

**Technical Report Series of DCGI, Volume 4, Year 2014**

Department of Computer Graphics and Interaction

Czech Technical University in Prague, CZ

Faculty of Electrical Engineering

Website: <http://dcgi.fel.cvut.cz>

Optimal Combination of Techniques  
in Multiple Importance Sampling

Vlastimil Havran and Mateu Sbert

## Authors' Addresses

Vlastimil Havran  
Department of Computer Graphics and Interaction  
Faculty of Electrical Engineering  
Czech Technical University in Prague  
Karlovo nám. 13, 12135 Praha 2  
Czech Republic  
E-mail: havran AT fel DOT cvut DOT cz  
Web: <http://dcgi.fel.cvut.cz/members/havran>

Mateu Sbert  
Department of Informàtica i Matemàtica Aplicada  
University of Girona  
Edifici P4, Campus Montilivi, 17071 Girona  
Spain  
E-mail: mateu AT ima DOT udg DOT edu  
Web: <http://ima.udg.edu/~mateu>

## Funding Acknowledgements

This work has been partially funded by the Czech Science Foundation under research programs P202/12/2413 and GA14-19213S and by the Grant Agency of the Czech Technical University in Prague, grant No. SGS13/214/OHK3/3T/13, and on the Spanish side by grant TIN2013-47276-C6-1-R from the Spanish Government and 2014 SGR 1232 from the Catalan Government and National Natural Science Foundation of China (Nos. 61372190 and 61331018).

Technical Report Series of DCGI, available at <http://dcgi.fel.cvut.cz/techreps>,  
Department of Computer Graphics and Interaction, Czech Technical University in Prague,  
Faculty of Electrical Engineering, Czech Republic.

ISSN 1805-6180

Copyright © Department of Computer Graphics and Interaction, Faculty of Electrical Engineering, Czech Technical University in Prague, August 2014.

## Abstract

Since its introduction by Veach, the Multiple Importance Sampling (MIS) technique has been widely used in Computer Graphics in many rendering algorithms. MIS is based on weighting several sampling techniques into a single estimator. When the mixing weights are taken such that the sample contributions are balanced, i.e. they are the same for all techniques, it becomes a balance heuristic. It has been used since its invention almost exclusively on equal sampling for all techniques, and until now the question whether unequal sampling can give better variance, has raised little interest, maybe due to its intrinsic difficulty and also due to the fact that good results were already obtained with equal sampling. The most interesting cases of the use of MIS in Computer Graphics, where an environment map is a particular case, correspond to the integral of a product of functions. Based on the properties of the balance heuristic MIS as a weighted mixture of distributions, weights proportional to the number of samples, we obtain for this kind of integral an implicit closed formula for the optimal sampling. We also take into account the cost of each sampling technique. Although this closed formula cannot be written in an explicit way, we outline an iterative procedure for obtaining the optimal values. To bypass the combinatorially growing cost of the iterative procedure, we introduce a sound heuristic approximation based on the optimal combination of two independent estimators with known variance. We validate our theory with the results from implementing 1-dimensional function examples and 2-dimensional examples of environment map illumination.

## Keywords

global illumination, rendering equation analysis, multiple importance sampling, Monte Carlo, statistics

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Related work</b>	<b>3</b>
<b>3</b>	<b>Importance sampling</b>	<b>5</b>
3.1	Importance sampling for product of functions . . . . .	5
3.2	Importance sampling in the rendering equation for illumination by an environment map . . . . .	6
3.3	Analysis for a constant environment map and arbitrary BRDF . . . . .	7
3.4	Analysis for a non-constant environment map and arbitrary BRDF . . . . .	7
<b>4</b>	<b>Multiple importance sampling with optimal weights</b>	<b>10</b>
4.1	Optimal coefficients for an equal sample cost . . . . .	10
4.2	Optimal coefficients for non-equal sample cost . . . . .	13
4.3	Heuristic for weights . . . . .	14
4.4	Application to illumination by an environment map . . . . .	16
<b>5</b>	<b>Results</b>	<b>17</b>
5.1	Domain 1D . . . . .	17
5.2	Domain 2D . . . . .	17
5.3	Limitations . . . . .	22
<b>6</b>	<b>Conclusion and future work</b>	<b>23</b>
	<b>Bibliography</b>	<b>25</b>
	<b>Appendix A - Recapitulation of basic notions of random variable and estimators</b>	<b>28</b>
	<b>Appendix B - BRDF properties</b>	<b>29</b>
	<b>Appendix C - One-dimensional example</b>	<b>30</b>
	<b>Appendix D - 2D analytical example with Lafortune-Phong BRDF model</b>	<b>31</b>
	<b>Appendix E - Online computation</b>	<b>32</b>
	<b>Appendix F - Rendered images</b>	<b>33</b>
	<b>Appendix G - Notation</b>	<b>35</b>

# 1 Introduction

Since its introduction by Veach and Guibas [VG95], Multiple Importance Sampling (MIS) technique has been widely used in Computer Graphics, particularly in rendering. Its usefulness and efficiency in generating nice images has even been recognized by a 2014 Hollywood Academy award to Eric Veach. MIS is based on combining several sampling techniques into a single estimator. Veach proved that, from the several MIS techniques that he presented, the balance heuristic gives the least variance, supposing that the number of samples for each technique is the same. Veach showed that the balance heuristic was the same as the classic Monte Carlo estimator, using a mixture of techniques, where the weights of each technique in the mixture are proportional to the number of samples taken from the technique. He used an equal number of samples for each technique, and the balance heuristic was almost exclusively used thereafter as a combination with an equal number of samples, or equivalently, as a mixture of distributions with equal weights.

To the best of our knowledge, the question whether sampling other than equal sampling in MIS can give better variance, has only been raised in a recent paper by Lu et al. [LPG13], which uses rough Taylor second order approximation around a weight of 1/2 for the environment map problem, combining the two sampling techniques, BRDF and a sampling environment map, and in [CSKA01], where the cost of sampling is also taken into consideration when minimizing the variance of the combination of stochastic iteration and bidirectional path tracing. However, neither paper exploited to the full the fact that the MIS balance heuristic is a mixture of distributions technique.

The most interesting cases in rendering algorithms are the integral of the product of the functions. Based on the properties of the mixture of distributions, and in the usual importance sampling Monte Carlo techniques, we will obtain for this kind of integral, and for any number of product functions, an implicit closed formula that the different weights have to fulfill for optimality. In other words, we give a recipe for the number of samples to take for each strategy in order to minimize the variance. We also take into account the cost of each sampling technique. The resulting formula cannot be written in an explicit way, but we outline an iterative procedure for obtaining the optimal values. As this iterative procedure suffers from combinatorial explosion for increasing the number of functions in the product, we introduce a sound heuristic approximation formula to compute the distribution of samples from the sampling functions. Through 1-dimensional function examples and 2-dimensional function examples with illumination by an environment map, we show examples of the efficiency of our approach.

This technical report is further structured as follows. In the next section we survey related work. In section 3 we describe importance sampling and its use in the rendering equation, together with optimal weights for a linear combination of estimators. In section 4 we present the main contribution. We derive the optimal weights for MIS, together with an approximation of these weights in a simpler formula. In section 5 we provide the results for the sampling algorithm on an example application, the multiplication of BRDF and an environment map, as defined by the rendering equation. The last section 6 contains the conclusions, and is followed up by appendices containing selected theoretical proofs and exemplary results.

## 2 Related work

We review here the MIS algorithms for rendering, light transport, and other related work.

**MIS algorithms in rendering.** Since its introduction by Veach and Guibas [VG95] MIS algorithms, mainly with a balance heuristic, have been used for rendering. They were first used for Bidirectional Path Tracing by Veach and Guibas [VG94] themselves. Since then, they have found numerous uses in global illumination rendering algorithms, including those with scattering, in particular the use of the balance heuristic. Examples of recent publications using MIS are the work by van Antwerpen [vA11], by Bouchard et al. [BIOP13], the Photon Beam Diffusion method by Habel et al. [HCJ13].

**MIS optimization.** Csonka et al. [CSKA01] used MIS to combine bidirectional path tracing and ray-bundle based stochastic iteration, generalizing MIS to a sequence of integrals. As the computation cost and the contribution to the final result of a single integral vary, they introduced the concept of cost in MIS, approximated the variance, and minimized it through a gradient search method and Lagrange multipliers, with the cost as a constraint.

Lu et al. [LPG13], using the fact that balance heuristic MIS is importance sampling with a mixture of densities, approximated the variance for the product of two functions representing BRDF and an environment map, using a second order Taylor expansion around a value of  $1/2$ , which corresponds to an equal number of samples for the two coefficients. From this expansion they obtained optimal coefficients, but the farther these coefficients were from the  $1/2$  value, the worse the approximation was. The cost was not included in the scheme.

**Adaptive sampling schemes inspired by MIS.** There has been progress in research on adaptive sampling schemes, including those based on MIS. Among the advances that have been made, we would like to point here to the older work of Hesterberg [Hes88] introducing defensive importance sampling, which limits the variance explosion of importance sampling. The review article by Owen and Zhou [OZ00] surveys the principles in mixture and multiple importance sampling at that time. More recently, Doc et al. [DGMR07] derive sufficient convergence conditions for adaptive mixtures. Cornet et al. [CMMR12] (applied to final gathering in [TOS10]) present optimal recycling of past simulations in iterative adaptive multiple importance sampling algorithms. This work was extended by Marin et al. [MPS12].

The practitioners in computer graphics deal with the problem of optimization for importance sampling by many approaches, including ad-hoc formulae. Even for a simple case such as illuminating an environment map with importance sampling is needed and is shown to be efficient by Colbert [CPF10]. Their proposed solution is valid only for a single preselected BRDF model restricting the parameters to a one- or two-dimensional environment map and a single environment map. In contrast we stay in our approach with simple statistical tools.

**Light Transport Analysis.** Light transport analysis has made major progress in the last decade, and there have been a number of papers related to our approach. Ramamoorthi and Hanrahan [RH01] presented the signal processing framework for inverse rendering. Ramamoorthi et al. [RH04] then studied reflection for the environment map and formulated the product of the environment map and BRDF as a convolution, for which they used the Fourier domain. Another approach was taken by Durand et al. [DHS<sup>+</sup>05], where Fourier analysis is applied for different components of light paths. Ramamoorthi et al. [RMB07] went on to extend their

former approach [RH01] to first-order and second order analysis of lighting. Ramamoorthi et al. [RAMN12] studied combination of statistics and frequency analysis of visibility for soft shadows, following the note of Durand [Dur11], where he studied the relation between variance and the power spectrum. More recently, Subr and Kautz [SK13] studied stochastic sampling strategies using Fourier analysis to estimate the bias from the spectrum. Recently, Belcour et al. [BSS<sup>+</sup>13] have also accelerated computation in the time-space domain such as motion blur and depth of field through adaptive sampling and reconstruction based on predicting the anisotropy and the bandwidth of the integrand.

**Other literature.** We will not refer here to all related publications, including those referred to as double and triple product integration in rendering, rendering with many lights methods [DKH<sup>+</sup>13], density estimation, precomputed radiance transfer [Ram09], and importance sampling schemes [Die11]. Although they are all interesting and related to our approach via the rendering equation, they attack the problem from different perspectives.

Our work differs significantly from the adaptive MIS schemes and other papers. We limit ourselves to the product of any number of functions, but we provide the closed formula for the optimal weights of deterministic sampling densities for MIS with the balance heuristic that minimize the variance of the function product, including cases with a different sampling cost. This is highly useful in global illumination algorithms solving a rendering equation that contains the product of the functions.

### 3 Importance sampling

In this section, we briefly recall importance sampling and its application to rendering algorithms. Then, as an example, we study direct illumination of the surface from environment maps.

#### 3.1 Importance sampling for product of functions

For the purposes of this paper, we first recall the idea of importance sampling. The reader can consult the Appendix A for the basic notions of random variable, estimator and variance, and indeed the classic references on Monte Carlo [RK08, KW86].

When solving an integral by Monte Carlo,  $I = \int_{\Omega} h(x)dx$ , we first have to convert the integral to an expected value,  $I = E[h(X)/f(X)] = \int_{\Omega} (h(x)/f(x)) f(x)dx$ , where  $f(x)$  is a probability density function (abbreviated as pdf), and then we consider the estimator

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N \frac{h(x_i)}{f(x_i)}, \tag{3.1}$$

where  $x_i$  are  $N$  samples drawn from according to pdf  $f(x)$ . Using index  $i$  we stress that the estimator depends on  $N$ . We have that  $E[\hat{I}] = E[h(X)/f(X)] = I$ , where the symbol "hat" over a variable  $X$  means an estimator of  $X$ . The more similar the pdf  $f(x)$  is to  $h(x)$ , the smaller the variance will be. This is called importance sampling.

An unbiased estimator of the variance  $V[h(X)/f(X)]$  will then be:

$$V[\widehat{h(X)/f(X)}] = \frac{\sum_{i=1}^N \left( \frac{h(x_i)}{f(x_i)} - \left( \sum_{i=1}^N \frac{h(x_i)}{f(x_i)} \right) / N \right)^2}{(N - 1)} \tag{3.2}$$

And remembering that an estimator is the average sum of  $N$  independent identically distributed variables, variance  $V[\hat{I}]$  is then:

$$V[\hat{I}] = \frac{1}{N} V[h(X)/f(X)] \approx \frac{\sum_{i=1}^N \left( \frac{h(x_i)}{f(x_i)} - \left( \sum_{i=1}^N \frac{h(x_i)}{f(x_i)} \right) / N \right)^2}{N \cdot (N - 1)} \tag{3.3}$$

For an unbiased estimator, the expected value of the Mean Square Error (MSE) is equal to the variance.

As a particular case, and for the purposes of this paper, suppose we want to solve by MC the integral  $I = \int_{\Omega} h_1(x)h_2(x)dx$  where  $h_1(x)$  and  $h_2(x)$  are non-negative within the domain of integration. We can choose to do importance sampling on  $h_2(x)$ , for example. We need to compute the normalization factor  $s_2 = \int_{\Omega} h_2(x)dx$ , and then we have the pdf  $\frac{1}{s_2}h_2(x)$ . We can write  $I = s_2 \int_{\Omega} h_1(x) (h_2(x)/s_2) dx = s_2 E[h_1(x)] = E[s_2 h_1(x)]$ , and then using the estimator eq. 3.1 to compute the integral we obtain

$$I \approx \hat{I} = s_2 \frac{1}{N} \sum_{i=1}^N h_1(x_i) \tag{3.4}$$



Observe that the variance would then be:

$$V[s_2 h_1(X)] = (s_2)^2 V[h_1(X)] \quad (3.5)$$

and using the definition of variance (see Appendix A, eq. 6.3) we get:

$$\begin{aligned} V[\hat{I}] &= V[s_2 h_1(X)] \\ &= (s_2)^2 \left( \int_{\Omega} h_1(x)^2 (h_2(x)/s_2) dx - (E[h_1(X)])^2 \right) \end{aligned} \quad (3.6)$$

$$= s_2 \int_{\Omega} h_1(x)^2 h_2(x) dx - I^2 \quad (3.7)$$

as  $s_2 E[h_1(X)] = I$ . An unbiased estimator of the variance  $V[\hat{I}]$  when the samples are drawn according to  $h_2(x)$  is:

$$\begin{aligned} \widehat{V[\hat{I}]} &= \frac{\sum_{i=1}^N \left( s_2 h_1(x_i) - \left( \sum_{i=1}^N s_2 h_1(x_i) \right) / N \right)^2}{(N-1)} \\ &= (s_2)^2 \frac{\sum_{i=1}^N \left( h_1(x_i) - \left( \sum_{i=1}^N h_1(x_i) \right) / N \right)^2}{(N-1)} \end{aligned} \quad (3.8)$$

We can generalize eq. 3.7 to the product of any number of non-negative functions  $I = \int \Pi_i h_i(x) dx$ :

$$V_k[\hat{I}] = s_k \int_{\Omega} \left( \prod_{i \neq k} h_i^2(x) \right) h_k(x) dx - I^2, \quad (3.9)$$

where by  $V_k[I]$  we mean the variance when doing importance sampling with respect to the  $h_k(x)$  function and  $s_k$  is the corresponding normalization constant. Observe that all these importance sampling estimators are unbiased, and thus their expected value is  $I = \int \Pi_i h_i(x) dx$ , as we can always restrict the integration domain to non-zero values of the product of functions, i.e., where  $\Pi_i h_i(x) \neq 0$ .

## 3.2 Importance sampling in the rendering equation for illumination by an environment map

We refer the reader to the Appendix B for the definition of BRDF. The rendering equation [Kaj86] expresses the radiance with zero self-emission from a surface with normal vector  $\vec{n}$  as:

$$L(x, \omega_o) = \int_{\Omega} L(x, \omega_i) f_r(x, \omega_i, \omega_o) (\omega_i \cdot \vec{n}) d\omega_i \quad (3.10)$$

Without loss of generality and for the purposes of studying the variance, we restrict our study to the case when the incoming illumination is represented by an environment map with radiant intensity  $R(\omega_i)$  given by unit  $[W/sr]$ . Then the eq. 3.10 simplifies to:

$$L(x, \omega_o) = \int_{\Omega} R(\omega_i) f_r(x, \omega_i, \omega_o) (\omega_i \cdot \vec{n}) d\omega_i \quad (3.11)$$

Suppose we want to solve this equation by MC importance sampling. When we do importance sampling on incoming radiant intensity  $R(\omega_i)$ , then the estimator for the outgoing radiance becomes:

$$L(x, \omega_o) \approx R^* \frac{1}{N} \sum_{i=1}^N f_r(x, \omega_i, \omega_o)(\omega_i \cdot \vec{n}), \quad (3.12)$$

where the integrated radiant intensity is computed as:

$$R^* = \int_{\Omega} R(\omega_i) d\omega_i \quad (3.13)$$

Another possibility is to do importance sampling on  $f_r$  times cosine. We end up in the estimator for radiance as

$$L(x, \omega_o) \approx a(x, \omega_o) \frac{1}{N} \sum_{i=1}^N R(\omega_i), \quad (3.14)$$

where  $a(x, \omega_o)$  is the albedo of BRDF defined as:

$$a(x, \omega_o) = \int_{\Omega} f_r(x, \omega_i, \omega_o) \cdot (\omega_i \cdot \vec{n}) d\omega_i \quad (3.15)$$

### 3.3 Analysis for a constant environment map and arbitrary BRDF

Considering  $R(\omega_i) = \text{const.} = R$  the estimator in eq. 3.12 simplifies to:

$$L(x, \omega_o) \approx 2\pi \frac{R}{N} \sum_{i=1}^N f_r(x, \omega_i, \omega_o) \cdot (\omega_i \cdot \vec{n}) \quad (3.16)$$

as for eq. 3.13 the  $R^* = 2\pi$  for constant incoming radiance  $R = 1$ . We can then estimate the variance  $V[L(x, \omega_o)]$  of estimator eq. 3.16 and hence the noise in the image.

For the second estimator, the outgoing radiance from eq. 3.14 becomes:

$$L(x, \omega_o) = R \int_{\Omega} f_r(x, \omega_i, \omega_o) (\omega_i \cdot \vec{n}) d\omega_i = R \cdot a(x, \omega_o) \quad (3.17)$$

### 3.4 Analysis for a non-constant environment map and arbitrary BRDF

In order to simplify the analysis, we consider full visibility here, which does not restrict the generality of the formulae below. If the environment map is not constant,  $L(x, \omega_o)$  along a primary ray for the point on an object illuminated by the environment map can be computed with the two importance sampling estimators defined above, eq. 3.12 and eq. 3.14. Let us call them  $L_1$  and  $L_2$ . They correspond respectively to the expected value  $E[R^* f_r(x, \omega_i, \omega_o)(\omega_i \cdot \vec{n})]$  when using as pdf  $\frac{R(\omega_i)}{R^*}$ , and to the expected value  $E[a(x, \omega_o) R(\omega_i)]$  when using as pdf  $\frac{f_r(x, \omega_i, \omega_o) \cdot (\omega_i \cdot \vec{n})}{a(x, \omega_o)}$ , with integrated radiant intensity  $R^*$  equal to:

$$R^* = \int_{\Omega} R(\omega_i) d\omega, \quad (3.18)$$

where  $R(\omega_i)$  is the radiant intensity of environment map at direction  $\omega_i$ . The two estimators in eq. 3.12 and eq. 3.14 then become:

$$L_1(\widehat{x, \omega_o}) = R^* \frac{1}{N} \sum_{i=1}^N f_r(x, \omega_i, \omega_o)(\omega_i \cdot \vec{n}) \quad (3.19)$$

$$L_2(\widehat{x, \omega_o}) = a(x, \omega_o) \frac{1}{N} \sum_{i=1}^N R(\omega_i) \quad (3.20)$$

The variances  $V_1 = V[R^* f_r(x, \omega_i, \omega_o)(\omega_i \cdot \vec{n})]$ ,  $V_2 = V[a(x, \omega_o)R(\omega_i)]$  can be estimated by eq. 3.5 and eq. 3.9.

Using eq. 3.7, their exact values are equal to:

$$V_1 = R^* \int (f_r(x, \omega_i, \omega_o))^2 (\omega_i \cdot \vec{n})^2 R(\omega_i) d\omega - (L(x, \omega_o))^2 \quad (3.21)$$

$$V_2 = a(x, \omega_o) \int (R(\omega_i))^2 f_r(x, \omega_i, \omega_o) \cdot (\omega_i \cdot \vec{n}) d\omega - (L(x, \omega_o))^2 \quad (3.22)$$

Suppose we now want to use  $N$  samples distributed among the two estimators,  $n_1 + n_2 = N$ . Any convex combination of  $L_1$  and  $L_2$ ,  $\alpha L_1 + (1 - \alpha)L_2$  is an unbiased estimator of  $L(x, \omega_o)$ , i.e.,  $E[\alpha L_1 + (1 - \alpha)L_2] = \alpha E[L_1] + (1 - \alpha)E[L_2] = L(x, \omega_o)$ , as  $E[L_1] = E[L_2] = L(x, \omega_o)$ . Its variance is equal to:

$$\begin{aligned} V[\alpha L_1 + (1 - \alpha)L_2] &= \alpha^2 V[L_1] + (1 - \alpha)^2 V[L_2] \\ &= \frac{\alpha^2}{n_1} V_1 + \frac{(1 - \alpha)^2}{n_2} V_2 \end{aligned} \quad (3.23)$$

When we fix  $\alpha = const$  then with the method of Lagrange multipliers we can obtain the values  $n_1$  and  $n_2$  for minimum variance:

$$n_1 = \frac{N \cdot \alpha \cdot \sqrt{V_1}}{\alpha \sqrt{V_1} + (1 - \alpha) \sqrt{V_2}} \quad \text{and} \quad n_2 = \frac{N \cdot (1 - \alpha) \cdot \sqrt{V_2}}{\alpha \sqrt{V_1} + (1 - \alpha) \sqrt{V_2}}. \quad (3.24)$$

It can easily be shown that a similar result is also valid for a convex combination of  $M$  independent estimators, i.e.,

$$n_k = \frac{N \cdot \alpha_k \cdot \sqrt{V_k}}{\sum_k^M \alpha_k \sqrt{V_k}}, \quad (3.25)$$

where  $V_k$  is the variance of estimator  $k$ . The optimal variance will then be

$$\sum_k^M \alpha_k^2 \frac{V_k}{n_k} = \sum_k^M \alpha_k \frac{\sqrt{V_k} \left( \sum_k^M \alpha_k \sqrt{V_k} \right)}{N} = \frac{\left( \sum_k^M \alpha_k \sqrt{V_k} \right)^2}{N} \quad (3.26)$$

If we try to find the  $\alpha_k$  coefficients that minimize eq. 3.26, the only trivial solution is that all variances are equal,  $V_k = V$ , with minimum variance value  $V/N$ .

If the  $n_k$  values are known, using again Lagrange multipliers we find that the optimal combination is:

$$\alpha_k = \frac{n_k / V_k}{\sum_k^M n_k / V_k} \quad (3.27)$$

Observe that in general, for  $M$  estimators,  $\sum_k^M n_k/V_k$  is the inverse of the harmonic mean (defined as  $H(\{x_i\}) = \frac{M}{\sum_i^M 1/x_i}$ ) of the  $n_k/V_k$  values times  $M$ , as the harmonic mean  $H(\{V_k/n_k\})$  would be:

$$H(\{V_k/n_k\}) = \frac{M}{\sum_k^M n_k/V_k} \quad (3.28)$$

Thus we get the formula weighting coefficients  $\alpha_k$ :

$$\alpha_k = \frac{H(\{V_k/n_k\})n_k}{MV_k} \quad (3.29)$$

The optimal variance  $V_{min}$  can be then derived to be equal to

$$V_{min} = \frac{1}{\sum_k^M n_k/V_k} = \frac{H(\{V_k/n_k\})}{M} \quad (3.30)$$

The optimal combination when the variances are known is presented for example in [GD59].

We can try to minimize eq. 3.30 for a fixed budget of samples, i.e.,  $\sum_k^M n_k = N$  fixed, but the result is simply that we sample only the estimator with less variance.

Let us try to optimize at the same time  $\alpha_k$  and  $n_k$ . Using the method of Lagrange multipliers we have to derivate with respect to  $\alpha_k$  and  $n_k$  and we get:

$$\sum_k^M \alpha_k^2 \frac{V_k}{n_k} + \lambda_1 \left( \sum_k^M n_k - N \right) + \lambda_2 \left( \sum_k^M \alpha_k - 1 \right) \quad (3.31)$$

We obtain for all  $k$  two equations

$$2\alpha_k \frac{V_k}{n_k} + \lambda_2 = 0 \quad \text{and} \quad -\alpha_k^2 \frac{V_k}{n_k^2} + \lambda_1 = 0 \quad (3.32)$$

From both equations we get:

$$\frac{\alpha_k}{n_k} = -\frac{2\lambda_1}{\lambda_2} = \frac{1}{N} \quad (3.33)$$

i.e., for all  $k$ ,  $\frac{\alpha_k}{n_k}$  is constant. Substituting in eq. 3.32 we see that there is only a trivial solution when for all  $k$  the  $V_k = V$  are the same. In this case the variance would be as expected:

$$\sum_k^M \alpha_k^2 \frac{V_k}{n_k} = \sum_k^M \alpha_k \frac{V}{N} = \frac{V}{N} \quad (3.34)$$

The conclusions so far are that, if you have a fixed combination of weighting coefficients  $\{\alpha_k\}$ , then use eq. 3.25 to compute the count of samples. If you have a fixed count of samples for each estimator  $\{n_k\}$ , then use eq. 3.27 to compute the weighting coefficients. In that case, if you know in advance which is the most efficient estimator, only sample that estimator. If not, you can use batches of samples to estimate the variance of the different estimators and then combine them using eq. 3.27.

## 4 Multiple importance sampling with optimal weights

Veach and Guibas [VG95] introduced the Multiple Importance Sampling estimator, which allowed different sampling techniques to be combined in a novel way. They proved that the optimal heuristic MIS case was when the weights were proportional to the number of samples, a technique that was named the balance heuristic. Given  $M$  sampling techniques (i.e., pdf's)  $f_k(x)$ , if we sample  $n_k$  samples  $x_{k,i}$  from each,  $\sum_k n_k = N$ , the balance heuristic MIS estimator for  $I = \int g(x)dx$  is given by the sum

$$\hat{I} = \sum_k^M \sum_i^{n_k} \frac{g(x_{k,i})}{\sum_j n_j f_j(x_{k,i})} \quad (4.1)$$

Veach later observed (see [Vea97, section 9.2.2.1]) that this estimator can be written, for the case of the balance heuristic, as a standard Monte Carlo estimator, where the pdf  $f(x)$  is given by a deterministic mixture of distributions

$$\begin{aligned} \hat{I} &= \frac{1}{N} \sum_k^M \sum_i^{n_k} \frac{g(x_{k,i})}{\sum_j (n_j/N) f_j(x_{k,i})} \\ &= \frac{1}{N} \sum_k^M \sum_i^{n_k} \frac{g(x_{k,i})}{\sum_j \alpha_j f_j(x_{k,i})} = \frac{1}{N} \sum_l^N \frac{g(x_l)}{\sum_j \alpha_j f_j(x_l)}, \end{aligned} \quad (4.2)$$

where in the samples we have dropped the subindex  $k$  for the sampled technique, and

$$f(x) = \sum_k^M \alpha_k f_k(x), \quad \alpha_k = n_k/N, \quad \sum_k^M \alpha_k = 1 \quad (4.3)$$

This is, the estimator is  $\hat{I} = E[g(x)]$  when doing importance sampling with  $f(x) = \sum_k^M \alpha_k f_k(x)$ .

We will next analyze the variance of balance heuristic MIS, or equivalently, the variance of a mixture of distributions. Our interest is in obtaining the coefficients  $\alpha_k$  that give minimum variance (example problem shown in Figure 4.1). Observe that these coefficients determine the number of samples to be taken for each technique.

### 4.1 Optimal coefficients for an equal sample cost

Without loss of generality, we take the integral  $\mu = \int g(x)f(x)dx = \int g(x) (\sum \alpha_k f_k(x)) dx = \sum \alpha_k \mu_k$ , which is the expected value of  $g(x)$  according to pdf  $f(x)$ , and for all  $M$  sampling strategies  $\mu_k = \int g(x)f_k(x)dx = E_k[g(x)]$  is the expected value of  $g(x)$  when sampling from

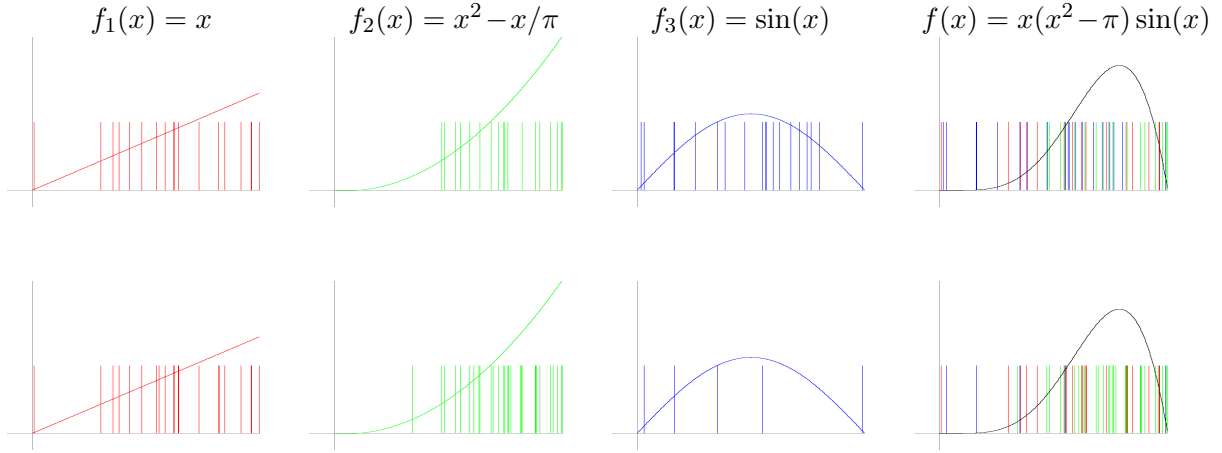


Figure 4.1: In the 1D example of a product of three functions on domain  $\langle 0, \pi \rangle$ , the count of samples can be taken (top row) equally  $\alpha_1 = \alpha_2 = \alpha_3 = \frac{1}{3}$ , (bottom row) non-equally  $\alpha_1 = 0.391, \alpha_2 = 0.520, \alpha_3 = 0.089$ , computed the supposed equal sample cost according to the new technique, to decrease the variance and improve the efficiency of the estimate of  $V[f(x)]$  by 138%. For illustrative purposes, 42 samples are shown in total.

$f_k(x)$ . The variance can be computed as:

$$\begin{aligned}
V[g(x)] &= \int g(x)^2 f(x) dx - \mu^2 & (4.4) \\
&= \int g(x)^2 \left( \sum_{k=1}^M \alpha_k f_k(x) \right) dx - \mu^2 \\
&= \sum_{k=1}^M \alpha_k \int g(x)^2 f_k(x) dx - \mu^2 \\
&= \sum_{k=1}^M \alpha_k (V_k[g(x)] + \mu_k^2) - \mu^2 \\
&= \sum_{k=1}^M \alpha_k (V_k[g(x)] + (\mu_k - \mu)^2)
\end{aligned}$$

The last equality comes from:  $\sum \alpha_k (-2\mu_k \mu + \mu^2) = -2\mu \sum \alpha_k \mu_k + \mu^2 = -\mu^2$ . Note that  $V_k[g(x)]$  is the variance of  $g(x)$  with respect to  $f_k(x)$  (samples are taken according to  $f_k(x)$ ).

Suppose we now want to compute the integral of the product of non-negative functions  $h_k(x)$  as  $I = \int \prod h_k(x) dx$ , for simplicity defined in the unit domain  $x \in \langle 0, 1 \rangle$ , using mixture function  $f(x) = \sum \alpha_k h_k(x) / s_k$ , where  $s_k$  are the normalization factors  $s_k = \int h_k(x) dx$ . We transform the integral into  $I = \int \frac{\prod h_k(x)}{\sum \alpha_k h_k(x) / s_k} (\sum \alpha_k h_k(x) / s_k) dx$  and thus  $I = E \left[ \frac{\prod h_k(x)}{\sum \alpha_k h_k(x) / s_k} \right]$ . Observe that  $g(x) = \frac{\prod h_k(x)}{\sum t_k h_k(x)}$  in eq. 4.4. When simplifying the notation by  $t_k = \alpha_k / s_k$  and using an auxiliary index  $j$  corresponding to  $k$  to distinguish between the sums, the variance of the

estimator is equal to:

$$\begin{aligned}
V \left[ \frac{\prod h_k(x)}{\sum t_k h_k(x)} \right] &= \sum_{j=1}^M \int \left( \frac{\prod h_k(x)}{\sum_k t_k h_k(x)} \right)^2 (t_j h_j(x)) dx - I^2 \\
&= \int \sum_{j=1}^M \frac{(\prod h_k(x))^2}{(\sum_k t_k h_k(x))^2} (t_j h_j(x)) dx - I^2 \\
&= \int \frac{\prod h_k^2(x) \left( \sum_{j=1}^M t_j h_j(x) \right)}{(\sum_k t_k h_k(x))^2} dx - I^2 \\
&= \int \frac{\prod h_k^2(x)}{\sum_k t_k h_k(x)} dx - I^2 \tag{4.5}
\end{aligned}$$

Suppose we now want to find the  $t_k$  that minimizes the variance (which will automatically give us optimal  $\alpha_k$  as  $s_k$  are fixed constants). We will use the Lagrange Multipliers method. For all  $j$  from 1 to  $M$ :

$$\frac{\partial \left( V \left[ \frac{\prod h_k(x)}{\sum t_k h_k(x)} \right] + \lambda (\sum_k t_k s_k - 1) \right)}{\partial t_j} = 0 \tag{4.6}$$

subjected to the constraint  $\sum_k t_k s_k = 1$  we have

$$\begin{aligned}
\frac{\partial V \left[ \frac{\prod h_k(x)}{\sum t_k h_k(x)} \right]}{\partial t_j} &= \frac{\partial \int \left( \frac{\prod h_k^2(x)}{\sum t_k h_k(x)} \right) dx - I^2}{\partial t_j} \\
&= \int \prod h_k^2(x) \frac{\partial}{\partial t_j} \left( \frac{1}{\sum t_k h_k(x)} \right) dx \\
&= - \int \frac{(\prod h_k^2(x)) h_j(x)}{(\sum t_k h_k(x))^2} dx \\
&= -s_j \int \frac{(\prod h_k^2(x)) (h_j(x)/s_j)}{(\sum t_k h_k(x))^2} dx \\
&= -s_j \left( E_j \left[ \left( \frac{\prod h_k(x)}{\sum t_k h_k(x)} \right)^2 \right] \right) \\
&= -s_j \left( V_j \left[ \frac{\prod h_k(x)}{\sum t_k h_k(x)} \right] + \mu_j^2 \right), \tag{4.7}
\end{aligned}$$

where  $V_j$  is the variance and  $E_j$  the expected value when sampling according to pdf  $h_j(x)/s_j$ . Thus

$$-s_j \left( V_j \left[ \frac{\prod h_k(x)}{\sum t_k h_k(x)} \right] + \mu_j^2 \right) + \lambda s_j = 0 \tag{4.8}$$

and then for all  $j$ ,

$$V_j \left[ \frac{\prod h_k(x)}{\sum t_k h_k(x)} \right] + \mu_j^2 = \lambda. \tag{4.9}$$

This means that the optimal combination of  $t_k$  (and thus  $\alpha_k$ ) is reached when all  $\lambda$  values are equal. Using eq. 4.4 we can see that the minimum variance will be

$$\begin{aligned} V \left[ \frac{\prod h_k(x)}{\sum t_k h_k(x)} \right] &= \sum \alpha_k \left( V_k \left[ \frac{\prod h_k(x)}{\sum t_k h_k(x)} \right] + \mu_k^2 \right) - \mu^2 \\ &= \sum \alpha_k \lambda - \mu^2 = \lambda - \mu^2 \end{aligned} \quad (4.10)$$

On the other hand, eq. 4.10 gives us the value of  $\lambda$ :

$$\lambda = V \left[ \frac{\prod h_k(x)}{\sum t_k h_k(x)} \right] + \mu^2 = E \left[ \left( \frac{\prod h_k(x)}{\sum t_k h_k(x)} \right)^2 \right] \quad (4.11)$$

We can easily check that when all  $h_k$  functions are constant (being thus  $h_k(x) = s_k$ ) except one, say  $h_m$ , for all  $j$  all expected values in eq. 4.9 are equal to  $\prod_k s_k^2$  for the combination  $\alpha_m = 1$  and  $\alpha_j = 0$  for  $j \neq m$ , thus we just have to sample from the non-constant function, as should be expected. Indeed, in this case the optimal variance is zero, as  $\mu = \prod_k s_k$  and  $\mu^2 = \lambda = \prod_k s_k^2$ .

Equation 4.9 implies that if all  $\lambda$  are equal, then for all pairs  $\{h_i(x), h_j(x)\}$  holds:

$$\int \left( \frac{\prod h_k(x)}{\sum t_k h_k(x)} \right)^2 (h_i(x)/s_i - h_j(x)/s_j) dx = 0 \quad (4.12)$$

This also tells us that in the trivial case all  $M$  functions  $h_k$  are the same, and any weighting scheme  $\{\alpha_k\}$  results in zero variance, as would be expected. Observe that, in general, the second factor in the integral,  $(h_i(x)/s_i - h_j(x)/s_j)$  can be positive or negative for different values of  $x$ , thus in general eq. 4.12 does not imply the particular case of equal functions.

A strategy for approximating the optimal  $\alpha_k$  values in general would run along the following lines. First, for each  $j$ , sample  $m_j$  samples from  $h_j(x)/s_j$ . For each  $j$ , with the  $m_j$  samples, compute eq. 4.9 for different combinations of  $\alpha_k$  values, using the estimator eq. 6.8. Observe that for each  $j$  you need to take the  $m_j$  samples only once and store them to test a possible reweighting scheme with any  $\{\alpha_k\}$ ,  $\sum \alpha_k = 1$ . For each different combination of  $\alpha_k$  values, the values of eq. 4.9 for all  $j$  are computed. The optimal  $\alpha_k$  values would be those that make all these values the most similar. This strategy could be organized by taking batches of samples to improve on the  $\{\alpha_k\}$  gradually.

## 4.2 Optimal coefficients for non-equal sample cost

The optimal  $\alpha_k$  values considered before do not take into account the possibility of different costs to sample for each pdf  $h_k(x)/s_k$ . Let us now consider this cost, and let  $c_k$  be this cost for one sample. The average cost per sample is  $\sum c_k \alpha_k = \sum c_k \alpha_k s_k / s_k = \sum c_k t_k s_k$ . Thus, taking into account the cost of sampling, the optimal  $\alpha_k$  values will come from minimizing the inverse of efficiency, the cost of the variance times, extending eq. 4.5

$$\left( \sum c_k t_k s_k \right) V \left[ \frac{\prod h_k(x)}{\sum t_k h_k(x)} \right] \quad (4.13)$$



subjected to the constraint  $\sum_k t_k s_k = 1$ . Using the Lagrange Multipliers method and auxiliary index  $j$  again, we get:

$$\frac{\partial \left( (\sum_k c_k t_k s_k) V \left[ \frac{\prod h_k(x)}{\sum_k t_k h_k(x)} \right] + \lambda (\sum_k t_k s_k - 1) \right)}{\partial t_j} = 0. \quad (4.14)$$

After derivation and factoring out normalization constants  $s_j$ , we obtain:

$$c_j V \left[ \frac{\prod h_k(x)}{\sum t_k h_k(x)} \right] - \left( \sum_k c_k t_k s_k \right) \left( V_j \left[ \frac{\prod h_k(x)}{\sum t_k h_k(x)} \right] + \mu_j^2 \right) = -\lambda. \quad (4.15)$$

which can also be written as

$$c_j V \left[ \frac{\prod h_k(x)}{\sum \frac{\alpha_k}{s_k} h_k(x)} \right] - \left( \sum_k c_k \alpha_k \right) E_j \left[ \left( \frac{\prod h_k(x)}{\sum \frac{\alpha_k}{s_k} h_k(x)} \right)^2 \right] = -\lambda \quad (4.16)$$

To minimize the variance in the given cost budget, and thus maximize the efficiency, for all  $j$  the quantity  $\lambda$  in eq. 4.16 has to be the same. By multiplying eq. 4.16 by  $\alpha_j$  summing over  $j$ , we can obtain the value of  $\lambda = (\sum_k c_k \alpha_k) \mu^2$ .

Observe that eq. 4.16 would express the same condition as in eq. 4.9 when all costs  $c_k = c$  are equal, as then the first term on the left side of eq. 4.16 does not depend on  $k$ , and  $\sum_k c_k \alpha_k$  reduces to  $c$ . Observe also that the value of  $\mu$  is independent of the  $\alpha_k$ 's, and that the expected values appearing in eq. 4.16 can be estimated in the same way as before for eq. 4.9, by sampling  $m_j$  samples for all  $j$  from  $h_j(x)/s_j$  pdf, while  $\mu^2$  can be estimated from estimating the  $\mu_k = E_k \left[ \frac{\prod h_k(x)}{\sum t_k h_k(x)} \right]$  and then from  $\mu^2 = (\sum_k \mu_k)^2$ .

Now we present a strategy for approximating the optimal  $\alpha_k$  coefficients. Assume sampling in  $B$  batches with an initial batch of  $N_0$  samples, so  $m_{j,0} = N_0/M$  (assuming batch  $b$  with  $N_b$  samples  $\sum_B N_b = N$  for the total number of samples  $N$ ). The sampling procedure for the  $b$ -th batch would be expressed by the Algorithm 1.

The batched procedure has time complexity for finding the optimal combination  $O(N)$  due to the inner loop in lines 5 to 12, which is computationally demanding; however, we find the optimal solution for the number of samples and the minimal variance of the estimate.

### 4.3 Heuristic for weights

The question now arises whether without reverting to the time consuming batching procedure presented in Algorithm 1 there is some heuristic that would directly approximate the optimal  $\alpha_k$  values. A hint is given by the equation 3.27. Thus we will consider the  $\alpha_k$  values inversely proportional to the variances of the estimators taking as pdf each  $h_k(x)/s_k$  in turn, respectively, i.e.,

$$\alpha_k = \frac{H(\{V_k\})}{M V_k}, \quad (4.17)$$

where  $V_k$  is from equation 3.9 computed as

$$V_k = s_k \int \left( \prod_{i \neq k} h_i(x) \right)^2 h_k(x) dx - I^2 \quad (4.18)$$

---

**Algorithm 1:** The sampling algorithm with optimal distribution of samples from  $M$  sampling strategies taking  $N$  samples in total in  $B$  batches, assuming  $U$  weighting combinations.

---

```

1 for  $b \leftarrow 1$  to  $B$  do
2   In the  $b$ -th batch for each  $j \in \langle 1, M \rangle$  select new  $m_{j,b}$  samples according to the
   distribution  $h_j(x)/s_j$ .
3   Compute total number of samples  $m'_j = \sum_{j=1}^b m_{j,b}$  taken for each  $h_j(x)$ .
4   Select  $U$  weighting combinations indexed by  $l$  with values  $\{\alpha_{k,l}\}$  so it gives the linear
   combination  $\sum_{k=1}^M \alpha_{k,l} = 1$  so that  $\alpha_{k,l} > 0$ , and set  $t_{k,l} = \alpha_{k,l}/s_k$ .
5   for  $l \leftarrow 1$  to  $U$  do
6     For  $l$ -th weighting combination
7     for  $j \leftarrow 1$  to  $M$  do
8       Compute  $\mu_{j,l} = E_j \left[ \frac{\prod h_k(x)}{\sum t_{k,l} h_k(x)} \right]$  and  $\nu_{j,l} = E_j \left[ \left( \frac{\prod h_k(x)}{\sum t_{k,l} h_k(x)} \right)^2 \right]$ , using all
       samples  $m'_j$  already taken for each  $h_j(x)$  for the previous sampling batches
       with estimators eq. 6.6 and eq. 6.8. (i.e. in eq. 6.6 and eq. 6.8
        $g(x) = \frac{\prod h_k(x)}{\sum t_{k,l} h_k(x)}$  and  $N = m'_j$ ).
9     Compute  $\mu_l^2 = \left( \sum_{j=1}^M \mu_{j,l} \right)^2$ .
10    for  $j \leftarrow 1$  to  $M$  do
11      Compute for the  $j$ -th sampling strategy and the  $l$ -th weighting combination
      the left side of eq. 4.16, let us call it  $\Lambda_{j,l}$ .
12      Normalize the vector  $\lambda_{j,l} = \frac{\Lambda_{j,l}}{\sum_j \Lambda_{j,l}}$ 
13    Select the index  $l_0$  such that the normalized vector given by  $\lambda_{j,l_0}$  has less distance to
    the uniform distribution. For example, select the vector with maximum entropy. The
    best combination is the combination corresponding to index  $l_0$ , i.e. the coefficients
     $\alpha_{k,l_0}$ .
14    Determine the number of samples  $N_{b+1}$  to be sampled in the next batch (e.g.
    uniformly so  $N_{b+1} \approx N/B$ ).
15    Compute the number of samples  $m_{j,b+1}$  for the next batch of  $N_{b+1}$  samples according
    to  $\alpha_{k,l_0}$  considering the number of already taken samples for each  $h_j(x)$  for all
    batches of samples already taken:  $\sum_{j=1}^M m_{j,b+1} = N_{b+1}$ .

```

---

Observe from eq. 4.17 that the less the variance of a technique is, the more we sample from that technique, remember from the definition 4.3, that the count of the samples is proportional to  $\alpha_k$ . Trivially, for any pair  $\{k, j\}$ ,  $V_k \neq 0, V_j \neq 0$ , we get  $\alpha_k/\alpha_j = V_k/V_j$ . Thus, when all variances  $V_k$  are equal so are all the  $\alpha_k$ .

Let us examine the limiting case, when the variance  $V_k$  is null. For this to happen, all functions  $h_{j \neq k}(x)$  have to be constant, as doing importance sampling then with pdf  $h_k(x)/s_k$  the integrand will be constant, and the variance  $V_k$  will be null. In that case, no other variance can be null (excluding the trivial case where  $h_k(x)$  is also constant), and, taking limits in eq. 4.17 for  $V_k \rightarrow 0$ ,  $\alpha_k = 1$ ,  $\alpha_{j \neq k} = 0$ .

The other limiting case, when some variance(s)  $V_m$  are very big with respect to the other variances, taking limits in eq. 4.17 for  $V_m \rightarrow \infty$ , we obtain  $\alpha_m = 0$ .

We can include the cost in a batching strategy to approximate the optimal  $\alpha_k$  coefficients. Suppose we include the cost  $c_k$  into equation for sampling efficiency as:

$$\alpha_k \propto \frac{1}{c_k \left( s_k \int \left( \prod_{i \neq k} h_i(x) \right)^2 h_k(x) dx - I^2 \right)} \quad (4.19)$$

then we extend similarly the formulae for weighting coefficients:

$$\alpha_k = \frac{H(\{c_k V_k\})}{M c_k V_k} \quad (4.20)$$

Indeed, the variances are not usually known in advance, so they have to be estimated using some batch of samples.

**Discussion.** Lu et al. [LPG13], using also the fact that balance heuristic MIS is importance sampling with a mixture of densities, approximate the variance for the product of two functions representing BRDF and an environment map, using a second order Taylor expansion around  $\alpha = 1/2$ . Thus the farther from this  $\alpha = 1/2$ , the worse the approximation. On the other hand, we have given here a closed formula for the variance of the product of any number of functions, a closed formula for its optimum, a procedure for approximating the optimal  $\alpha$  coefficients that is valid for any range of  $\alpha$  values, and heuristic formulae for the optimal  $\alpha$  values.

## 4.4 Application to illumination by an environment map

Applying the results of the previous sections to environment map illumination computation, the chosen weights for sampling the environment map and BRDF respectively, would be, considering eqs. 3.21 and 3.22:

$$\alpha_1 \propto \frac{1}{V_1} = \frac{1}{R^* \int (f_r(x, \omega_i, \omega_o))^2 (\omega_i \cdot \vec{n})^2 R(\omega_i) d\omega - (L(x, \omega_o))^2}$$

$$\alpha_2 \propto \frac{1}{V_2} = \frac{1}{a(x, \omega_o) \int (R(\omega_i))^2 f_r(x, \omega_i, \omega_o) \cdot (\omega_i \cdot \vec{n}) d\omega - (L(x, \omega_o))^2}$$

i.e.  $\alpha_1 = \frac{H(\{V_1, V_2\})}{2V_1}$  and  $\alpha_2 = \frac{H(\{V_1, V_2\})}{2V_2}$

and taking into account the cost of sampling:

$$\alpha_1 = \frac{H(\{c_1 V_1, c_2 V_2\})}{2c_1 V_1}, \alpha_2 = \frac{H(\{c_1 V_1, c_2 V_2\})}{2c_2 V_2}$$

## 5 Results

Without lack of generality, we tested the proposed adaptive sampling algorithm on 1D and 2D functions. We keep our analysis as simple as possible and restrict the numerical evaluation to the implementation of environment map illumination even, although MIS has been used in hundreds of applications in published papers for more involved problems, which are beyond the scope of this paper.

### 5.1 Domain 1D

For the 1D domain, we show an example with the product of three functions that is shown in Figure 4.1. The efficiency improvement for equal costs of the three sampling strategies when taking 1000 samples, where 200 samples were pilot samples, was +29%, in comparison with equal count sampling from all three strategies. The theoretical analysis is presented in Appendix C.

### 5.2 Domain 2D

For the 2D domain, we prepared two tests showing the results of our approach on the problem of illumination by an environment map. The first test is an analytical derivation of formulas in Mathematica, and corresponds to the illumination of an object with the Lafortune-Phong BRDF model with an environment map simplified to the function  $\cos \theta$ . It is described in Appendix D. The improvement in efficiency is by +138% compared to the equally distributed samples for the same cost of sampling from both strategies. For uneven costs of sampling from both strategies, where the ratio of costs was measured from C++ implementation of the algorithm as 1 to 10, the improvement in efficiency is +1462%.

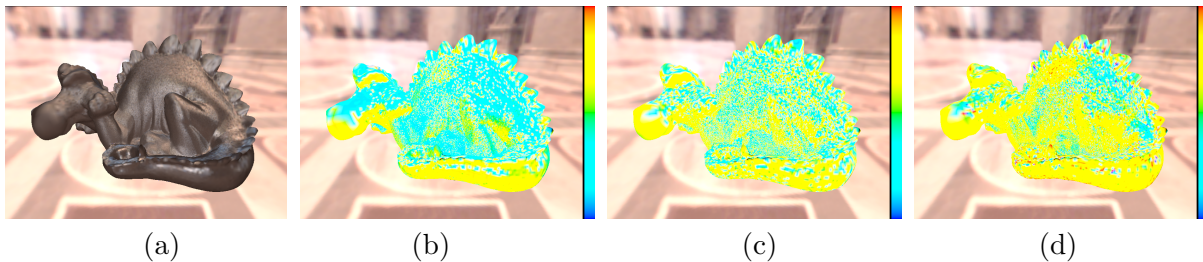


Figure 5.1: Per pixel sampling of an object covered by spatially varying BRDF (a) rendered image with 200 samples, (b) optimal selection of samples found by a combinatorial search in  $O(N^2)$  time, (c) our approximation to the exact values computed in  $O(1)$  time, (d) the previous method of Lu et al. [2013]. The distribution of samples is shown: deep blue color indicates sampling only from the environment map, red color only from BRDF.

The second test in the 2D domain was implemented in C++ for real environment map data and a broad range of BRDF values. The sampling is computed independently for each pixel with a different set of random values. The texture on an object is computed pixel by pixel and

is finally rendered by the OpenGL application. To demonstrate the results on a wide range of surface reflectances, we set up a scene with spatially varying BRDF properties to show many results in a single image. Using a single BRDF over a complex shape is much more restrictive and depicts a single situation even if surface normal is changing. We show an example of a computed image on a more complex 3D shape in Figure 5.1. Our algorithm also works in all the cases including complex 3D shapes, as samples for all the pixels are taken independently.

Our extended scene setup consists of a single rectangle or 3D objects that is put into the middle of a scene represented by an environment map. We used three environment maps with different ratios of variance and squared mean, which we call *City hall*, (low  $\sigma^2/\mu^2$ ), *Grace Cathedral* (moderate  $\sigma^2/\mu^2$ ), *St’Peters Dome* (high  $\sigma^2/\mu^2$ ). We have used the physically corrected Phong-Lafortune BRDF model [LW94b]  $f_r(\omega_o, \omega_i) = \rho_d/\pi + \rho_s \frac{n+2}{2\pi} \cdot \cos^n(\alpha)$ . We have been changing the exponent and the ratio between diffuse and specular albedo along the rectangle, while the albedo of the BRDF is set to constant. We have chosen this BRDF model since it allows for analytical importance sampling. The diffuse albedo  $\rho_d$  is interpolated in axis  $y$  on a rectangle by function  $\rho_d(y) = (1 - y)^2$  for range  $y \in \langle 0, 1 \rangle$ , the specular albedo is  $\rho_s = 1 - \rho_d$  (vertically, closest to the viewer, bottom side of image  $\rho_d = 0$  and the top side of image  $\rho_d = 1$  in Figure 5.2(a)). To depict interesting cases, we use mapping of the specular exponent of the Lafortune-Phong model shown in Figure 5.2(b) for  $x \in \langle 0, 1 \rangle$  using this formula  $n(x) = -1 - 0.111211/(0.1 \cdot x^{0.2} - 0.101101)$  that gives range  $\langle 0.1, 100 \rangle$ . The Figure 5.2 shows a visualization of varying the properties of Phong’s model with pseudocolor used for visualization, including the computed variance BRDF along the rectangle. Pseudo-color based on a rainbow is used for visualization, and the values are standardized between zero (blue) and one (red). The Figure 5.3 shows the converged rendered textures for the same rectangle with varying BRDF together, as illuminated by three environment maps in Figure 5.4. In addition, on the right side we also show images and a texture atlas for a more complex object, *Phlegmatic Dragon*, illuminated by *St’Peters Dome*, as shown in Figure 5.1.

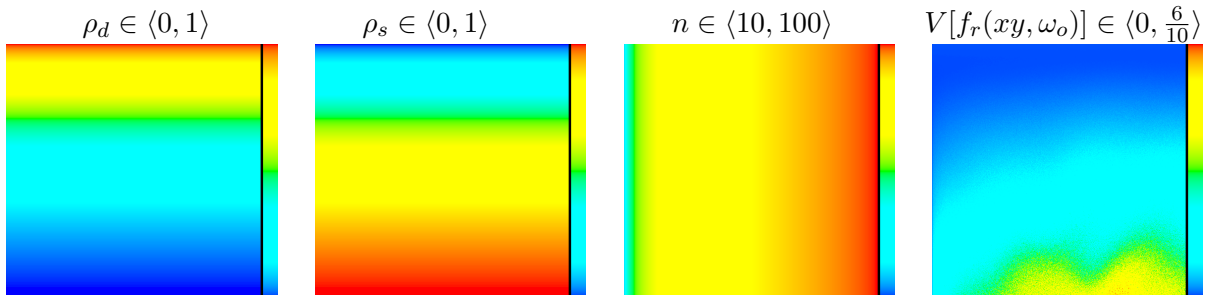


Figure 5.2: Lafortune-Phong BRDF model  $f_r(\omega_o, \omega_i) = \rho_d/\pi + \rho_s \frac{n+2}{2\pi} \cdot \cos^n(\alpha)$  spatially varying (left) diffuse albedo  $\rho_d$  in range  $\langle 0, 1 \rangle$ , (middle left) specular albedo  $\rho_s$  (albedo  $\rho_d + \rho_s = 1$ ), (middle right) specular exponent  $n$  in range  $\langle 0.1, 100 \rangle$ , (right) variance of this spatially varying BRDF example computed for 10,000 samples shown in range  $\langle 0, \frac{6}{10} \rangle$  when sampled from a constant environment map.

We have tested the methods for moderate number of samples  $n = 200$  to show visible differences and evaluated RMSE against the reference. The properties of three used environment maps are given in Table 5.1. We have normalized the power of all environment maps (the power after normalization is one) to make the results from rendered images mutually comparable. We

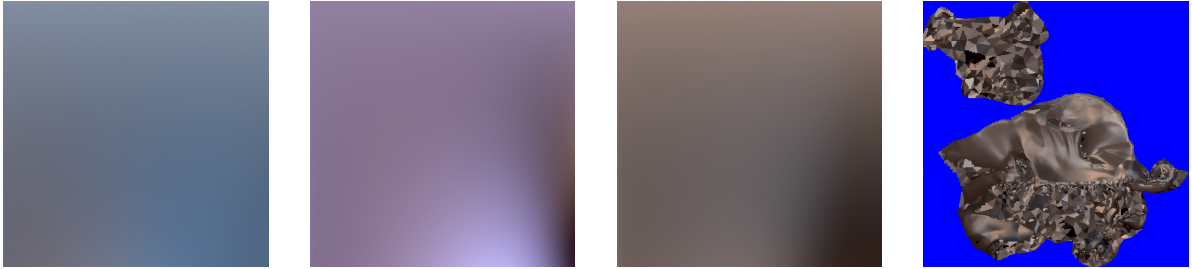


Figure 5.3: Converged results for texture on rectangle with varying BRDF and environment map for rectangle scene (left) illuminated by *City hall*, (middle left) by *Grace Cathedral*, (middle right) by *St'Peters Dome*, (right) texture on Phlegmatic dragon illuminated *St'Peters Dome*.

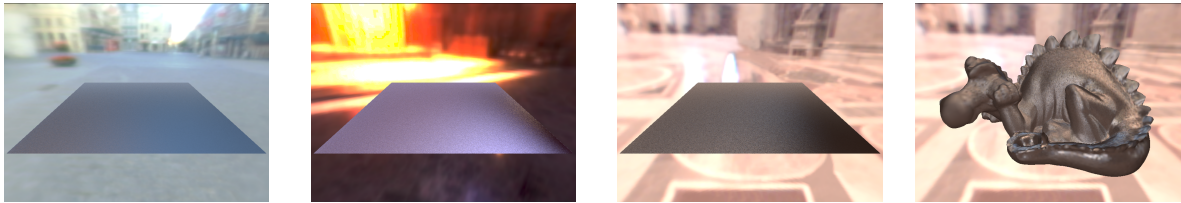


Figure 5.4: Showing the texture positioned in environment map, rendered by our adaptive sampling algorithm for 200 samples, (left) rectangle illuminated by *City hall* (middle left), rectangle illuminated *Grace Cathedral*, (middle right) rectangle illuminated by *St'Peters Dome*, (right) Phlegmatic dragon illuminated by *St'Peters Dome*.

also show the variances including the one of north (upper) hemisphere as the normal of the rectangle's surface is oriented towards this part of environment map.

The numerical results as RMSE against the converged reference images are summarized in Table 5.2. For our algorithm we set the number of samples in the pilot stage to 20%, where 20 (10%) samples were sampled according to BRDF and another 20 samples (10%) according to an environment map in the pilot stage of sampling. The computation after taking the pilot sampling was then organized in the next 4 sampling stages, taking 40 samples in each stage. Before each stage, the number of samples for this stage was computed using eq. 4.20 (for the same cost of sampling from both strategies so eq. 4.17) projected it to the end of the next sampling stage, taking into consideration all the history of sampling, including the pilot stage.

	<i>City hall</i>	<i>Grace Cathedral</i>	<i>St'Peters Dome</i>
$\sigma_S^2/\mu_S^2$	0.016	0.219	1.438
$\sigma_\Omega^2/\mu_\Omega^2$	1.613	26.159	154.727

Table 5.1: Properties of three environment maps. The symbol  $S$  stands for the whole sphere, symbol  $\Omega$  for upper hemisphere when computed value  $\sigma^2/\mu^2$ .

Although our algorithm selects the number of samples according to a heuristic, it decreases the variance both against standard MIS, taking an equal count of samples ( $\alpha_1 = \alpha_2 = 0.5$ ), and

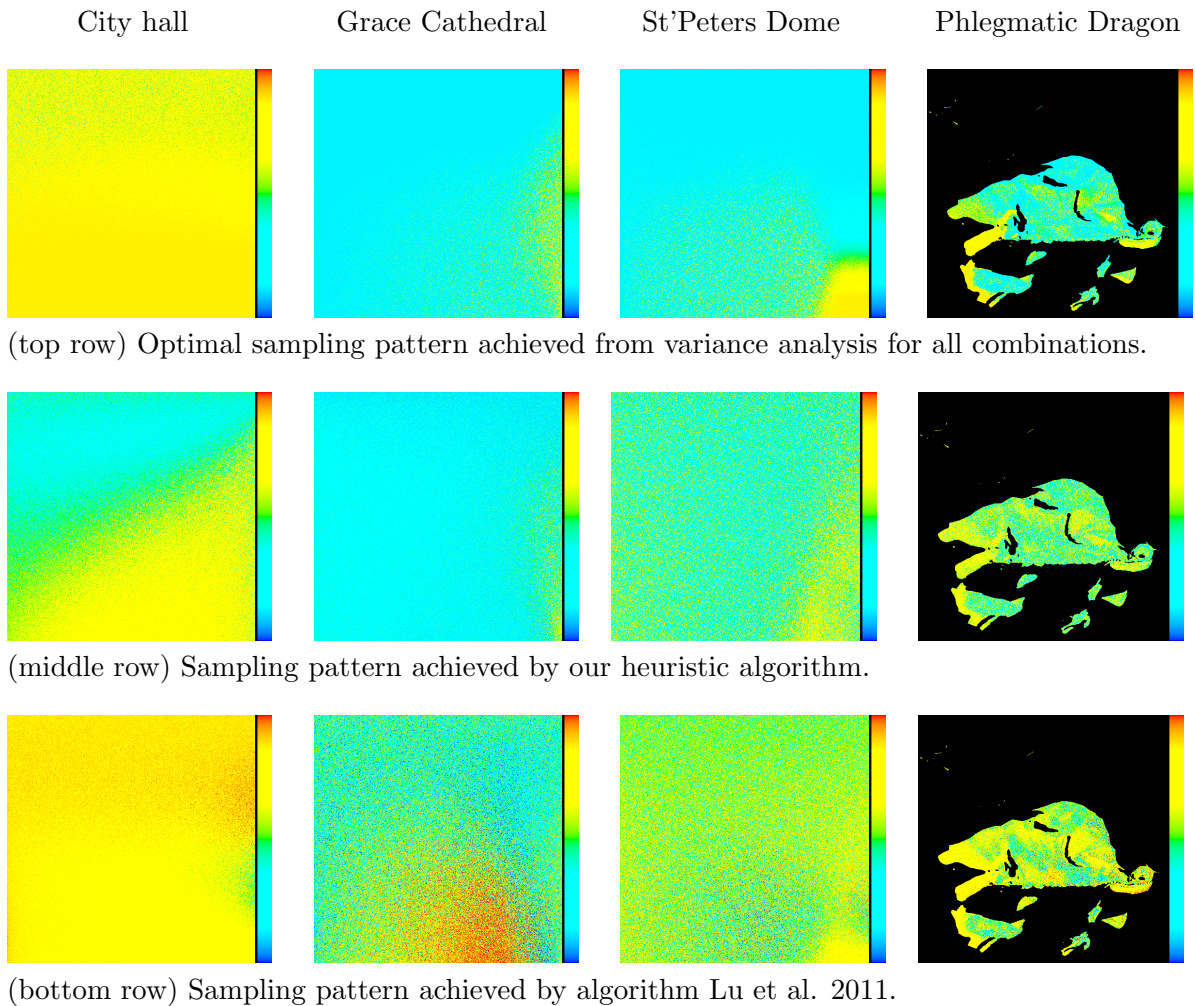


Figure 5.5: The sampling patterns produced by adaptive sampling algorithms for four data sets, samples distribution between environment map (blue) and BRDF (red). (Top row) statistically optimal sampling pattern for 200 samples, (middle row) sampling pattern found by our heuristic in Section 4.3, (bottom row) sampling pattern by the method of Lu et al. [2013]. Columns: (left) rectangle texture illuminated by *City hall*, (middle left) rectangle texture illuminated by *Grace Cathedral*, (middle right) rectangle texture illuminated *St'Peters Dome*, (right) Phlegmatic dragon illuminated by *St'Peters Dome*, with rendered images corresponding to Figure 5.1.

also against the recent paper by Lu et al. [LPG13], except the last setting. The optimal scheme was considered as taking an equal number of samples in the pilot stage of sampling, where in the pilot stage of sampling we took 20% (40) of all samples. Then by the combinatorial search the best possible weighting scheme was found by minimizing the estimated variance from all the 160 combinations at the end of sampling, which corresponds to eq. 4.16 (so in our case for the same cost of sampling eq. 4.11). This explains why in the case of *St'Peters Dome* and sampling according to environment map the variance is even lower (no samples from the  $BRDF \cdot \cos(\theta)$  were taken). However, any adaptive sampling algorithm has to take at least some samples from all sampled functions before it starts prioritize some sampling strategy.

	<i>City hall</i>	<i>Grace Cath.</i>	<i>St'Peters Dome</i>	<i>St'Peters Dome</i> Phlegmatic Dragon
	rectangle	rectangle	rectangle	
Sampling scheme	RMSE×1000			
Environment map	14.79	16.26	8.70	9.03
$f_r(x, \omega_v) \cos(\theta)$	7.87	45.42	88.04	53.64
MIS50-50 [VG95]	7.37	24.26	12.03	11.36
[LPG13]	7.25	25.32	13.86	13.14
Our scheme (eq. 4.17)	7.20	20.55	11.65	11.48
Optimal scheme (eq. 4.9)	7.12	21.15	11.25	12.69

Table 5.2: Numerical results for four rendered scenarios using two geometric data sets (rectangle and Phlegmatic Dragon 3D model) and three environment maps, the cost of sampling from both strategies is the same. The error of computed images for six sampling schemes for taking 200 samples are shown. Values of RMSE×1000 computed from RGB across the whole image against the reference converged image for each sampling scheme is reported (the lower value of RMSE the better sampling scheme).

We can observe in the images in Figure 5.5 that it is difficult to predict how to reach the optimal distribution of samples in advance. Our technique using not only the samples in the pilot stage but all samples taken up to the beginning of some sampling batch attempts to optimize the count of samples to the end of sampling. In this way, it closely approaches the optimal sampling distribution between strategies that has the lowest estimated variance.

The computation costs for standard MIS algorithm with equal counts and the proposed adaptive sampling MIS algorithm are almost the same. This is because the computation of variance and the mean in the new sampling scheme, and computing heuristically by eq. 4.20 how many samples there will be taken in the next stage of sampling presents only a negligible overhead in the whole algorithm, in comparison with the cost of sampling. This cost (in seconds) of sampling from BRDF ( $f_r(x, \omega_v) \cos(\theta)$ ) and sampling from the environment map was almost equal in our tests ( $cost_{BRDF} = 0.8cost_{EM}$ ), as we stored the precomputed lights to an array in preprocessing and randomized their selection during sampling. In this way, we avoided a double binary search (required for importance sampling by inverse transform method) and decreased the cost of sampling from the environment map. An equal cost of sampling from both strategies is the worst case for our scheme, but we still achieve an improvement. When a binary search for inverse transform method has to be done, in our implementation the cost of sampling according to BRDF is about five times less than the cost of sampling from environment map ( $cost_{BRDF} = 0.21cost_{EM}$ ), the efficiency improvements were then even higher.

**Discussion.** Other adaptive sampling schemes with MIS are possible, but our proposed method uses almost no memory and almost no computational overhead, and in addition it is accurate with respect to the theory presented in the previous section. The potential improvement in sampling efficiency for another application of a sampling scheme for the product of functions is therefore much higher for those sampling schemes where the sampling costs for



the functions in product differ. Different sampling costs can be included in a relatively simple way, as we have shown. The increase in the number of samples can even improve results of our technique as then the reliability of variance estimate is higher.

### 5.3 Limitations

We have not included the visibility in the variance analysis of the reflectance equation, for given example of environment map illumination, but it can be seen as another function in the product. The problem is that importance sampling can be done according to the visibility only using general methods such as Metropolis. This could be extremely costly: it may not pay off to sample according to the visibility in dependence on the sampling cost. Our analysis gives a lower bound of variance assuming a full visibility (so, trivially, the computed variance of visibility function is zero),

This technique requires an estimate of the number of samples, and also an estimate of the variances given by eqs. 4.11 or 4.16. This is computationally only a small overhead, but a reliable estimate requires more than just ten or twenty samples. This is the same as in the work of Lu et al. [LPG13], where the minimum number of samples in the pilot stage is 64, or in any other adaptive techniques [MPS12, CMMR12, DGMR07].

We show in Appendix E that the samples need not be stored to the memory for reweighting. The evaluation of variances and estimated mean can be organized by online and robust computation [Knu97] of the mean and variance, so that the results of reweighting the stored data are the same as for online computation.

## 6 Conclusion and future work

We have analyzed the Multiple Importance Sampling (MIS) estimator with balance heuristic. Using the fact that balance heuristic MIS is importance sampling with a mixture of densities, and considering the particular case of interest in rendering the integral of a product of functions, we have given for its MIS estimator a closed formula for the variance, a closed formula for the optimal weights, i.e., the distribution of sampling that optimizes the variance, a batch procedure to obtain them and, finally, a heuristic formula to approximate the optimal weights directly. We have also considered the cost of sampling in our formulation, and in this case we optimize the efficiency of the estimator. We have demonstrated our results with experiments on 1D functions and 2D functions, where as the application example we have taken the direct illumination of an object with spatially varying BRDF lit by an environment map. We have shown that the improvement in the sampling efficiency can be from negligible for some function settings to very high, reaching a factor of five to ten for the same running time compared to previous algorithms.

In future we plan to test our results in other scenarios, e.g. illumination by VPLs for a given random walk for many lights methods. In general, the scheme is easily applicable to many existing algorithms that use MIS and require a moderate to high number of samples.

# Acknowledgements

We would like to thank Paul Debevec for the environment maps (<http://www.pauldebevec.com/Probes/>) and the Phlegmatic Dragon project for the geometry (<http://dcgi.felk.cvut.cz/cgg/eg07/index.php?page=dragon>). This research has been supported by the Czech Science Foundation under projects No. P202/12/2413 and GA14-19213S and by the Grant Agency of the Czech Technical University in Prague, grant No. SGS13/214/OHK3/3T/13, and on Spanish side by grant TIN2013-47276-C6-1-R from the Spanish Government and by grant 2014 SGR 1232 from the Catalan Government and National Natural Science Foundation of China (Nos. 61372190 and 61331018).

# Bibliography

- [BIOP13] G. Bouchard, J.-C. Iehl, V. Ostromoukhov, and P. Poulin. Improving Robustness of Monte-Carlo Global Illumination with Directional Regularization. In ACM, editor, *SIGGRAPH-ASIA*, pages 22:1–22:4, December 2013.
- [BSS<sup>+</sup>13] L. Belcour, C. Soler, K. Subr, N. Holzschuch, and F. Durand. 5D Covariance Tracing for Efficient Defocus and Motion Blur. *ACM Trans. Graph.*, 32(3):31:1–31:18, July 2013.
- [CMMR12] J.-M. Cornuet, J.-M. Marin, A. Mira, and C. P. Robert. Adaptive Multiple Importance Sampling. *Scandinavian Journal of Statistics*, 39(4):798–812, 2012.
- [CPF10] M. Colbert, S. Premoze, and G. Francois. Importance Sampling for Production Rendering. In *ACM SIGGRAPH 2010 Course Notes*. ACM SIGGRAPH, 2010.
- [CSKA01] F. Csonka, L. Szirmay-Kalos, and G. Antal. Cost-Driven Multiple Importance Sampling for Monte-Carlo Rendering. Technical Report TR-186-2-01-19, Institute of Computer Graphics and Algorithms, Vienna University of Technology, October 2001.
- [DGMR07] R. Douc, A. Guillin, J.-M. Marin, and C. P. Robert. Convergence of Adaptive Mixtures of Importance Sampling Schemes. *The Annals of Statistics*, 35(1):420–448, 02 2007.
- [DHS<sup>+</sup>05] F. Durand, N. Holzschuch, C. Soler, E. Chan, and F. X. Sillion. A Frequency Analysis of Light Transport. *ACM Trans. Graph.*, 24(3):1115–1126, July 2005.
- [Die11] Dietger van Antwerpen. A Survey of Importance Sampling Applications in Unbiased Physically Based Rendering, June 2011. (unpublished manuscript).
- [DKH<sup>+</sup>13] C. Dachsbacher, J. Křivánek, M. Hašan, A. Arbree, B. Walter, and J. Novák. Scalable Realistic Rendering with Many-Light Methods. *Computer Graphics Forum*, 32(1), 2013.
- [Dur11] F. Durand. A Frequency Analysis of Monte-Carlo and other Numerical Integration Schemes. Technical Report MIT-CSAIL-TR-2011-052, Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02139 USA, December 2011.
- [GD59] F. A. Graybillk and R. Deal. Combining Unbiased Estimators. *Biometrics*, 15:543–550, Dec 1959.
- [HCJ13] R. Habel, P. H. Christensen, and W. Jarosz. Photon Beam Diffusion: A Hybrid Monte Carlo Method for Subsurface Scattering. *Computer Graphics Forum (Proceedings of EGSR 2013)*, 32(4):27–37, June 2013.

- [Hes88] T. C. Hesterberg. *Advances in Importance Sampling*. PhD thesis, Stanford University, 1988.
- [Kaj86] J. T. Kajiya. The Rendering Equation. In D. C. Evans and R. J. Athay, editors, *Computer Graphics (SIGGRAPH '86 Proceedings)*, volume 20, pages 143–150, August 1986.
- [Knu97] D. E. Knuth. *The Art of Computer Programming, Volume 2 (3rd Ed.): Seminumerical Algorithms*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1997.
- [KW86] M. Kalos and P. Whitlock. *Monte Carlo Methods: Basics*. Monte Carlo Methods. J. Wiley & Sons, 1986.
- [LPG13] H. Lu, R. Pacanowski, and X. Granier. Second-Order Approximation for Variance Reduction in Multiple Importance Sampling. *Computer Graphics Forum*, 32(7):131–136, 2013.
- [LW94a] E. P. Lafortune and Y. D. Willems. Using the Modified Phong BRDF for Physically Based Rendering. Technical Report CW197, Department of Computer Science, Katholieke Universiteit Leuven, Leuven, Belgium, November 1994.
- [LW94b] E. P. Lafortune and Y. D. Willems. Using the Modified Phong Reflectance Model for Physically Based Rendering. Technical Report report CW197, Dept. of Computer Science, K.U.Leuven, 1994.
- [MPS12] J.-M. Marin, P. Pudlo, and M. Sedki. Consistency of the Adaptive Multiple Importance Sampling. *Preprint arXiv:1211.2548*, November 2012.
- [NRH<sup>+</sup>77] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis. Geometric Considerations and Nomenclature for Reflectance. Monograph 161, National Bureau of Standards (US), October 1977.
- [OZ00] A. Owen and Y. Zhou. Safe and Effective Importance Sampling. *Journal of the American Statistical Association*, 95:135–143, 2000.
- [Ram09] R. Ramamoorthi. Precomputation-Based Rendering. *Found. Trends. Comput. Graph. Vis.*, 3(4):281–369, April 2009.
- [RAMN12] R. Ramamoorthi, J. Anderson, M. Meyer, and D. Nowrouzezahrai. A Theory of Monte Carlo Visibility Sampling. *ACM Trans. Graph.*, 31(5):121:1–121:16, September 2012.
- [RH01] R. Ramamoorthi and P. Hanrahan. A Signal-Processing Framework for Inverse Rendering. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques, SIGGRAPH '01*, pages 117–128, New York, NY, USA, 2001. ACM.
- [RH04] R. Ramamoorthi and P. Hanrahan. A Signal-Processing Framework for Reflection. *ACM Trans. Graph.*, 23(4):1004–1042, October 2004.

- [RK08] R. Rubinstein and D. Kroese. *Simulation and the Monte Carlo Method*. Wiley Series in Probability and Statistics. Wiley, 2008.
- [RMB07] R. Ramamoorthi, D. Mahajan, and P. Belhumeur. A First-order Analysis of Lighting, Shading, and Shadows. *ACM Trans. Graph.*, 26(1), January 2007.
- [SK13] K. Subr and J. Kautz. Fourier Analysis of Stochastic Sampling Strategies for Assessing Bias and Variance in Integration. *ACM Trans. Graph.*, 32(4):128:1–128:12, July 2013.
- [TOS10] Y. Tokuyoshi, S. Ogaki, and S. Sebastian. Final Gathering Using Adaptive Multiple Importance Sampling. In *ACM SIGGRAPH ASIA 2010 Posters*, SA '10, pages 47:1–47:1, New York, NY, USA, 2010. ACM.
- [vA11] D. van Antwerpen. Improving SIMD Efficiency for Parallel Monte Carlo Light Transport on the GPU. In *Proceedings of the ACM SIGGRAPH Symposium on High Performance Graphics*, HPG '11, pages 41–50, New York, NY, USA, 2011. ACM.
- [Vea97] E. Veach. *Robust Monte Carlo Methods for Light Transport Simulation*. PhD thesis, Stanford University, 1997.
- [VG94] E. Veach and L. J. Guibas. Bidirectional Estimators for Light Transport. In *Proceedings of the 5th Eurographics Workshop on Rendering*, pages 147–162, 1994.
- [VG95] E. Veach and L. J. Guibas. Optimally Combining Sampling Techniques for Monte Carlo Rendering. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, SIGGRAPH '95, pages 419–428, New York, NY, USA, 1995. ACM.

## Appendix A - Recapitulation of basic notions of random variable and estimators

We recall here the basic statistics and Monte Carlo estimation concepts that are needed in this paper. The reader can consult the classic references on Monte Carlo [RK08, KW86].

**Random Variable.** Given a random variable  $X$  over a domain  $\Omega$ , distributed according to probability distribution function pdf  $f(X)$ , its mean or expected value  $E[X]$  is the integral  $\int_{\Omega} xf(x)dx$ . The expected value of a function  $g(X)$  of random variable  $X$  (which in turn is a random variable) is  $E[g(X)] = \int_{\Omega} g(x)f(x)dx$ . The expected value of a sum of random variables is the sum of expected values, and the product with a constant  $k$  is the constant times the expected value, i.e.,  $E[k \sum g_i(X_i)] = k \sum E[g_i(X_i)]$ . We will use below the particular case when  $k = 1/N$ , and all random variables are identically distributed. Then we have:

$$\begin{aligned} E\left[\frac{1}{N} \sum_1^N g(X_i)\right] &= \frac{1}{N} \sum_1^N E[g(X_i)] \\ &= \frac{1}{N} NE[g(X)] = E[g(X)] \end{aligned} \quad (6.1)$$

**Variance.** The variance  $V[X]$  of distribution  $X$  is defined as

$$\begin{aligned} V[X] &= E[(X - E[X])^2] = E[X^2] - (E[X])^2 \\ &= \int_{\Omega} x^2 f(x)dx - (E[X])^2. \end{aligned} \quad (6.2)$$

In general, the variance  $V[g(X)]$  of function  $g(X)$  is defined as

$$\begin{aligned} V[g(X)] &= E[(g(X) - E[g(X)])^2] = E[g(X)^2] - (E[g(X)])^2 \\ &= \int_{\Omega} g(x)^2 f(x)dx - (E[g(X)])^2 \end{aligned} \quad (6.3)$$

The variance of a sum of independent random variables is the sum of the variances, and the product with a constant  $k$  is the squared constant times the variance, i.e.,  $V[k \sum g_i(X_i)] = k^2 \sum V[g_i(X_i)]$ . Thus, if all random variables are independent and identically distributed, then:

$$\begin{aligned} V\left[\frac{1}{N} \sum_1^N g(X_i)\right] &= \frac{1}{N^2} \sum_1^N V[g(X_i)] = \frac{1}{N^2} NV[g(X)] \\ &= \frac{1}{N} V[g(X)] \end{aligned} \quad (6.4)$$

An estimator of a quantity is a random variable such that its expected value is equal to that the quantity. A *biased estimator* is when the expected value differs by a *bias* from this quantity. An *unbiased estimation* of the mean value  $E[g(X)]$  is obtained by drawing  $N$  samples according to  $X$ :

$$E[g(X)] \approx \frac{1}{N} \sum_1^N g(x_i) \quad (6.5)$$

Observe that the estimator is just the average of the sum of  $N$  random variables identically distributed:

$$E[\widehat{g(X)}] = \frac{1}{N} \sum_1^N g(X_i). \quad (6.6)$$

Clearly  $E[E[\widehat{g(X)}]] = E[g(X)]$ . We use a "hat" symbol to differentiate the estimator from the quantity being estimated. In the definition of successive estimators we will use the samples directly, instead of the random variables that they are sampled from.

An *unbiased estimator* of the variance  $V[g(X)]$  is as follows:

$$V[\widehat{g(X)}] = \frac{\sum_{i=1}^N \left( g(x_i) - \left( \sum_{i=1}^N g(x_i) \right) / N \right)^2}{N - 1} \quad (6.7)$$

But observe that an estimator for  $E[g(X)^2]$  is

$$E[\widehat{g^2(X)}] = \frac{1}{N} \sum_{i=1}^N g^2(x_i) \quad (6.8)$$

## Appendix B - BRDF properties

BRDF (i.e. the bidirectional reflectance distribution function) is defined by Nicodemus [NRH<sup>+</sup>77] by means of the reflectance equation:

$$f_r(x, \omega_o, \omega_i) = \frac{dL(x \rightarrow \omega_o)}{L(x \leftarrow \omega_i) \cdot \cos(\theta_i) d\omega_i} = \frac{dL(x \rightarrow \omega_o)}{L(x \leftarrow \omega_i) \cdot (\omega_i \cdot \vec{n}) d\omega_i} \quad (6.9)$$

In this definition it fulfills Helmholtz reciprocity, given as:  $f_r(x, \omega_o, \omega_i) = f_r(x, \omega_i, \omega_o)$

Observe that  $f_r(x, \omega_o, \omega_i)$  is not in principle a probability density function (as it does not integrate to 1 with incoming directions  $\omega_i$ , but see below), it is unitless (i.e. unit is [sr<sup>-1</sup>], but this is still unitless as a steradian is unitless). To obtain a probability density function (pdf) from the BRDF, it should be redefined using the reflectance equation. We can define it in the following way, for fixed  $\omega_o$  and integrating over incoming directions  $\omega_i$ .

$$\text{pdf}(x, \omega_o, \omega_i) = \frac{1}{a(x, \omega_o)} \cdot f_r(x, \omega_i, \omega_o) \cdot (\omega_i \cdot \vec{n}), \quad (6.10)$$

where albedo  $a(x, \omega_o)$  for fixed outgoing direction  $\omega_o$  is defined as:

$$a(x, \omega_o) = \int_{\Omega} f_r(x, \omega_i, \omega_o) \cdot (\omega_i \cdot \vec{n}) d\omega_i \quad (6.11)$$

And then:

$$\int_{\Omega} \text{pdf}(x, \omega_o, \omega_i) d\omega_i = \frac{1}{a(x, \omega_o)} \int_{\Omega} f_r(x, \omega_i, \omega_o) \cdot (\omega_i \cdot \vec{n}) d\omega_i = 1 \quad (6.12)$$



## Appendix C - One-dimensional example

Suppose we want to solve the integral  $I = \int_0^\pi x \left(x^2 - \frac{x}{\pi}\right) \sin(x) dx = 10.2884$  by MIS sampling on functions  $x$ ,  $\left(x^2 - \frac{x}{\pi}\right)$  and  $\sin(x)$  respectively. We first find the normalization constants:  $\int_0^\pi x dx = \pi^2/2$ ,  $\int_0^\pi \left(x^2 - \frac{x}{\pi}\right) dx = 8.764$ ,  $\int_0^\pi \sin(x) dx = 2$ . Then we find the three variances when doing importance sampling with all three pdfs, respectively:

$$V_1 = \pi^2/2 \int_0^\pi x \left(x^2 - \frac{x}{\pi}\right)^2 \sin^2(x) dx - I^2 = 29.7928$$

$$V_2 = 8.764 \int_0^\pi x^2 \left(x^2 - \frac{x}{\pi}\right) \sin^2(x) dx - I^2 = 23.4828$$

$$V_3 = 2 \int_0^\pi x^2 \left(x^2 - \frac{x}{\pi}\right)^2 \sin(x) dx - I^2 = 123.896$$

Thus, according to eq. 4.18 and applying eq. 3.29, with all  $n_i = 1$ , we get, with  $H(\{V_1, V_2, V_3\}) = 35.6206$ ,

$$\begin{aligned} \alpha_1 &= \frac{H(\{V_1, V_2, V_3\})}{3V_1} = 0.398538 \\ \alpha_2 &= \frac{H(\{V_1, V_2, V_3\})}{3V_2} = 0.505627 \\ \alpha_3 &= \frac{H(\{V_1, V_2, V_3\})}{3V_3} = 0.0958351 \end{aligned}$$

Remember now from eq. 4.5 that the variance of MIS is given by the following expression:

$$V(\alpha_1, \alpha_2, \alpha_3) = \int_0^\pi \frac{x^2 \left(x^2 - \frac{x}{\pi}\right)^2 \sin^2(x)}{\frac{2\alpha_1 x}{\pi^2} + \frac{\alpha_2}{8.764} \left(x^2 - \frac{x}{\pi}\right) + \frac{1}{2}\alpha_3 \sin(x)} dx - I^2$$

Substituting the  $\alpha$  values found above, we have that  $V(0.398538, 0.505627, 0.0958351) = 25.803$ . On the other hand,  $V(1/3, 1/3, 1/3) = 33.4113$ , and there is a gain of 30%. By iterating over all possible  $\alpha_i$  values with step 0.01 we can find that the minimum variance value is 23.247, very near to our heuristically computed variance 25.803. We can check for heuristically computed weighting coefficients  $\{0.398538, 0.505627, 0.0958351\}$  above the value of eq. 4.11:

$$\int_0^\pi \frac{x^2 \left(x^2 - \frac{x}{\pi}\right)^2 \sin^2(x)}{\frac{2\alpha_1 x}{\pi^2} + \frac{\alpha_2}{8.764} \left(x^2 - \frac{x}{\pi}\right) + \frac{1}{2}\alpha_3 \sin(x)} dx = 131.654$$

and the three variances according to eq. 4.9:

$$\int_0^\pi \frac{x^2 \left(x^2 - \frac{x}{\pi}\right)^2 \sin^2(x)}{\left(\frac{2\alpha_1 x}{\pi^2} + \frac{\alpha_2}{8.764} \left(x^2 - \frac{x}{\pi}\right) + \frac{1}{2}\alpha_3 \sin(x)\right)^2} \frac{2x}{\pi^2} dx = 128.466$$

$$\int_0^\pi \frac{x^2 \left(x^2 - \frac{x}{\pi}\right)^2 \sin^2(x)}{\left(\frac{2\alpha_1 x}{\pi^2} + \frac{\alpha_2}{8.764} \left(x^2 - \frac{x}{\pi}\right) + \frac{1}{2}\alpha_3 \sin(x)\right)^2} 8.764 \left(x^2 - \frac{x}{\pi}\right) dx = 137.427$$

$$\int_0^\pi \frac{x^2 \left(x^2 - \frac{x}{\pi}\right)^2 \sin^2(x)}{\left(\frac{2\alpha_1 x}{\pi^2} + \frac{\alpha_2}{8.764} \left(x^2 - \frac{x}{\pi}\right) + \frac{1}{2}\alpha_3 \sin(x)\right)^2} \frac{1}{2} \sin(x) dx = 114.445$$

Remember that for the optimal case all four values should be equal. We can see that the discrepancy is relatively small for the heuristically obtained  $\alpha$  values. On the other hand, for  $\alpha$  values  $\{1/3, 1/3, 1/3\}$  the integrals evaluate to 139.262 and  $\{143.513, 162.005, 112.267\}$  with a bigger discrepancy.

Let us now introduce the sampling costs. Suppose that the relative costs are  $c_1 = 1$ ,  $c_2 = 6.24$ ,  $c_3 = 3.28$ . These costs are taken from measurement of an implementation in C++. Using eq. 4.20, the new optimal values of weighting coefficients are found to be  $\{0.783361, 0.159209, 0.0574306\}$ . Although the variance now increases a little  $V(0.783361, 0.159209, 0.0574306) = 28.9247$ , the relative efficiency increases from 1.2947 for equal sampling costs to 2.06141, when taking into account the different sampling costs. This presents doubling of sampling efficiency with a negligible computation overhead.

## Appendix D - 2D analytical example with Lafortune-Phong BRDF model

Let us consider here the formula for the physically-based variant of the Phong model presented by Lafortune [LW94a]. We assume that the whole lobe is above the surface. For simplicity, we will assume the case when the outgoing direction  $\omega_o$  is along the surface normal.

$$f_r(\omega_o, \omega_i) = \rho_d/\pi + \rho_s \frac{n+2}{2\pi} \cos^n(\theta) \quad (6.13)$$

Let us also consider an environment map given by  $R(\omega) = R \cdot \cos^k(\theta)$ , where  $R$  and  $k$  are constants. As  $d\omega = \sin\theta d\theta d\phi$  the outgoing radiance  $L(\omega_o)$  is then given by the integral:

$$L(\omega_o) = R \int_0^{2\pi} \int_0^{\pi/2} \left( \rho_d/\pi + \rho_s \frac{n+2}{2\pi} \cos^n(\theta) \right) \cos\theta \cos^k(\theta) \sin\theta d\theta d\phi \quad (6.14)$$

and integrating over  $\phi$

$$L(\omega_o) = 2\pi R \int_0^{\pi/2} \left( \rho_d/\pi + \rho_s \frac{n+2}{2\pi} \cos^n(\theta) \right) \cos^k(\theta) \cos\theta \sin\theta d\theta \quad (6.15)$$

Changing variables  $x = \cos\theta$  (so  $dx = -\sin\theta d\theta$ ) we get by substitution:

$$L(\omega_o) = 2\pi R \int_0^1 \left( \left( \rho_d/\pi + \rho_s \frac{n+2}{2\pi} x^n \right) x \right) x^k dx \quad (6.16)$$

We thus have the integral of the product  $h_1(x)h_2(x)$ , where  $h_1(x)$  corresponds to the BRDF times cosine

$$h_1(x) = \left( \rho_d/\pi + \rho_s \frac{n+2}{2\pi} x^n \right) x,$$

and  $h_2(x) = x^k$  is the environment map.

We can then find the optimal combination and the efficiency with respect to equal coefficients. For typical values,  $n = 5$ ,  $\rho_d = \rho_s = 0.5$ , the relative efficiency for the equal sampling cost is 2.38, with optimal coefficients  $\alpha_1 = 0.89$ ,  $\alpha_2 = 0.11$ . If we suppose the environment map is very costly to sample, for example  $c_1 = 1, c_2 = 10$ , then the relative efficiency is 15.62, with optimal coefficients  $\alpha_1 = 0.99$ ,  $\alpha_2 = 0.01$ .

Solving the integral, we get then value for outgoing radiance equal to:

$$\begin{aligned} L(\omega_o) &= 2\pi R \int_0^1 \left( (\rho_d/\pi + \rho_s \frac{n+2}{2\pi} x^n) x \right) x^k dx \\ &= R \left[ \left( 2\rho_d \frac{x^{k+2}}{k+2} + \rho_s (n+2) \frac{x^{n+k+2}}{n+k+2} \right) \right]_0^1 \\ &= R \left( \frac{2\rho_d}{k+2} + \rho_s \frac{n+2}{n+k+2} \right) \end{aligned} \quad (6.17)$$

This is useful for debugging of a real implementation, when we compare the results obtained by numerical integration against the analytically computed values for arbitrary  $\rho_d$ ,  $\rho_s$ ,  $n$ , and  $k$ .

## Appendix E - Online computation

Instead of using the two-pass algorithm, the online algorithm is often used for computing both the sample and the population variance, e.g. Knuth's one-line algorithm [Knu97] for the sample variance:

$$V[X] \approx \frac{\sum_{i=1}^N \left( g(x_i) - \frac{1}{N} \sum_{i=1}^N g(x_i) \right)^2}{N-1} \quad (6.18)$$

$$= \frac{1}{N-1} \sum_{i=1}^N g^2(x_i) - \frac{N}{N-1} \left( \frac{1}{N} \sum_{i=1}^N g(x_i) \right)^2 \quad (6.19)$$

The algorithm for the unbiased estimator of the sample variance and mean for function  $g(x)$  given random samples  $x_i$  is then as follows:

```

i = 0;
mean = 0;
M2 = 0;
for each g(x_i) in data:
    i = i + 1;
    delta = g(x_i) - mean;
    mean = mean + delta/i;
    M2 = M2 + delta.(g(x_i) - mean);

```

```

end
// It returns the sample variance and the mean value
return {variance = M2/(i - 1), mean}

```

On the practical side, it is easy to compute both the mean and the variance incrementally with almost zero storage. This is also suited for a stream-based computational model architecture such as GPUs. It is necessary to use this variance computation algorithm in order to make the algorithm for estimating the integral efficient. Its computation is trivial, since:  $V[X] = E[X^2] - E[X]^2$ , so computing  $E[X^2] = V[X] + E[X]^2$ . Further, multiplication by factor  $\frac{N-1}{N}$  is required to get the unbiased estimate of  $E[X^2]$  in the form of eq. 6.8, when the online one pass algorithm above is used to compute population variance instead of sample variance.

## Appendix F - Rendered images

We show the rendered images by different sampling strategies including the technique presented here in the results section in Figure 6.1 for 200 samples for sampling methods (a) according to  $\text{BRDF} \cdot \cos(\theta)$ , (b) according to environment map, (c) multiple importance sampling with the balance heuristic  $\alpha_1 = \alpha_2 = 1/2$  from both  $\text{BRDF} \cdot \cos(\theta)$  and environment map, (d) the method of Lu et al. 2013 computing  $\alpha$  by Taylor expansions around  $1/2$ , (e) our MIS method for equal sample cost, (f) the optimal MIS method that combinatorially explores all weighting combinations from 160 combinations (considering to take 20% samples in a pilot stage), and (g) the MIS ( $\alpha_1 = \alpha_2 = 1/2$ ) the converged image for  $10^5$  samples. The numerical results as RMSE values are shown in Table 5.2.

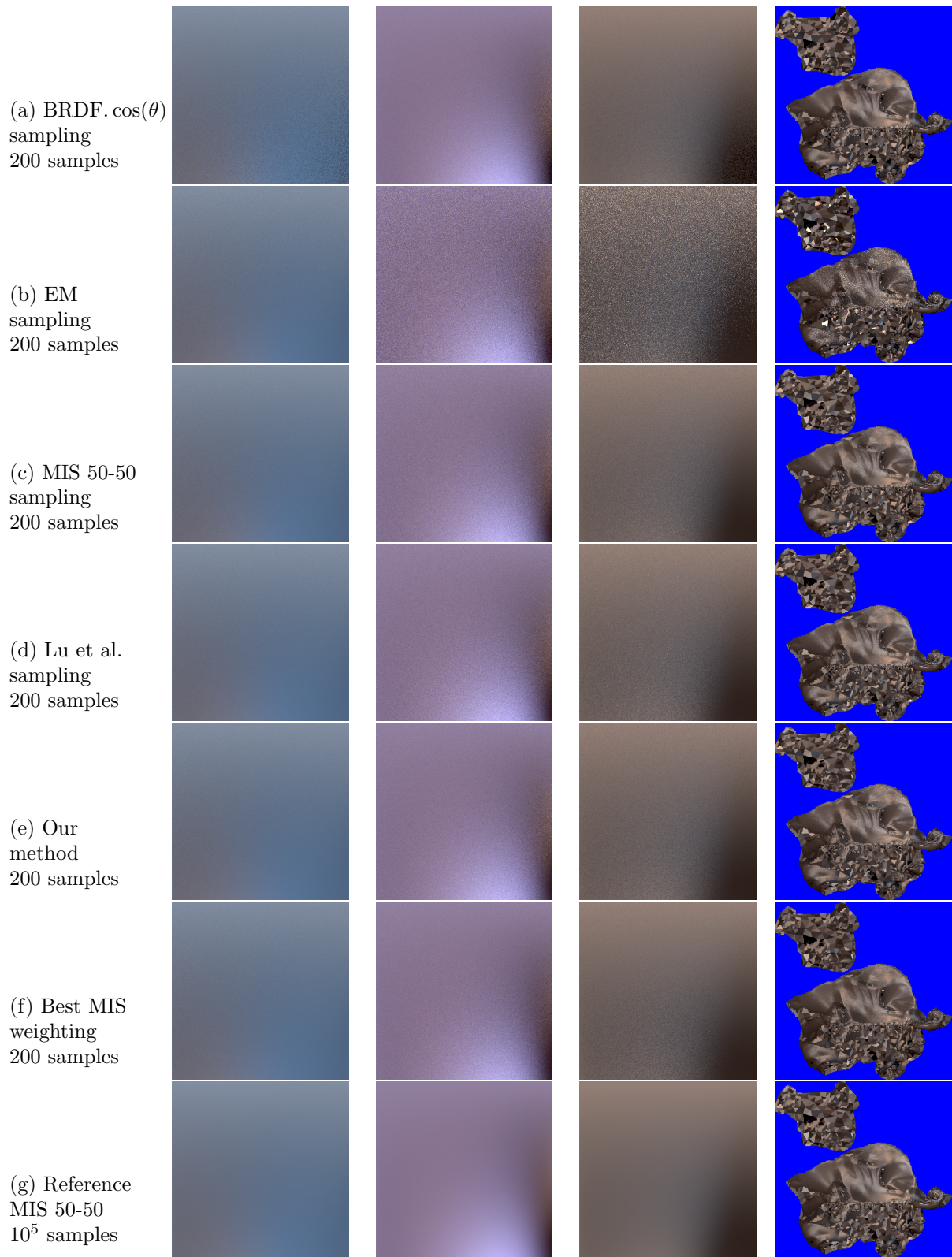


Figure 6.1: The rendered images as rectangle textures for 200 samples for tested algorithms.

## Appendix G - Notation

$x$  ... point in space  
 $X$  ... function domain such as  $f(x)$  or for integration  $dx$   
 $n, m$  ... number of samples  
 $X, Y$  ... distributions such as random variables  
 $f_r(x, \omega_i, \omega_o)$  ... bidirectional reflectance distribution function (BRDF)  
 $f(X)$  ... function corresponding to pdf  
 $g(X)$  ... function of random variable  $X$   
 $N$  ... number of samples  
 $n_k$  ... number of samples for  $k$ -th sampling strategy  
 $\vec{n}$  ... normal vector  
 $E[X]$  ... mean of distribution  $X$   
 $V[X]$  ... variance of distribution  $X$   
 $E[g(X)]$  ... mean of function  $g(X)$   
 $V[g(X)]$  ... variance of function  $g(X)$   
 $\sigma$  ... standard deviation  
 $E_j[g(X)]$  ... mean of function  $g(X)$  when samples are drawn from according to  $h_j(x)$   
 $V_j[g(X)]$  ... variance of function  $g(X)$  when sampling are drawn from according to  $h_j(x)$   
pdf ... probability density function  
 $I$  ... integral of a function  
 $f(x)$  ... function to be used as pdf for importance sampling  
 $x_i$  ... the  $i$ -th sample drawn  
 $h(x), h_k(x)$  ... function being integrated  
 $s$  ... normalization factor of sampling function  $h(x)$ ,  $s = \int h(x)dx$   
 $k, j, l$  ... subindex in  $\sum$  and  $\Pi$ .  
 $\mu$  ... mean value of estimated variable  
 $\mu_k, \mu_j$  ... mean value of the estimated variable when sampled according to  $h_k(x)$   
 $c_k$  ... cost of a single sample for  $k$ -th sampling strategy  
 $\alpha$  ... coefficient of mixture of two functions  $\alpha \in \langle 0, 1 \rangle$   
 $\alpha_k$  ... ratio of samples from the  $k$ -th sampling strategy  $\sum_k \alpha_k = 1$   
 $t_k$  ... auxiliary variable for the  $k$ -th sampling strategy  $t_k = \alpha_k/s_k$   
 $\lambda$  ... the auxiliary variable for the Lagrange multipliers method  
 $\omega$  ... direction vector  
 $d$  ... delta for integration such as  $dx$   
 $\Omega$  ... the hemispherical domain  
 $S$  ... the spherical domain  
 $i$  ... notation of the incoming direction for  $\omega_i$   
 $o$  ... notation of the outgoing direction for  $\omega_o$   
 $L(x, \omega_o) [\frac{W}{m^2 sr}]$  ... (outgoing) radiance from point  $x$  in direction  $\omega_o$   
 $R(\omega_i) [\frac{W}{sr}]$  ... (incoming) radiant intensity in direction  $\omega_i$   
 $\theta$  ... angle between normal  $\vec{n}$  and some direction  $\omega$   
 $U$  ... number of mixture combinations tested in the algorithm  
 $b$  ... index of the  $b$ -th sampling batch  
 $B$  ... number of batches  
 $\Lambda$  ... the equalized quantity in sampling algorithm



**Technical Report Series of DCGI, Volume 4, Year 2014**

Department of Computer Graphics and Interaction

Czech Technical University in Prague, CZ

Faculty of Electrical Engineering

Website: <http://dcgi.fel.cvut.cz>

---

Karlovo nám. 13  
121 35 Praha 2  
Czech Republic

Tel: (+420) 2 2435 7557  
Fax: (+420) 2 2435 7556