A statistical framework for comparing importance sampling methods, and an application to rectangular lights.

Pixar Tech Memo, # 14-01 version 2

Leonid Pekelis , Christophe Hery

February 12, 2014

Figure 1: Three of the 8 scene geometries analyzed to estimate $\mathbb{E}_\rho$ - expectation over pixel space. The rest are shown in figure 4.

1 Abstract

Importance sampling methods for ray tracing are numerous. One reason is due to the lack of a robust, quantitative method for comparison. This tech memo takes strides to provide a framework for comparing sampling methods on fair ground. We define a simple, mathematical notion of strictly preferring one method to another. We demonstrate how to apply the framework in comparing three methods from sampling light rays from a rectangular light source. One method, Angular Stratification, is new to the literature.

2 Introduction

We have a number of ways to importance sample light rays from a rectangular light. The lack of cohesion on the best way stems from the fact that solid angle sampling fits naturally into an angular basis while polygons generally are described in Euclidean. It isn't natural to measure a box with circles. This discussion stems back to Arvo '95 [2], who proposed a method for sampling spherical triangles.

In this memo, we compare three prevailing methods: *Angular Approximation, Angular Stratification,* and *Spherical Rectangles*.
Section 3 will describe each of these in turn, as well as a general fix for numerical issues caused by glancing angles.

A general cause of multiple sampling methods is the absence of a unified framework for comparing them. Without consistent and reasonable metrics there is no reason to believe anyone's method is better than your own. Furthermore, authors can only make relative comparisons. The phrase "variance reduced by half" carries different weight if the competing method is 1 unit away from the global optimum, or 100. This is playing a particularly cruel version of darts - with the point values changing in scale and the center constantly moving.

Section 4 presents an optimization argument, which derives a framework for comparing the 3 methods above. Where possible we keep calculations general enough to extend to other situations, or, at the very least, note areas of current assumption and simplification.

Section 5 gives the results of statistical analysis, applying the framework to rectangular light sampling.

3 Rectangular Light Sampling

Importance sampling reduces variance of Monte Carlo estimators. For ray tracing, this means cleaner images with fewer rays per polygon. For light source sampling, it has been found that distributions approximating the solid angle measure go a long way in this direction. If the light source is circular, this distribution is possible to compute, and analytical answers are available, at least in terms of tabulated functions, see e.g. [4].

Unfortunately projecting a polygon onto a unit sphere isn't as easy. In this section I'll review three approaches to solid angle sampling a rectangle from a reference point - *Angular approximation, Angular stratification*, and *Spherical rectangles*. In the following, I will abbreviate method names as AA, AS and SR.

Establishing a little notation, take a rectangle R described by it's bottom-left corner $a_0 = (x_0, y_0, z_0)$, width w and height h. Coordinates are fixed by a reference point $p = (0, 0, 0)$ oriented such that the z-axis remains fixed for any point in R. Let ω denote solid angle measure, and θ and γ be the horizontal and vertical angles formed by the coordinates of R and p.

3.1 Angular approximation

The first method seeks tractability by combining Euclidean sampling with solid angle approximations from the circular case. The procedure can be broken down into three main

steps:

1. Step over equi-area boxes - $B_k = (x_0 + n\delta, y_0 + n\delta) \times (x_0 + (n+1)\delta, y_0 + (n+1)\delta)$.
2. Jitter inside the box - $p_k \sim Uniform(B_k)$.
3. Approximate area by circular solid angle of the stratum - $\omega(B_k) \approx \omega(C_k)$ where C_k is a circle with center $a_0 + (n + .5)\delta(1, 1, 0)$ and radius $\delta/2$, assuming p lies on the normal line set to C_k's center. Expressions for $\omega(C_k)$ are straightforward and readily available online.

This procedure is attractive because it requires few lines of code, and is computationally efficient. It also gives the right answer as $\delta \to 0$. Or put in other words, the error caused by the approximation decreases with the number of samples. On the other hand, the approximation necessarily introduces error, hampering convergence rate.

3.2 Angular stratification

A second method, proposed by Christophe Hery, takes a mechanical interpretation of proportional angular measure. This is achieved by sequentially marching over equiangular rectangles of R, front-loading the burden of angular measure as shown in the following steps

1. Step over equiangular rectangles, $B_k = (x_k, x_{k+1}) \times (y_k, y_{k_1})$ such that $\theta_k = \theta/n$ and $\gamma_k = \gamma/n$ for $k = 0, \ldots, n-1$, where θ_k and γ_k are the horizontal and vertical angles formed by the coordinates of B_k and p.
2. Jitter inside the box - $p_k \sim Uniform(B_k)$.
3. The solid angle of B_k is easily seen to be $(x_{k+1} - x_k)(y_{k+1} - y_k)/\left\|p_k - p\right\|^2$.

As this is a novel method presented in this work, we include pseudocode in the Appendix. Note that while we do not follow this implementation below, it is possible to sample $\theta_k \sim U(\theta)$ and $\gamma_k \sim U(\gamma)$ directly when stratification is not available, e.g with progressive rendering. In this case the Euclidean distribution of the horizontal component is

$$pdf_W(\theta_k) = \frac{\sin(a + \theta_k)^2}{\theta ||W_1|| \sin(a)}$$

where W_1 is a vector from p to the center of the left side of R, and a denotes the angle formed by W_1 and the horizontal axis of the plane defined by R. The vertical distribution is congruent and independent, so the overall sampling distribution becomes

$$pdf_{AM}(\theta_k, \gamma_k) = pdf_W(\theta_k) pdf_H(\gamma_k) \frac{\left\|p - p_k\right\|^2}{\cos(\theta')}. \tag{1}$$

While not quite as simple as angular approximation, the computational burden is light, and judicious use of trigonometric identities make it even more so. All the computations in this method are also exact. A downside is that equiangular rectangles are not required to be well behaved, and therefore sampling in some of these regions may cause numerical issues.

3.3 Spherical rectangles

Finally, in a recent paper by Urena, Fajardo and King [5], the authors propose an extension of [2] to parametrizing a spherical rectangle. In particular, they describe a procedure to calculate a map $f : [0,1]^2 \to Q$, with Q the region obtained by projecting R onto a unit sphere about p, such that for any $U \subseteq [0,1]^2$, $\mathcal{A}(U) \propto \omega(f(U))$, with $\mathcal{A}$ the planar area measure.

Sampling can then be performed by taking two independent uniform draws $u = (u_1, u_2)$, applying the map f, and projecting back onto R. Both a detailed derivation and pseudocode can be found in the cited paper.

It should be noted that the map f is a theoretical generalization of the 2nd method, which holds specifically for angular stratification. Hence we predict that these two methods will exhibit similar characteristics.

As a final point, we found that *Spherical rectangles* had numerical instability when $\mathcal{A}(Q)$ was very small. As shown in figure 5 (method 4), these errors persisted even after we removed errors due to glancing angles (see section 3.4 for details). A relatively straightforward fix comes from applying a limiting argument to equations (6)-(8) in [5], and boils down to taking the first component of f as identity when $\mathcal{A}(Q)$ is below float precision.

3.4 Fixing numerical errors at glancing angles

During initial testing, we noticed that all three methods exhibited instability for glancing angles - cases where the point p is almost in the plane defined by R, call it $L(R)$.

This isn't very surprising as these cases are exactly the ones where we see a phase transition in solid angle measure. This is most readily seen by examining the equations for solid angle measure of a disk approximation to R (denoted by C), and noting that either $\omega(C) \to 0$ or ∞ as p approaches $L(R)$ at a point outside or inside C. See, for example, equations (19) and (28) in [4]. This also suggests an explanation for why numerical errors get worse for very large or very small area lights.

Our fix leverages the same approximation used in method 1. A straightforward computation shows that the solid angle measure of a disk when p lies on a normal line to the center of C is

$$\omega(C) = 2\pi \left(1 - \frac{1}{\sqrt{r^2 + 1}}\right) \tag{2}$$

where $r = r_c / \|p - C\|^2$, the ratio of C's radius to distance from p. Since importance sampling estimates are normalized by the inverse of (2), cases where $r \to 0$ will be exactly those where light transport approximations explode, causing the visual hot spots in figure 2. We mitigate these errors by taming the curvature of $\omega(C)$ close to 0, applying a 1st order Taylor Approximation to (2) in all cases where $r < \epsilon$. For all three methods above, this leads to a particularly simple sampling distribution in edge cases,

$$pdf(p_k) = \frac{\|p - p_k\|^2}{\cos(\theta')wh}$$

The effect of this fix is most prominent for small area lights, since it's easier to find cases where r is small. The top two images of figure 2 show a slice of a cornell box rendered with one rectangular light in the ceiling and rendered using method AS, without and with our fix. Errors for the other two methods were at least as egregious. Yet even this does not remove numerical error at glancing angles completely. The bottom row of figure 2 shows a comparison of rendered images using the fixed methods when the size of the light is large relative to the scene. We find that method SR is most adept at dealing with extreme glancing angles when the light is large.

With these results in mind, we decided to not consider glancing angles in the statistical analysis of variance and cost in section 5. In particular, for the results of the next section to hold we assume that our sampling methods are unbiased, a notion that loses tractability with persistent and varied numerical errors.

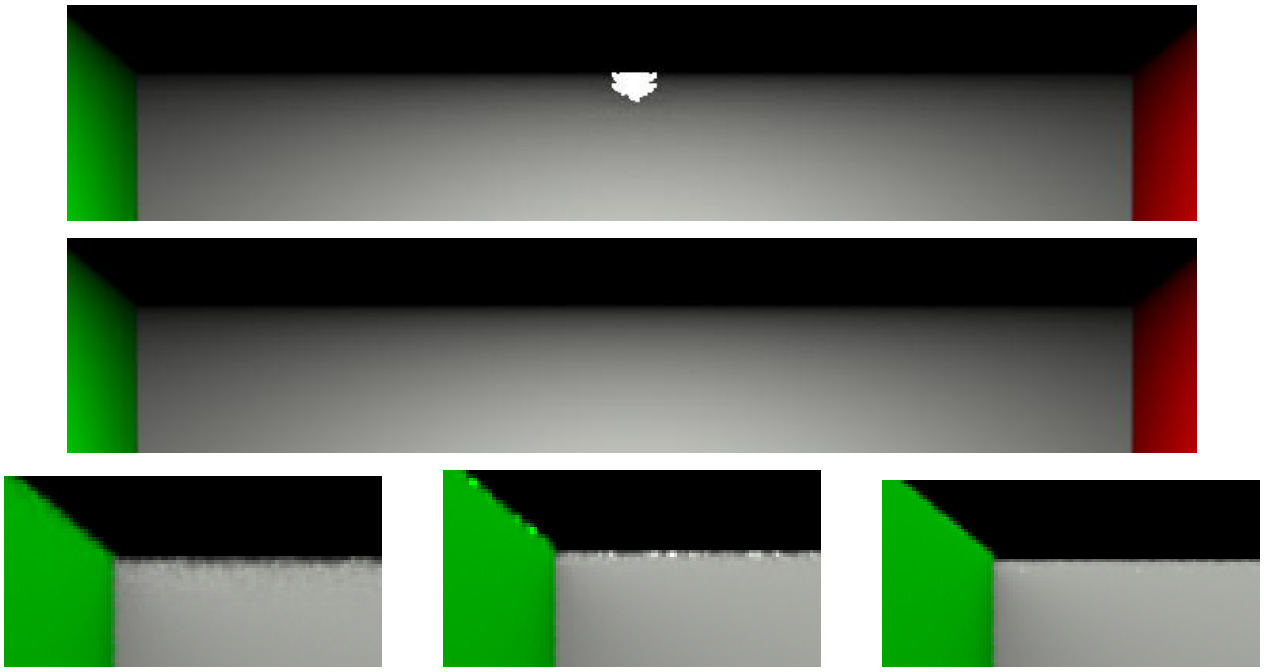

Figure 2: Errors caused by glancing angles. All images show cropped sections of the same cornell box as in figure 4. The top image highlights numerical integration error with a very small area light in the ceiling, the `tiny` light as described in Section 5. The second image shows the effect of using our fix. The next three images show persisting numerical error for methods AA, AS and SR with a very large area light, `size49` light in Section 5. Method SR is most adept at dealing with errors in these circumstances.

4 Framework for method comparison

In this section, we attempt to answer in the positive the question - can we utilize importance sampling theory to evaluate and compare sampling methods in physically-based ray tracing? In setting up some notation, we quickly review the light transport problem. For a given point, p, in a 3 dimensional scene, we evaluate the emitted radiance in direction ω_r by evaluating the integral

$$L(p, \omega_r) \int_\Omega S(\omega, \omega_r; p) V(\omega; p) L(p, \omega) cos(\theta) d\omega$$

where S is the BRDF, potentially dependent on p, V a visibility indicator, and L the incoming radiance from direction ω. Since this function is for the most part computationally intractable, we estimate an importance sampled approximation, $\hat{L}(p, \omega_r)$. To render an image, the light transport problem is solved across polygons, call them δ, and then their contributions are aggregated together for each pixel on a camera lens, call these ρ. Slightly abusing notation, the radiance expressed at each screen pixel, $L(\rho)$, is approximated by

$\hat{L}(\rho) = A_\Delta \left[\hat{L}(\delta, \omega(\rho)) \right]$

with A_Δ denoting an aggregator over polygons. An optimization one may seek to achieve in production rendering is

$$\min_n C(n) \quad s.t. \quad \sum_\rho D\left(L(\rho), \hat{L}(\rho)\right) < \epsilon, \tag{3}$$

for a given error tolerance $\epsilon > 0$, a fixed importance sampling method, C the cost of evaluating n samples and D some suitable error function.

The following lemma shows how this optimization can motivate comparison across methods.

Lemma 4.1. *Let $\mathcal{M}$ be some set of sampling methods for evaluating $L(p, \omega_r)$ such that for $m \in \mathcal{M}$ the sampling cost is proportional to the number of samples,*

$$C_m(n) = c_m n$$

, and

$$D_{m,n} \equiv \mathbb{E}_\rho D\left(L(\rho), \hat{L}_m(\rho)\right) = a_m n^{-b_m} \tag{4}$$

, exponential in number of samples. Furthermore assume we wish to optimize (3) over $\mathcal{M}$ independent of the particular rendered image. Then for any two methods m_1 and m_2, we strictly prefer m_1 to m_2 for all sample sizes $n > n^*$ where

$$n^* = \inf \left\{ n \ : \ \frac{D_{m_1,n}}{D_{m_2,n}} \leq \left(\frac{n-1}{n} \right)^{b_{m_1}} \left(\frac{\mathbb{E}_\rho[c_{m_2}]}{\mathbb{E}_\rho[c_{m_1}]} \right)^{b_{m_2}} \right\},$$

and $\mathbb{E}_\rho$ is an expectation over all pixels arising from possible scene geometries. We adhere to convention that the infimum of an empty set is ∞.

Proof. Relegated to the appendix for brevity, which also contains the definition of *strict preference*.

□

The expectation $\mathbb{E}_\rho$ arises from the assumption that we don't expect cost and convergence rate to be exactly the same across different scene geometries. Since what we really care about is the final visual representation of each pixel, we can pass all this through to variation among pixels. Note that this may mean that the pertinent range of variation is much smaller than

"all possible scene geometries." Finally note that as consequence of the above lemma, for any two methods of equal cost, n^* becomes the smallest sample size such that the ratio of convergence rate is below 1 times a correction factor for integer constrained sample sizes.

One procedure using lemma 4.1 for comparing importance sampling methods is

1. Estimate $D_{m,n}$ and c_m across different methods, m, sample sizes, n, and pixel combinations, ρ, among a suitable test bank of images.
2. Average over pixel combinations to uncover image independent estimate of signal.
3. Estimate n^* for any pairwise comparison methods.

A statement we would have liked to make is that for D L2 loss we have $D_{m,n} = I\!E_\rho Var\hat{L}_m(\rho)$, and in fact, under the additional assumption that $\mathcal{A}_\Delta$ is a linear operator we can. Unfortunately this would imply that $b_m \equiv 1$ and results such as those in figure 6 suggest that the connection between L2 convergence and importance sampling variance is not as straightforward for image rendering. Hence we stay clear of such terminology in the sequel, and prefer convergence rate or error.

With this framework in tow, we turn to evaluating and comparing the three rectangular light sampling methods.

5 Statistical Analysis

All tests were performed on a 12 core machine with Intel Xeon X5660 CPUs and 23.4GB of RAM, and RenderMan 18.0 rendering system. The scene used to compare our 3 methods was a cornell box with a fancy ball in the center and one rectangular light source, shown in 4. The dimensions of each rendered image were 1000 x 1000 pixels.

Our experiment varies three factors: sampling method, number of light samples - 20 values in the range $[1, 1024]$, and variations of scene geometry. We assume a way scene geometry affects performance is through the distribution of solid angle measures across polygons in the scene. In other words, some sampling methods may be better able to handle a high concentration of small solid angles than others. From this, we defined 8 scene variations to test across, described as follows,

`close` Baseline size (16x16 in) moved halfway down the box on the z-axis.

`tiny` A .1x.1in light with VERY high intensity.

`sizex` Rectangles with side lengths `x` in $\{4, 9, 16, 25, 36, 49\}$ with total flux kept constant.

Note that we tacitly assume these 8 variations adequately cover the variation in $I\!E_\rho$. The Appendix contains an example rendered image under each scene variation.

6 Cost

According to Lemma 4.1, we wish to estimate the per-sample cost of each method, averaged over method-dependent geometries. Since we have potentially many significant levels, and even more interactions, I used a BIC all-subsets regression to penalize inclusion of unnecessary variables. Furthermore, as many of the coefficients are close to one another in magnitude, we performed a series of likelihood ratio tests to test equivalence of variables. The following main conclusions arise from the analysis

- small and large lights both seem to have lower startup costs than average sized lights,
- cost per sample goes up fairly linearly with the size of the light for AA, while AS does not have higher per sample cost for large lights,
- and SR has about 10% per sample cost increase, mitigated somewhat for larger lights.

We find that, averaging over scene variations, there is no statistically significant difference in cost between AA and AR methods, while SR has roughly a 6.9 % greater run time than first two. Hence in the notation of section 4, we take $c_{AA}/c_{AS} = 1$ and $c_{SR}/c_i = 1.069$ for $i = AA, AS$.

Table 2 in the appendix gives full results from regression. Included variables with p value greater than 0.001, our threshold for significance accounting for some multiple testing slack, are not reported.

7 Convergence Rate

We estimate convergence by $L2$ distance in RGB space from a reference image, averaged over pixels. The reference image is relative, method dependent, and can be assumed a good empirical approximation to letting $n \to \infty$ for a given scene geometry and sampling method. Figure 5 in the Appendix gives an overview of all data points. In the figure, "Mdist 4" corresponds to method SR without a correction small solid angles (see 3.3 for a discussion). We wanted to demonstrate the impact of the correction on convergence rate, in particular among different scene geometries.

The average convergence rate is found by first averaging across scenes, producing an estimate

$$\hat{D}_{n,m} \approx \frac{1}{8} \sum_{g} \hat{V}_{n,m,g},$$

and then linearly regressing the log transform on inverse number of samples, $\log(\hat{D}_{n,m}) \sim \beta_0 - \beta_1 \log(n)$. Figure 6 in the Appendix shows the superiority of this fit to one assuming unitary convergence rate, i.e $\beta_1 \equiv 1$. Two noteworthy secondary effects found in figure 5 are the apparent acceleration of convergence rate for method AA, and greater variability of SR about it's fit. These effects support our claims that errors from approximating rectangles by circular solid angles decrease with sample size, and SR is more prone to scene dependent errors.

A resulting estimate of average convergence rate is

$$\hat{D}_m(n) = e^{\hat{\beta}_0} n^{-\hat{\beta}_1}.$$

Note that it is reasonable to assume that $D_m(n) \to 0$ since our experiment is explicitly method unbiased. Table 1 gives results, while figure 3(left) plots the fitted estimates over an interesting range of samples. Our results are consistent with the qualitative hypotheses from section 3.

To compare the two exact methods we estimate the ratios $r_{in} = D_{i,n}/D_{1,n}$. The resulting fits, shown in figure 3(right) give us a more precise answer to the question of when does each method overtake angular approximation. Following section 4, we want to find the smallest number of samples that reject the null hypothesis that $r_{in} \geq c_i$ with $c_2 = 1$ and $c_3 = (C_{AA}/C_{SR})^{b_1} \approx .93$. Noting that we assume a monotonically decreasing convergence rate, the p-value for each n can be computed as

$$p_{in} = P(\hat{r}_{in} - C_i < Z_\alpha \mid r_{in} \geq C_i)$$

with Z_α the α-quatile of a standard normal variate, and for each i, $\hat{n}$ is the first n such that $p_{in} < \alpha$. This controls the false positive rate - the chance that the true $n^* > \hat{n}$ - at level α. The dashed lines in figure 3(right) show $\hat{n}$ for both methods. We conclude that rendered images with more than 8 and 30 light samples makes AS and SR preferred to AA, respectively.

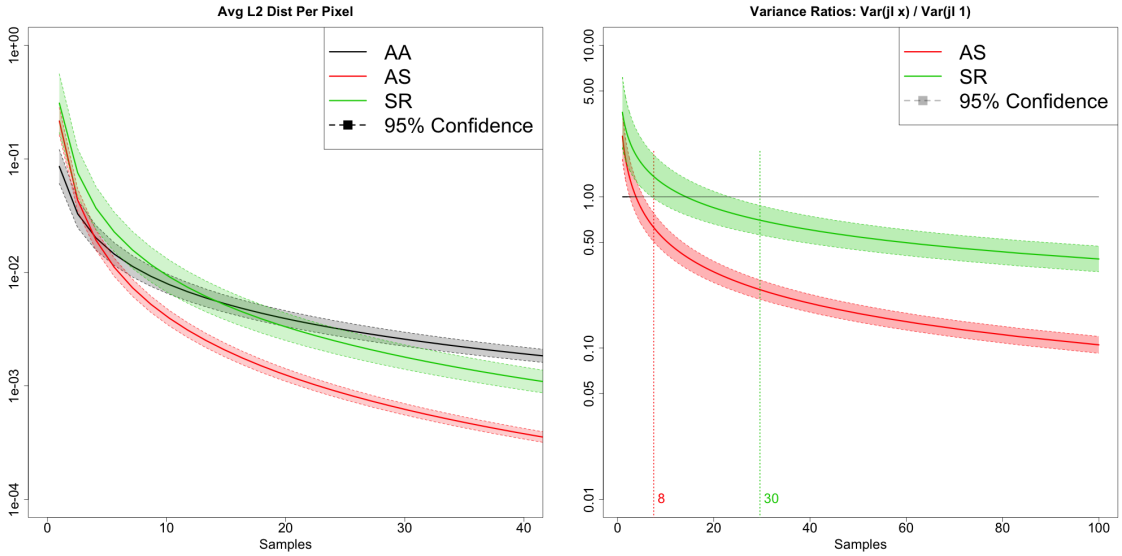

Figure 3: Results of convergence rate estimates. Left figure - estimated convergence rate, Right figure - estimated convergence ratios. Dotted lines denote minimum number of samples to prefer a method to AA, with preference as defined in the proof of lemma 4.1. Abbreviations are AA - Angular Approximation, AS - Angular Stratification, SR - Spherical Rectangles

8 Conclusion

In this technical memo, we sought to achieve two objectives: propose a framework for comparing importance sampling methods in ray tracing, fairly compare 3 competing methods for sampling from a rectangular light source.

The authors feel we have achieved modest success at both. Starting with the second, we find that while Angular Approximation gains an early advantage in convergence, possibly due to its computational simplicity, both Angular Stratification and Spherical Rectangles attain better convergence with relatively few light samples. These transitions happen well within practical bounds for industry rendering. Computational cost is nearly identical for Angular Approximation and Angular Stratification and about 7 % more for Spherical Rectangles. Using our framework, we conclude that for rendered images with more than 5 light samples, Angular Approximation is optimal.

Our success in the first objective is decidedly murkier. In addition to cost and convergence, we found that method SR is better able to handle numerical error at extreme glancing angles. Assumptions to simplify calculations forced us to concentrate solely on error free regions in the results above. We also do not consider variability of estimates about their mean, which may indicate certain methods are more robust to scene geometry. On the other hand, it was the framework, through it's limitations, that cast light on these additional results.

At the least, we hope that we have demonstrated the value of taking a formal approach to comparing sampling methods in ray tracing, and that the details are not as straightforward as might seem at first glance. Besides the more apparent next steps of expanding a comparative framework to account for bias and scene dependent variance, the authors feel a perhaps even more pressing need is to give a standardized answer to: *how do scene geometries vary*?

As there are standard collections of test cases for machine learning [1] and face recognition [3], we feel it would be useful to find a similar collection for sampling methods in global illumination.

References

[1] J Alcalá, A Fernández, J Luengo, J Derrac, S García, L Sánchez, and F Herrera. Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing*, 2010.

[2] James Arvo. Stratified sampling of spherical triangles. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 437–438. ACM, 1995.

[3] Ralph Gross. Face databases. In *Handbook of Face Recognition*, pages 301–327. Springer, 2005.

[4] F Paxton. Solid angle calculation for a circular disk. *Review of Scientific Instruments*, 30(4):254–258, 1959.

[5] Carlos Ureña, Marcos Fajardo, and Alan King. An area-preserving parametrization for spherical rectangles. In *Computer Graphics Forum*, volume 32, pages 59–66. Wiley Online Library, 2013.

9 Appendix

Proof of Lemma 4.1. To enforce an *image independent* method, we consider the sum in (3) to be over all possible pixels from different scene geometries. Standard limiting arguments allow us to transfer the average of this sum to an expectation,

$$\frac{1}{|\rho|}\sum_{\rho} D\left(L(\rho), \hat{L}(\rho)\right) \to I\!E_{\rho} D\left(L(\rho), \hat{L}(\rho)\right)$$

.

A similar argument applies to the cost factor, $c_m \to I\!E_{\rho}[c_m]$. Let the expected error and cost parameters for methods 1 and 2 be denoted $E_i(n)$ and c_i for $i = 1, 2$, and $n_i(\epsilon)$ be defined as the minimum number of samples such that $E_i(n) < \epsilon$.

We define **strict preference** by the following: Consider any $\epsilon > 0$ such that $n_1(\epsilon) \wedge n_2(\epsilon) > n^*$, then method 1 is strictly preferred to method 2 if

$$c_1 n_1(\epsilon) < c_2 n_2(\epsilon)$$

For the remainder, we fix ϵ and reduce notational complexity by taking $n_i \equiv n_i(\epsilon)$. It follows from assumption of proportional cost and exponentially decreasing and unbiased error that

$$c_1 n_1 = c_2 n_2 \left[\left(\frac{E_2(n_2)}{E_1(n_1)} \right)^{\delta_2^{-1}} \left(\frac{c_1}{c_2} \right) \right]$$

and that

$$E_2(n_2) < \epsilon < E_1(n_1 - 1) = \left(\frac{n_1 - 1}{n_1} \right)^{-\delta_1} E_1(n_1).$$

Combining these results and considering the definition of n^* finishes the proof. □

	Coef (β_0)	Rate (β_1)
M1	0.086 (0.081 , 0.091)	-1.03 (-1.028 , -1.033)
M2	0.216 (0.208 , 0.224)	-1.72 (-1.718 , -1.722)
M3	0.31 (0.26 , 0.371)	-1.515 (-1.507 , -1.523)

Table 1: Variance regression coefficient estimates. Method coding is `M = 1-4`, $AA = 1$, $AS = 2$, $SR = 3$, and variables are β_0 coefficient and β_1 rate of convergence.

Algorithm 1 Pseudocode for rectangular light sampling method - Angular Stratification (AS).

▷ Objective: Compute random angular increments θ_n and γ_n. Update pdf accordingly.
▷ Pre-defined variables: U_{vec} and V_{vec} are horizontal and vertical vectors describing the light source. U_{norm} and V_{norm} are unit vectors in the same directions. p is a point in the scene. $lightCenterDir$ is a vector from p to the center of the rectangular light, R. And ξ_1 and ξ_2 are uniform random draws in $[0, 1]$.
vector $W_1 = lightCenterDir - 0.5 * U_{vec}$
vector $W_2 = W_1 + U_{vec}$
vector $H_1 = lightCenterDir - 0.5 * V_{vec}$
vector $H_2 = H_1 + V_{vec}$
float $LW_1^2 = W_1 \cdot W_1$
float $LW_1 = \sqrt{LW_1^2}$
▷ . . . do the same for W_2, H_1 and H_2
float $cos_X = -W_1 \cdot U_{norm}/LW_1$
float $sin_X = \sqrt{1.0 - cos_X * cox_X}$
▷ . . . do the same for cos_Y and sin_Y
float $dx = 1.0/numSamplesX$
float $dy = 1.0/numSamplesY$
float $\theta = \text{ACOS}((LW_1^2 + LW_2^2 - R_{width}^2) * 0.5/(LW_1 * LW_2))$ ▷ law of cosines
float $\theta_n = \theta * dx$
float $tan_W = \text{TAN}(\theta_n)$
▷ . . . do the same to calculate horizontal increment γ_n
▷ March over equi-angular rectangles. First loop over x coordinates.
float $x_1 = 0.0$
float $tan_{x1} = 0.0$
for $(int\ x = 0; x < numSamplesX; x+ = 1)$ **do**
    float $tan_{x2} = (tan_{x1} + tan_W)/(1.0 - tan_{x1} * tan_W)$
    float $x_2 = LW_1 * tan_{x2}/(sin_X + tan_{x2} * cos_X)$
    float $L_x = x_2 - x_1$ ▷ Calculate length of step in horizontal direction
    ▷ Now loop over y coordinates
    float $y_1 = 0.0$
    float $tan_{y1} = 0.0$
    for $(int\ y = 0; y < numSamplesY; y+ = 1)$ **do**
        ▷ . . . do the same for y_2 and L_y
        float $\delta_x = (x_1 + \xi_1 * L_x)/R_{width}$ ▷ jitter over R_k
        float $\delta_y = (y_1 + \xi_2 * L_y)/R_{height}$
        $y_1 = y_2$
        $tan_{y1} = tan_{y2}$
        point $p_i = a_0 + \delta_x * U_{vec} + \delta_y * V_{vec}$ ▷ point on the light
        vector $lightDir = p_i - p$ ▷ direction towards the light
        float $\|p - p_i\|^2 = lightDir \cdot lightDir$
        vector $lightDirNorm = lightDir/\sqrt{\|p - p_i\|^2}$
        float $cos_{\theta'} = -W_{norm} \cdot lightDirNorm$
        float $lightPdf = \|p - p_i\|^2 / (cos_{\theta'} * L_x * L_y * numSamples)$ ▷ solid angle measure
    end for
    $x_1 = x_2$
    $tan_{x1} = tan_{x2}$
end for

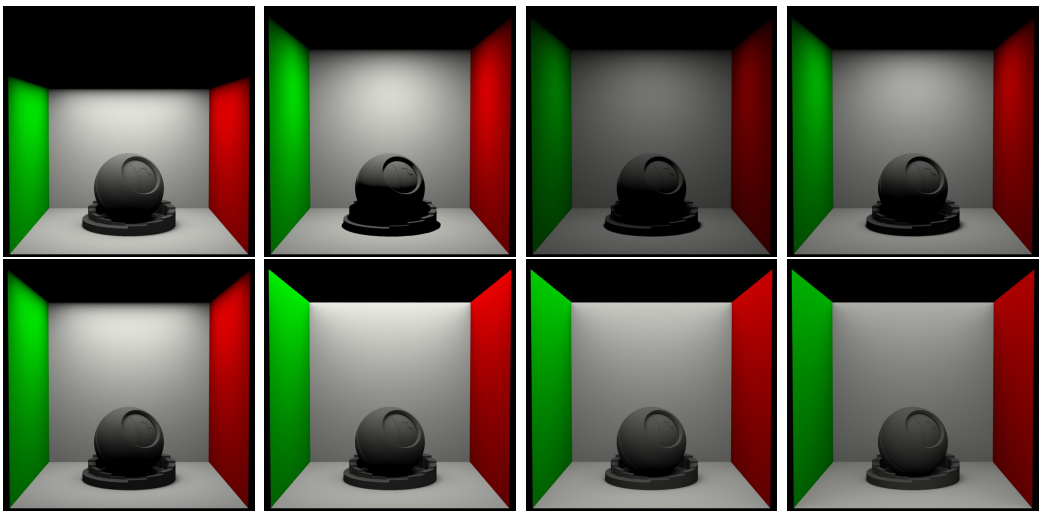

Figure 4: The 8 scene geometries analyzed, from left to right: `close`, `tiny`, `size4`, `size9`, `size16`, `size25`, `size36`, `size49`.

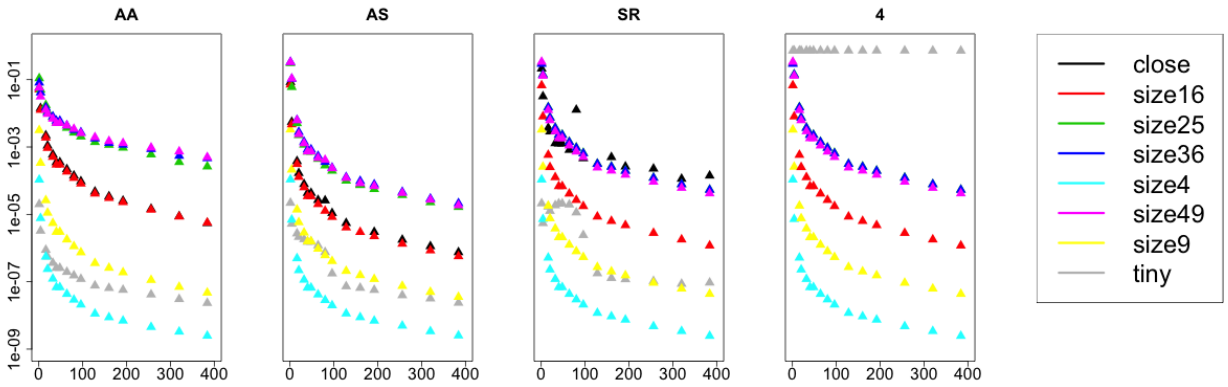

Figure 5: Pixel averages of L2 Distance from reference by importance sampling distribution and scene geometry. The far right image shows convergence for method SR without a fix for small solid angles and hence includes numerical error. Color code is *black - close, red - size16, green - size25, purple - size36, teal - size4, pink - size49, yellow - size9, grey - tiny*

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	32.6503	0.1352	241.55	0.0000e+00
Samples	0.1320	0.0005	274.84	0.0000e+00
size36TRUE	-1.3228	0.1860	-7.11	3.3147e-12
size4TRUE	-3.6569	0.1860	-19.66	1.2929e-66
size49TRUE	-1.4887	0.1860	-8.00	6.4048e-15
size9TRUE	-3.4746	0.1860	-18.68	1.4402e-61
tinyTRUE	-1.9234	0.1860	-10.34	3.7417e-23
M2TRUE	0.7689	0.1593	4.83	1.7615e-06
M3TRUE	0.9683	0.1593	6.08	2.1659e-09
M4TRUE	0.9374	0.1646	5.69	1.9485e-08
Samples:closeTRUE	-0.0041	0.0008	-5.20	2.7179e-07
Samples:size25TRUE	0.0036	0.0005	7.65	8.0947e-14
Samples:size36TRUE	0.0062	0.0008	7.37	5.6260e-13
Samples:size4TRUE	-0.0074	0.0006	-13.00	3.4991e-34
Samples:size49TRUE	0.0057	0.0008	6.76	3.3408e-11
Samples:size9TRUE	-0.0037	0.0006	-6.12	1.7098e-09
Samples:tinyTRUE	-0.0101	0.0006	-17.73	1.0318e-56
Samples:M3TRUE	0.0110	0.0005	22.04	4.8011e-79
Samples:M4TRUE	0.0112	0.0005	20.95	2.5582e-73
Samples:size36TRUE:M2TRUE	-0.0047	0.0010	-4.64	4.3670e-06
Samples:size49TRUE:M2TRUE	-0.0073	0.0010	-7.16	2.4800e-12
Samples:M3TRUE:closeTRUE	-0.0065	0.0010	-6.37	3.7394e-10
Samples:size36TRUE:M3TRUE	-0.0080	0.0010	-7.84	2.0986e-14
Samples:size49TRUE:M3TRUE	-0.0056	0.0010	-5.53	4.7459e-08
Samples:size36TRUE:M4TRUE	-0.0037	0.0010	-3.59	3.5935e-04
Samples:size49TRUE:M4TRUE	-0.0095	0.0010	-9.22	5.2427e-19

Table 2: Regression Results for Cost Analysis. Dependent variable is total render run time in seconds. The intercept here corresponds to the initial cost of our baseline condition - method 1 with a 16in light source - about 33 seconds. The baseline case costs an additional 0.13 seconds per light sample to render the image. Method coding is `M = 1-4`, $AA = 1$, $AS = 2$, $SR = 3$, and method 4 is SR without the fixes in sections 3.3 and 3.4. We find that our fixes have little overall effect on runtime.

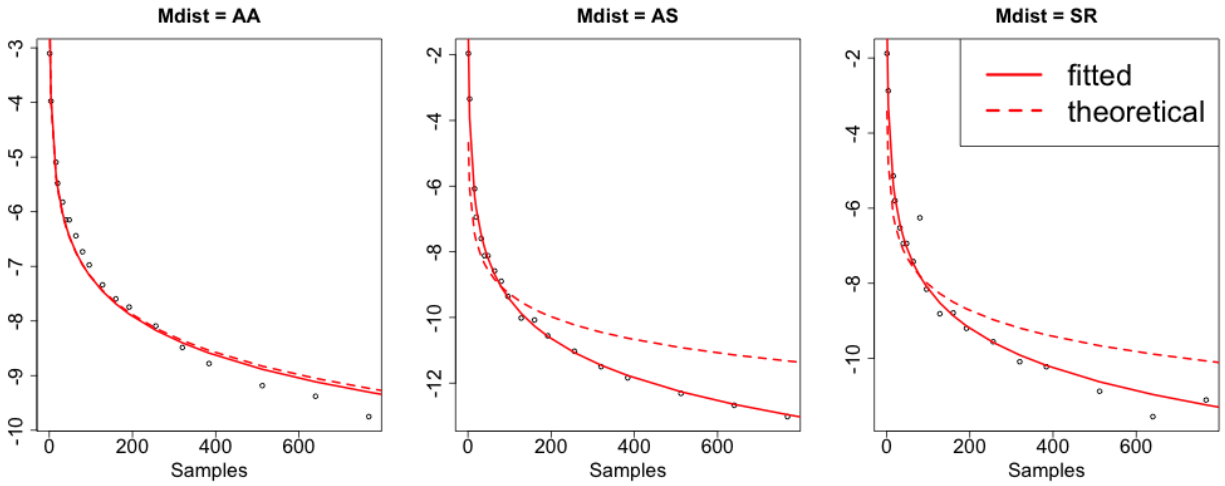

Figure 6: Fits of convergence rates to averages of L2 distances over scene geometries. Theoretical estimates assume the rate $b_m = 1$ in (4). The fits are more accurate than theoretical assumptions. Interesting features are the variability of estimates for SR, and apparent acceleration of rate for AA. The latter is seen in the bias of the AA fit at large samples. In all likelihood the rate smoothly increases with n as the error in approximation decreases (see section 3.1 for further discussion). The bias occurs since we assumed b_m constant in all fits.