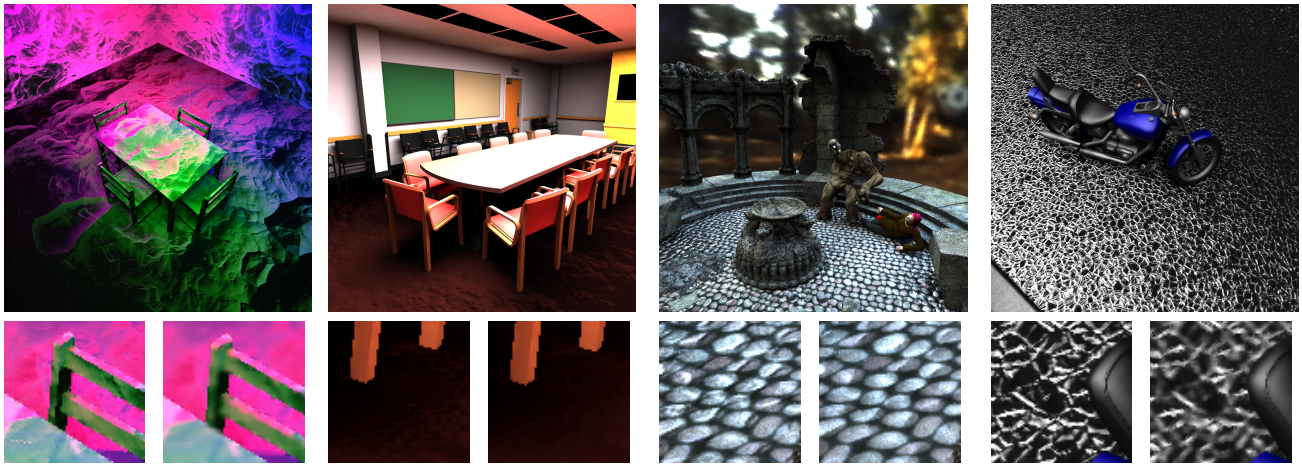
Screen Space Spherical Harmonics Filters for Instant Global Illumination

Benjamin Segovia Ingo Wald

Intel Labs



(a) SciFi-ERW6 (b) Conference Room (c) Temple (d) Night Motorclycle

Figure 1: The top row shows four 1024×1024 pictures using 4×4 interleaved sampling patterns and the *incoming* radiance filter presented in this paper (9 spherical harmonics coefficients). The bottom row provides close-ups. Our technique is used on the left pictures whereas the right pictures are computed using a classical filter on the *ir*radiance. (a) shows a non-realistic scene using three colored light sources and a high-frequency, complex normal map. (b) shows the conference room with a normal mapped floor. A global illumination algorithm (Instant Radiosity) was also used. (c) presents another application using a sampled light probe while (d) shows a motorcycle on a wet floor directly illuminated by two area light sources. In all cases, filtering the incoming radiance with our technique provides much better and less blurry results than simply filtering irradiance.

Abstract

We present a general technique that allows for appling the instant global illumination (IGI) technique to scenes with glossy BRDFs and high-frequency geometric features like bump and normal maps. Like IGI, we use $n \times m$ interleaved sampling patterns to reduce the lighting computations by a factor of $n \times m$, and filter incoming illumination in screen space. Unlike IGI, we filter (per-pixel) the *incoming radiance field* rather than the normal-dependent *irradiance*. To represent the incoming radiance field, we use spherical harmonics functions whose coefficients will then be filtered in screen space. This provides us with directional information on the incident illumination, which then allows to apply it to non-diffuse BRDFs, or on surfaces with strongly varying normals such as normal, bump, or displacement mapped surfaces.

1 Introduction

In computer graphics, lighting computations that are both fast and accurate remain a challenge. Even if we are getting closer to a real time refresh of physically based solutions in simple cases (for diffuse-only scenes and easy-to-find relevant light paths), non-interactive simulation times are often required to achieve high-quality results.

In this paper, we present a rendering technique that aims for the following features:

- be cheap enough for interactive use;
- handle direct as well as indirect lighting;
- preserve details for both direct and indirect “hard” shadows;
- preserve high frequencies on the surfaces of the objects due to, for example, normal or displacement maps.

Some of these goals can be achieved with the “Instant Global Illumination” (IGI) technique [Wald et al. 2002]. IGI is conceptually simple: by replacing the whole radiance field with point lights, it allows for handling both direct and indirect lighting as well as soft and hard lighting effects. In combination with fast screen space filters to reduce the computation overhead, IGI is also fast, and there are, by now, multiple real time implementations using either coherent ray tracing or GPU based rasterization approaches [Laine et al. 2007; Ritschel et al. 2008]. Even interactive frame rates in current games have already been reported [Shishkovtsov 2005].

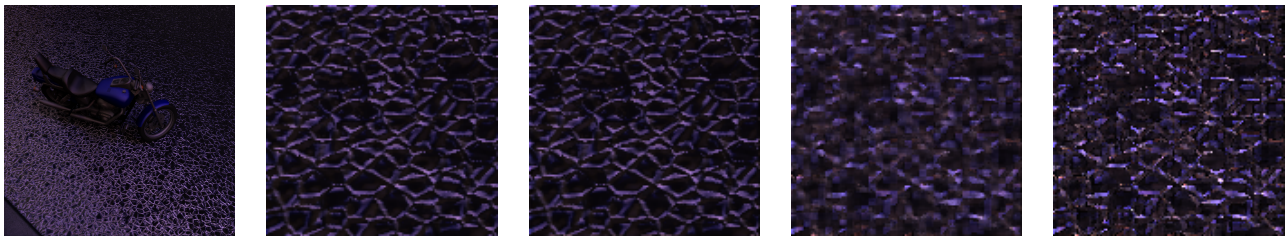
However, as shown on Figure 2, the original IGI method performs a screen-space filter on the *irradiance* seen through each pixel, which limits it to mostly-diffuse BRDFs and, in particular, to surfaces whose normals do vary but slowly (since the irradiance depends on the normal). Thus, traditional IGI is incompatible with highly detailed surfaces such as, for example, normal, bump, or displacement mapped surfaces. Indeed, once surface features become smaller than the filter applied to the irradiance, these features are destroyed.

Inspired by works on environment map rendering [Ramamoorthi and Hanrahan 2002] and radiance volumes [Nijasure et al. 2003], we realize that spherical harmonics (SHs) allow us to compactly store the entire *incident radiance field* per pixel, to filter this instead of filtering irradiance, and to apply this filtered radiance field to non-diffuse BRDFs as well as surfaces with high-frequency geometric details.

2 Background

Our approach is based on three commonly used strategies:

- Monte-Carlo integration based on virtual point lights to *compute* the incoming radiance field;
- Spherical harmonics to *represent* this radiance field;



(a) (b) (c) (d) (e)

Figure 2: Filter quality comparison with 8×8 interleaved sampling. As shown by (a), the motorcycle is lit here by a light probe. (b) shows the ground truth results using a brute force approach with no interleaved sampling and no filter at all. (c) provides the result obtained using our spherical harmonics filter. (d) shows the result obtained while filtering the irradiance and using a too aggressive filter while (e) shows how too strict filtering criteria also lead to poor quality while directly filtering the irradiance.

- A discontinuity-aware screen-space *filtering* step to remove Monte-Carlo noise from the final image, and to provide a smooth (though biased) result.

2.1 The Rendering Equation

Like every rendering algorithm, we want to solve the rendering equation [Kajiya 1986], which expresses the reflected radiance L_r at any point $\mathbf{x}$ according to the incident radiance field L_i convolved with the surface BRDF $\rho_{\mathbf{x}}$ at $\mathbf{x}$. If $\Omega_{\mathbf{x}}$ is the hemisphere above $\mathbf{x}$:

$$L_r(\mathbf{x}, \omega_r) \quad = \quad \int_{\Omega_{\mathbf{x}}} L_i(\mathbf{x}, \omega_i)\ \hat{\rho}_{\mathbf{x}}(\omega_r, \omega_i)\ d\omega_i,$$

where $\hat{\rho}_{\mathbf{x}}(\omega_r, \omega_i) = \rho_{\mathbf{x}}(\omega_r, \omega_i)\cos(\theta_i)$ is the *projected* BRDF $\hat{\rho}$.

2.2 Instant Radiosity

Instant Radiosity [Keller 1997] replaces the radiance field by a set of hemispherical *Virtual Point Lights* (VPLs), with the illumination at any point $\mathbf{x}$ given by the radiance received from these VPLs. Representing the whole incoming radiance field via VPLs generalizes both direct and indirect illumination, as well as whether the illumination originated on point or area light sources. Generating the VPLs is simple and cheap, and can be re-done every frame.

Using the same samples (i.e. VPLs) for every pixel, Instant Radiosity is a *dependent sampling* technique (see Figure 3a). As a result, rather than random per-pixel noise Instant Radiosity produces a discretization error that varies smoothly over the image. It also produces highly coherent visibility computations that are amenable to evaluation with either GPUs or coherent ray tracers.

2.3 Interleaved Sampling & Discontinuity Filtering

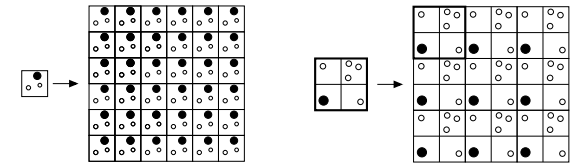
Instead of using exactly the same samples (i.e. VPLs) for the entire image, Interleaved Sampling (IS) [Keller and Heidrich 2001] (see Figure 3) uses a set of $n \times m$ uncorrelated sample sets. By regularly alternating among these different sets from pixel to pixel, every $n \times m^{th}$ pixel will have the same sample set, but the correllation of neighboring pixels is broken up. Since more samples are considered, this reduces the overall discretization error in the image [Keller and Heidrich 2001] but results in some structured noise patterns visible in the final picture (see Figure 3c).

Discontinuity Filtering (DF) [Keller 1998] is a technique that filters a noisy per-pixel signal—such as the irradiance on the surface seen through that pixel—in a continuity-aware way (i.e. it does not filter across geometric discontinuities). Though more general, DF is particularly interesting when combined with IS: if the discontinuity filter is exactly as wide as the interleaving grid ($n \times m$ pixels), continuous regions again consider all $n \times m$ sample sets, effectively removing the structured noise (see Figure 3d). Thus, the combination of IS and DF provides images that are smooth, and whose quality is roughly comparable with traditional instant radiosity images computed with $n \times m$ as many samples.

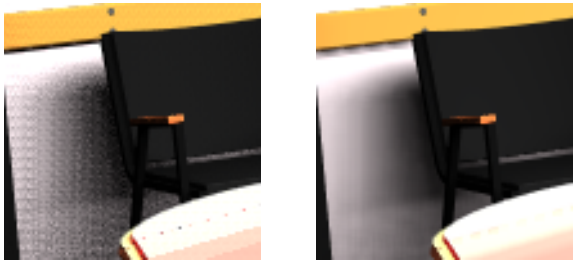
The combination of instant radiosity, interleaved sampling, and a matching discontinuity filter on the irradiance is exactly what instant global illumination is, and it already fulfills most of the goals outlined above. However, by design it has two shortcomings. The need to filter the irradiance field in screen space requires that each pixel separately stores both irradiance and BRDF information (the BRDF must not get filtered), which are then combined after filtering. Since irradiance does not contain directional information, this only makes sense for mostly diffuse BRDFs; glossy surfaces require extra work, and are often not supported.

Second, the discontinuity filter by design should filter only in smooth regions, and neighboring pixels with strongly varying normals should not be filtered. This effectively disables filtering in regions with high-frequency normal detail and consequently, exposing the IS patterns as shown on Figure 2e. Simply filtering the irradiance across varying shading normals does not help, either: since irradiance $E(\mathbf{x}) = \int_{\Omega_{\mathbf{x}}} L_i(\mathbf{x}, \omega_i)\cos\theta_i d\omega_i$ depends on the surface normal, a filter that is independent of the normal produces strong blurring of the illumination (see Figure 2d). Thus, filtering irradiance leaves one with two bad choices: filtering too much and therefore destroying all high frequency details, and using stricter filtering criteria that lead to disturbing IS artifacts.

The core idea of our method is to improve the filtering step by storing and filtering the per-pixel *incoming radiance field* instead of the per-pixel *irradiance*, which allows for filtering across varying surface normals, and for using non-diffuse BRDFs (which require directional information about the incoming lighting).



(a) Standard Dependent Sampling (b) 2×2 Interleaved Sampling



(c) Interleaved Sampling + No Filter (d) Interleaved Sampling + Filter

Figure 3: Interleaved Sampling and Discontinuity Buffering. (a) shows a standard sampling pattern. All pixels use the same sampling pattern. (b) shows a 2×2 interleaved sampling pattern. Non-correlated samples are dispatched over 2×2 pixels. (c) shows the application of interleaved sampling to illumination without using discontinuity filtering and (d) shows the same application using it.

2.4 Incoming Radiance Field with Spherical Harmonics (SH)

To represent the incoming radiance field, we use spherical harmonics basis functions. Spherical harmonics Y_l^m define an orthonormal basis over the spherical directions Ω. A function f defined on Ω may be projected onto the SH basis such that f may be directly expressed as a weighted sum of the SH functions

$$\forall l \in \mathbf{N}, \forall m \in [-l,l], \quad f_l^m = \int_\Omega Y_l^m(\omega) f(\omega) d\omega$$

and

$$f(\omega) = \sum_{l=0}^{+\infty} \sum_{m=-l}^{l} f_l^m Y_l^m(\omega)$$

SH functions allow to compactly represent a function defined on the sphere by a set of coefficients. For this reason, they are commonly used in computer graphics to compress radiance fields and BRDFs.

The orthonormality of the basis also provides an easy way to compute integrals of function products. Indeed, given f and g defined on Ω by their respective SH coefficients f_l^m and g_l^m, we have:

$$\int_\Omega f(\omega) g(\omega) d\omega = \sum_{l=0}^{+\infty} \sum_{m=-l}^{l} f_l^m g_l^m$$

This property naturally leads to decompose Equation 1 into two parts: on the one hand, the projected BRDF $\hat{\rho}$ and the other hand, the incoming radiance term L_i. This decomposition is actually the core of many rendering algorithms using spherical harmonics and the core of our method, which we expose in the next section.

3 Algorithm

Our algorithm proposes a four step approach:

1. Just as in IGI, generate $n \times m$ uncorrelated sets of VPLs representing both direct and indirect lighting;
2. Like IGI, for each pixel inside a $n \times m$ pattern, use the associated VPLs to compute the incident radiance; however, instead of storing irradiance and diffuse material component per pixel, project the incident radiance into an SH basis and store the resulting SH coefficients as well as the full projected BRDF information (possible in an SH representation, too);
3. In the discontinuity filtering stage, filter—inside each $n \times m$ pixel block—the SH coefficients, yielding a smooth incident radiance field with directional information;
4. For each pixel, convolve the projected BRDF and the incident radiance field by computing the dot product between their respective coefficients.

3.1 Decomposing the Rendering Equation

As we use spherical harmonics to represent projected BRDFs and incoming radiance fields, we therefore consider their respective coefficients L_l^m and B_l^m such that:

$$L_l^m = \int_S Y_l^m(\omega) L(\omega) d\omega \tag{1}$$

and

$$B_l^m(\omega_r) = \int_S Y_l^m(\omega) \hat{\rho}(\omega_r, \omega) d\omega \tag{2}$$

As explained in Section 2.4, the rendering equation may be directly expressed as a dot product:

$$L_r(\omega_r) = \sum_{l=0}^{+\infty} \sum_{m=-l}^{l} L_l^m B_l^m(\omega_r).$$

Though our method works with arbitrary-degree SH basis, similar applications have shown quadratic harmonics ($l = 2$) to provide satisfactory results for representing not-too-glossy BRDFs and lighting [Ramamoorthi and Hanrahan 2002], so we use quadratic SHs (i.e. 9 SH coefficients), too.

3.2 Incident Radiance Field Coefficients L_l^m

We first replace the physical light sources and the indirect lighting contributions by $n \times m$ sets of Virtual Point Lights (which may be directional if we are sampling a sky). Using these VPLs, we evaluate the per-pixel spherical harmonics coefficients. This step is basically a Monte-Carlo integration method since we directly compute the coefficients by replacing Equation 1 by the estimator $Y_l^m(\omega)L(\omega)/p(\omega)$ where ω is a random variable with density p. In every other aspect, this step is identical to classical IGI.

3.3 Projected BDRF Coefficients B_l^m

As specified by Equation 2, the BRDF coefficients *depend* on the outgoing direction ω_r. We follow [Nijasure et al. 2003] by choosing radially symmetric BDRFs such as Lambertian or Phong models. This allows a fast and simple implementation. Using the rotational property and the orthogonality of spherical harmonics, we are actually able to re-express the SH coefficients of the BRDF such that they only depend on $\theta_i = \cos^{-1}(\mathbf{C} \cdot \omega_c)$ where $\mathbf{C}$ is a central direction chosen specifically for each BRDF.

Hence, once reparameterized, the BDRF becomes independent of the outgoing direction ω_r and its SH coefficients are therefore also expressed independently of ω_r by:

$$B_l^m = \sqrt{\frac{4\pi}{2l+1}}\, \hat{\rho}_l\, Y_l^m(\omega_c)$$

where $\hat{\rho}_l$ only depends on the BRDF and is given by:

$$\hat{\rho}_l = 2\pi \int_0^{\frac{\pi}{2}} \hat{\rho}(\theta_i) Y_l^0(\omega_i) d\omega_i.$$

3.3.1 Lambertian BRDFs

The central vector for a Lambertian BRDF is the normal vector such that the Lambertian BRDF is given by:

$$\hat{\rho}(\theta) = \frac{k_d}{\pi} \cos\theta$$

where θ is the angle between the incoming vector and the normal and k_d is the diffuse reflection. ρ_l is then given by:

$$\rho_l = \frac{k_d}{pi} \begin{cases} \frac{\sqrt{\pi}}{2} & \text{if } l = 0 \\ \sqrt{\frac{\pi}{3}} & \text{if } l = 1 \\ 2\pi \cdot \sqrt{\frac{2l+1}{4\pi}} \frac{(-1)^{\frac{l}{2}-1} \cdot l!}{(l+2)\cdot(l-1)\cdot 2^l \cdot \left(\frac{l}{2}!\right)^2} & \text{if } l \geq 2 \text{ and } l \text{ even} \\ 0 & \text{otherwise} \end{cases}$$

3.3.2 Phong BRDFs

The central vector is here the reflected vector such that the Phong BRDF is given by:

$$\hat{\rho}(\theta) = k_s \frac{n+1}{2\pi} \cdot \cos\theta^n$$

where θ is the angle between the incoming vector and the reflected vector, k_s is the specular reflection and n is the Phong exponent. ρ_l is then given by:

$$\rho_l = k_s \sqrt{\frac{2l+1}{4\pi}} \begin{cases} \frac{(n+1)(n-1)(n-3)\ldots(n-l+2)}{(n+l+1)(n+l-1)\ldots(n+2)} & \text{if l is odd} \\ \frac{n(n-2)\ldots(n-l+2)}{(n+l+1)(n+l-1)\ldots(n+3)} & \text{if l is even} \end{cases}$$

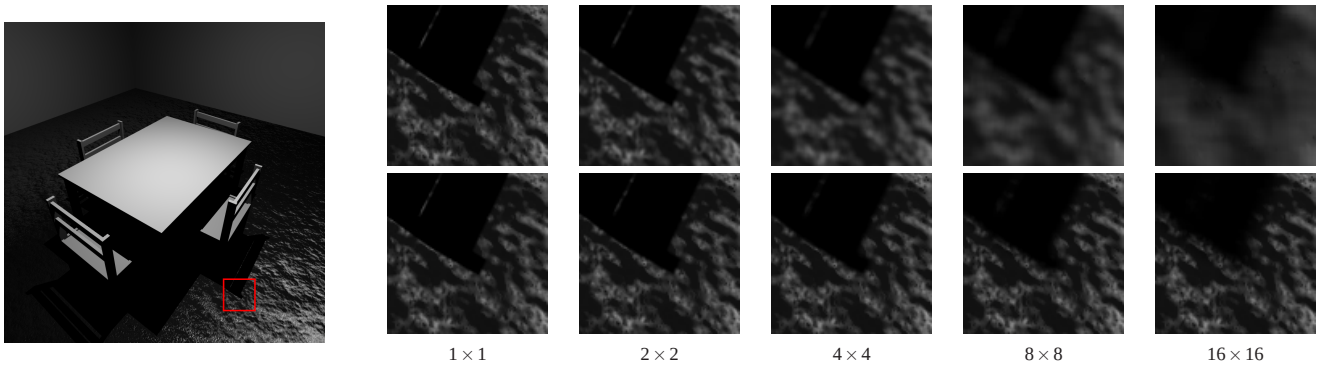


Figure 4: Comparison to classical IGI on a bump-mapped surface, for different filter sizes. Top: Directly filtering irradiance across varying normals leads to severe blurring and loss of detail. Bottom: filtering the incoming radiance field (represented using spherical harmonics) allows to robustly filters across varying normals while preserving geometric detail. As a side result, the technique is more robust with larger filter kernels.

4 Implementation and Results

To evaluate our technique, we have implemented it on top of an existing, prototypical IGI implementation. The modifications required for our method are simple and straightforward, and have been implemented in a single day. A real-time implementation based on either a GPU or coherent ray tracing based implementation is not available yet, but integrating our method into such a system should be rather straightforward.

4.1 Performance Impact

With a real-time implementation not yet being available, we cannot measure our method's actual performance impact. However, the biggest cost factor in any IGI implementation is tracing shadow rays to the VPLs, and this part of the algorithm is totally unaffected by our technique.

Compared to classical IGI, we have to store 9 RGB SH coefficients per pixel rather than a single RGB irradiance value, as well as somewhat more BRDF data than only the diffuse component; other information like normal, mesh ID, depth, etc have to be stored by both variants. The per-pixel BRDF could also be stored in SH terms; for memory reasons however we currently store the BRDF parameters instead, and convert it to SH form only during the filtering stage. In total, having to store this additional information increases the per-pixel storage requirements by about $2-3\times$, depending on actual data formats (eg, whether the BRDF or SH data are stored in floats, bytes, or halves).

Projecting the incident radiance into SH basis is straightforward, and its cost should be insignificant compared to tracing the corresponding rays. Filtering the SH coefficients and performing the dot product between BRDF and incident radiance coefficients is slightly more costly primarily because there are now 9 RGB SH coefficients rather than a single RGB value. For this, too, the additional cost should be less than tracing a single shadow ray.

4.2 Quality Impact for Diffuse Scenes

To validate our algorithm, we first compared it to traditional Instant Global Illumination for scenes which only have diffuse materials and no high frequency geometric details. For such scenes, the main difference is that we project both incident radiance and BRDF into a spherical harmonics basis; so for all intents and purposes for such scenes our technique behaves exactly like classical IGI.

In particular, note that our method also uses spherical harmonics, it does *not* make the same approximations and assumptions that precomputed radiance transfer [Sloan et al. 2002] techniques do: instead of interpolating precomputed information from mesh vertices, we evaluate the incident radiance field at the exact surface locations where the surface samples are shaded, and perform exact visibility queries at those locations. We also perform all lighting computations on-the-fly, and retain no information from frame to frame, so the method can be used in fully dynamic environments.

Being a variant of IGI, there is a slight amount of blurring of illumination due to the screen-space discontinuity filtering (by filtering across neighboring pixels), but this is usually not noticeable, and exactly the same as in classical IGI. Consequently, for diffuse and low-frequency scenes the main difference is that we project both incident radiance and BRDF into a spherical harmonics basis.

As shown by Ramamoorthi and Hanrahan [2001], using 9 SH coefficients captures *illuminated* Lambertian BRDFs with over 99% accuracy for most incoming radiance fields. Therefore, for the smoothly varying diffuse parts of the scenes presented in the article, classical IGI and our approach provide indistinguishable results.

4.3 Scenes with High Frequency Details

Unlike classical IGI, our method can also handle non-diffuse BRDFs as well as normal, bump, or displacement mapped surfaces (or other scenes with high frequency normal variations). To demonstrate this, we have created four sample scenes that are shown in Figure 1. All scenes contain glossy normal mapped surfaces (some procedurally generated, some from textures). As shown in the bottom row, using our technique preserves the high frequency details whereas directly filtering the irradiance leads to heavily blurred information. "Directly filtering irradiance" in this context means that we disabled the normal test in classical IGI; otherwise normal-mapped regions are usually left unfiltered, exhibit strongly disturbing interleaved sampling artifacts.

This is also shown in Figure 2: whereas classical IGI produces either interleaved sampling artifacts (Figure 2e) or blurring (Figure 2d), our method filters across normal variations without blurring, resulting in images that are smooth yet preserve detail (Figure 2c). Figure 4 also provides a more intuitive understanding of the method's results: using a large filter kernel will blur shadow boundaries just like classical IGI will, but unlike classical IGI illumination detail arising from high frequency surfaces are preserved. This effect is particularly apparent during animations, as can be seen in the accompanying video.

Eventually, this should also allow for higher filter widths (i.e. larger n and m) than used in classical IGI. This, in turn, allows for achieving the same image quality with fewer VPLs, and consequently higher performance. This, however, is an incidental side result that we will not investigate any deeper.

4.4 Limitations

The biggest drawback of our method is that having to store more data per pixel increases both bandwidth and compute cost – but not by intolerable amounts. Also, very spiky BRDFs and illumination features (hard shadows) may become slightly blurred after project-

ing to an SH basis.

In addition, though our method solves some of IGIs most notorious problems, some other IGI limitations still remain: Scenes with highly improbable light transport paths and/or lots of light sources may produce "bad" VPL sets that require additional handling. Transparency and specular effects like reflections and refractions lead to each pixel "seeing" multiple surface samples, so filtering a single sample per pixel is not sufficient any more (one possible approach is sketched in [Wald et al. 2002], but can break down in some cases). While IGI usually produces smooth *illumination*, directly visible geometry can still produce visible aliasing. Our method can filter across surfaces with high normal variations, but still cannot filter across "real" discontinuities like silhouettes, so thin geometric objects may still be noisier than larger objects. Finally, IGI cannot efficiently handle some features such as caustics.

5 Conclusion

We have presented a simple technique that allows for using Instant Global Illumination for scenes with glossy and high frequency surfaces while offering the same results for diffuse scenes with almost no extra limitations. In particular, our method introduces no new restrictions or trade-offs, and can be easily used with dynamic scenes, dynamic lighting, direct and indirect illumination, area as well as point, directional, and environment light sources, etc.

Some limitations (see previous Section) still exist, but for real-world applications like video games these are far less of a problem than not being able to handle glossy or normal-mapped surfaces—which our technique is now able to handle. With this, we believe IGI—once augmented with our technique—is a viable candidate to drive video game like applications with full per-frame direct and indirect lighting.

In a next logical step, we plan on implementing our technique on top of a real-time instant global illumination implementation that prototypically already exists. Research-wise, we believe the biggest remaining problem for game-like applications is to efficiently support many light sources, in particular if their importance varies for different pixels.

References

KAJIYA, J. T. 1986. The Rendering Equation. In *Computer Graphics (Proceedings of ACM SIGGRAPH)*, D. C. Evans and R. J. Athay, Eds., vol. 20, 143–150.

KELLER, A., AND HEIDRICH, W. 2001. Interleaved Sampling. *Proceedings of the 12th Eurographics Workshop on Rendering*, 269–276.

KELLER, A. 1997. Instant Radiosity. *Computer Graphics, Proceedings of SIGGRAPH '97*, 49–56.

KELLER, A. 1998. *Quasi-Monte Carlo Methods for Realistic Image Synthesis*. PhD thesis, University of Kaiserslautern.

LAINE, S., SARANSAARI, H., KONTKANEN, J., LEHTINEN, J., AND AILA, T. 2007. Incremental Instant Radiosity for Real-Time Indirect Illumination. In *Proceedings of Eurographics Symposium on Rendering 2007*, 277–286.

NIJASURE, M., PATTANAIK, S., AND GOEL, V. 2003. Interactive Global Illumination in Dynamic Environments Using Commodity Graphics Hardware. In *Proceedings of the 11th Pacific Conference on Computer Graphics and Applications*, 450.

RAMAMOORTHI, R., AND HANRAHAN, P. 2001. An Efficient Representation of Irradiance Environment Maps. *Computer Graphics (Proceedings of ACM SIGGRAPH)*, 497–500.

RAMAMOORTHI, R., AND HANRAHAN, P. 2002. Frequency Space Environment Map Rendering. *Computer Graphics (Proceedings of ACM SIGGRAPH)*, 517–526.

RITSCHEL, T., GROSCH, T., KIM, M. H., SEIDEL, H.-P., DACHSBACHER, C., AND KAUTZ, J. 2008. Imperfect Shadow Maps for Efficient Computation of Indirect Illumination. In *SIGGRAPH Asia'08 Conference Proceedings*, 1–8.

SHISHKOVTSOV, O. 2005. *GPU Gems 2*. Addison Wesley, ch. Deferred Shading in S.T.A.L.K.E.R.

SLOAN, P.-P., KAUTZ, J., AND SNYDER, J. 2002. Precomputed Radiance Transfer for Real-Time Rendering in Dynamic, Low-Frequency Lighting Environments. In *Computer Graphics (Proceedings of ACM SIGGRAPH)*, 527–536.

WALD, I., KOLLIG, T., BENTHIN, C., KELLER, A., AND SLUSALLEK, P. 2002. Interactive Global Illumination using Fast Ray Tracing. *Rendering Techniques*, 15–24. (Proceedings of the 13th Eurographics Workshop on Rendering).