann. Inst. Statist. Math. 37*, 1, 461–472.

VEACH, E. 1998. Robust monte carlo methods for light transport simulation. Ph.D. thesis, Stanford University. AA19837162.

VEACH, E. AND GUIBAS, L. 1994. Bidirectional estimators for light transport. In *Proceedings of the Eurographics Rendering Workshop.* 147–162.

VORBA, J. 2011. Bidirectional photon mapping. In *Proceedings of the 15th Central European Seminar on Computer Graphics.* 25–32.

WALTER, B. J. 1998. Density estimation techniques for global illumination. Ph.D. thesis, AA19900037, Cornell University, Ithaca, NY.

WAND, M. P. AND JONES, M. C. 1994. Multivariate plug-in bandwidth selection. *Comput. Statist. 9*, 97–116.

WOLVERTON, C. AND WAGNER, T. 1969. Recursive estimates of probability densities. *IEEE Trans. Syst. Sci. Cybernet. 5*, 3, 246–247.

WONG, K. W. AND WANG, W. 2005. Adaptive density estimation using an orthogonal series for global illumination. *Comput. Graph. 29*, 5, 738–755.

YAMATO, H. 1971. Sequential estimation of a continuous probability density function and mode. *Bull. Math. Statist. 14*, 1–12.

Received April 2012; revised November 2012; accepted November 2012