Supplemental Document of "A Path Space Extension for Robust Light Transport Simulation"

Toshiya Hachisuka[1,3] Jacopo Pantaleoni[2] Henrik Wann Jensen[3]
[1]Aarhus University [2]NVIDIA Research [3]UC San Diego

A Derivations of Path Space Extension

A.1 Original Path Integral Formulation

We first summarize the path integral formulation by Veach [Veach 1998]. Using the path integral formulation, light transport simulation can be expressed as:

$$I_j = \int_\Omega f_j(\bar{x}) d\mu(\bar{x}), \tag{1}$$

where j is an index to the pixel (or measurement), f_j is the measurement contribution function, μ is a measure of paths, and Ω is the set of transport paths of all lengths.

We evaluate this integral using Monte Carlo integration with a random path $\bar{X}$ with the probability density $p(\bar{X})$ by:

$$E\left[\frac{f_j(\bar{X})}{p(\bar{X})}\right] = \int_\Omega \frac{f_j(\bar{x})}{p(\bar{x})} p(\bar{x}) d\mu(\bar{x}) = I_j. \tag{2}$$

Each term is defined as follows.

Path: Given a set of vertices on surfaces $x_i \in \mathcal{M}$ for each i, a path $\bar{x}$ of length k is defined in the form

$$\bar{x} = x_0 x_1 \dots x_k, \tag{3}$$

where $1 < k < \infty$. By convention, x_0 is at a light source.

Space: Let Ω_k represent the paths of length k, then the path space Ω is defined as

$$\Omega = \bigcup_{k=1}^{\infty} \Omega_k. \tag{4}$$

Ω thus represents the set of paths of all lengths.

Measure: A measure μ_k on a path of length k is a product measure of the standard area measure A at all the k vertices as follows:

$$d\mu_k(\bar{x}) = dA(x_0) dA(x_1) \cdots dA(x_k), \tag{5}$$

which can also be written as

$$\mu_k = \underbrace{A \times \cdots \times A}_{\text{k + 1 times}} = A_{k+1}. \tag{6}$$

The measure μ on the space Ω is the sum of the measures of the paths of each length:

$$\mu(D) = \sum_{k=1}^{\infty} \mu_k(D \cap \Omega_k), \tag{7}$$

where $D \subset \Omega_k$ is a set of paths.

Measurement Contribution Function: The measurement contribution function f_j can be derived from the three-point form of the rendering equation and the measurement equation:

$$L(y \to x) = L_e(y \to x) + \int_{\mathcal{M}} L(z \to y) f_s(z \to y \to x) G(z \leftrightarrow y) dA(z), \tag{8}$$

$$I_j = \int_{\mathcal{M}^2} W_e^{(j)}(x \to e) L(x \to e) G(x \leftrightarrow e) dA(x) dA(e). \tag{9}$$

To define the measurement contribution function, we recursively expand the above equation:

$$\begin{aligned} I_j = &\sum_{i=1}^{\infty} \int_{\mathcal{M}^{k+1}} L_e(x_0 \to x_1) G(x_0 \leftrightarrow x_1) \\ &\cdot \prod_{p=1}^{k-1} f_s(x_{p-1} \to x_p \to x_{p+1}) G(x_p \leftrightarrow x_{p+1}) \\ &\cdot W_e^{(j)}(x_{k-1} \to x_k) dA(x_0) \cdots dA(x_k) = \\ &\int_{\mathcal{M}^2} L_e(x_0 \to x_1) G(x_0 \leftrightarrow x_1) W_e^{(j)}(x_0 \to x_1) dA(x_0) dA(x_1) \\ &+ \int_{\mathcal{M}^3} L_e(x_0 \to x_1) G(x_0 \leftrightarrow x_1) f_s(x_0 \to x_1 \to x_2) \\ &\cdot G(x_1 \leftrightarrow x_2) W_e^{(j)}(x_1 \to x_2) dA(x_0) dA(x_1) dA(x_2) \\ &+ \cdots . \end{aligned} \tag{10}$$

As noted by Veach, f_j is defined by extracting the term from the recursive expansion of the above equation. For example, given a path $\bar{x} = x_0 x_1 x_2$, we have

$$\begin{aligned} f_j(\bar{x}) = &L_e(x_0 \to x_1) G(x_0 \leftrightarrow x_1) f_s(x_0 \to x_1 \to x_2) \\ &\cdot G(x_1 \leftrightarrow x_2) W_e^{(j)}(x_1 \to x_2) \end{aligned} \tag{11}$$

Probability Density: To evaluate the path integral formulation using Monte Carlo integration, we need to define the probability density $p(\bar{x})$ of a path $\bar{x}$. Since μ is a product of the area measure, we have

$$p(\bar{x}) = \frac{dP}{d\mu}(\bar{x}) = \frac{dP}{d\mu}(x_0 \dots x_k) = \prod_{i=0}^{k} \frac{dP}{dA}(x_i) \tag{12}$$

where each $\frac{dP}{dA}(x_i)$ can be computed either directly or by using the relationship $\frac{dP}{dA}(x_i) = \frac{dP}{d\sigma}(\omega_o) \frac{d\sigma(\omega_o)}{dA(x_i)}$ if we sample a vertex x' from x in the direction $\omega_o = \overbrace{x' - x}$ with local path sampling.

A.2 Path Space Extension

We extend the original path integral formulation to include density estimation. Similar to the original path integral formulation, we formulate the above integral as an integral over our *extended* path space Ω' with the extended measurement contribution function f'_j:

$$I'_j = \int_{\Omega'} f'_j(\bar{x}') d\mu'(\bar{x}'), \tag{13}$$

and we want to evaluate the integral using Monte Carlo integration with random paths $\bar{X}'$ the probability density $p'(\bar{X})$ by:

$$E\left[\frac{f'_j(\bar{X}')}{p'(\bar{X}')}\right] = \int_{\Omega'} \frac{f'_j(\bar{x}')}{p'(\bar{x}')} p'(\bar{x}') d\mu'(\bar{x}') = I'_j. \tag{14}$$

In the following, we use symbols with primes for our extended definition.

Path: Using vertices $x_i, x'_i \in \mathcal{M}$, an extended path $\bar{x}'$ of length k is defined in the form

$$\bar{x}' = x_0 x_1 x'_1 x_2 \dots x_{k-1} x'_{k-1} x_k, \tag{15}$$

where $1 < k < \infty$. The difference is that an extended path has an additional vertex x'_i at each intermediate vertex x_i. An extended path is "disconnected" in the sense of light transport path, where the location of disconnection can happen between $x_i, x'_i \in \bar{x}'$ for each i except for $i = 0$ and $i = k$.

Space: We define Ω'_k as the space of the paths of length k and the entire path space Ω' is defined as

$$\Omega' = \bigcup_{k=1}^{\infty} \Omega'_k. \tag{16}$$

Since each extended path $\bar{x}' \in \Omega'_k$ contains $k-1$ additional vertices $x'_i \in \mathcal{M}$ compared to the original path, we have an extended path space as

$$\Omega'_k = \Omega_k \cup \mathcal{M}^{k-1}. \tag{17}$$

Measure: Similar to the original formulation, a measure μ'_k on a path of length k is a product measure of the standard area measure A at all the vertices:

$$\begin{aligned} d\mu'_k(\bar{x}) &= dA(x_0)dA(x_1)dA(x'_1)\cdots dA(x_k) \\ \mu'_k &= \underbrace{A \times \cdots \times A}_{\text{2k times}} = A_{2k} \\ \mu'(D) &= \sum_{k=1}^{\infty} \mu'_k(D \cap \Omega'_k). \end{aligned} \tag{18}$$

Based on this definition, we can see that a measure in our extended space has extended to

$$\mu'_k = \mu_k \times A_{k-1}. \tag{19}$$

Measurement Contribution Function: We can derive the extended measurement contribution function based on the integral formulation of density estimation [Silverman 1986] at each intermediate vertex;

$$\begin{aligned} &L(y \to x) = L_e(y \to x) + \\ &\int_{\mathcal{M}} \left(\int_{\mathcal{M}} L(z \to w) K(w, y) G(z \leftrightarrow w) dA(w) \right) f_s(z \to y \to x) dA(z) \\ &= L_e(y \to x) + \\ &\int_{\mathcal{M}^2} L(z \to w) K(w, y) f_s(z \to y \to x) G(z \leftrightarrow w) dA(w) dA(z) \end{aligned} \tag{20}$$

$$I_j = \int_{\mathcal{M}^2} W_e^{(j)}(x \to e) L(x \to e) G(x \leftrightarrow e) dA(x) dA(e), \tag{21}$$

where $K(w, y)$ is a kernel function for density estimation.

The difference from the original light transport equation is that density estimation adds a convolution of the incoming radiance distribution by $K(w, y)$ as $\int_{\mathcal{M}} L(z \to w) G(z \leftrightarrow w) K(w, y) dA(w)$. The measurement equation stays the same. The path is disconnected at w, y in the sense of light transport path.

Similar to the original path integral formulation, we recursively expand the above equation as

$$\begin{aligned} I'_j = &\sum_{i=1}^{\infty} \int_{\mathcal{M}^{2k}} L_e(x_0 \to x_1) G(x_0 \leftrightarrow x_1) \\ &\cdot \prod_{p=1}^{k-1} K(x_p, x'_p) f_s(x_{p-1} \to x'_p \to x_{p+1}) G(x'_p \leftrightarrow x_{p+1}) \\ &\cdot W_e^{(j)}(x_{k-1} \to x_k) dA(x_0) dA(x_1) dA(x'_1) \cdots dA(x_k) = \\ &\int_{\mathcal{M}^2} L_e(x_0 \to x_1) G(x_0 \leftrightarrow x_1) W_e^{(j)}(x_0 \to x_1) dA(x_0) dA(x_1) \\ &+ \int_{\mathcal{M}^4} L_e(x_0 \to x_1) G(x_0 \leftrightarrow x_1) K(x_1, x'_1) f_s(x_0 \to x'_1 \to x_2) \\ &\cdot G(x'_1 \leftrightarrow x_2) W_e^{(j)}(x'_1 \to x_2) dA(x_0) dA(x_1) dA(x'_1) dA(x_2) \\ &+ \cdots . \end{aligned} \tag{22}$$

f'_j is defined by extracting the term from the recursive expansion of the above equation. For example, given a path $\bar{x}' = x_0 x_1 x'_1 x_2$, we have

$$\begin{aligned} f'_j(\bar{x}') = &L_e(x_0 \to x_1) G(x_0 \leftrightarrow x_1) K(x_1, x'_1) \\ &\cdot f_s(x_0 \to x'_1 \to x_2) G(x'_1 \leftrightarrow x_2) W_e^{(j)}(x'_1 \to x_2) \end{aligned} \tag{23}$$

As we describe below, we perform density estimation or vertex perturbation at x_1, x'_1 to generate this path.

Probability Density: The probability density $p(\bar{x}')$ of an extended path $\bar{x}'$ is defined with respect to a product of the area measure;

$$p(\bar{x}) = \frac{dP}{d\mu}(\bar{x}) = \frac{dP}{d\mu}(x_0 \dots x_k) = \prod_{i=0}^{k} \frac{dP}{dA}(x_i, x'_i). \tag{24}$$

The important difference is that local path sampling does not define how to relate two disconnected vertices $x_i, x'_i \in \bar{x}$ in an extended path.

Depending on how we generate an extended path, we have two cases. For the sake of brevity, we assume that there is only one disconnection at $x_q, x'_q \in \bar{x}$ in the path in the following. This is true in our application in this paper.

The first case is that x_q and x'_q are generated by independent statistical processes. An example is that we generate a light path that ends at x_q and an eye path that ends at x'_q by independent local path sampling, as in standard photon density estimation. In this case, the probability density of the entire extended path is simply a product of two probability densities:

$$\begin{aligned} p(\bar{x}') &= \prod_{i=0}^{q-1} \frac{dP}{dA}(x_i) \cdot \frac{dP}{dA}(x_q, x'_q) \prod_{i=q+1}^{k} \frac{dP}{dA}(x_i) \\ &= \prod_{i=0}^{q} \frac{dP}{dA}(x_i) \cdot \frac{dP}{dA}(x'_q) \prod_{i=q+1}^{k} \frac{dP}{dA}(x_i). \end{aligned} \tag{25}$$

This equation defines the probability density of a path in photon density estimation. Note that there is no influence due to the density estimation kernel. The density estimation kernel affects the measurement contribution function instead.

The second case is that both x_q and x'_q are generated by the same statistical process. Traditional local path sampling does not specify how to generate x'_q from x_q (or x_q from x'_q). We thus introduce a vertex perturbation as we described in the paper - x'_q is generated

by perturbing x_q. Assuming a circular support with the radius r for perturbation, this is formalized as:

$$\begin{aligned} p(\bar{x}') &= \prod_{i=0}^{q-1} \frac{dP}{dA}(x_i) \cdot \frac{dP}{dA}(x_q, x_q') \prod_{i=q+1}^{k} \frac{dP}{dA}(x_i) \\ &= \prod_{i=0}^{q} \frac{dP}{dA}(x_i) \cdot \frac{dP}{dA}(x_q'|x_q) \cdot \prod_{i=q+1}^{k} \frac{dP}{dA}(x_i) \\ &= \prod_{i=0}^{q} \frac{dP}{dA}(x_i) \cdot \frac{1}{\pi r^2} \cdot \prod_{i=q+1}^{k} \frac{dP}{dA}(x_i). \end{aligned} \tag{26}$$

By comparing with the original path integral formulation, we see that this case has an extra factor of $\frac{1}{\pi r^2}$:

$$p(\bar{x}') = \frac{1}{\pi r^2} \prod_{i=0}^{q} \frac{dP}{dA}(x_i) \prod_{i=q+1}^{k} \frac{dP}{dA}(x_i) = \frac{1}{\pi r^2} p(\bar{x}), \tag{27}$$

which is the reason why we need to multiply $\frac{1}{\pi r^2}$ in the probability density of an extended Monte Carlo path integration.

Reduction to Other Formulations: Our extended path integral formulation reduces to other existing formulations by changing the kernel function. For example, by defining $K(w, y) = \delta(||w - y||)$, we obtain the original path integral formulation. We can think of regular Monte Carlo path integration as performing vertex perturbation using the delta function at all the vertices. All the factors due to the delta function then cancel out exactly, leaving the original ratio of the measurement contribution function and the probability density.

In this paper, we handle a special case

$$K(x_i, x_i') = \begin{cases} \frac{1}{\pi r^2} & \text{for } ||x_i - x_i'|| < r \\ 0 & \text{otherwise,} \end{cases} \tag{28}$$

where we perform density estimation or vertex perturbation only at x_i, x_i' and $K(w, y) = \delta(||w - y||)$ otherwise.

One possible generalization is to consider non-delta kernel functions at all the vertices, which corresponds to performing multiple density estimation or vertex perturbation. Since our focus in this paper is to combine density estimation and Monte Carlo path integration, this generalization is left as future work.

For non-delta kernels, a solution to the extended path integral formulation is different from the solution to the original path integral formulation. Therefore, our estimate is biased when we compared to the original solution. However, for the limit of delta kernels, we have

$$\begin{aligned} &L(y \to x) = L_e(y \to x) + \\ &\int_{\mathcal{M}} \left(\int_{\mathcal{M}} L(z \to w) \delta(w, y) G(z \leftrightarrow w) dA(w) \right) f_s(z \to y \to x) dA(z) \\ &= L_e(y \to x) + \\ &\int_{\mathcal{M}} L(z \to y) f_s(z \to y \to x) G(z \leftrightarrow y) dA(z) \end{aligned} \tag{29}$$

which is equal to the original light transport equation. Progressive density estimation achieves this condition with an infinite number of samples, thus our estimate is consistent.

B Bias-Aware Multiple Importance Sampling

B.1 Problem Settings

In order to take bias into account in multiple importance sampling, we consider a biased estimator as an unbiased estimator of a biased solution. This makes it possible to characterize bias as a result of modifications to an original integrand. We denote such modifications by the ith sampling technique as $b_i(x)$. Following the same notation as Veach's, the jth sample from the ith technique has the following contribution:

$$F_{i,j} = \frac{w_i(X_{i,j})(f(X_{i,j}) + b_i(X_{i,j}))}{p_i(X_{i,j})}, \tag{30}$$

where $w_i(X_{i,j})$ is the weight function, $f(X_{i,j})$ is the integrand, $b_i(X_{i,j})$ is the modifications to the original integrand that introduces bias, and $p_i(X_{i,j})$ is the probability density function. The expected (and potentially biased) value from the ith technique is then given by

$$\mu_i = E[F_{i,j}] = \int_\Omega w_i(x)(f(x) + b_i(x)) \mathrm{d}\mu(x). \tag{31}$$

Note that $b_i(X_{i,j})$ is not bias itself, but rather the contribution to the bias from the sample $X_{i,j}$. We can describe bias from the ith technique as the difference between the biased solution (i.e. μ_i) and the correct solution;

$$B[F_{i,j}] = E[F_{i,j}] - \int_\Omega w_i(x) f(x) \mathrm{d}\mu(x) = \int_\Omega w_i(x) b_i(x) \mathrm{d}\mu(x). \tag{32}$$

Our goal is to find a weighting strategy which has expected error that is not arbitrarily far away from the truly optimal weighting strategy. This is also what the original balance heuristic achieves. Our contribution is a derivation that shows necessary modifications for the original balance heuristic to keep this optimality in combination with a biased technique.

The goal of the derivation of the original balance heuristic is to minimize variance of the solution. This is because the error of unbiased Monte Carlo techniques is solely characterized by variance. However, in our setting, we also need to take bias into account. We therefore minimize the squared error based on *bias-variance decomposition*:

$$E[(F - \hat{\mu})^2] = V[F] + B[F]^2, \tag{33}$$

where $\hat{\mu}$ is the correct solution, $V[F]$ is variance, and $B[F]$ is bias. Minimizing squared error including bias in general, however, is a very challenging task. This is because bias is a systematic error that happens because of various reasons, and it is often difficult to define general characteristics of bias in order to perform any theoretical analysis on quantities including bias. We therefore make a couple of assumptions that are often reasonable in rendering.

First, we only consider the case where we have one nth biased technique in addition to other $n - 1$ unbiased techniques. This can be true in our method if we restrict ourselves to consider only one photon density estimation technique. Second, we assume that the contribution to bias from each sample is constant. This is also reasonable in photon density estimation as we only consider neighboring photons which tend to cause similar error within each radiance estimate. We can thus set $b_i(x) = 0$ for $i \neq n$ and $b_n(x) = b_n$ for $i = n$. Notice that we are overloading the notation of b_n for readability. We can then expand $E[(F - \hat{\mu})^2]$ as follows:

$$\begin{aligned} E[(F - \hat{\mu})^2] &= V\left[\sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} F_{i,j} \right] + B\left[\sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} F_{i,j} \right]^2 \\ &= \int_\Omega \left(\sum_{i=1}^{n-1} \frac{w_i^2(x) f^2(x)}{n_i p_i(x)} + \frac{w_n^2(x)(f(x) + b_n)^2}{n_n p_n(x)} \right) \mathrm{d}\mu(x) \\ &+ \left(\int_\Omega w_n(x) b_n \mathrm{d}\mu(x) \right)^2 - \sum_{i=1}^{n} \frac{1}{n_i} \mu_i^2, \end{aligned} \tag{34}$$

In the following derivations, we will show how to minimize the sum of the first two terms. The last term $\sum_{i=1}^{n} \frac{1}{n_i}\mu_i^2$ has the same bound as the original derivation by Veach thus the last term is independent from weighting functions.

B.2 Minimizing Squared Error

Even after we made some simplifying assumptions, Equation (34) is still difficult to minimize with respect to w_i, since bias introduced the integral term $\left(\int_\Omega w_n(x)b_n \mathrm{d}\mu(x)\right)^2$. In order to yield the optimal w_i including this term, it seems that we need to solve an integral equation which is often intractable. However, we can show that minimizing a point-wise expression,

$$\sum_{i=1}^{n-1} \frac{w_i^2(x)}{n_i p_i(x)} + \frac{w_n^2(x)(1+r_n(x))^2}{n_n p_n(x)} + A r_n^2(x) w_n^2(x) \qquad (35)$$

indeed suffices to minimize the full expression of $D[F]$ in Equation (34), where we defined $r_n(x) = \frac{b_n}{f(x)}$ and $A = \int_\Omega \mathrm{d}\mu(x)$ for readability.

We first start by trying to find an alternative expression for $\left(\int_\Omega w_n(x)b_n \mathrm{d}\mu(x)\right)^2$. In order to obtain such an expression, we use the Cauchy-Schwarz inequality and consider the bound of this term;

$$\left(\int_\Omega w_n(x)b_n \mathrm{d}\mu(x)\right)^2 \le A \int_\Omega w_n^2(x) b_n^2 \mathrm{d}\mu(x). \qquad (36)$$

Minimizing the bound in general does not minimize the original term since the bound might not have the same extrema as the original function. In our case, however, the bound and the function happen to have extrema at exactly the same points since

$$\frac{\partial}{\partial w_n}\left(\int_\Omega w_n(x)b_n \mathrm{d}\mu(x)\right)^2 = A\frac{\partial}{\partial w_n}\int_\Omega w_n^2(x) b_n^2 \mathrm{d}\mu(x). \qquad (37)$$

Note that this is possible because of our assumptions on bias. Since the latter is the upper bound of the function, minimizing the bound also minimizes the function given the fact that they have extrema at the same points.

Furthermore, since the sums of two functions $f(x) + g(x)$ and $f(x) + h(x)$ have the same extrema if $\frac{dg}{dx} = \frac{dh}{dx}$, minimizing

$$\int_\Omega \left(\sum_{i=1}^{n-1} \frac{w_i^2(x) f^2(x)}{n_i p_i(x)} + \frac{w_n^2(x)(f(x)+b_n)^2}{n_n p_n(x)} + A w_n^2(x) b_n^2 \right) \mathrm{d}\mu(x) \qquad (38)$$

is equivalent to minimizing the corresponding sums in Equation (34). We can yield Equation 35 by dividing the integrand of this equation by $f^2(x)$.

B.3 Bias-Aware Balance Heuristic

Using the method of Lagrange multipliers, the minimum value of Equation 35 is attained when all $n+1$ partial derivatives (n derivatives for w_i and one for λ) of the expression

$$\sum_{i=1}^{n-1} \frac{w_i^2}{n_i p_i} + \frac{w_n^2(1+r_n)^2}{n_n p_n} + A r_n^2 w_n^2 + \lambda\left(\sum_{i=1}^{n} w_i - 1\right) \qquad (39)$$

are zero. Note that we dropped the notation (x) similar to the Veach's derivation since this is a point-wise minimization of the function.

The solution to this equation yields

$$\hat{w}_i(\vec{x}) = \frac{n_i p_i'(\vec{x})}{\sum_{k=1}^{M} n_k p_k'(\vec{x})}, \qquad (40)$$

where

$$p_i'(\vec{x}) = \begin{cases} p_i(\vec{x}) & (i \neq n) \\ p_n(\vec{x}) \frac{1}{(1+r_n)^2 + n_n p_n(\vec{x}) A r_n^2} & (i = n), \end{cases} \qquad (41)$$

A is a constant, and r_n is the relative magnitude of the contribution of the bias to the sampled value. Here the nth technique is biased. Note that if there is no bias $r_n = 0$, we obtain $p_i' = p_i$ and Equation 40 turns into the original balance heuristic. Note also that taking an infinite number of samples $\lim N \to \infty$ turns the weight for a biased method into zero, which makes the combined estimate converge to the correct solution even if $r_n \neq 0$.

Using this bias-aware balance heuristic as a combination strategy, the resulting estimator $\hat{F_B}$ satisfies the following inequality:

$$\mathrm{Error}[\hat{F_B}]^2 - \mathrm{Error}[F]^2 \le \left(\frac{1}{\min_i n_i} - \frac{1}{\sum_i n_i}\right)\mu^2. \qquad (42)$$

Notice the difference from the optimality claim of the original balance heuristic. This inequality is defined with the operator Error that returns error which includes both bias and variance. Similar to "variance gap", we call the left hand side as "error gap", which is the difference of errors between the provably good combination and any other combination.

Unfortunately, we cannot use this provably good strategy in practice. In order to use this strategy, one would have to evaluate the magnitude of bias relative to the sampled value, r_i. Even if we had a method to estimate r_i, the provably good weighting strategy would require the additional ability of estimating r_i of samples which were *not even sampled* by a biased technique.

We propose one practical solution to this problem, which is to use the original balance heuristic in combination with progressive photon mapping [Hachisuka et al. 2008]. Since bias in progressive photon mapping is guaranteed to converge to zero [Knaus and Zwicker 2011] as we add more samples, the difference between the original balance heuristic and the bias-aware balance heuristic (Equation 40) are expected to converge to zero at an infinite number of samples. The challenge however is that we still would like to pursue a provably good combination with any number of samples. In the following subsections, we describe a condition on this approach that keeps the resulting combination provably good.

C Error Gap of the Balance Heuristic

In order to analyze the influence of bias, we first look at the consequence of using the original balance heuristic by ignoring bias in biased estimators. In any biased estimator, error is characterized by the following *bias-variance decomposition*:

$$\mathrm{Error}[F]^2 = \mathrm{Var}[F] + \mathrm{Bias}[F]^2. \qquad (43)$$

We then look at the error gap (*not* the variance gap) of the original balance heuristic

$$\begin{aligned} &\mathrm{Error}[\hat{F}]^2 - \mathrm{Error}[F]^2 \\ &= \mathrm{Var}[\hat{F}] + \mathrm{Bias}[\hat{F}]^2 - \mathrm{Var}[F] - \mathrm{Bias}[F]^2 \\ &= \mathrm{Var}[\hat{F}] - \mathrm{Var}[F] + \mathrm{Bias}[\hat{F}]^2 - \mathrm{Bias}[F]^2. \end{aligned} \qquad (44)$$

We thus obtain

$$\begin{aligned}
&\mathrm{Error}[\hat{F}]^2 - \mathrm{Error}[F]^2 \\
&\leq \left(\frac{1}{\min_i n_i} - \frac{1}{\sum_i n_i}\right)\mu^2 + \mathrm{Bias}[\hat{F}]^2 - \mathrm{Bias}[F]^2 \\
&\leq \left(\frac{1}{\min_i n_i} - \frac{1}{\sum_i n_i}\right)\mu^2 + \mathrm{Bias}[\hat{F}]^2. \qquad (45)
\end{aligned}$$

Therefore, the error gap of the original balance heuristic under the presence of biased estimator is bounded by the original bound plus the additional term due to bias. Comparing this inequality and the inequality in Equation 42, using the original balance heuristic can be further away from the truly optimal (unknown) combination than the bias-aware balance heuristic by the additional term $\mathrm{Bias}[\hat{F}]^2$. This result shows that, depending on how bias changes according to the number of samples, the balance heuristic can be arbitrary away from a provably good strategy under the presence of a biased estimator.

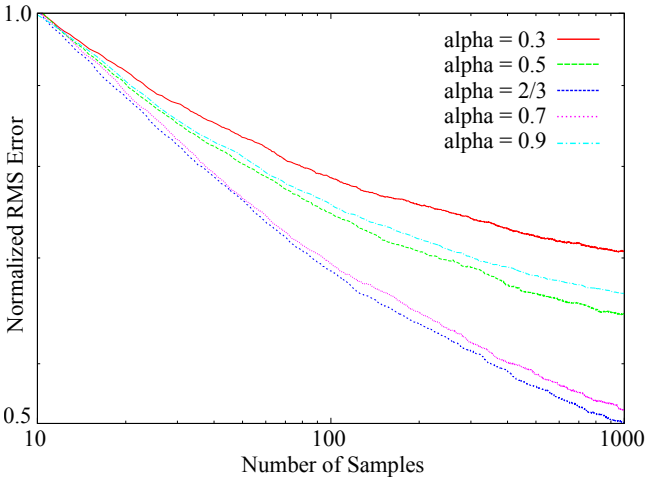

Figure 1: *Error gaps due to different α values. The graph plots RMS errors of the rendered images of the torus scene with different values for the alpha parameter. RMS errors are normalized such that the graph shows the difference in convergence rates. As predicted by our theoretical analysis, $\alpha = 2/3$ gives us the fastest convergence rate.*

D Condition for a Provably Good Strategy

As we mentioned earlier, we use progressive photon density estimation [Hachisuka et al. 2008] with the hope that the original balance heuristic is still close to the provably good combination of the bias-aware balance heuristic. We show that setting the alpha parameter of progressive photon density estimation to $2/3$ can indeed achieve such a combination with the original balance heuristic in the Veach's sense [1995].

Knaus and Zwicker [2011] showed that the asymptotic convergence rates of bias and variance in progressive density estimation are

$$\mathrm{Var} = O\left(\frac{1}{n^\alpha}\right) \qquad \mathrm{Bias} = O\left(\frac{1}{n^{1-\alpha}}\right), \qquad (46)$$

where α is the parameter that controls the reduction rate of the radius in progressive density estimation. Substituting this result into Equation 45 yields

$$\mathrm{Error}[\hat{F}]^2 - \mathrm{Error}[F]^2 \leq \left(\frac{1}{n_n^\alpha} - \frac{1}{\sum_i n_i}\right)\mu^2 + \frac{C}{n_n^{2(1-\alpha)}}, \qquad (47)$$

where C is a constant. We used $\min_i n_i = n_n^\alpha$ by considering the fact that the variance of progressive photon mapping converges at the rate of $O\left(\frac{1}{n^\alpha}\right)$. We take the effect of slower convergence rate into account by replacing n_n by n_n^α. Note that this does not affect the derivation of the bias-aware balance heuristic since the derivations do not try to achieve the optimal distribution of the number of samples.

Our goal is to find conditions such that

$$\left(\frac{1}{n_n^\alpha} - \frac{1}{\sum_i n_i}\right)\mu^2 + \frac{C}{n_n^{2(1-\alpha)}} \approx \left(\frac{1}{n_n^\alpha} - \frac{1}{\sum_i n_i}\right)\mu^2 \qquad (48)$$

for large enough $N = \sum_i n_i$. Note that the right hand side also uses the equation $\min_i n_i = n_n^\alpha$ since we are now combining progressive photon density estimation and Monte Carlo path integration.

Now, consider the difference between the convergence rates of the bounds of the error gap in the original balance heuristic and the bias-aware balance heuristic:

$$\begin{aligned}
\left(\frac{1}{n_n^\alpha} - \frac{1}{\sum_i n_i}\right)\mu^2 + \frac{C}{n_n^{2(1-\alpha)}} &\in O\left(\frac{1}{n_n^\alpha}\right) + O\left(\frac{1}{n_n^{2(1-\alpha)}}\right) \\
\left(\frac{1}{n_n^\alpha} - \frac{1}{\sum_i n_i}\right)\mu^2 &\in O\left(\frac{1}{n_n^\alpha}\right). \qquad (49)
\end{aligned}$$

The difference in convergence rates of the two bounds is minimized at $\alpha = 2/3$, which is the solution for $\alpha = 2(1-\alpha)$. In other words, using $\alpha = 2/3$ makes sure that the bound of the error gap from any other combination strategies reduces with the convergence rate of the bias-aware balance heuristic. The resulting error gap is

$$\begin{aligned}
\mathrm{Error}[\hat{F}]^2 - \mathrm{Error}[F]^2 &\leq \left(\frac{1}{n_n^\alpha} - \frac{1}{\sum_i n_i}\right)\mu^2 + \frac{C}{n_n^{2(1-\alpha)}} \\
&\leq (C+1)\left(\frac{1}{n_n^\alpha} - \frac{1}{\sum_i n_i}\right)\mu^2 \qquad (50)
\end{aligned}$$

Note that any other values of the alpha parameter makes the bound arbitrary away from above with given N. Figure 1 shows the results of a numerical experiment that confirms our theory. We have found that using the alpha value other than $2/3$ results in slower convergence rates. The condition $\alpha = 2/3$ is not only theoretically critical, but also practically important.

References

HACHISUKA, T., OGAKI, S., AND JENSEN, H. W. 2008. Progressive photon mapping. In *ACM SIGGRAPH Asia 2008 Papers*, 130:1–130:8.

KNAUS, C., AND ZWICKER, M. 2011. Progressive photon mapping: A probabilistic approach. *ACM Transaction on Graphics 30* (May), 25:1–25:13.

SILVERMAN, B. W. 1986. *Density estimation for statistics and data analysis*. Chapman and Hall.

VEACH, E., AND GUIBAS, L. J. 1995. Optimally combining sampling techniques for Monte Carlo rendering. In *ACM SIGGRAPH 1995 Papers*, 419–428.

VEACH, E. 1998. *Robust Monte Carlo methods for light transport simulation*. PhD thesis, Stanford University, Stanford, CA, USA. AAI9837162.