Towards Interactive Global Illumination Effects via Sequential Monte Carlo Adaptation

Vincent Pegoraro[1] Carson Brownlee[1] Peter S. Shirley[1,2] Steven G. Parker[1,2]

[1]University of Utah [2]NVIDIA Corporation

ABSTRACT

This paper presents a novel method that effectively combines both control variates and importance sampling in a sequential Monte Carlo context while handling general single-bounce global illumination effects. The radiance estimates computed during the rendering process are cached in an adaptive per-pixel structure that defines dynamic predicate functions for both variance reduction techniques and guarantees well-behaved PDFs, yielding continually increasing efficiencies thanks to a marginal computational overhead. While remaining unbiased, the technique is effective within a single pass as both estimation and caching are done online, exploiting the coherency in illumination while being independent of the actual scene representation. The method is relatively easy to implement and to tune via a single parameter, and we demonstrate its practical benefits with important gains in convergence rate and applications to both off-line and progressive interactive rendering.

Index Terms: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Ray-tracing

1 INTRODUCTION

Global illumination effects are a key component to the plausible depiction of an environment and the ability to efficiently render these phenomena has considerable scientific implications. Although realism is often of concern to the movie and gaming industries of which common tools include 1-bounce indirect lighting, such interest also emerged in the visualization community where ambient occlusion shading was shown to provide enhanced perceptual cues by better conveying depth information and spatial relationships [16].

Despite their low order of convergence, Monte Carlo methods are a very general and robust technique for stochastically estimating multi-dimensional integrals and have consequently been heavily used in path-tracing to render complex global illumination effects. Several techniques were developed to reduce the variance of such estimates including importance sampling, control variates and (ir)radiance caching, sometimes trading noise for bias perceptually less noticeable in order to yield plausible renderings with practical computation times. Integrating the product of the cached radiance and the BRDF must in general be done on the fly via resampling. While realistic for a few coefficients, this becomes prohibitive for refined representations making it hard to predict whether the reduction in variance will actually overcome the computational overhead.

Building on the previous concepts, this paper presents a novel method for single-bounce illumination exploiting the coherency of the integrand on displayed surfaces induced by the correlation of primary rays. Radiance estimates computed during rendering are cached in a per-pixel data structure designed as a directional grid of adaptive resolution. This caching scheme provides dynamically refined representations of two predicate functions allowing both control variates and importance sampling to be used in a sequential Monte Carlo context. This context allows for increases in the *order* of convergence (not just a constant noise reduction factor) of the estimation process. Since each new estimate is evaluated according to a fixed snapshot of the two predicates, no bias is introduced, while allowing the functions to evolve between samples.

This document starts by providing an overview of the related work and theoretical background. The method is then presented followed by both quantitative and qualitative results along with a discussion of its limitations.

2 RELATED WORK

The pioneer work of Ward et al. [35] on irradiance caching introduced an octree structure guided by the density of rays rather than by the actual geometry. In this approximation model for diffuse surfaces, irradiance estimates are computed by interpolating previous records if available. Otherwise, a new irradiance record is estimated by sampling the hemisphere and cached. Ward et al. [34] later refined the algorithm by using gradient information to compute better interpolations. Smyk et al. [30] then proposed placing irradiance samples more effectively by influencing their distribution according to the estimated speed of illumination change. Tabellion et al. [32] improved the original method using a modified irradiance gradient caching technique combined with an approximate lighting model. Christensen et al. [7] concurrently introduced a tiled 3D MIP map (fixed-size grids nested in an adaptive octree) storing irradiance values estimated from a photon tracing pre-pass, which are interpolated and combined with final gathering during rendering.

Instead of caching irradiance, Chiu et al. [6] presented a fixed-size 3D data structure initialized through a particle trace where each cell contains a 2D directional field approximating the incoming radiance. Neighboring "light volumes" are then spatially averaged to determine the local irradiance approximation. Greger et al. [15] later described a more refined ambient term computation by storing the radiances in a geometry-dependent bi-level grid of which cells contain a fixed-size radiance field. The irradiance distribution function is pre-computed for a pre-defined set of directions and approximated during rendering by averaging the records. Christensen et al. [8] proposed a hierarchical representation of the incoming radiance based on Haar wavelets. The method relies on a final gathering step and a secondary system is solved to guide importance-driven refinement. All these techniques represent the incoming radiance as a discontinuous piecewise constant function.

Arikan et al. [1] subsequently proposed a structure initialized via a photon-map-based pre-pass only accounting for distant contributions of which incident radiance is expected to be smooth. The latter are approximated using spherical harmonics and combined with nearby contributions evaluated using a visibility heuristic during rendering. Křivánek et al. [24] also made use of (hemi)spherical harmonics to approximate the diffuse term. Their method was refined in [23] by reducing the artifacts due to the (ir)radiance caching schemes based on perceptual criteria, trading correctness for smoothness. Spherical harmonics are well-suited to represent radiance but their global support requires all coefficients to be recomputed at each update. Also, evaluating their integral implies a linear-cost dot product prohibitive for large numbers of coefficients.

Considering volume properties, Blasi et al. [4] proposed to send rays from the light sources during a first pass, computing the probability of interception at each step and storing the fraction of energy scattered isotropically in the corresponding voxels. The energy stored is accumulated during the rendering pass where camera-rays travel straight through the scene, yielding better estimates while mainly benefiting phase functions with a dominant isotropy. Based on radiosity, they [5] subsequently proposed to store data only on the envelope of a medium during the first pass, which prevents the observer from being inside the media and requires an increased sampling rate for accurate renderings.

Other methods focused on using the cached values to guide the sampling process more efficiently. When accurate computations are needed, Jensen [20] proposed to use the samples in the photon map for importance sampling rather than directly computing estimates from them. Dutré et al. [9] extended Lepage's method [26] to 2D using a k-D tree, modifying the support of equally sampled bins rather than the sampling probability of predefined bins, hence allowing for flexible stepwise probability density functions (PDFs). The hierarchical nature of the algorithm subsequently described in [10] is achieved by looking up neighboring piecewise constant cells of fixed grids if the current cell has not converged yet, which is particularly inefficient during the initialization process. Instead, Pietrek et al. [28] used Haar wavelets to represent adaptive hemispherical PDFs defined per surface patch. Their results showed little impact of the order of the representation on the variance reduction for importance sampling purposes. Also, Hey et al. [19] proposed to compute the PDFs as the sum of the footprint of each particle. In order to guarantee non-zero stepwise representations, the PDF values are customarily artificially clamped to a minimal threshold.

Lafortune et al. [25] proposed to cache radiance values in a duotricenary tree (the 5D extension of an octree) refined based on the density of primary samples. In addition to guiding the sampling process, the stepwise fixed-grid hemispherical representations built via resampling are also used as control variates. Besides the undesirable discontinuous nature of the resulting integrand, this double usage actually does not yield to any further variance reduction as discussed in the next section. Also, while they reported reductions in variance but with unaffected convergence rates, the linear cost of resampling induced large computational overheads. For each sample set drawn from the BRDF distribution, Szécsi et al. [31] proposed to linearly combine via a variance-minimizing weight a classical importance sampled estimate and one using static approximations to direct light sources or environment maps as control variates. This consistent (asymptotically unbiased) method privileges each individual technique where it performs best, but the chosen PDF does not correlate with the actual control variates integrand and the techniques consequently do not directly benefit one another. Instead, Fan et al. [12] proposed an unbiased method defining a mixture as a weighted sum of components, using one function for the BRDF and one for each light source. While fixed coefficients drive the sampling distribution, a linear system is assembled and solved to compute the mixture coefficients defining the control variates.

Smyk et al. [29] and Gautron et al. [13] also investigated ways to adapt such caching methods to the temporal change of incoming radiance. Instead, sequential methods focusing on adaptation during the rendering process itself recently received some attention in the graphics community and the works of Fan [11] and Ghosh et al. [14] showed promising applications of this framework.

3 Theoretical Background

This section provides an overview of the related theoretical background, including variance reduction techniques for Monte Carlo integration and the main concepts of radiative energy transfer. The reader is referred to classic texts for further details.

3.1 Monte Carlo Integration

Monte Carlo methods are a general and robust technique for stochastically evaluating multi-dimensional integrals. The basic method computes the integral F of a function f on a domain D as $F = \|D\|\bar{f}$, the mean value $\bar{f}$ of f being evaluated as the sample mean. To reduce the variance of the estimates, several techniques were developed [18, 22].

The control variates method assumes the knowledge of a function g approximating the integrand f and of which integral G can be analytically computed. If $f-g$ is nearly constant, the variance will be reduced as the original integral is rewritten

$$F = \int_D f(\vec{x})d\vec{x} = \int_D [f(\vec{x}) - g(\vec{x})]\,d\vec{x} + G. \tag{1}$$

Instead, importance sampling assumes a normalized PDF $p \geq 0$ correlated with f and such that $p \neq 0$ whenever $f \neq 0$. If f/p is nearly constant (ideally F), the variance will be reduced. Defining a continuous random variable $\vec{X}$ distributed according to p and the expectation E, the original integral is reformulated as

$$F = \int_D f(\vec{x})d\vec{x} = \int_D \frac{f(\vec{x})}{p(\vec{x})}p(\vec{x})d\vec{x} = E\left[\frac{f(\vec{X})}{p(\vec{X})}\right]. \tag{2}$$

Both techniques can be combined [27] by estimating the integral term of equation 1 using importance sampling, yielding the unbiased estimator $\widehat{F}$ of standard deviation $\sigma[\widehat{F}]$, reading for N samples

$$F = E\left[\frac{f(\vec{X}) - g(\vec{X})}{p(\vec{X})}\right] + G \Rightarrow \widehat{F} = \frac{1}{N}\sum_{i=1}^{N}\frac{f(\vec{x_i}) - g(\vec{x_i})}{p(\vec{x_i})} + G \tag{3}$$

$$\sigma\left[\widehat{F}\right] = \sqrt{\frac{1}{N}V\left[\frac{f(\vec{X}) - g(\vec{X})}{p(\vec{X})}\right]} = \frac{1}{N^{\frac{1}{2}}}\sigma\left[\frac{f(\vec{X}) - g(\vec{X})}{p(\vec{X})}\right]. \tag{4}$$

Equation 3 shows that p should now resemble $f-g$ rather than f. As the sign of the integrand $f-g$ might here vary while $p \geq 0$ must hold, an alternative is to correlate p with $|f-g|$ instead [2]. Also, if g is proportional to p, the previous estimator becomes identical to the one with importance sampling alone. This implies that if a function is used for importance sampling, using it as a control variate as well will not yield any further variance reduction [33].

Equation 4 shows that when using static predicate functions p (as when importance sampling from the BRDF) and g (as when $g = 0$ and $p = 1/\|D\|$ for basic Monte Carlo integration), the method exhibits an order of converge of $1/2$, meaning that n^2 times as many samples are necessary to reduce the expected error by $1/n$. In this context, these techniques yield a reduction of variance if $V[(f-g)/p] < V[f\|D\|]$ (both constant with respect to N) corresponding to a vertical translation on a logarithmic scale of the convergence curves shown in figure 4 (division by a constant factor on a linear scale). To affect the slope of these curves, i.e. the convergence rate, sequential Monte Carlo methods are adequate.

In Markov chain Monte Carlo (MCMC) methods, the next state solely depends on the present state, i.e. every future state is conditionally independent of every prior state. Sequential Monte Carlo (SMC) methods split the computation in stages such that the estimator in a subsequent stage is adapted based on the information gained during previous stages in the sequence. While this dependent sampling may appear to introduce bias, it can be proven that the result is unbiased and that the method can considerably increase the rate of convergence of the estimation process [17]. This can be illustrated by assuming adaptive predicates g and p such that $V[(f-g)/p]$ decreases with an order 2α with respect to $V[f]$. The standard deviation then becomes

$$\sigma\left[\widehat{F}\right] = \sqrt{\frac{1}{N}\frac{1}{N^{2\alpha}}V\left[f(\vec{X})\right]} = \frac{1}{N^{\frac{1}{2}+\alpha}}\sigma\left[f(\vec{X})\right]. \tag{5}$$

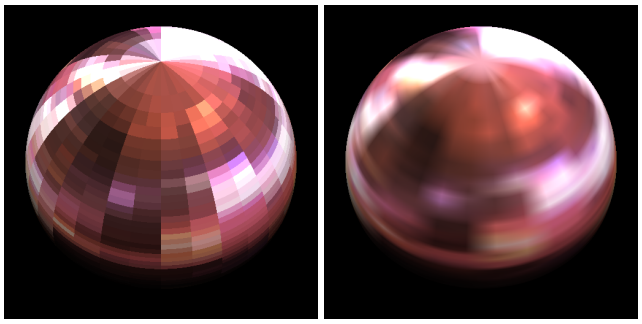

Figure 1: Per-color-channel coefficients stored in each directional cell of a radiance cache (left) and 2D B-spline reconstruction of the weighted incoming radiance defining predicate g (right).

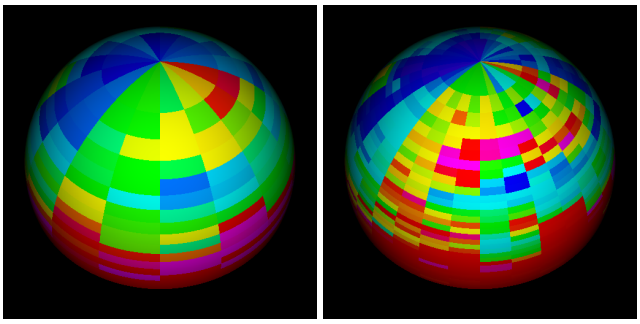

Figure 2: Color-mapped scalar values of predicate p at 2 different refinement stages.

3.2 Light Transport

The evolution of (spectral) radiance L as light interacts with a surface is defined by the Rendering Equation (RE). For a given position in space $\vec{x}$ and direction $\vec{\omega}$, this equation reads [21]

$$L(\vec{x},\vec{\omega}) = L_e(\vec{x},\vec{\omega}) + \int_{2\pi} L(\vec{x},\vec{\omega}_i)\beta(\vec{\omega},\vec{\omega}_i,\vec{n})\ \vec{\omega}_i\cdot\vec{n}\ d\vec{\omega}_i. \quad (6)$$

where L_e is the emitted radiance, $\vec{n}$ the surface normal, and β the bidirectional reflectance distribution function (BRDF) which must satisfy $\int_{2\pi}\beta(\vec{\omega},\vec{\omega}_i,\vec{n})\ \vec{\omega}_i\cdot\vec{n}\ d\vec{\omega}_i \leq 1$ to ensure energy conservation.

4 SMC Adaptation for Rendering

This section describes how to carry the evaluation of integral 6 in a sequential Monte Carlo context using both control variates and importance sampling. We introduce a per-pixel 2D data structure allowing for efficient estimations of integrals over solid angles in which the samples computed during rendering are cached. The following subsections detail the caching schemes for each variance reduction technique as well as the adaptive refinement strategy before explaining how this information is used for estimate evaluations.

4.1 Caching for Control Variates

The representation defining predicate g should provide low-cost read/write access and efficient computation of G. B-splines meet both criteria as their basis functions have local support and their integral evaluates to a simple sum of coefficients regardless of their order, except at the domain boundaries. For this property to hold in 2D, we regularly partition the normalized hemispherical coordinates $s = \phi/2\pi$ and $t = 1 - cos(\theta)$ to yield uniform solid angles.

Control variates lead to the new integrand $f - g$ of which properties must be analyzed in correlation with the complexity of evaluating g. While order 0 B-splines are the cheapest (involving 1 coefficient), their piecewise constant representation artificially introduces undesirable high-frequency discontinuities in the integrand, therefore decreasing the potential benefit of the method. Order 1 B-splines (piecewise linear) consequently provide higher quality estimates for a modest overhead (4 coefficients) while remaining natural interpolants. B-splines of order 2 (piecewise quadratic) and higher obviously entail a higher cost while being smoother and less tight to the control points as the support of the basis functions increases, usually yielding lower quality estimates. Order 1 B-splines are consequently most suitable and a grid representation allows for efficient interpolation.

We exploit the periodicity in s and introduce in t two polar values computed as the average of the boundary coefficients at $t = 0$ and $t = 1$ respectively. The first allows to eliminate discontinuities at the pole when reconstructing g and to regularize the top boundary with respect to integration. The bottom boundary is implied as zero (due to the cosine term in the integrand) and the second polar value allows for an efficient computation of G which evaluates to its simple weighted subtraction to a hemispherical average. As shown in figure 1, each directional cell holds a color of which channels represent the coefficients of the 2D B-splines defining predicate g. Whenever a new sample is estimated, its color is averaged with the corresponding cell's coefficients while incrementing its counter of cached records C which determines the respective weights $1/(C+1)$ and $C/(C+1)$. The hemispherical and polar averages are maintained and updated at each write operation, allowing the constant time computation of both g and G during estimations.

4.2 Caching for Importance Sampling

For efficiency reasons, the resolution used to represent predicate p is set to be the same as the one for g. Drawing samples from a given PDF can be done by inverting its cumulative distribution function (CDF) defined as its partial integral. This favors cheap low-orders while continuity is not crucial here. Order 0 B-splines are therefore adequate. In addition to the radiance coefficients and records counter, each cell contains a scalar estimate of the value of $|f - g|$ over the associated solid angle as shown in figure 2. When a new sample of f is added to a cell, the value of g is determined and $f - g$ computed. Since the latter is a color, a scalar PDF sample is generated by averaging the absolute values of its channels and merged with the cell's PDF coefficient.

To make the sampling process inexpensive, each cache maintains a logical tree of partial sums [33] similar in spirit to a Huffman tree, stored in a flat array of size $2N-1$ with N being the number of cells. Each node of this complete binary tree holds the sum of its 2 children, starting with the cells' values of p as the leaves up until the root holding the sum of all PDF coefficients. While write operations need to traverse the $log_2(2N)$ nodes of a branch, the space of basis functions can now be sampled in log time given a random number. Normalization is achieved by multiplying the random number by the value of the root node. If the random quantity is greater than the value of the first child of the current node, its PDF value is subtracted from the quantity and the second child becomes the current node, the latter being set to the first child otherwise. The process is recursively repeated as to traverse an entire branch until a cell is reached and a random direction is drawn from the linear CDF.

4.3 Adaptive Refinement

The proposed data structure provides an adaptive representation permanently refining in correlation with the current records population. It is initialized as a screen-size buffer of radiance caches with a single initial cell of which radiance B-spline coefficients, PDF value and records counter default to zero. Since the PDF is not relevant at this stage, a uniform directional sampling strategy is used. For each primary ray traced, the cache corresponding to the target pixel is

identified in constant time while the direction of the secondary ray determines the cache's cell which should be updated. If the refinement criterion is met, the resolution of the cache is doubled in both polar and azimuthal coordinates while duplicating previous records to preserve the data repartition. The cells' records counters of the cache are then divided by the dimensionality of the split, i.e. 4. This effectively reduces the weight of ancient coarse records and allows future locally relevant samples to be more influential. Inheritance is enforced by preventing the counters from being rounded down to zero which would cause a new record to overwrite rather than being merged with ancestral information. While each cell of the radiance cache has to be processed, the linear cost of refining is however not prohibitive as its frequency of occurrence is low compared to other read/write accesses.

This inheritance strategy allows a PDF to always contain a portion of its ancestors. By prohibiting the refinement of the initial cells until their PDF coefficient is non-zero, all PDFs are guaranteed to be non-zero as well. This allows the PDFs to tend freely towards zero where needed while remaining implicitly well-behaved without the need for an artificial bound as in previous approaches.

4.4 Refinement Criterion

The refinement criterion is defined as a threshold on the average value of the records counters also maintained in each radiance cache to yield a constant time access. Such criterion will adaptively promote deeper refinement based on the density of rays while controlling the inertia of the system. Decreasing it will increase the versatility of the caches requiring a smaller population before refining. This induces predicate functions quickly morphing into the target functions, yielding improved convergence rates and lower variance.

However, if the threshold is too low, the caches might evolve while being under-populated and yield unreliable predicates generating estimates of increased variance. Hence, the optimal criterion is the lowest one guaranteeing that the structure contains meaningful information before refining. In our experiments, it was determined empirically by conducting a few trial-and-error tests on down-sampled images or on the fly during interactive sessions.

4.5 Minimizing Variance

When no information is available yet about the radiance term of equation 6, importance sampling the integrand based on the product of the BRDF and cosine term often yields lower variance than a uniform sampling strategy as used by the radiance caches. Because it is correlated with f but not with $|f-g|$, it is not desirable to use such PDF with the control variates as a substitute to the PDF of a cache. This consequently yields two estimators: a classical MCMC importance sampling estimator preferable upon start-up, and a SMC estimator most efficient at higher population levels. In order to use the estimator performing best given the current population, each radiance cache is associated to a variance tracker.

Whenever a secondary ray needs to be traced, the tracker indicates which estimator yields lower variance at the given sampling stage and a direction is drawn from the corresponding PDF to compute a new sample. From this single sample, two estimates are evaluated and fed to the tracker computing and aggregating their respective sample variance based on an accumulated integral estimate. While the variance of the first estimator is constant, the dynamic nature of g and p induces the variance of the second estimator to vary throughout the sampling process. The weights used to update its estimated variance are consequently computed by clamping the total number of estimates to a threshold value, hence making recent variance estimates more influential than older ones. Because the evolution of the estimator's variance is directly correlated to its inertia, the threshold value and refinement criterion should be correlated as well. In practice, our experiments revealed that good results were obtained by setting them to an identical value.

4.6 Estimate Evaluation

For each estimation of integral 6, the radiance cache and variance tracker corresponding to the pixel being rendered are identified. The tracker indicates the estimator from which to derive a sample direction. In case of the MCMC estimator, the PDF associated to the BRDF provides a means of defining a 2D sample direction on the hemisphere, classically by inverting its CDF. In case of the SMC estimator, the tree of partial PDF sums allows to importance sample a direction of associated p. A newly ray-traced radiance value is then evaluated and multiplied by the BRDF and cosine value to yield a new sample f from which two estimates are computed. For the MCMC estimator, this is achieved by simply dividing the sample by its PDF value. For the SMC estimator, the term G used for control variates is directly read from the cache and the integrand approximation g for the given direction is computed from the B-spline coefficients. This term is then subtracted from f and the result is divided by p and added to G to form the final low-variance estimate. While both are necessary to approximate the variance of the estimators, only the estimate initially indicated by the tracker does contribute to the integral evaluation, the other being biased as its PDF value does not correlate with the actual sample distribution.

4.7 Pseudo-Code

Figure 3 provides a high-level pseudo-code illustration of the integration of the various steps individually presented. Lines 3 to 9 correspond to importance sampling as described in sections 4.2 and 4.6, lines 12 and 13 to the control variates step from section 4.1 and lines 14 and 15 to the actual estimation process from section 4.6. Finally, line 16 updates the variance estimates based on section 4.5, line 17 populates the structure as explained in sections 4.1 and 4.2, while line 18 corresponds to the refinement step from section 4.3. For clarity sake, the BRDF term here implicitly refers to the product of the actual BRDF and the cosine term (i.e. the dot product).

```
1.  EstimateHemisphericalIntegral()
2.  (cache, tracker) = ray.GetPixelData();
3.  estimator = tracker.GetEstimator();
4.  if (estimator == SMC)
5.    (direction, p_SMC) = cache.GetSampleDirection();
6.    p_MCMC = brdf.GetPDF(direction);
7.  else
8.    (direction, p_MCMC) = brdf.GetSampleDirection();
9.    p_SMC = cache.GetPDF(direction);
10. radiance = TraceRay(position, direction);
11. f = radiance * brdf.GetValue(direction);
12. G = cache.GetIntegral();
13. g = cache.GetValue(direction);
14. estimate_SMC = G + (f - g) / p_SMC;
15. estimate_MCMC = f / p_MCMC;
16. tracker.Update(estimate_SMC, estimate_MCMC);
17. cache.AddRecord(direction, f);
18. if (cache.CriterionIsMet()) cache.Refine();
19. return (estimator == SMC) ? estimate_SMC : estimate_MCMC;
```

Figure 3: Pseudo-code for the integral estimation

5 Results

In order to demonstrate the convergence characteristics of the method, we experimented with a test-bed consisting of a plane illuminated by a gradient background for which an analytical solution was derived. The quantitative results are shown in figure 4 where the number of samples per pixel on the abscissa increases by a factor 4. The slopes of the root mean squared error (RMSE) curves il-

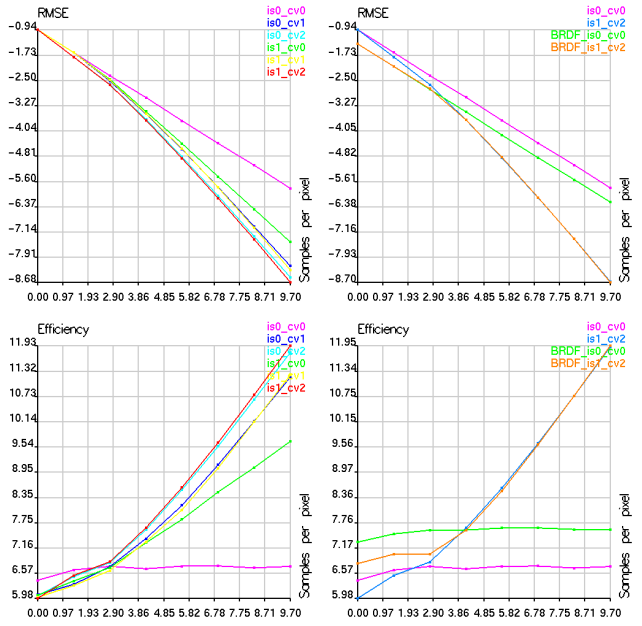

Figure 4: Logarithmic plots of root mean squared error and efficiency versus number of samples per pixel for several strategies: no control variate (cv0), piecewise constant control variates (cv1), piecewise linear control variates (cv2), no importance sampling (is0) and piecewise constant PDFs (is1), where is1_cv2 corresponds to the proposed SMC method and is0_cv0 to MCMC. Results are shown for an initial uniform sampling strategy (left) and compared to a combination with BRDF importance sampling (right).

lustrate the 0.5 convergence rate of MCMC integration compared to the higher order of SMC here reaching 0.94 ($\sim 88\%$ gain). Considering the rightmost vertices of the BRDF-based graphs, the MCMC approach would require about 138 times as many samples in order to reach the error level achieved by the SMC method which in contrast only requires a 79% overhead (factor of 1.79) in computational time, yielding a 77X speed-up. The graphs also illustrate the impact of the variance trackers determining which of the MCMC BRDF importance sampling estimator (BRDF_is0_cv0) or the SMC estimator (is1_cv2) performs best at the current population level as to yield the minimum variance of the two (BRDF_is1_cv2).

While it is constant for MCMC integration, the efficiency $((variance * cost)^{-1})$ of the SMC method keeps increasing with the sampling rate. Due to the smooth illumination, control variates here provide most of the gain whereas importance sampling has a more prominent impact in complex scenes. This illustrates their respective strengths. While control variates approximate well smooth variations, importance sampling performs better for higher frequencies by focusing on strong contributions. When combined, control variates allow importance sampling to focus on hard features rather than smooth high contributions, yielding increased efficiencies. The impact of the variance trackers is here again illustrated. Because both initially yield similar results but at different costs, the SMC estimator alone (is1_cv2) is slightly less efficient than a MCMC uniform sampling estimator (is0_cv0). By minimizing variance at a marginal cost, the trackers allow the combined estimator (BRDF_is1_cv2) to yield reduced initial efficiency loss compared to the MCMC BRDF importance sampling estimator (BRDF_is0_cv0) while preserving the substantial gains at higher sampling rates.

Table 1 details the memory requirements of the data structure for various qualitative experiments generated at a resolution of 512x512 pixels on an Intel Xeon 3.00GHz processor desktop with

Figure	5	6	7
Criterion	8	8	16
Depth	4	3	4
Cells	256	64	256
Memory	9 KB	2 KB	9 KB

Table 1: Characteristics of the radiance caches after rendering the listed figures, including the value used for the refinement criterion, the refinement depth reached, and the corresponding number of cells and peak memory usage (using double-precision).

2GB of RAM. The figures were generated by processing each pixel at a time, allowing the memory associated with a radiance cache to be deallocated once the corresponding pixel has been rendered. The results are shown next to their associated error images.

Figure 5 shows a scene rendered with 1-bounce global illumination and where indirect lighting is mainly reflected by the floor. Control variates alone adequately capture and reduce variance from this large source but have little impact on evaluating finer contributions. While importance sampling alone also reasonably reduces variance in some places, it introduces several largely under/over-estimated pixels scattered in the image. However, when combining both techniques, control variates mainly handle the smooth illumination from the floor and allow importance sampling to focus on higher frequency signals, reinforcing the statement made earlier. Because the resulting sampling strategy differs from the one with importance sampling alone, both computational overheads and variations in path-space will impact the overall rendering cost.

Figure 6 shows an application to ambient occlusion shading. While being less substantial due to the relatively lower sampling rate, the superiority of the combination of both variance reduction techniques in the SMC framework is here again illustrated. Although the method ideally relies on purely static integrands, the coherency in the scene allows pixel-space jittering to be used without compromising convergence. Visible artifacts only appear here as a few high variance pixels along the edges of the light polygons due to the strong uncorrelation between the integrand on a light's surface and the one on the ceiling behind it.

Figure 7 shows an application to environment map illumination with a Phong BRDF model. Due to the higher specularity of the material and the details of the environment map, SMC importance sampling alone here performs better than SMC control variates alone, while their combination exploits their respective strengths. Although the proposed SMC method effectively reduces variance on the left side of the image, MCMC BRDF importance sampling performs better on the right side where the illumination distribution coincides with the specular lobe. This illustrates the limitations of the method for specular materials as discussed in section 6.

Finally, as each estimation accesses a single radiance cache associated to the pixel being rendered, the method is readily suitable to a pixel-based parallel implementation and was integrated to the open source interactive ray-tracing system Manta [3]. The technique was used as part of a progressive rendering framework where the image is iteratively refined whenever the user stops moving the camera and reset otherwise. Because all pixels are rendered in parallel rather than sequentially, all radiance caches must here be maintained simultaneously. To control memory allocation, an upper bound on their maximal depth was set. Figure 8 shows the progressive convergence of images generated during an interactive session on 4 Dual Core Opteron 2.4GHz processors. Due to the overhead of the method, the resulting frame rates are lower than the ones obtained with MCMC, allowing the latter to trace more samples per unit of time. The initially slightly noisier images illustrate the marginal efficiency loss of the technique at low population levels. Its efficiency however rapidly grows as the convergence rate increases, ultimately reducing variance more effectively.

6 Discussion and Future Work

While remaining statistically correct, the quality of the estimates will gradually degrade as the specularity of the BRDF increases. In addition to the initial sample distribution causing an unbalanced population of the structure, such BDRF will strongly shape the integrand and require the caches to be heavily populated before being chosen by the trackers. Further investigation is required to alleviate this limitation while preserving the efficiency of the method.

Although the primary focus of this paper is the efficiency of the method rather than its memory requirements, the use of an appropriate data compaction scheme could be considered. Note that the quantities reported can be trivially halved by using a single-precision floating point representation rather than double as was done for consistency reasons with the particular ray-tracer used.

Even though setting the refinement criterion requires little effort, further investigation is also needed to determine an optimal formulation which adapts to the local complexity in lighting rather than being global to the scene. Moreover, while the method showed to be beneficial in a progressive rendering context, future directions of research should explore ways of increasing efficiency at low population levels in order to reach true interactivity.

7 Conclusion

We have presented a novel method which effectively combines both control variates and importance sampling in a symbiotic sequential Monte Carlo context. While handling general single-bounce global illumination effects, the method yields continually increasing efficiencies thanks to a modest computational overhead achieved by exploiting the correlation of the primary rays and implicitly guarantees non-zero PDFs via its inheritance strategy.

A main advantage is that both estimation and caching are done online, allowing the sampling process to be driven by both visual importance and features of interest in the scene while remaining unbiased. The algorithm exploits the coherency in illumination of the latter while being independent of its actual representation. The technique is also relatively easy to implement in a general Monte Carlo ray-tracer and easy to tune via a single refinement parameter.

In addition to important gains in the convergence rate, the quantitative and qualitative results showed that this combined model outperforms the individual variance reduction techniques on which it is based. The method consequently appears as a promising step towards interactively rendering global illumination effects via self-tuning estimators that learn to become effective based on the information previously collected during the rendering process itself.

Acknowledgements

This research was supported by the U.S. Department of Energy through the Center for the Simulation of Accidental Fires and Explosions, under grant W-7405-ENG-48. The authors wish to thank Dave Edwards, Thiago Ize and Ingo Wald for helpful discussions. Sponza and Sibenik models courtesy of Marko Dabrovic at RNA Studios, Conference Room model courtesy of Greg Ward at Lawrence Berkeley Laboratory, David model courtesy of the Stanford Digital Michelangelo Project, and Grace Cathedral environment map courtesy of Paul Debevec's Light Probe Image Gallery.

References

[1] O. Arikan, D. A. Forsyth, and J. F. O'Brien. Fast and Detailed Approximate Global Illumination by Irradiance Decomposition. In *SIGGRAPH*, pages 1108–1114, 2005.

[2] P. Bekaert. *Hierarchical and Stochastic Algorithms for Radiosity*. PhD thesis, Leuven, Belgium, 1999.

[3] J. Bigler, A. Stephens, and S. G. Parker. Design for Parallel Interactive Ray Tracing Systems. In *IRT*, pages 187–196, 2006.

[4] P. Blasi, B. L. Saëc, and C. Schlick. A Rendering Algorithm for Discrete Volume Density Objects. *CGF*, 12(3):201–210, 1993.

[5] P. Blasi, B. L. Saëc, and C. Schlick. An Importance Driven Monte-Carlo Solution to the Global Illumination Problem. In *EGWR*, pages 177–187, 1994.

[6] K. Chiu, K. Zimmerman, and P. Shirley. The Light Volume: an Aid to Rendering Complex Environments. In *EGWR*, pages 1–10, 1996.

[7] P. H. Christensen and D. Batali. An Irradiance Atlas for Global Illumination in Complex Production Scenes. In *EGSR*, 2004. p133-141.

[8] P. H. Christensen, E. J. Stollnitz, D. H. Salesin, and T. D. DeRose. Global Illumination of Glossy Environments Using Wavelets and Importance. *Transactions on Graphics*, 15(1):37–71, 1996.

[9] P. Dutré and Y. D. Willems. Importance-driven Monte Carlo Light Tracing. In *EGWR*, pages 185–194, 1994.

[10] P. Dutré and Y. D. Willems. Potential-driven Monte Carlo Particle Tracing for Diffuse Environments with Adaptive Probability Functions. In *Eurographics Workshop on Rendering*, pages 306–315, 1995.

[11] S. Fan. *Sequential Monte Carlo Methods for Physically-Based Rendering*. PhD thesis, University of Wisconsin-Madison, USA, 2006.

[12] S. Fan, S. Chenney, B. Hu, K.-W. Tsui, and Y.-C. Lai. Optimizing Control Variate Estimators for Rendering. *CGF*, 25(3):351–358, 2006.

[13] P. Gautron, K. Bouatouch, and S. Pattanaik. Temporal Radiance Caching. In *SIGGRAPH Sketches*, page 171, 2006.

[14] A. Ghosh, A. Doucet, and W. Heidrich. Sequential Sampling for Dynamic Environment Map Illumination. In *EGSR*, 2006. p115-126.

[15] G. Greger, P. Shirley, P. M. Hubbard, and D. P. Greenberg. The Irradiance Volume. *Computer Graphics and App.*, 18(2):32–43, 1998.

[16] C. P. Gribble and S. G. Parker. Enhancing Interactive Particle Visualization with Advanced Shading Models. In *APGV*, 2006. p111-118.

[17] J. H. Halton. Sequential Monte Carlo. *Cambridge Philosophical Society*, 58:57–78, 1962.

[18] J. M. Hammersley and D. C. Handscomb. *Monte Carlo Methods*. Chapman and Hall, 1964.

[19] H. Hey and W. Purgathofer. Importance Sampling with Hemispherical Particle Footprints. In *SCCG*, pages 107–114, 2002.

[20] H. W. Jensen. Importance Driven Path Tracing Using the Photon Map. In *EGWR*, pages 326–335, 1995.

[21] J. T. Kajiya. The Rendering Equation. *SIGGRAPH*, 1986. p143-150.

[22] M. H. Kalos and P. A. Whitlock. *Monte Carlo Methods, Volume I: Basics*. Wiley, 1986.

[23] J. Křivánek, K. Bouatouch, S. N. Pattanaik, and J. Zára. Making Radiance and Irradiance Caching Practical: Adaptive Caching and Neighbor Clamping. In *EGSR*, pages 127–138, 2006.

[24] J. Křivánek and P. Gautron. Radiance Caching for Efficient Global Illumination Computation. *TVCG*, 11(5):550–561, 2005.

[25] E. P. Lafortune and Y. D. Willems. A 5D Tree to Reduce the Variance of Monte Carlo Ray Tracing. In *EGWR*, pages 11–20, 1995.

[26] G. P. Lepage. A New Algorithm for Adaptive Multidimensional Integration. *Journal of Computational Physics*, 27:192–203, 1978.

[27] A. Owen and Y. Zhou. Safe and Effective Importance Sampling. *Journal of the American Statistical Association*, 95(449):135–143, 2000.

[28] G. Pietrek and I. Peter. Adaptive Wavelet Densities for Monte Carlo Ray Tracing. In *WSCG*, pages 217–224, 1999.

[29] M. Smyk, S. I. Kinuwaki, R. Durikovic, and K. Myszkowski. Temporally Coherent Irradiance Caching for High Quality Animation Rendering. In *Eurographics*, pages 401–412, 2005.

[30] M. Smyk and K. Myszkowski. Quality Improvement for Indirect Illumination Interpolation. In *ICCVG*, pages 685–692, 2002.

[31] L. Szécsi, M. Sbert, and L. Szirmay-Kalos. Combined Correlated and Importance Sampling in Direct Light Source Computation and Environment Mapping. *CGF*, 23(3):585–593, 2004.

[32] E. Tabellion and A. Lamorlette. An Approximate Global Illumination System for Computer Generated Films. *TOG*, 23(3):469–476, 2004.

[33] E. Veach. *Robust Monte Carlo Methods for Light Transport Simulation*. PhD thesis, Stanford University, USA, 1997.

[34] G. J. Ward and P. Heckbert. Irradiance Gradients. In *EGWR*, pages 85–98, 1992.

[35] G. J. Ward, F. M. Rubinstein, and R. D. Clear. A Ray Tracing Solution for Diffuse Interreflection. In *SIGGRAPH*, pages 85–92, 1988.

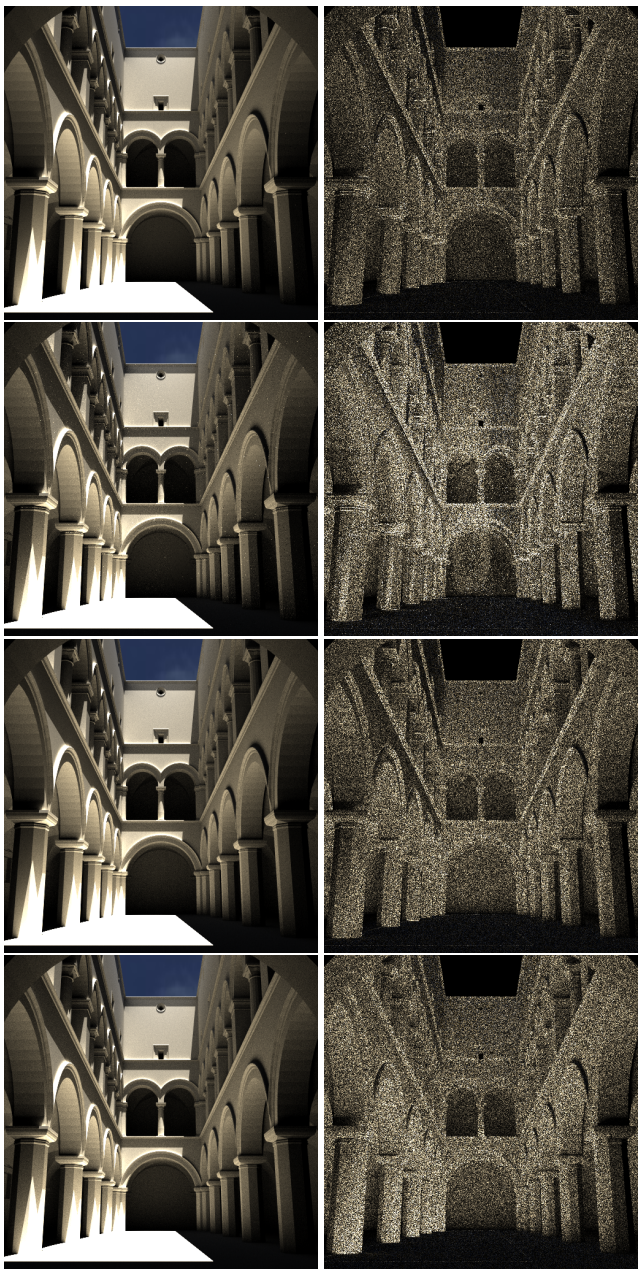

Figure 5: The Sponza Atrium illuminated by a directional light with 1-bounce global illumination, rendered using (from top to bottom) SMC importance sampling and control variates (1024 spp), SMC importance sampling alone (1066 spp), SMC control variates alone (1059 spp), MCMC BRDF importance sampling (1144 spp) all in 2.6 hours.

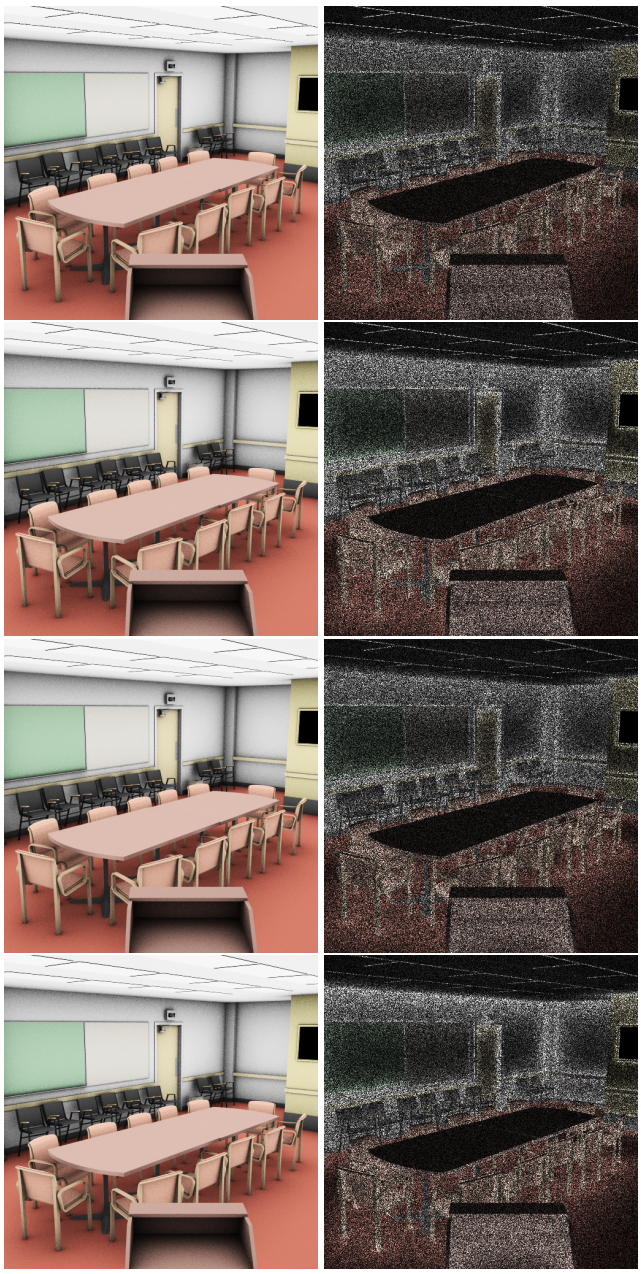

Figure 6: The Conference Room shaded with ambient occlusion, rendered using (from top to bottom) SMC importance sampling and control variates (256 spp), SMC importance sampling alone (261 spp), SMC control variates alone (258 spp), and MCMC BRDF importance sampling (274 spp) all in 43 minutes.

Figure 7: David with a Phong BRDF illuminated by the Grace Cathedral environment map, rendered using (from top to bottom) SMC importance sampling and control variates (1024 spp), SMC importance sampling alone (1074 spp), SMC control variates alone (1026 spp), and MCMC BRDF importance sampling (1203 spp) all in 1.8 hours.

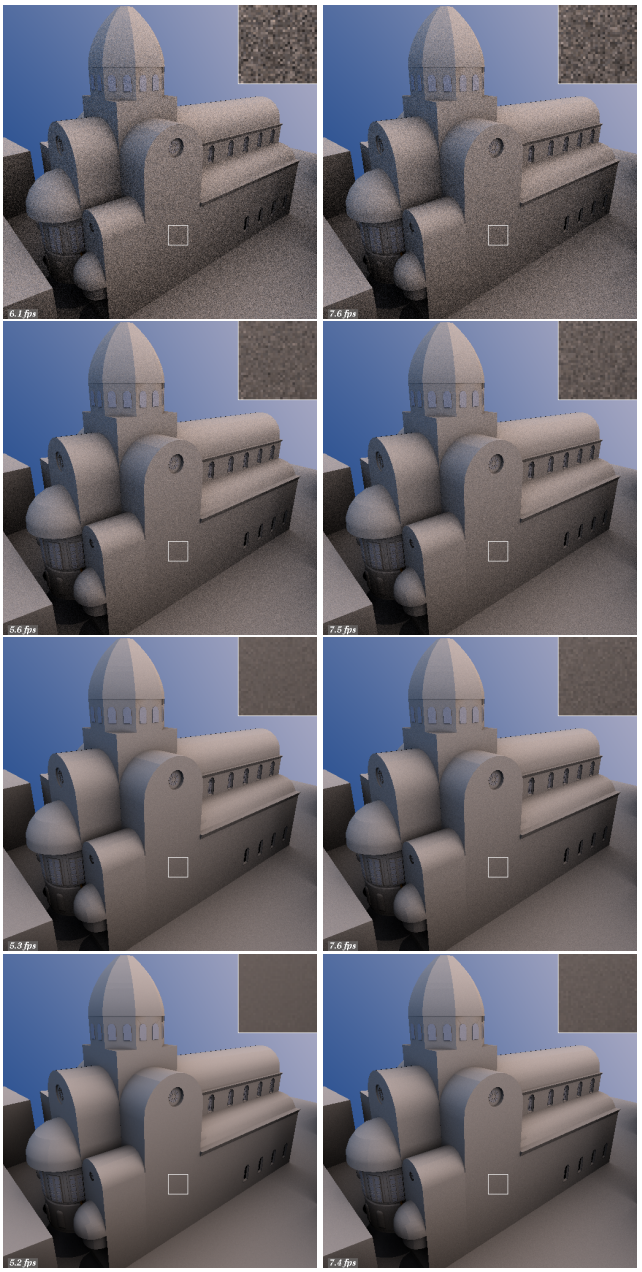

Figure 8: Screenshots of the Sibenik Cathedral during an interactive session rendered progressively using SMC importance sampling and control variates (left) and MCMC BRDF importance sampling (right) by keeping the camera steady (from top to bottom) for 4 seconds, 15 seconds, 1 minute and 4 minutes.