Eurographics Symposium on Rendering (2004)
A. Keller, H. W. Jensen (Editors)

Spherical Harmonic Gradients for Mid-Range Illumination

Thomas Annen[1] Jan Kautz[2] Frédo Durand[2] Hans-Peter Seidel[1]

[1]MPI Informatik
Saarbrücken, Germany

[2]Massachusetts Institute of Technology – CSAIL
Cambridge, USA

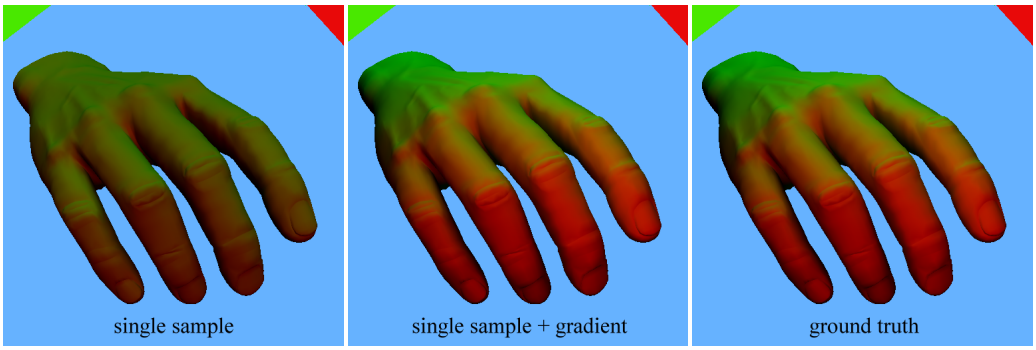


Figure 1: *From left to right. A hand model shaded using only one sample at the center. The middle image shows the same model shaded using one sample and the analytical gradient. The last image shows a reference rendering where the incident radiance field is sampled per-vertex.*

Abstract

Spherical harmonics are often used for compact description of incident radiance in low-frequency but distant lighting environments. For interaction with nearby emitters, computing the incident radiance at the center of an object only is not sufficient. Previous techniques then require expensive sampling of the incident radiance field at many points distributed over the object. Our technique alleviates this costly requirement using a first-order Taylor expansion of the spherical-harmonic lighting coefficients around a point. We propose an interpolation scheme based on these gradients requiring far fewer samples (one is often sufficient). We show that the gradient of the incident-radiance spherical harmonics can be computed for little additional cost compared to the coefficients alone. We introduce a semi-analytical formula to calculate this gradient at run-time and describe how a simple vertex shader can interpolate the shading. The interpolated representation of the incident radiance can be used with any low-frequency light-transfer technique.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Color, Shading, Shadowing and Texture

1. Introduction

In recent years, several methods have been proposed that permit the usage of global incident lighting in real-time rendering [SKS02, NRH03, SHHS03]. These approaches represent the incident radiance in spherical harmonics (SH). They, however, assume distant lighting. As a result, they are incapable of rendering scenes with mid-range lighting effects without visual error (Figure 1).

To alleviate this problem, Sloan et al. [SKS02] sample the incident light field at multiple points over the object. While this was shown to be a possible solution, the high computational cost of multiple sampling leaves room for improvement.

In this paper, we propose to compute a first-order Taylor expansion of the spherical harmonic coefficients around a sampling point. We show how the gradient of the incident radiance (represented in SH) can be computed for little additional cost compared to the coefficients alone. A semi-analytical formula is introduced to calculate this gradient at run-time. The incident radiance can now be extrapolated to different positions around the original sample lo-

cation. In case of multiple samples, the interpolation quality is greatly improved, thus requiring less samples. Extrapolation/interpolation can be easily performed in a vertex shader on the GPU. The extrapolated/interpolated incident radiance can then used together with any radiance transfer technique, e.g. [RH01, SKS02].

2. Previous Work

Our work uses the same framework as the recent *precomputed radiance transfer* technique [SKS02]. This approach permits the illumination of objects with low-frequency incident lighting represented in spherical harmonics [Edm60]. The object can either be diffuse [SKS02] or glossy [KSS02, SHHS03, LK03]. Rendering can be performed in real-time, but requires precomputing the transfer for self-shadowing and other global illumination effects.

Precomputed radiance transfer is limited to distant illumination, unless multiple incident radiance samples are taken and interpolated [SKS02]. We improve on this by computing the gradient of the spherical harmonics coefficients around a sample point. This enables extrapolation of the incident radiance to other points in space, which in turn can be used to improve interpolation of multiple samples.

Our technique is similar in spirit to the *irradiance gradients* for ray-tracing by Ward and Heckbert [WH92]. They propose to compute gradients of the view-independent irradiance at various sample points in order to improve interpolation. This involves a translational gradient (for the change of position) as well as rotational gradient (for the change of relative surface orientation). In contrast to their work, we choose to compute a gradient for the whole sphere of incident radiance (given in SH), independent of any incident surface orientation. Therefore we only need a translational gradient, but not the rotational gradient. However, our gradient is higher-dimensional, as we encode the directional radiance information through the vector of lighting coefficients. Essentially, we trade dependencies on the receiver orientation for a more comprehensive directional treatment of incident radiance. Furthermore, we focus on real-time rendering, whereas their application area was offline rendering.

Irradiance gradients were generalized by Arvo [Arv94], who additionally accounted for occlusions. As Ward and Heckbert, we have to decided to neglect occlusion changes. This is motivated by implementation robustness and simplicity goals, and is justified by the use of low-frequency incident illumination, which is not very susceptible to visibility changes.

3. Review: Shading with Spherical Harmonics

Computing exit radiance at a diffuse surface is usually computed by the following integral:

$$L_{\mathbf{p}} \;=\; \int_{\Omega} I_{\mathbf{p}}(\omega)\cdot V^*_{\mathbf{p}}(\omega)\,d\omega, \qquad (1)$$

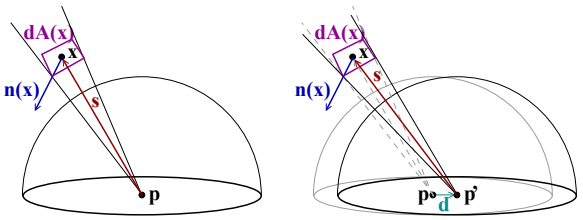


Figure 2: *At a point* $\mathbf{p}$*, we see point* $\mathbf{x}$ *in direction* $\mathbf{s} := \mathbf{x} - \mathbf{p}$*. The point* $\mathbf{x}$ *has a differential solid area* $dA(\mathbf{x})$*. When we move from* $\mathbf{p}$ *to* $\mathbf{p}'$ *along* $\mathbf{d}$*, the direction* $\mathbf{s}$ *changes, as well as the angle between* $\mathbf{s}$ *and* $\mathbf{n}$*.*

$$V^*_{\mathbf{p}}(\omega) \;=\; V_{\mathbf{p}}(\omega)(\mathbf{n}_{\mathbf{p}}\cdot\omega), \qquad (2)$$

where $L_{\mathbf{p}}$ is the exitant radiance at point $\mathbf{p}$, $I_{\mathbf{p}}$ is the incident radiance at $\mathbf{p}$, $V_{\mathbf{p}}$ is the visibility function at $\mathbf{p}$, $V^*_{\mathbf{p}}$ is the cosine-weighted visbility, and $\mathbf{n}_{\mathbf{p}}$ is the normal at $\mathbf{p}$. Integration is performed over all directions ω.

This is an expensive integral, and it needs to be computed at every point $\mathbf{p}$ on an object. Recently, techniques [SKS02, SHHS03] were introduced to speed up the computation of this integral under the following assumptions: Lighting is assumed to be low-frequency, the object is static, and the illumination is distant (i.e., $I_{\mathbf{p}}(\omega) = I(\omega)$ is the same for all $\mathbf{p}$).

If we now project $I(\omega)$ and $V^*_{\mathbf{p}}(\omega)$ into the spherical harmonics (SH) basis y_i, we get two coefficient vectors $\mathbf{I} = (c_0, c_1, \ldots)^T$ and $\mathbf{V}^*_{\mathbf{p}}$, and exitant radiance can then be easily computed with a dot-product [SKS02]:

$$L_{\mathbf{p}} = \mathbf{I}\cdot\mathbf{V}^*_{\mathbf{p}}. \qquad (3)$$

Multiple Samples. For close- and mid-range illumination, the incident radiance does change at every $\mathbf{p}$. Hence, the above distant-illumination assumption is not valid.

To alleviate this problem, Sloan et al. [SKS02] propose to take multiple incident radiance samples at several $\mathbf{p}_j$ and interpolate between the $\mathbf{I}_{\mathbf{p}_j}$. We also propose to use multiple samples in these cases, but increase interpolation quality using our gradient-based technique.

4. Spherical Harmonic Gradients

Consider the configuration visualized in Figure 2. A single sample of incident radiance is taken at point $\mathbf{p}$. We want to estimate the incident radiance $I_{\mathbf{p}'}(\omega)$ at a new point $\mathbf{p}' = \mathbf{p} + \mathbf{d}$, where $\mathbf{d}$ is simply the translation vector. More precisely, we want the coefficient vector $\mathbf{I}_{\mathbf{p}'}$.

We approximate the coefficients $\mathbf{I}_{\mathbf{p}'} = (c'_0, c'_1, \ldots)^T$ at point $\mathbf{p}'$ with a first-order Taylor expansion:

$$c'_i = c_i + (\nabla c_i \cdot \mathbf{d}). \qquad (4)$$

As we will show in the next section, the gradients ∇c_i can be computed efficiently and allows for fast rendering.

Assumptions. Before we derive the gradient formulation, we introduce the assumptions we make. We neglect specular surfaces in the environment and assume that exitant radiance at a 3D point is independent of direction. This assumption is made only for the environment contribution, and the illuminated object can be specular.

Furthermore, we assume that visibility does not change when we move away from $\mathbf{p}$. That is, the set of visible points remains unchanged. Although this assumption almost never holds for real scenes, artifacts are likely to be negligible, since we consider only low-frequency incident illumination. If visibility changes are dramatic and artifacts might occur, multiple samples and gradient-based interpolation can be used to circumvent potential problems (see Section 6).

4.1. Gradient

We need to compute the gradient of the coefficients c_i. Projecting the incident radiance into spherical harmonics is done by integrating it against the SH basis functions:

$$c_i = \int_\Omega y_i(\omega)\ I(\omega)\ d\omega. \tag{5}$$

Similar to the rendering equation, this formula can be written as an integral over the sphere of directions Ω or over visible scene surfaces $\mathcal{S}$. For gradient computation, this choice will influence which terms of the integrand need to be derived and which ones are constant. Both approaches yield similar algebraic complexity, but different numerical integration. In the surface-based case, differential surface elements are assumed fixed and we need to derive the various angular terms (see Figure 2). In contrast, in the angle integration, we assume that angles are fixed but the derivatives of radiance over the scene surfaces needs to be computed. The former approach is simpler because it makes a direct use of the environment-map-sampled representation of the surface elements and because the angular terms are simple to derive, as shown below.

We thus rewrite this integral so that integration is performed over visible surfaces $\mathcal{S}$. We first define the non-unit vector $\mathbf{s} = (\mathbf{x} - \mathbf{p})$, where $\mathbf{x}$ is a point on a visible surface, substituting it into Equation 5:

$$c_i = \int_\mathcal{S} y_i(\frac{\mathbf{s}}{||\mathbf{s}||})\ I(\mathbf{x})\ d\mathbf{s}. \tag{6}$$

The main difference is that the incident radiance now depends on the points $\mathbf{x}$ of the visible surfaces $\mathcal{S}$. This is useful for taking the gradient, since the radiance leaving $\mathbf{x}$ remains the same, even when the sample location $\mathbf{p}$ moves to $\mathbf{p}'$.

The measure $d\mathbf{s}$ can be rewritten as the differential surface area dA at $\mathbf{x}$ weighted by the squared distance to that surface point and by the angle between the surface normal $\mathbf{n}(\mathbf{x})$ and the direction $\mathbf{s}$ towards $\mathbf{p}$ [CW93]:

$$d\mathbf{s} = dA\ \frac{\mathbf{n}(\mathbf{x}) \cdot (\mathbf{s}/||\mathbf{s}||)}{||\mathbf{s}||^2}. \tag{7}$$

We explain how to compute dA in practice in the next section. We then write the gradient of c_i:

$$\nabla c_i = -\int_\mathcal{S} \nabla \left(y_i(\frac{\mathbf{s}}{||\mathbf{s}||}) \frac{\mathbf{n}(\mathbf{x}) \cdot (\mathbf{s}/||\mathbf{s}||)}{||\mathbf{s}||^2} \right)\ I(\mathbf{x})\ dA \tag{8}$$

Note, that we take the gradient in $\mathbf{s}$ instead of $\mathbf{p}$, since it is easier to write. The only difference between the two gradients is the sign (when $\mathbf{p}$ moves along $\mathbf{d}$, then $\mathbf{s}$ moves along $-\mathbf{d}$). Note also, that because we use the integration on surfaces, the incident radiance $I(\mathbf{x})$ does not vary under translation of the incident point.

To simplify notations we define the geometric term

$$g(\mathbf{s}) = \frac{\mathbf{n}(\mathbf{x}) \cdot (\mathbf{s}/||\mathbf{s}||)}{||\mathbf{s}||^2} \tag{9}$$

and Equation 8 becomes

$$\nabla c_i = -\int_\mathcal{S} \nabla \left(y_i(\frac{\mathbf{s}}{||\mathbf{s}||}) \cdot g(\mathbf{s}) \right) I(\mathbf{x})\ dA, \tag{10}$$

which we expand using the derivative of a product to

$$\nabla c_i = -\int_\mathcal{S} \left(\nabla y_i(\frac{\mathbf{s}}{||\mathbf{s}||}) \cdot g(\mathbf{s}) + y_i(\frac{\mathbf{s}}{||\mathbf{s}||}) \cdot \nabla g(\mathbf{s}) \right) I(\mathbf{x})\ dA. \tag{11}$$

The gradients of the geometric term is simply:

$$\nabla g(\mathbf{s}) = \frac{\mathbf{n}(\mathbf{x})}{||\mathbf{s}||^3} - 3\mathbf{s}\ \frac{(\mathbf{n}(\mathbf{x}) \cdot \mathbf{s})}{||\mathbf{s}||^5}. \tag{12}$$

The gradient of the spherical harmonics y_i can be easily derived analytically. It is convenient to use the Cartesian formulation [VMK88], as we compute the translational gradient in Cartesian coordinates. The recursion on Legendre polynomials can also be exploited. We provide C code for these terms online.

Note, that the integrand of Equation 11 is analytical. The actual integration is performed numerically, as some quantities ($I(\mathbf{x})$, the normal $\mathbf{n}(\mathbf{x})$, and dA) come from sampled representations (cube maps in our case).

4.2. Discussion

The gradient formulation in Equation 11 has the additional advantage that one can tabulate the translational gradient of the spherical harmonics for each texel location. By applying the chain-rule and factoring $1/||\mathbf{s}||^2$ outside this term, the integrand can be precomputed for unit directions ω and does not need to be recomputed.

In this paper we only show examples of objects with diffuse surfaces and precomputed radiance transfer. Glossy surfaces [SKS02, SHHS03] as well as non-shadowed objects

[RH01] can be incorporated the same way. Just the final dot-product (Equation 3) is replaced with a different operation.

5. SH Coefficient Extrapolation using the Gradient

Rendering an object using the SH gradient, involves the following steps.

First, we pick a single sample location (usually the centroid of an object) and render a cube map containing exit radiances $I(\mathbf{x})$ of the surrounding objects/emitters. Additionally, we read back the depth-buffer, as we need to know the distance r from $\mathbf{p}$ to $\mathbf{x}$ for each texel. In a separate pass we render the emitters again but with color-coded normals, which we also read back yielding $\mathbf{n}(\mathbf{x})$.

We go over each texel t of the cube maps and analytically evaluate the integrand. The numerical integration over all texels provides the gradients ∇c_i. The term dA can be computed from the read-back data as follows:

$$dA = d\omega_t \frac{r^2}{\mathbf{n}(\mathbf{x}) \cdot \omega}, \tag{13}$$

where $d\omega_t$ is the solid angle of the texel t, ω is the unit direction from $\mathbf{p}$ towards $\mathbf{x}$, and r is the distance from $\mathbf{p}$ to $\mathbf{x}$.

At each vertex of the object, we now compute the estimated coefficients:

$$c_i' = c_i + (\nabla c_i \cdot \mathbf{d}). \tag{14}$$

These coefficients are then used to compute the final exitant radiance using the dot-product between the coefficients c_i' and the cosine-weighted visibility coefficient vector $\mathbf{V}_\mathbf{p}^*$ (see Equation 3).

We currently compute the gradients on the CPU. The new coefficients c_i' as well as the final dot-product are computed in a vertex shader on the GPU.

6. SH Coefficient Interpolation using the Gradient

If light is emitted from nearby or overlapping sources, a single extrapolated sample might not be sufficient due to parallax changes. Such a scenario is sketched in Figure 3. If SH coefficients are computed only at object $\mathbf{O_i}$'s center, then vertex $\mathbf{v}$ would not be shaded properly, because object $\mathbf{O_{i+2}}$ is becomes visible from this $\mathbf{v}$. Shading for objects with larger extend requires interpolation between multiple sample points in order to achieve faithful results [WH92].

In such a case, we choose N sample points p_j over the object (using Sloan et al.'s method [SKS02]). The incident radiance coefficients c_i^j as well as the gradient ∇c_i^j is computed for each sample point p_j.

At each vertex of the object, we compute the estimated coefficients as a weighted sum of all extrapolated samples:

$$c_i' = \sum_j w_j \left(c_i^j + \nabla c_i^j \cdot \mathbf{d}_j \right), \tag{15}$$

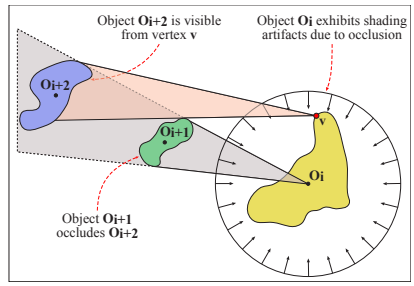


Figure 3: *Due to visibility changes, a previously hidden object becomes visible.*

where w_j is a weight based on the distance between the vertex position $\mathbf{v}$ and the sample point position $\mathbf{p}_j$. We use a simplified version of Ward and Heckbert's formula [WH92], which was also employed by Sloan et al. [SKS02]:

$$w_j = \frac{(1/\|\mathbf{v} - \mathbf{p}_j\|)^b}{\sum_i (1/\|\mathbf{v} - \mathbf{p}_i\|)^b}. \tag{16}$$

If only a few samples are used, then rendering can be done in a shader on the GPU.

Note, that the gradient-based interpolation scheme could also be used to improve interpolation of the irradiance volumes technique of Greger et al. [GSHG98].

7. Results

For all the results, presented in this section we have used a 3 GHz PC, equipped with 2 GB RAM and an NVIDIA Quadro FX 3000 (NV35) graphics board. The viewport size for all renderings was 512×512. The size of a single cube map face, into which we read back the necessary data from the frame buffer, was 64×64.

The strength of the gradient method can be illustrated by moving an object below a small and local area light, see Figure 4. In this example, we compute the original incident lighting sample and the gradient only once at the center of the head in the middle image. We then move the head from left to right underneath the light. Shading is computed with our gradient method based on the single sample. Shading responds correctly to the change of location.

Figure 5 shows a bird model lit by three different colored area light sources. It compares a single sample with the gradient-based technique and ground truth. The same comparison is made in Figure 1. These comparisons highlights the quality improvement due to our method. One can see that our gradient method produces results very similar to the actual ground truth.

In Figure 6, we show three cube maps: cube map (a) is the original sample; cube map (b) shows extrapolated incident

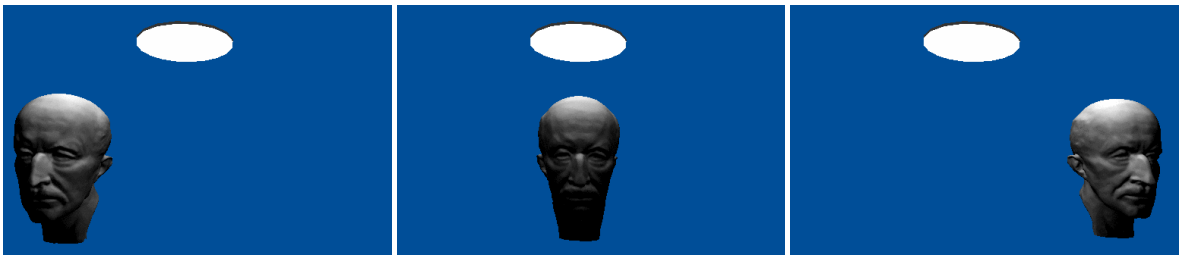


Figure 4: *Max-Planck model moved underneath a local emitter from left to right. The incident lighting is only sampled once at the center of the head (middle image).*

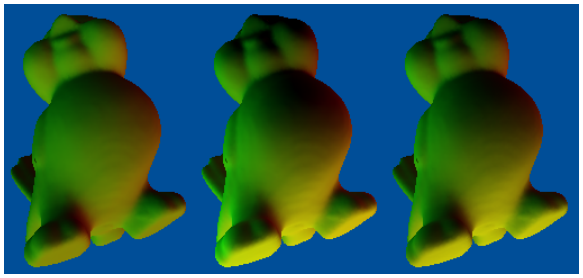


Figure 5: *Bird model rendered with only a single sample (no gradient), single sample with gradient, and ground truth.*

lighting using the gradient method; cube map (c) is ground truth. As can be seen, maps (b) and (c) are very similar.

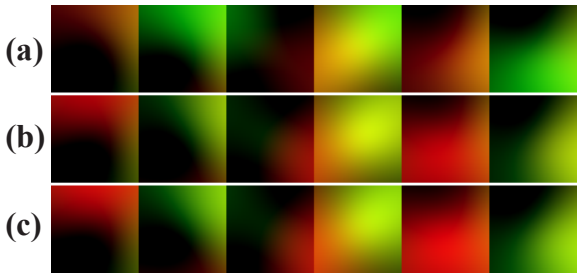


Figure 6: *Reconstruction of incident radiance at a vertex* **v** *of the bird model. (a) original incident radiance sample at center of the bird, (b) reconstructed sample at* **v** *(original sample plus the gradient) (c) reference incident radiance rendered* **v**.

In Table 1, we list the rendering times for the previous technique (no gradient) and our proposed technique (with gradient). Performance is very similar for both methods, but quality is increased by using our method.

In Figure 7, a tooth model is illuminated by partially overlapping emitters. Different renderings compare extrapolation of a single sample against interpolation of eight samples.

Model	#vertices	single sample	with gradient
Planck	25K	23.8	19.7
Bird	32K	8.4	7.9
Tooth	2.5K	40.3	34.5

Table 1: *Rendering times in frames per second for different models and different techniques.*

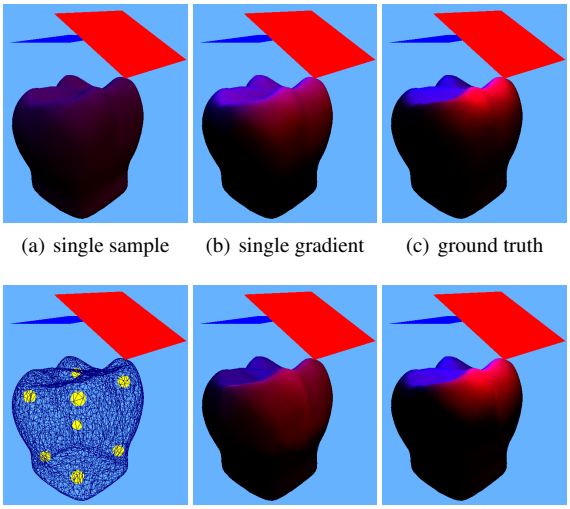
Rendering (b) uses a single sample with a gradient. It does similarly well as eight samples without gradients (see (e)). The only noticeable difference is the top of the tooth. Here visibility changes should make it mainly blueish, but the extrapolated single sample is purple. Interpolation of eight samples with gradients (f) produces virtually the same result as the reference image.

7.1. Discussion

In our experience, a single sample using the gradient is often sufficient. Only for very nearby emitters—where "nearby" means closer than the distance from the sample position to the object's bounding box (empirically determined)—we have found that one needs multiple samples. In these cases the $1/r^2$ falloff is very important and cannot be modeled by our first-order approximation.

In the case of overlapping emitters and occlusion changes, multiple samples are needed as well. We have only experimented with a fixed number of samples for now. For all the cases we have tried, up to 8 samples seem to be sufficient.

Figure 8 depicts the RMS error between ground truth and extrapolated incident radiance from a single sample (tooth example, see Figure 7). The error is shown for the red color channel only. The red error curve includes the blue emitter as a blocker, whereas the green error curve is computed with the blue emitter removed from the scene (no occlusion changes). As expected, the error is higher when occlusion changes occur, but not significantly.



(a) single sample (b) single gradient (c) ground truth

(d) eight samples (e) eight samples (f) eight gradients

Figure 7: *Comparison of a single sample versus multiple samples and gradients versus no gradients: (a) A single sample without gradient, (b) A single sample with gradient, (c) Reference model, (d) Eight sample locations, (e) Interpolation of eight samples without gradient, (f) Interpolation of eight samples using the gradients.*

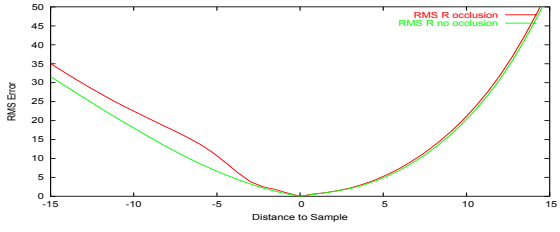


Figure 8: *RMS error of extrapolated radiance for the tooth example. The red curve includes occlusion changes, the green curve does not (blue blocker removed from scene).*

8. Conclusions

We have presented a method for efficient shading of objects that are lit with nearby area sources. We extrapolate/interpolate incident radiance based on a first-order approximation. To this end, we have derived an analytic formula for the gradient, which can be efficiently computed at run-time.

Using this additional information, we achieve results close to a per-vertex reference while the additional cost is small. Simple extrapolation has limitations in terms of accuracy, namely when extrapolation is done at locations far from the initial sample position. Interpolation of multiple samples maintains good image quality in these cases.

We would like to implement a robust and automatic scheme for selecting sample positions in dynamic scenes. New samples should be generated or removed based on the spatial relationship between objects.

Furthermore, we would like to experiment with higher-order Taylor expansions. This should allow even better quality without taking multiple samples of the incident radiance. It would be interesting to find out at which order the computation of the Taylor expansion becomes more expensive than taking multiple samples.

Acknowledgments

We would like to thank Paul Debevec for the light probes and Paul Green for proof-reading the paper. This work has been sponsored in part by an NSF CISE Research Infrastructure Award (EIA-9802220) and the German Academic Exchange Service (DAAD).

References

[Arv94] ARVO J.: The Irradiance Jacobian for Partially Occluded Polyhedral Sources. In *Proceedings of SIGGRAPH 94* (July 1994), Computer Graphics Proceedings, Annual Conference Series, pp. 343–350.

[CW93] COHEN M., WALLACE J.: *Radiosity and Realistic Image Synthesis*. Academic Press Professional, Cambridge, MA, 1993.

[Edm60] EDMONDS A.: *Angular Momentum in Quantum Mechanics*. Princeton University, Princeton, NJ, 1960.

[GSHG98] GREGER G., SHIRLEY P., HUBBARD P., GREENBERG D.: The Irradiance Volume. *IEEE Computer Graphics And Applications 18*, 2 (1998), 32–43.

[KSS02] KAUTZ J., SLOAN P.-P., SNYDER J.: Fast, Arbitrary BRDF Shading for Low-Frequency Lighting Using Spherical Harmonics. In *13th Eurographics Workshop on Rendering* (June 2002), pp. 301–308.

[LK03] LEHTINEN J., KAUTZ J.: Matrix Radiance Transfer. In *Proceedings of the 2003 symposium on Interactive 3D graphics* (2003), pp. 59–64.

[NRH03] NG R., RAMAMOORTHI R., HANRAHAN P.: All-Frequency Shadows Using Non-linear Wavelet Lighting Approximation. *ACM Transactions on Graphics 22*, 3 (2003), 376–381.

[RH01] RAMAMOORTHI R., HANRAHAN P.: An Efficient Representation for Irradiance Environment Maps. In *Proceedings of ACM SIGGRAPH 2001* (August 2001), ACM Press, pp. 497–500.

[SHHS03] SLOAN P.-P., HALL J., HART J., SNYDER J.: Clustered Principal Components for Precomputed Radiance Transfer. *ACM Transactions on Graphics 22*, 3 (2003), 382–391.

[SKS02] SLOAN P.-P., KAUTZ J., SNYDER J.: Precomputed Radiance Transfer for Real-Time Rendering in Dynamic, Low-Frequency Lighting Environments. *ACM Transactions on Graphics 21*, 3 (2002), 527–536.

[VMK88] VARSHALOVICH D., MOKSALEV A., KHERSONSKII V.: *Quantum Theory of Angular Momentum*. World Scientific Publishing Co., Singapore, 1988.

[WH92] WARD G., HECKBERT P.: Irradiance Gradients. In *Third Eurographics Workshop on Rendering* (1992), pp. 85–98.