DOI: 10.1111/cgf.15049
EUROGRAPHICS 2024 / A. Bermano and E. Kalogerakis
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 43 (2024), Number 2

Enhancing Spatiotemporal Resampling with a Novel MIS Weight

Xingyue Pan [1,2] Jiaxuan Zhang[2] Jiancong Huang[2] and Ligang Liu[†1]

[1]University of Science and Technology of China
[2]LightSpeed Studios, China

Figure 1: *We introduce the symmetric ratio weight and visibility optimizations, which are designed to effectively reduce energy loss and noise. We conduct a series of tests on the* BISTRO *[Lum17], and* ZERO-DAY *[Win19] scenes, utilizing the RTXDI framework. Our approach is compared with a baseline that employs the generalized pairwise MIS as proposed in Lin et al.'s work. Our method outperforms the baseline in the* BISTRO *scene, achieving a Symmetric Mean Absolute Percentage Error (SMAPE) of 9.72% for the denoised result and 37.16% for the noisy version, compared to the baseline's 12.31% and 37.85%, respectively. Similarly, in the* ZERO-DAY *scene, our method have achieved SMAPEs of 22.5% and 29.6% for the denoised and noisy results, while the baseline achieved 29.9% and 37.6%, respectively.*

Abstract

In real-time rendering, optimizing the sampling of large-scale candidates is crucial. The spatiotemporal reservoir resampling (ReSTIR) method provides an effective approach for handling large candidate samples, while the Generalized Resampled Importance Sampling (GRIS) theory provides a general framework for resampling algorithms. However, we have observed that when using the generalized multiple importance sampling (MIS) weight in previous work during spatiotemporal reuse, variances gradually amplify in the candidate domain when there are significant differences. To address this issue, we propose a new MIS weight suitable for resampling that blends samples from different sampling domains, ensuring convergence of results as the proportion of non-canonical samples increases. Additionally, we apply this weight to temporal resampling to reduce noise caused by scene changes or jitter. Our method effectively reduces energy loss in the biased version of ReSTIR DI while incurring no additional overhead, and it also suppresses artifacts caused by a high proportion of temporal samples. As a result, our approach leads to lower variance in the sampling results.

CCS Concepts

• ***Computing methodologies*** → *Rendering; Ray tracing; Monte Carlo algorithms;*

† The corresponding author (e-mail: lgliu@ustc.edu.cn).

1. Introduction

Real-time rendering of complex scenes with millions of dynamic light sources and indirect illumination has been a long-standing challenge in computer graphics. The advent of hardware-accelerated ray tracing in modern GPUs has opened up new possibilities for achieving high-quality rendering at interactive frame rates. However, even with these advancements, the limited number of rays that can be traced per pixel in real-time applications still poses a significant challenge for path tracing, especially when combined with state-of-the-art denoising algorithms.

Recently, spatiotemporal reservoir resampling (ReSTIR [BWP*20, WKL*23]) has demonstrated promising results in rendering scenes with millions of light sources using a minimal number of shadow rays per pixel. ReSTIR achieves efficient sampling by combining samples from different sampling domains. Subsequently, ReSTIR has been extended to compute global illumination, volume rendering and path tracing, among others.

Our method addresses the limitations of Multiple Importance Sampling (MIS) weights within the framework of Generalized Resampled Importance Sampling (GRIS). In scenarios where the radiance distribution exhibits significant spatial variations, spatial reuse may introduce additional variance. To mitigate this issue, we introduce a novel symmetric ratio MIS weight, which effectively improves the situation, particularly in scenes with significant spatial variations due to factors such as geometry edges, normal maps, spatially varying materials, and shadow edges.

Furthermore, we apply our MIS weight to temporal reuse, enhancing the sensitivity of the results to changes and efficiently sharing information about important paths that contribute to lighting across both spatial and temporal dimensions. Our improvements do not incur additional overhead and achieve better results in multiple scenarios compared to the original method.

Our contributions can be summarized as follows:

- We propose a novel MIS approach within the ReSTIR pipeline, effectively resolving the issue caused by combining neighborhood samples that exhibit significant differences in target radiance distribution compared to the center sample.
- We explore the impact of different Multiple Importance Sampling (MIS) weights on the final variance, offering valuable insights into the selection of MIS weights for improved rendering quality.

2. Related Work

The field of rendering algorithms has witnessed significant advancements, with techniques such as sampling importance resampling (SIR) [Rub87] and multiple importance sampling (MIS) [VG95] being widely adopted. Talbot [TCE05] proposed Resampling Importance Sampling (RIS), which utilizing resampling to accelerate the process of light transport rendering. Subsequently, ReSTIR [BWP*20] introduced weighted reservoir sampling [Cha82] to perform RIS in a streaming fashion. It also incorporates spatiotemporal reuse to obtain additional candidates for resampling.

ReSTIR is an efficient Monte Carlo algorithm initially proposed for computing direct illumination and later extended to compute global illumination [OLK*21], volume rendering [LWY21], path tracing [LKB*22], and differentiable rendering [CSN*23]. To reduce the variance of ReSTIR, there are currently three main approaches.

Engineering optimizations, such as adjusting the pipeline of ReSTIR to achieve more efficient light computation [WP21]. Another optimization involves storing the reservoir in a hash grid built entirely on the GPU, making it more cache-friendly during reuse, as demonstrated by world-space spatiotemporal reservoir reuse [Boi21].

Enhancing the efficiency of sampling, particularly reducing the discrepancy between the initial sampling distribution and the radiance distribution, represents another avenue for optimization. Grid-based reservoirs [BJW21] proposes a technique that involves pre-sampling local lights and storing them in a uniform grid for initial sampling. Additionally, traditional sampling techniques like path guiding and precomputed PDFs [VKv*14, HEV*16, MGN17, MMR*19] can be considered to optimize the initial sampling process.

More effective blending of spatiotemporal samples. ReSTIR [BWP*20] rejects neighborhoods with significant differences by considering factors such as depth and normal thresholds, and combines the neighborhoods using MIS weights similar to those in [TCE05]. Another approach is to estimate the differences in target PDFs between different reservoirs [Tok23] to reject neighborhood candidates, but requires an additional round of resampling to estimate the PDFs. Lin et al. [LKB*22] extended the RIS theory and provided more effective pairwise MIS weights. Our method focuses primarily on this aspect and introduces a novel symmetric ratio MIS weight that is more suitable for some RIS scenarios, without any additional time or space overhead.

Optimal MIS weight. Multiple Importance Sampling (MIS) [VG95] provides an estimator for combining Monte Carlo integrals, robustly blending contributions from different sampling domains, and has been extended to bidirectional path tracing [Vea98]. Numerous works have focused on investigating how to choose more appropriate MIS weights. For instance, [LPG13] optimized the allocation of samples between BSDF and environment maps by estimating variances. [SHSK18] proposed a balance heuristic estimator. [KVG*19] theoretically analyzed the variance of optimal MIS weights and established a connection between the new weighting functions and control variate. [HGS23] extended the theory using a control variate to address challenges in real-world rendering applications. Previous work primarily focused on general MIS optimization, while our approach introduces MIS optimization specifically tailored for RIS.

3. Background

We assume that readers already have a basic understanding of importance sampling concepts, such as importance sampling, multiple importance sampling, and resampling. In this section, we will begin by introducing the ReSTIR algorithm [BWP*20, LKB*22, WKL*23] and then discuss how *Multiple Importance Sampling* (MIS) plays a role in enhancing the effectiveness of ReSTIR.

3.1. ReSTIR

ReSTIR [BWP*20] is a spatiotemporal reservoir resampling method, extending *Resampled Importance Sampling* (RIS) in [TCE05]. Sampling from a complex distribution of lights can be effectively achieved by ReSTIR, that generates samples consistent with the lighting distribution using a two-pass algorithm: generating samples for reservoirs, and streaming combined candidates from spatiotemporal reservoirs by resampling.

To solve the lighting equation

$$f_o = \int_\Omega f_i(x)\,\mathrm{d}x = \int_\Omega L_e \cdot \rho \cdot G, \tag{1}$$

where L_e represents radiance, ρ represents the Bidirectional Scattering Distribution Function (BSDF), and G represents the geometric term, we often employ Monte Carlo sampling. In this context, resampled importance sampling plays a significant role.

Resampled Importance Sampling (RIS) generates candidates x_i from a source PDF p_i (e.g. $p_i \propto L_e$) and resamples them with a probability of w_i proportional to the target function $\hat{p}$ (e.g. $\hat{p} \propto L_e \cdot \rho \cdot G$) to obtain the sample Y. Lin et al. proposed candidate weight w_i can be expressed as

$$w_i = m_i(T_i(x_i))\,\hat{p}(T_i(x_i))\,W_i \cdot \left|\frac{\partial T_i}{\partial x_i}\right|, \tag{2}$$

where m_i is the weight of Multiple Importance Sampling (MIS), m_i should satisfy $\forall x \in \Omega$, $\sum_{i=1}^{M} m_i(x) = 1$, with M denoting the count of distinct sampling strategies employed. In the next section, we will provide a detailed explanation of m's specific form. T_i is a shift mapping from domain Ω_i of x_i to the integral domain Ω. W_i is an unbiased contribution weight, $W_i = 1/p_i(x_i)$ when generating samples from a tractable PDF p_i. The integral can be approximated unbiasedly using

$$\int_\Omega f(x)\,\mathrm{d}x = E[f(Y)\,W_Y] \approx f(Y)\,W_Y, \tag{3}$$

where

$$W_Y = \frac{1}{\hat{p}(Y)} \sum_{i=1}^{M} w_i \tag{4}$$

is an unbiased contribution weight for estimating the reciprocal PDF of the resampled sample. An unbiased contribution weight means that for a random variable $X \in \Omega$,

$$E[f(X)\,W] = \int_{\mathrm{supp}(X)} f(x)\,\mathrm{d}x \tag{5}$$

for any integrable function $f : \Omega \to \mathbb{R}$, the integral is naturally limited to where $p > 0$, i.e., $\mathrm{supp}(X)$, where p represents the probability distribution of the random variable X.

The ReSTIR algorithm generates initial $k = 1, \ldots, n$ samples from a sequence of tractable source PDFs p_i for every reservoir i, and resamples using the unbiased contribution weight $W_{ik}(x_{ik}) = 1/p_i(x_{ik})$. Since $\Omega_{ik} = \Omega_i$, $T_{ik} = T_i$, $p_{ik} = p_i$. the resampling weight for generating initial samples is given by $w_{ik}(x_{ik}) = (1/M) \cdot (\hat{p}(x_{ik})/p_i(x_{ik}))$.

The core of ReSTIR is spatiotemporal reuse. Since the PDF of reservoir samples is intractable, we use the unbiased contribution weight from Eq. (4) as the candidate weight in equation Eq. (2). While reusing samples from the spatiotemporal reservoir, ReSTIR effectively maintains the unbiased nature of the sampling results.

3.2. MIS Weights in Resampling

Even when using a target function that is consistent with the distribution of the function being integrated, i.e. $\hat{p} = f$, the variance remains the same as importance sampling with the same number of samples. The resampling process itself does not reduce variance.

The effectiveness of resampling lies in the unbiased reuse of spatiotemporal samples, which exponentially increases the effective number of samples. Therefore, the crucial factor that affects the variance in resampling is how to combine samples from different sampling domains. The simplest approach is to use a constant weight, considering each sample equally important.

Constant MIS weight with $1/M$ is defined as

$$m_i(y) = \frac{1}{M}. \tag{6}$$

Although weights of $1/M$ are optimal in many instances (such as when different candidates are independently and identically distributed), Lin et al. [LKB*22] pointed out that this is typically not the case when the candidates are not identically distributed.

[TCE05] suggested the inclusion of Multiple Importance Sampling (MIS) weights during the resampling process. [LKB*22] extended this to GRIS and defined the symbol $\hat{p}_{\leftarrow i}$ to represent "$\hat{p}$ from i".

$$\hat{p}_{\leftarrow i}(y) = \begin{cases} \hat{p}_i(T_i^{-1}(y)) \left|(\partial T_i^{-1})/(\partial Y_i)\right|, & \text{if } y \in T_i(\mathrm{supp}(X_i)), \\ 0, & \textit{otherwise}. \end{cases} \tag{7}$$

Generalized balanced heuristic MIS weight is defined as

$$m_i(y) = \frac{\hat{p}_{\leftarrow i}(y)}{\sum_{j=1}^{M} \hat{p}_{\leftarrow j}(y)}. \tag{8}$$

They choose the balance heuristic MIS weight, which was initially introduced by Talbot et al. [TCE05] and later extended to neighborhood combining in resampling by Bitterli et al. [BWP*20]. However, it is unfortunate that the computation complexity of this weight is $O(M^2)$, making it unsuitable for ReSTIR when using a streaming sampler. [BWP*20] proposed another multiple importance sampling method.

Constant MIS weight with contribution follows the same form as balanced heuristic. Specifically, it first performs resampling using a constant weight, given by

$$W_Y = \frac{c_i(Y)}{m_i(Y)} \cdot \frac{1}{\hat{p}(Y)} \sum_{i=1}^{M} w_i, \tag{9}$$

where c_i is only related to the final selected sample Y from M resamplings. m_i can be a constant function, such as $1/M$. By evaluating the weight only once for the selected sample, this approach reduces the complexity to $O(M)$.

To ensure more accurate weights for each sample, Lin et al. [LKB*22] divided the samples into canonical samples and non-canonical samples.

Definition 1 (Canonical Sample) An input sample $X_i \in \Omega_i$ is canonical if its domain is Ω, it uses the identity shift map $T_i(x) = x$, uses $\hat{p}_i = \hat{p}$, and covers $\operatorname{supp}\hat{p}$ (i.e., $\operatorname{supp}\hat{p} \subset \operatorname{supp} X_i$).

Let R denote the set of canonical samples, and $|R|$ represent the number of canonical samples in each resampling. Lin et al. [LKB*22] introduced a pairwise MIS weight.

Generalized pairwise MIS is defined as

$$m_i(y) = \begin{cases} \dfrac{1}{M-|R|} \displaystyle\sum_{j \notin R} \dfrac{\hat{p}(y)}{|R|\hat{p}(y) + (M-|R|)\,\hat{p}_{\leftarrow j}(y)}, & \text{if } y \in R, \\ \dfrac{\hat{p}_{\leftarrow i}(y)}{|R|\hat{p}(y) + (M-|R|)\,\hat{p}_{\leftarrow i}(y)}, & \text{if } y \notin R. \end{cases} \tag{10}$$

Lin et al. [LKB*22] extended Talbot's method to the pairwise case. This approach can be seen as replacing $\sum_{j \notin R} \hat{p}_{\leftarrow j}(y)$ with $(M-|R|)\hat{p}_{\leftarrow i}(y)$ for non-canonical samples. This method reduces the computational complexity of MIS weights from $O(M^2)$ to $O(M \cdot |R|)$. Typically, there is only one canonical sample in a single reuse, which can be seen as reducing the complexity to $O(M)$. And in temporal resampling, it is same as generalized Tablot weight, because both $M-|R|$ and $|R|$ equals to 1. Additionally, in Lin et al. [LKB*22], another "defensive" MIS weight is also proposed.

Generalized pairwise MIS (defensive case) is defined as

$$m_i(y) = \begin{cases} \dfrac{1}{M} + \dfrac{1}{M} \displaystyle\sum_{j \notin R} \dfrac{\hat{p}(y)}{|R|\hat{p}(y) + (M-|R|)\,\hat{p}_{\leftarrow j}((y)}, & \text{if } y \in R, \\ \dfrac{M-|R|}{M} \dfrac{\hat{p}_{\leftarrow i}(y)}{|R|\hat{p}(y) + (M-|R|)\,\hat{p}_{\leftarrow i}(y)}, & \text{if } y \notin R. \end{cases} \tag{11}$$

This method heuristically reduces the weight of non-canonical samples. Specifically, it multiplies the original weight of non-canonical samples by a coefficient of $(M-|R|)/M = 1-|R|/M$. This appropriately reduces the weight of non-canonical samples, with the proportion of non-canonical samples decreasing as the number of canonical samples increases.

3.3. Comparison of Different MIS Weights

In this section, we provide an intuitive explanation for the motivation behind introducing a new MIS weight, and present convergence analysis and variance analysis.

MIS is used to combine samples from different sampling probability functions. [VG95] suggested that it is possible to combine samples from lighting L_e and BSDF ρ sampling, both of which are important components contributing to the lighting function. They dominate in rough or smooth surfaces, respectively. This mixture can be seen as combining samples with different strengths and weaknesses (see Figure 2). However, the situation is different during the resampling process. In the ReSTIR pipeline, we have a sampling distribution that closely resembles the target function, which is then combined with samples from neighboring regions. However, for those neighbor samples whose radiance distribution significantly differs from the original, we refer to them as "unreliable neighbors". We aim to assign lower weights to these unreliable neighbors when they are combined, rather than increasing their

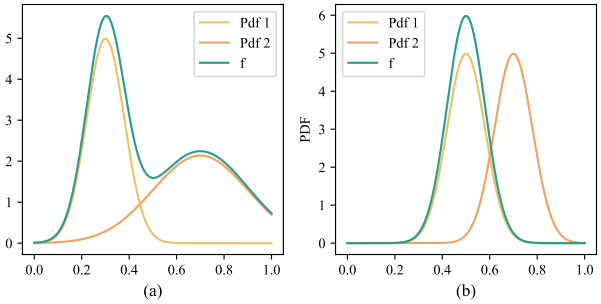

Figure 2: *Different sampling PDFs have distinct advantages in the integration process in (a), and MIS yields results with lower variance reduction. However, in the ReSTIR pipeline, as the resampling distribution converges to the target PDF, the distribution generated by non-canonical reservoirs is always inferior to the canonical distribution. Using a balance heuristic mixture may increase the variance.*

MIS weight as $\hat{p}$ increased. [Vea98, VG95] proposed the power heuristic to address this situation.

Generalize power heuristic is defined as

$$m_i(y) = \frac{p_{\leftarrow i}(y)^{\beta}}{\sum_{j=1}^{N} p_{\leftarrow j}(y)^{\beta}}. \tag{12}$$

However, the power heuristic alone still cannot effectively suppress the noise introduced by combining samples from different domains. In cases where the neighborhood samples have significantly larger values, i.e., $\hat{p}_{\leftarrow i}(y) \gg \hat{p}(y)$, the weights $m_i(y)$ in balanced heuristic MIS weight Eq. (9), power MIS (Equation 12), and pairwise MIS weight (Equation 10) are approximately equal to 1. To illustrate this, we used two normal distributions, p_1 and p_2, to simulate a reuse scenario, where $p_1 = p$ represents the canonical weight. We randomly generated samples x_1 and x_2 from the distributions, and selected the samples with weights $w_i(x_i) = m_i(x_i)\,p(x_i)\,/\,p_i(x_i)$. The resulting distribution is shown in Figure 3. It can be observed that using constant, balanced heuristic, or power heuristic all lead to the final distribution being influenced by p_2, with power heuristic being the most affected. In the following sections, we propose our symmetric ratio MIS weight (shown in Figure 3 the fourth column), which tends to ensure that the distribution after reuse closely resembles the distribution of the target function, without being significantly influenced by unreliable neighborhood distributions.

4. Method

4.1. Symmetric Ratio Weight

We define the difference between two functions as D,

$$D[f,g](x) = \min\left(\frac{f(x)}{g(x)}, \frac{g(x)}{f(x)}\right), \tag{13}$$

where $D[\cdot,\cdot]$ represents the difference between two functions, with range of $[0,1]$. The larger the difference, the smaller the corre-

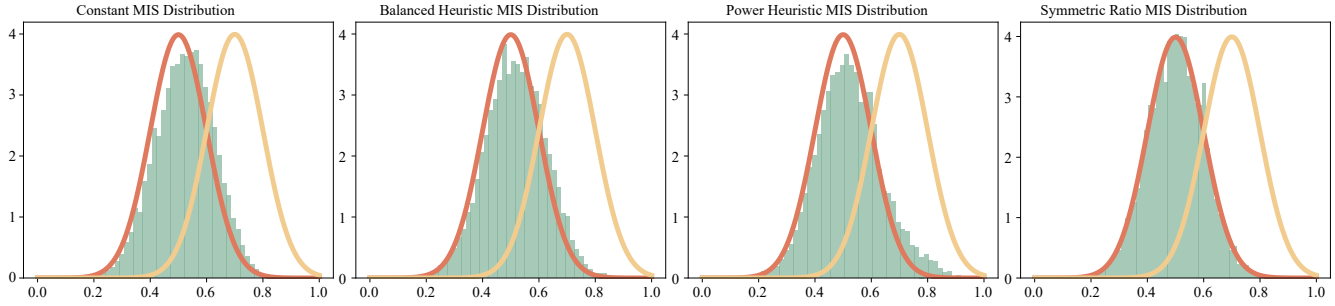

Figure 3: *Light green bar chart represents probability distribution results after a single combining using different MIS weights with power $\beta = 10$ for power heuristic and symmetric ratio MIS. The red function represents the canonical distribution, and the yellow function represents the neighborhood distribution. The leftmost column shows the result using constant MIS, the second column shows the balance heuristic MIS, the third column shows power heuristic MIS, and the rightmost column shows our proposed symmetric ratio MIS. It can be observed that as the power increases, the distribution of the power weight p_Y, deviates more from the original distribution. However, our symmetric ratio weight produces results that are very close to the original distribution.*

sponding symmetric ratio weight. We define the corresponding MIS weight named **symmetric ratio weight** as follows:

$$m_i(y) = \frac{D[\hat{p}, \hat{p}_{\leftarrow i}(y)]^\beta}{\sum_{j=1}^{N} D[\hat{p}, \hat{p}_{\leftarrow j}(y)]^\beta}. \tag{14}$$

Additionally, we extend this concept to the pairwise form:

$$m_i(y) = \begin{cases} \dfrac{1}{M-|R|} \displaystyle\sum_{j \notin R} \dfrac{1}{|R| + (M-|R|) D[\hat{p}, \hat{p}_{\leftarrow j}]^\beta(y)}, & \text{if } y \in R, \\ \dfrac{D[\hat{p}, \hat{p}_{\leftarrow i}](y)^\beta}{|R| + (M-|R|) D[\hat{p}, \hat{p}_{\leftarrow i}](y)^\beta}, & \text{if } y \notin R. \end{cases} \tag{15}$$

Our symmetric ratio weight can be seen as determining the weights of neighborhood samples based on the differences observed in the canonical sample. This approach helps suppress the occurrence probability of high-weight samples Y in the neighborhood, even when they have low canonical weights. This phenomenon often leads to artifacts in ReSTIR, like boiling.

Furthermore, as the exponent β increases, the power weight fails to correctly reject low-weight samples that have a high-weight in neighbor, which can result in convergence to results that deviate significantly from the target distribution, as shown in Figure 3 (c). However, with our symmetric ratio weight, even when reusing samples with significant differences, the probability p_Y of the reused samples remains similar to the original target PDF as Figure 3 (d).

We also propose a more defensive pairwise weight, which converges under more relaxed conditions, called **Asymmetric ratio Weight**:

$$m_i(y) = \begin{cases} \dfrac{D[\hat{p}, \hat{p}_{\leftarrow i}]^\beta(y)}{|R| + (M-|R|) D[\hat{p}, \hat{p}_{\leftarrow j}](y)^\beta}, & \text{if } \hat{p}(y) \le \hat{p}_{\leftarrow i}(y), \\ \dfrac{D[\hat{p}, \hat{p}_{\leftarrow i}](y)^\beta}{M}, & \text{if } \hat{p}(y) > \hat{p}_{\leftarrow i}(y), \end{cases} \tag{16}$$

when $y \notin R$. And $m_i(y) = \left(1 - \sum_{j \notin R} m_j(y)\right) / |R|$, when $y \in R$. This approach assigns higher weights to canonical samples, ensuring that when the number of samples M and the number of canonical samples $|R|$ increase, the convergence condition in GRIS is relaxed. We provide detailed explanations in the next subsection.

4.2. Variance Analysis

In this section, we analyze the variance of RIS and provide an explanation for the rationale behind defining our symmetric ratio MIS weight.

Assuming that each sample X_i in the reservoir is reasonably distributed for integrating $\hat{p}_i$, i.e., $\exists C_i$, s.t. $\hat{p}_i W_i < C_i$, where W_i is the unbiased contribution weight of x_i, then w_i has an upper bound:

$$w_i \le \frac{C_i}{|R|}. \tag{17}$$

Lin et al. had proven that the balanced heuristic MIS Eq. (9), pairwise MIS Eq. (10), and pairwise defensive MIS Eq. (11) hold. Similarly, our symmetric and asymmetric ratio weights also satisfy this condition, as follows:

Theorem 1 (Convergence of Symmetric Ratio Weight) If there exists C_i such that $\hat{p}_i W_i < C_i$, the power $\beta \ge 1$, then the symmetric ratio weight satisfies $w_i(y) \le \frac{C_i}{|R|}$.

Theorem 2 (Convergence of Asymmetric Ratio Weight) If there exists C_i such that $\hat{p}_i W_i < C_i$, the power $\beta \ge 1$, then the asymmetric ratio weight satisfies $w_i(y) \le \frac{C_i}{M}$, when $y \notin R$; $w_i(y) \le \frac{C_i}{|R|}$, when $y \in R$.

For independent samples, if the condition $w_i \le C_i/|R|$ holds, and there exists a common upper bound C, such that $C_i \le C$, then $\mathrm{Var}[\sum_{i=1}^{M} w_i] = \sum_{i=1}^{M} \mathrm{Var}[w_i]$ have

$$\mathrm{Var}[\sum_{i=1}^{M} w_i] \le \sum_{i=1}^{M} \frac{C_i^2}{4|R|^2} = \frac{C_i^2 M}{4|R|^2}, \tag{18}$$

by applying Popoviciu's inequality, We observe that for

$\mathrm{Var}[\sum_{i=1}^{M} w_i]$ to converge, it is necessary for $M/|R|^2 \to 0$, which implies that $|R|$ should grow faster than $\sqrt{M}$. However, our asymmetric ratio weight only requires $|R| \to \infty$ for convergence.

Theorem 3 (Variance of Asymmetric ratio Weight) For asymmetric ratio weight, $\mathrm{Var}[\sum_{i=1}^{M} w_i]$ converges to 0, when $|R| \to \infty$.

In other words, when using the asymmetric ratio weight, we do not need to ensure that $|R|$ grows faster than $\sqrt{M}$ in order to guarantee convergence of the variance.

The proofs of Theorems 1-3 will be provided in Appendix A.

4.3. Optimal Visibility Sampling

Employing the symmetric and asymmetric ratio weight tends to make the probability distribution p_Y after reuse more akin to the canonical PDF. We have also faced an additional problem that introduces substantial noise and deviation from the canonical distribution when implementing ReSTIR [BWP*20, OLK*21] visibility noise.

In practical applications of RIS, to reduce the cost of visibility tests, [BWP*20, OLK*21] dis not perform visibility tests for every resampling. Typically, they use unshadowed lighting $\hat{p} = L_e \cdot \rho \cdot G$ as the target function. During the initial sampling, ray-tracing is performed only after M resampling iterations, and the weights W of invisible candidates are set to 0. The same operation is performed after spatial reuse.

Figure 4: *Comparison of results with and without visibility optimization. The left image shows the overall comparison, while the right image provides a detailed view. The result with visibility optimization exhibits smoother shadows, while the result without visibility optimization shows significant black noise with zero luminance.*

In fact, setting the weights W of invisible candidates to 0 can be seen as using shadowed lighting $\hat{p} = L_e \cdot \rho \cdot G \cdot V$ as the target function and combining the candidates into an empty reservoir. Since we assume that samples from the canonical reservoir have already passed the visibility test, we perform the visibility test earlier by first resampling the spatial samples after spatial reuse, then tracing visibility rays for the resampled results, and finally combining the neighbor candidates with the canonical candidates, as shown in Algorithm 1.

Like the biased version of ReSTIR DI, our method is not an unbiased algorithm. Maintaining unbiasedness during spatial reuse requires an additional $M-1$ ray tracing operations. However, overall, our improvements yield a lower bias compared to the original method without incurring additional overhead. Comparative results will be presented in the following section.

Data: Reservoirs after temporal resue
Result: Spatial reuse result

Function `Spatial Reuse(`*centerReservoir*`)` **:**
  Reservoir r;
  for $i \leftarrow 1$ **to** M **do**
    Reservoir s;
    $s \leftarrow$ loadNeighborReservoir(canonicalReservoir);
    r.combine(s);
  end
  if *TraceVisibility(surface, r)* **:**
    $r.W \leftarrow 0$;
  r.combine(canonicalReservoir);
  return r;

Algorithm 1: Spatial reuse with optimal visibility sampling

5. Result and Discussion

We implement our method on ReSTIR DI [BWP*20], ReSTIR GI [OLK*21] and ReSTIR PT [LKB*22] for evaluation. We integrate our method into the RTXDI framework provided by NVIDIA (`https://github.com/NVIDIAGameWorks/RTXDI`) and compare the results with RTXDI biased version. Since our approach does not introduce additional performance overhead, the comparison of results primarily focuses on the error analysis, without including the performance analysis.

We recommend using asymmetric ratio weight when there is a significant variance or bias introduced during reuse, such as with higher temporal samples. However, in most cases, we recommend using the symmetric ratio weight. In our tests, we find that using symmetric ratio weight and asymmetric ratio weight yield similar results in spatial reuse. Therefore, we maintain consistency by using the same MIS weight for both spatial and temporal reuse in our tests. Additionally, when using the symmetric ratio weight or asymmetric ratio weight, we find that as β increases to 3, the variance decreases for spatial reuse. However, as β continues to increase, the weights of non-canonical samples gradually decrease, resulting in an increase in variance at sampling points where there is a significant difference from the canonical distribution in most neighborhoods. So we choose appropriate parameters of $\beta = 3$ for spatial reuse and $\beta = 1$ for temporal reuse, considering that temporal samples have a higher weight M, such as $M = 20$. For the generalized pairwise weight, we select $\beta = 1$ to align with the original GRIS, using a balance MIS instead of a power MIS.

We present the results of our method in different views of multiple scenes in Figure 5, comparing the outcomes of ReSTIR DI and

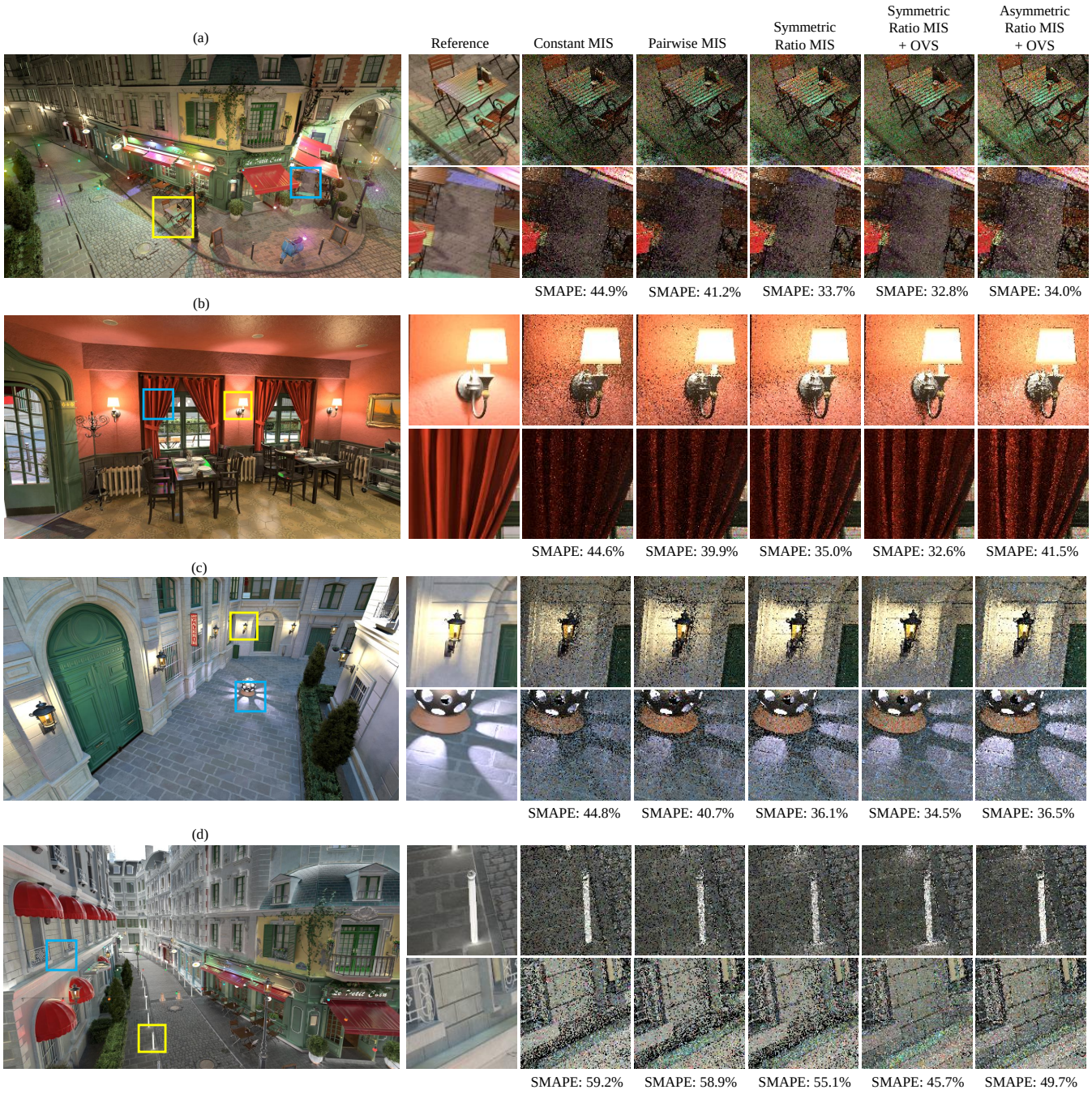

Figure 5: *Comparison of the results with different MIS weight. (a), (b) and (c) shows the results of ReSTIR DI. (d) shows the result with only global illumination. "OVS" means with the optimal visibility samplinig.*

ReSTIR GI, with parameter temporal samples $M = 20$, and one spatial reuse sample. Additionally, we evaluate our method on the BISTRO scene, which consists of 20,638 emissive triangles, and the ZERO-DAY scene, which contains 10,973 dynamic emissive triangles. In Figure 5 (a), we demonstrate that our approach produces less blurring at shadow boundaries. Figure 5 (b), (c), and (d) highlight the effectiveness of our optimal visibility sampling (OVS) in handling high-frequency noise at shadow edges. Near the red curtains in Figure 5 (b), our method exhibits lower variance, while

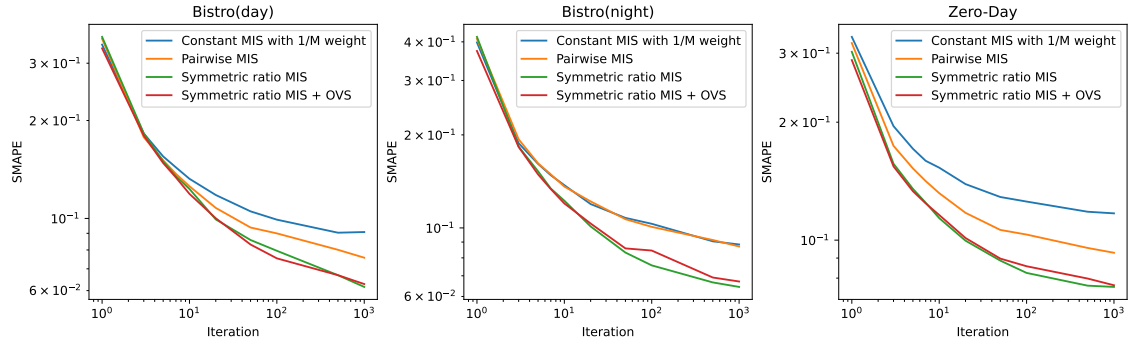

Figure 6: *We evaluate the convergence rate of the SMAPE using different MIS weights as the number of accumulated frames increased. Our proposed method consistently outperformed the original constant weight approach in terms of convergence. The figures above illustrate the SMAPE in the* BISTRO*(day),* BISTRO*(night), and* ZERO-DAY *scenes.*

Figure 5 (c) showcases reduced energy loss compared to ReSTIR DI biased version.

Figure 6 demonstrates the convergence of our method as the number of accumulated frames increases. It is worth noting that our method incurs no additional overhead, making this comparison fair. Our approach achieves significantly faster convergence and lower accumulated bias compared to the generalized pairwise MIS and constant MIS methods.

We also test our method within the ReSTIR PT framework. In the hybrid mode, the results of our symmetric ratio MIS were not significantly different from those obtained with the original pairwise MIS approach. However, in the reconnection mode, by increasing the number of spatial reuse iterations, the differences became more pronounced. Specifically, we employ 5 spatial rounds with 6 neighbor candidates each, as shown in Figure 7. With a higher number of spatial reusing iterations, the overall variance is noticeably reduced. Given that the contribution from neighborhood candidates became more substantial, using the original method could lead to large-area artifacts caused by infrequently occurring neighborhood samples, such as in corners or on metallic teapots. Our method, on the other hand, suppresses unreasonable neighborhood reuse, thereby avoiding such artifacts.

Candidates from the temporal domain are not always canonical. Due to changes in the scene or jitter, the sample space of the previous frame can differ significantly from the current frame. Therefore, it is necessary to use our symmetric ratio weight approach to handle such cases. Our method also reduces the variance caused by temporal differences, shown in Figure 8.

Jitter is commonly used for anti-aliasing. Due to jitter, the surface within a pixel can vary between different surfaces in adjacent frames. Even if the scene remains unchanged, temporal samples cannot be considered as canonical samples when jitter is enabled. In fact, due to the significant weight of the temporal component, jitter introduces more variance than anticipated, especially in scenes with large differences in local lighting functions, resulting in additional noise.

Figure 9 shows a comparison of the results of a smooth metal sphere under ambient lighting, using 4k environment maps of Derelict Overpass and Adamas Place Bridge. The top row utilizes the default parameters of RTXDI, with temporal samples of $M = 20$, and spatial reuse with one neighbor, and the boiling filter turned off. The bottom row uses a higher temporal sample ratio, with temporal samples of $M = 100$, while keeping the other parameters the same as in the first row. Since the initial sampling includes BRDF sampling, an initial sample of the canonical sample already approximates the target function distribution. However, samples from neighbors may have significant differences in distribution compared to the canonical samples, which increase variance when spatial reuse is applied. Additionally, due to the influence of jitter, the surface information of the current frame differs from the previous frame, and a high temporal sample ratio, such as in (a2), also increase variance. The original pairwise MIS cannot control the proportion of inappropriate non-canonical samples, whereas our method avoids this issue. In both temporal reuse and spatial reuse in (d1) and (d2), we utilize the symmetric ratio MIS, which effectively controls the high variance caused by significant differences in non-canonical samples.

Boiling is often considered to occur when low-probability samples are encountered. In methods like pairwise MIS in Eq. (10), the convergence of the resampling distribution p_Y depends on the ratio of the number of canonical samples $|R|$ to the total number of samples M. This means that as the temporal samples M increases, in regions with significant differences in the neighborhood lighting function, especially when there is a large difference in weights between samples with low probability of being sampled, a high-value weight quickly propagate to the neighborhood, resulting in boiling, as shown in Figure 10. The phenomenon of boiling is particularly pronounced when using a larger number of temporal samples. In ReSTIR, the boiling filter is heuristically used to limit samples with low sampling probabilities. However, this method also introduce bias by sacrificing energy.

For global illumination, similar situations also arise in the propagation of Jacobians and the reuse of smoother surfaces. For low-probability samples with larger Jacobians, significant variance can

Figure 7: *A comparison of results obtained by applying different MIS within the ReSTIR PT framework [LKB*22]. The right side of the figure highlights the details of the metallic teapot and the corner of the wall.*

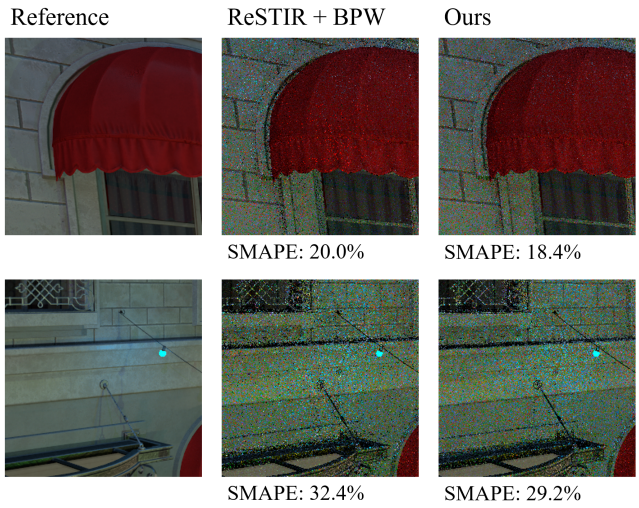

Figure 8: *Comparison of our symmetric ratio weight and generalized pairwise MIS in a dynamic lighting scene. Our method exhibits lower variance near the just illuminated blue light source and also performs better for highlights near the red sunshade.*

be introduced, especially near wall corners where Jacobians may exhibit substantial differences. In ReSTIR GI [OLK*21], the upper bound of the Jacobian is limited to avoid excessive variance. This approach is similar to the boiling filter, as it enforces a maximum weight for the samples, but comes at the cost of energy loss. In our comparative experiments, we refrain from using this trick to ensure unbiased results, as shown in Figure 11.

Bias. In real-time rendering, we do not trace additional rays to check the visibility of the current selection sample in each neighborhood, resulting in slightly darker results compared to the ground truth. By comparing the depth and normal values and setting a threshold for their differences, ReSTIR [BWP*20] heuristically rejects neighborhoods with significant discrepancies to reduce bias and variance.

[LKB*22] introduced the defensive pairwise MIS Eq. (11) heuristic to limit the weights of non-canonical samples from unreliable neighbors, while our method automatically controls the weights from unreliable neighborhoods.

We test different MIS methods in the BISTRO scene, as shown in Figure 12. Subfigure (b) represents ReSTIR with spatial reuse using a constant MIS of $1/M$. Subfigure (c) shows the pairwise MIS as defined in Eq. (10). Subfigure (d) demonstrates the use of our symmetric ratio pairwise MIS with a power of $\beta = 1$. Subfigure (e) represents the asymmetric ratio pairwise MIS used in both spatiotemporal reuse. We observe that our symmetric ratio MIS significantly reduces bias, especially in the case of asymmetric ratio pairwise MIS.

6. Conclusions and Future Work

We believe that our method will effectively avoid the influence of unreliable neighborhoods in regions with low initial sampling variance and demonstrate improved performance as the overall sampling efficiency increases. Our approach automatically balances the weights introduced by reusing candidates, preventing the introduction of excessive noise and bias that hinder convergence.

One of the limitations is that when there is a significant difference between the initial resample result and the target distribution, our conservative reuse of neighborhoods may result in a less significant reduction in variance compared to using uniform weights. To address this issue, we increase the number of temporal samples used in resampling, which has proven effective as our method does not exhibit degradation, such as boiling and GI artifacts shown in the figures above.

In future work, we plan to dynamically select MIS weights and corresponding powers based on surface properties such as roughness and metallic parameters. Additionally, since our method is applicable to any form of resampling, we extend its application to other resampling combining techniques. Another potential avenue of research is to utilize our symmetric ratio function, which is used to assess neighborhood differences, in denoising algorithms to effectively control the introduced bias while removing noise.

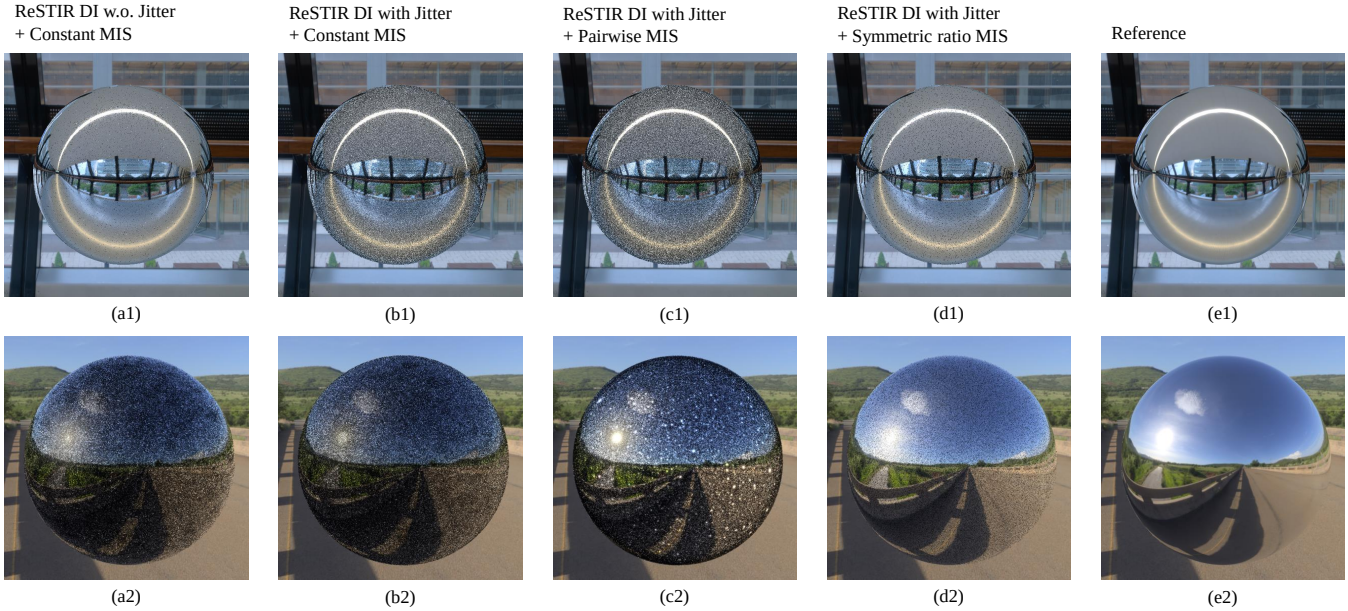

Figure 9: *Comparison of results for a smooth metal sphere. The upper row shows the temporal samples affected by jitter, with temporal samples of $M = 20$, while the lower row uses temporal samples of $M = 100$. With a high M, our symmetric ratio weight ensures a smaller variance in the final results by preserving the distribution of the combined samples close to the target distribution.*

Figure 10: *In the* BISTRO *scene, when using a large number of temporal samples $M = 1000$ and disabling the boiling filter. Bitterli et al. [BWP*20] introduce several artifacts like "boiling" on the ground and walls. This results in some regions appearing overly bright due to spatial reuse, while the overall image appears darker. In the "BPW" approach, we utilize the ray-traced bias correction mode for temporal reuse and pairwise MIS for spatial reuse. In both "DPW" and "ADPW", we employ symmetric ratio pairwise weight or asymmetric ratio pairwise weight for temporal and spatial reuse, respectively. Our method effectively suppress the boiling effect and reduce bias, particularly in ADPW.*

On the other hand, research into weights of temporal samples presents a promising area of study. For instance, [PFJ23] concentrated on computing adaptive temporal weights as opposed to fixed ones. The consideration of gradients or the application of rational weights from our framework to ascertain temporal weights also represents an important aspect of future work.

Acknowledgments

We express our gratitude to the anonymous reviewers for their invaluable comments and suggestions that have significantly improved our research. This work is supported by the National Key R&D Program of China (2022YFB3303400) and the National Natural Science Foundation of China (62025207). We also acknowledge the Amazon Lumberyard team for providing the

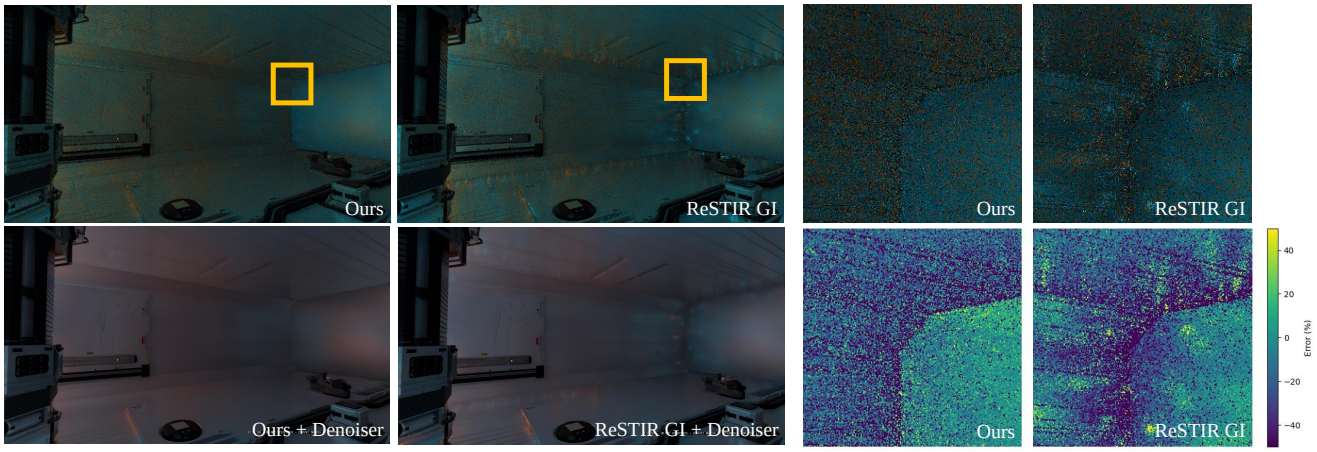

Figure 11: *The images presented depict the occurrence of cone-shaped artifacts at the wall corners in the* ZERO-DAY *scene when using* $M = 300$ *temporal samples and disabling the boiling filter. These artifacts are overly bright and are present in the original ReSTIR GI. However, our proposed asymmetric ratio weight effectively suppresses the occurrence of such artifacts.*

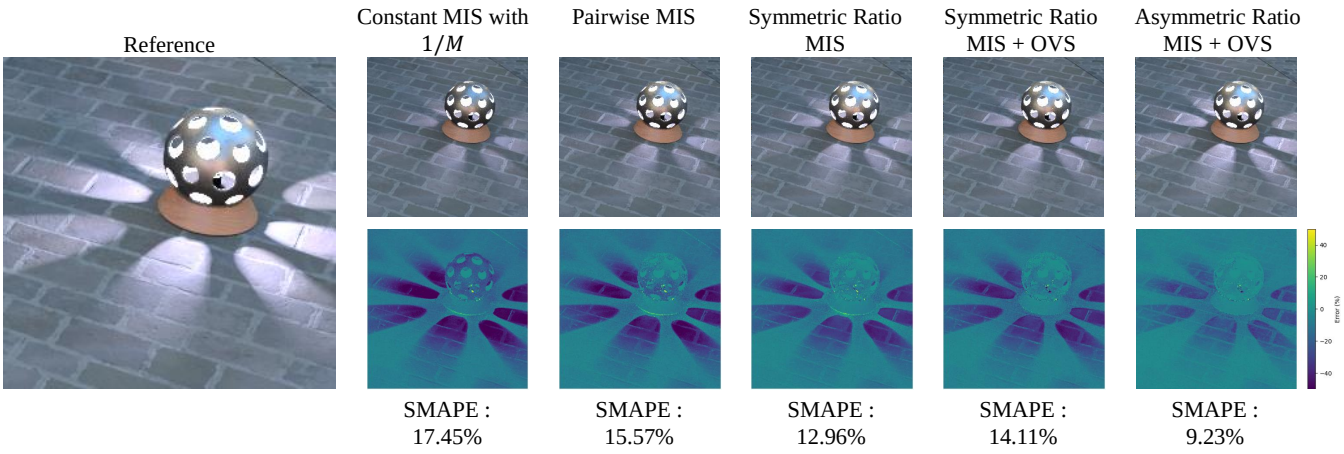

Figure 12: *The images presented depict the comparison of scene convergence when using different MIS weights for accumulating multiple frames in a static scene. The following visualization shows the error graph. In spatial reuse, bias is introduced due to the lack of testing the visibility of selected samples in each neighborhood. However, our proposed symmetric ratio weight, especially when utilizing the asymmetric ratio weight, significantly reduces this bias compared to the original method.*

BISTRO scene, and M. Winkelmann and K. Anderson for their contribution to the ZERO-DAY scene. This work was mainly done by Xingyue Pan during her internship at Lightspeed Studio.

References

[BJW21] BOKSANSKY J., JUKARAINEN P., WYMAN C.: Rendering many lights with grid-based reservoirs. In *Ray Tracing Gems II*, Marrs A., Shirley P., Wald I., (Eds.). APress, August 2021, pp. 351–365. doi:10.1007/978-1-4842-7185-8_23. 2

[Boi21] BOISSÉ G.: World-space spatiotemporal reservoir reuse for ray-traced global illumination. In *SIGGRAPH Asia 2021 Technical Communications* (New York, NY, USA, 2021), SA '21, Association for Computing Machinery. doi:10.1145/3478512.3488613. 2

[BWP*20] BITTERLI B., WYMAN C., PHARR M., SHIRLEY P., LEFOHN A., JAROSZ W.: Spatiotemporal reservoir resampling for real-time ray tracing with dynamic direct lighting. *ACM Trans. Graph. 39*, 4 (aug 2020). doi:10.1145/3386569.3392481. 2, 3, 6, 9, 10

[Cha82] CHAO M. T.: A general purpose unequal probability sampling plan. *Biometrika 69*, 3 (12 1982), 653–656. arXiv:https://academic.oup.com/biomet/article-pdf/69/3/653/591311/69-3-653.pdf, doi:10.1093/biomet/69.3.653. 2

[CSN*23] CHANG W., SIVARAM V., NOWROUZEZAHRAI D., HACHISUKA T., RAMAMOORTHI R., LI T.-M.: Parameter-space restir for differentiable and inverse rendering. In *ACM*

SIGGRAPH 2023 Conference Proceedings (New York, NY, USA, 2023), SIGGRAPH '23, Association for Computing Machinery. doi:10.1145/3588432.3591512. 2

[HEV*16] HERHOLZ S., ELEK O., VORBA J., LENSCH H., KŘIVÁNEK J.: Product importance sampling for light transport path guiding. *Computer Graphics Forum 35*, 4 (2016), 67–77. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12950, doi:https://doi.org/10.1111/cgf.12950. 2

[HGS23] HUA Q., GRITTMANN P., SLUSALLEK P.: Revisiting controlled mixture sampling for rendering applications. *ACM Trans. Graph. 42*, 4 (jul 2023). doi:10.1145/3592435. 2

[KVG*19] KONDAPANENI I., VEVODA P., GRITTMANN P., SKŘIVAN T., SLUSALLEK P., KŘIVÁNEK J.: Optimal multiple importance sampling. *ACM Trans. Graph. 38*, 4 (jul 2019). doi:10.1145/3306346.3323009. 2

[LKB*22] LIN D., KETTUNEN M., BITTERLI B., PANTALEONI J., YUKSEL C., WYMAN C.: Generalized resampled importance sampling: Foundations of restir. *ACM Trans. Graph. 41*, 4 (jul 2022). doi:10.1145/3528223.3530158. 2, 3, 4, 6, 9

[LPG13] LU H., PACANOWSKI R., GRANIER X.: Second-order approximation for variance reduction in multiple importance sampling. *Computer Graphics Forum 32*, 7 (2013), 131–136. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12220, doi:https://doi.org/10.1111/cgf.12220. 2

[Lum17] LUMBERYARD A.: Amazon lumberyard bistro, open research content archive (orca), July 2017. URL: http://developer.nvidia.com/orca/amazon-lumberyard-bistro. 1

[LWY21] LIN D., WYMAN C., YUKSEL C.: Fast volume rendering with spatiotemporal reservoir resampling. *ACM Trans. Graph. 40*, 6 (dec 2021). doi:10.1145/3478513.3480499. 2

[MGN17] MÜLLER T., GROSS M., NOVÁK J.: Practical path guiding for efficient light-transport simulation. *Computer Graphics Forum 36*, 4 (2017), 91–100. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13227, doi:https://doi.org/10.1111/cgf.13227. 2

[MMR*19] MÜLLER T., MCWILLIAMS B., ROUSSELLE F., GROSS M., NOVÁK J.: Neural importance sampling. *ACM Trans. Graph. 38*, 5 (oct 2019). doi:10.1145/3341156. 2

[OLK*21] OUYANG Y., LIU S., KETTUNEN M., PHARR M., PANTALEONI J.: Restir gi: Path resampling for real-time path tracing. *Computer Graphics Forum 40*, 8 (2021), 17–29. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.14378, doi:https://doi.org/10.1111/cgf.14378. 2, 6, 9

[PFJ23] PHILIPPI H., FRISVAD J. R., JENSEN H. W.: Practical Temporal and Stereoscopic Filtering for Real-time Ray Tracing. In *Eurographics Symposium on Rendering* (2023), Ritschel T., Weidlich A., (Eds.), The Eurographics Association. doi:10.2312/sr.20231129. 10

[Rub87] RUBIN D. B.: Comment. *Journal of the American Statistical Association 82*, 398 (1987), 543–546. doi:10.1080/01621459.1987.10478461. 2

[SHSK18] SBERT M., HAVRAN V., SZIRMAY-KALOS L.: Multiple importance sampling revisited: breaking the bounds. *EURASIP Journal on Advances in Signal Processing 2018*, 1 (2018), 15. doi:10.1186/s13634-018-0531-2. 2

[TCE05] TALBOT J., CLINE D., EGBERT P.: Importance Resampling for Global Illumination. In *Eurographics Symposium on Rendering (2005)* (2005), Bala K., Dutre P., (Eds.), The Eurographics Association. doi:10.2312/EGWR/EGSR05/139-146. 2, 3

[Tok23] TOKUYOSHI Y.: Efficient spatial resampling using the pdf similarity. *Proc. ACM Comput. Graph. Interact. Tech. 6*, 1 (may 2023). doi:10.1145/3585501. 2

[Vea98] VEACH E.: *Robust Monte Carlo Methods for Light Transport Simulation*. PhD thesis, Stanford, CA, USA, 1998. AAI9837162. 2, 4

[VG95] VEACH E., GUIBAS L. J.: Optimally combining sampling techniques for monte carlo rendering. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1995), SIGGRAPH '95, Association for Computing Machinery, p. 419–428. doi:10.1145/218380.218498. 2, 4

[VKv*14] VORBA J., KARLÍK O., ŠIK M., RITSCHEL T., KŘIVÁNEK J.: On-line learning of parametric mixture models for light transport simulation. *ACM Trans. Graph. 33*, 4 (jul 2014). doi:10.1145/2601097.2601203. 2

[Win19] WINKELMANN M.: Zero-day, open research content archive (orca), November 2019. URL: https://developer.nvidia.com/orca/beeple-zero-day. 1

[WKL*23] WYMAN C., KETTUNEN M., LIN D., BITTERLI B., YUKSEL C., JAROSZ W., KOZLOWSKI P.: A gentle introduction to restir path reuse in real-time. In *ACM SIGGRAPH 2023 Courses* (New York, NY, USA, 2023), SIGGRAPH '23, Association for Computing Machinery. URL: https://doi.org/10.1145/3587423.3595511, doi:10.1145/3587423.3595511. 2

[WP21] WYMAN C., PANTELEEV A.: Rearchitecting Spatiotemporal Resampling for Production. In *High-Performance Graphics - Symposium Papers* (2021), Binder N., Ritschel T., (Eds.), The Eurographics Association. doi:10.2312/hpg.20211281. 2

Appendix A: Proofs of Theorems

Theorem 1 (Convergence of Symmetric Ratio Weight). If there exists C_i such that $\hat{p}_i W_i < C_i$, the power $\beta \geq 1$, then the symmetric ratio weight satisfies $w_i \leq \frac{C_i}{|R|}$.

Proof

For $y \notin R$, let us denote $\alpha = \frac{M-|R|}{|R|}$,

$$\begin{aligned} w_i(y) &= m_i(y) \cdot \hat{p}(y) W_i \cdot \left| \frac{\partial y}{\partial x_i} \right| \\ &= \frac{1}{|R|} \frac{D[\hat{p}, \hat{p}_{\leftarrow i}](y)^\beta}{1 + \alpha D[\hat{p}, \hat{p}_{\leftarrow j}](y)^\beta} \cdot \hat{p}(y) W_i \cdot \left| \frac{\partial y}{\partial x_i} \right|, \end{aligned} \tag{19}$$

When $\hat{p}_{\leftarrow i}(y) < \hat{p}(y)$, we have that

$$\begin{aligned} &\frac{\partial}{\partial \beta} \left(\frac{\hat{p}_{\leftarrow i}(y)^\beta}{\hat{p}(y)^\beta + \alpha \hat{p}_{\leftarrow i}(y)^\beta} \right) \\ &= \frac{(\ln(\hat{p}_{\leftarrow i}(y)) - \ln(\hat{p}(y))) \hat{p}_{\leftarrow i}(y)^\beta \hat{p}^\beta}{(\hat{p}(y)^\beta + \alpha \hat{p}_{\leftarrow i}(y)^\beta)^2} \leq 0, \end{aligned} \tag{20}$$

so

$$\begin{aligned} w_i(y) &= \frac{1}{|R|} \frac{\hat{p}_{\leftarrow i}(y)^\beta}{\hat{p}(y)^\beta + \alpha \hat{p}_{\leftarrow i}(y)^\beta} \cdot \hat{p}(y) W_i \cdot \left| \frac{\partial y}{\partial x_i} \right| \\ &\leq \frac{1}{|R|} \frac{\hat{p}_{\leftarrow i}(y)}{\hat{p}(y) + \alpha \hat{p}_{\leftarrow i}(y)} \cdot \hat{p}(y) W_i \cdot \left| \frac{\partial y}{\partial x_i} \right| \\ &= \frac{\hat{p}(y)}{|R| \hat{p}(y) + (M - |R|) \hat{p}_{\leftarrow i}(y)} \cdot \hat{p}_{\leftarrow i}(y) W_i \cdot \left| \frac{\partial y}{\partial x_i} \right| \\ &\leq \frac{C_i}{|R|}. \end{aligned} \tag{21}$$

When $\hat{p}_{\leftarrow i}(y) \geq \hat{p}(y)$,

$$\begin{aligned} w_i(y) &= \frac{1}{|R|} \frac{\hat{p}(y)^\beta}{\hat{p}_{\leftarrow i}(y)^\beta + \alpha \hat{p}(y)^\beta} \cdot \hat{p}(y) W_i \cdot \left| \frac{\partial y}{\partial x_i} \right| \\ &\leq \frac{1}{|R|} \frac{\hat{p}(y)^\beta}{\hat{p}(y)^\beta + \alpha \hat{p}(y)^\beta} \cdot \hat{p}(y) W_i \cdot \left| \frac{\partial y}{\partial x_i} \right| \\ &= \frac{1}{M} \cdot \hat{p}(y) W_i \cdot \left| \frac{\partial y}{\partial x_i} \right| \\ &\leq \frac{1}{M} \cdot \hat{p}_{\leftarrow i}(y) \left| \frac{\partial y}{\partial x_i} \right| \cdot W_i \\ &\leq \frac{C_i}{M}. \end{aligned} \tag{22}$$

For $y \in R$,

$$\begin{aligned} w_i(y) &= \frac{1}{(M-|R|)\cdot|R|} \sum_{j \notin R} \frac{1}{1 + \alpha D[\hat{p}, \hat{p}_{\leftarrow j}](y)^\beta} \cdot \hat{p}(y) W_i \cdot \left| \frac{\partial y}{\partial x_i} \right| \\ &= \frac{1}{(M-|R|)\cdot|R|} (M-|R|) \cdot \hat{p}(y) W_i \\ &\leq \frac{C_i}{|R|}. \end{aligned} \tag{23}$$

□

Theorem 2 (Convergence of Asymmetric Ratio Weight). If there exists C_i such that $\hat{p}_i W_i < C_i$, the power $\beta \geq 1$, then the asymmetric ratio weight satisfies $w_i(y) \leq \frac{C_i}{M}$, when $y \notin R$, $w_i(y) \leq \frac{C_i}{|R|}$, when $y \in R$.

Proof
We have proven that $w_i(y) \leq C_i/|R|$, when $y \in R$, and $w_i(y) \leq C_i/M$, when $y \notin R$ and $\hat{p}_{\leftarrow i}(y) \geq \hat{p}(y)$.
It remains to show that $w_i(y) \leq C_i/M$, when $y \notin R$ and $\hat{p}_{\leftarrow i}(y) < \hat{p}(y)$. In this case,

$$\begin{aligned} w_i(y) &= \frac{\hat{p}_{\leftarrow i}(y)^\beta}{M \cdot \hat{p}(y)^\beta} \cdot \hat{p}(y) W_i \cdot \left| \frac{\partial y}{\partial x_i} \right| \\ &\leq \frac{\hat{p}_{\leftarrow i}(y)}{M \cdot \hat{p}(y)} \cdot \hat{p}(y) W_i \cdot \left| \frac{\partial y}{\partial x_i} \right| \\ &= \frac{\hat{p}(y)}{M \cdot \hat{p}(y)} \cdot \hat{p}_{\leftarrow i}(y) W_i \cdot \left| \frac{\partial y}{\partial x_i} \right| \\ &\leq \frac{C_i}{M} \end{aligned} \tag{24}$$

□

Theorem 3 (Variance of Asymmetric Ratio Weight). For independent samples, the variance $\mathrm{Var}[\sum_{i=1}^M w_i]$ using asymmetric ratio weight convergence to 0, when $|R| \to \infty$.

Proof

$$\begin{aligned} Var[\sum_{i=1}^M w_i] &= \sum_{i=1}^M \mathrm{Var}[w_i] \\ &= \sum_{y_i \in R} \mathrm{Var}[w_i] + \sum_{y_i \notin R} \mathrm{Var}[w_i] \\ &\leq |R| \frac{C^2}{4|R|^2} + (M-|R|) \frac{C^2}{4M^2} \\ &= \frac{C^2}{4} \left(\frac{1}{|R|} + \frac{M-|R|}{M^2} \right) \end{aligned} \tag{25}$$

It is evident that the variance converges to 0 as $|R| \to \infty$. □