# Real-time Radiance Caching using Chrominance Compression

K. Vardis G. Papaioannou A. Gkaravelis Dept. of Informatics, Athens University of Economics and Business, Greece



**Figure 1**. Our method retains the directional resolution of the luminance component of a radiance field, while suppressing the chrominance (left). This leads to efficient radiance cache throughput, while preserving details such as indirect shadows (middle). High compression can be exploited in the real-time rendering of large or detailed environments (right).

#### Abstract

This paper introduces the idea of expressing the radiance field in luminance/chrominance values and encoding the directional chrominance in lower detail. Reducing the spherical harmonics coefficients for the chrominance components allows the storage of luminance in higher order spherical harmonics in the same memory budget resulting in finer representation of intensity transitions. We combine the radiance field chrominance compression with an optimized cache population scheme, by generating cache points only at locations, which are guaranteed to contribute to the reconstructed surface irradiance. These computation and storage savings allow the use of higher-order spherical harmonics representation to sufficiently capture and reconstruct the directionality of diffuse irradiance, while maintaining fast and customizable performance. We exploit this radiance representation in a low-cost real-time radiance caching scheme, with support for arbitrary light bounces and view-independent indirect occlusion and showcase the improvements in highly complex and dynamic environments. Furthermore, our general qualitative evaluation indicates benefits for offline rendering application as well.

### 1. Introduction

Real-time radiance caching methods typically store information about the radiance field on a lattice of cache locations within the scene in order to improve efficiency. When spherical harmonics (SH) or other basis functions in the spherical domain are used for representing the function of the radiance field, the storage required for the coefficients of the truncated series to sufficiently approximate the original signal can be quite large. Additionally, if in-scattering due to transport of indirect lighting through participating media is absent, storing radiance for all grid points in the volume results in wasted computations. Calculating and sampling the radiance field in empty space, i.e. in locations that are not near the geometry and do not contribute to the lighting of nearby surfaces, can directly result in wasted computations or even incorrectly sampled radiance.

Inspired by the recent work on chrominance sub-sampling for frame buffer compression [Mavridis and Papaioannou 2012], *Compressed Radiance Caching* (CRC) introduces the conversion of incident radiance to the YCoCg color space before projecting this directional signal to the spherical harmonics basis. The chrominance components are encoded using lower order spherical harmonics than the luminance channel, thus reducing the required storage without significantly compromising the visual quality (see Figure 1). The chrominance components compression minimizes the storage, the bandwidth requirements and the number of texture fetches, as well as improves the texture cache coherency for real-time global illumination calculations.

Furthermore, in contrast to previous volume-based caching methods, we store incident radiance only near those surfaces, where indirect illumination is gathered for the final shading, but still benefit from the data access and interpolation mechanisms of a volumetric representation.

We exploit the storage and texture access savings of the radiance field compression by implementing a very fast and stable diffuse indirect lighting approach that extends the *Radiance Hints* method [Papaioannou 2011]. The improved method supports full view-independent indirect shadows in all indirect inter-reflections and in combination with a geometry-driven cache point selection, achieves real-time performance for arbitrarily complex scenes. More specifically, our contributions are the following:

- Introduction of *directional chrominance sub-sampling* for radiance field compression, reducing the cache storage and access cost (see Section 4). We apply this to real-time diffuse global illumination and demonstrate the resulting performance improvement.
- Optimized positioning of radiance cache points in a volumetric grid, storing the radiance field only near locations where the irradiance is going to be evaluated (see Section 5). For single bounce indirect illumination, this corresponds to

enabling cache points only near points visible to the camera. Due to a special *occupancy dilation* procedure, no view-dependent artifacts occur.

• *View-independent* approximate indirect shadowing for all light transport events, based on a binary geometry volume, which is already constructed for the purpose of the cache point occupancy determination (see Section 6).

While the last contribution essentially improves the specific indirect lighting method (Radiance Hints), both the chrominance compression and the optimized cache point positioning are generic contributions and can be applied to other radiance caching techniques.

# 2. Background

We summarize here previous work directly related to this paper and provide details on the Radiance Hints method, which our algorithm is based on. A more general overview of real-time global illumination methods can be found in existing surveys such as the one by Ritschel et al. [2012].

## 2.1. Related Work

**Real-time GI**. With interactive applications in mind, the prevailing approaches for approximating diffuse global illumination are based either on the concept of *instant radiosity* (IR) or on some type of *radiance caching*. Instant radiosity [Keller 1997] methods address inter-reflections by tracing photons from the light sources and placing virtual point lights (VPLs) representing the outgoing flux or radiosity at the intersected geometry locations, thus, replacing the gathering irradiance integral operation by direct lighting from these VPLs.

A typical GPU-friendly example of IR is Reflective Shadow Maps (RSMs) [Dachsbacher and Stamminger 2005]. RSMs exploit the rasterization procedure for the generation of shadow maps to create the VPLs, without explicitly tracing them in the scene. RSMs are implemented as multi-channel G-buffers with the depth, reflected flux, normal vector and position of each geometry sample registered in it. In the original method and most of its variants that followed, when shading the geometry, RSM samples are used as VPLs, but their contribution is gathered without taking scene occlusion into account. Imperfect Shadow Maps by Ritchel et. al. [2008] handle indirect occlusion for complex scenes by rendering a point-based shadow map for each VPL drawn from an RSM.

**Caching techniques**. Storing irradiance values in space in the form of a radial function was introduced by Greger et al. [1998] in order to compute diffuse global illumination on semi-dynamic environments. The key idea is that instead of storing irradiance values at surfaces locations, irradiance was stored in a bilevel volumetric grid, where each grid location contained directional irradiance values in the form of a radial function, as an approximation of the irradiance of an oriented surface at that location in the environment. Krivanek et al. [2008] extended Irradiance Caching to Radiance Caching to compute the indirect glossy and diffuse terms of low frequency BRDF's using hemispherical harmonics.

The radiance field can be cached in any location in space and Nijasure et al. [2005] store the radiance field using spherical harmonics at the vertices of a 3D grid. For each bounce, surface radiance is estimated by rendering cube maps at the center of each voxel. This method supports multiple bounces and indirect occlusion, but requires a large number of draw calls for the generation of cube maps. Kaplanyan et al. [2011] use a grid to iteratively propagate light from cell to cell in the entire volume, handling occlusion using a rough volumetric representation of the occupancy and surface normal direction by a view-dependent injection of points into a geometry volume buffer. The caching is performed at multiple cascades to handle large environments and update operations are amortized across multiple frames.

Crassin et al. [2011] proposed an approximate volume-based cone tracing for 1bounce indirect diffuse and specular illumination, where direct lighting is injected in a hierarchical volume and gathered at the shaded points by cone marching, exploiting a hierarchical prefiltering of the illumination on the volume for fast evaluation of the cone integral. Their method is generic and capable of detailed results at interactive rates, provided that a small part of the volume is dynamically updated.

**Compression in the YCoCg space**. The YCoCg transform [Malvar and Sullivan 2003] was introduced in the H.264 video compression format and decomposes an image to luminance (Y) and orange/green chrominance components (Co, Cg). It has been used since in diverse graphics applications such as texture compression (YCoCg-DXT5) and framebuffer compression [Mavridis and Papaioannou 2012]. Its effectiveness stems from the fact that the human visual system is less sensitive to spatial changes of colour than luminance, which can be exploited to sub-sample and store colour in a smaller resolution. We show later in the text how we exploit the above transformation to compress the radiance signal in the spherical domain. In fact, any other luminance and color offset transformation can be applied as well, but we chose YCoCg due to its superior decorrelation properties.

## 2.2. Radiance Hints

The Radiance Hints method (RH) [Papaioannou 2011], which this paper extends, combines the grid-based radiance caching of Nijasure et al. with the use of RSM sampling to generate the radiance field at each cache location, thus dispensing with the cubemap rendering.

A reflective shadow map is built for each light source that contributes to indirect

lighting. Incoming radiance from RSM samples is measured and an approximate radiance field is encoded as spherical harmonics coefficients and stored at each cache point laid out on a 3D grid. The process is repeated for all contributing lights and the respective radiance field coefficients are accumulated.

First-bounce indirect shadowing is supported optionally using a view-dependent probabilistic attenuation scheme. In the case of multiple bounces, a second lattice is created for interleaved radiance field update and the new radiance field is estimated by direct energy exchange among the cache points. Secondary bounce visibility is statistically approximated.

**RSM sampling**. The radiance field for the first bounce of indirect illumination is estimated by sampling the information stored in the RSM G-buffer of the light sources. Each cell in the shadow map represents a small area light source or a *pixel light* that illuminates the scene as if an actual light source was placed at that particular location. The radiance function (radiance field)  $L(P, \omega)$  at cache point *P* can be represented using the spherical harmonics basis functions  $Y_i^m(\hat{\omega})$  of degree *l* and order *m* as:

$$L(P,\hat{\omega}) \approx \sum_{l=0}^{n} \sum_{m=-l}^{l} \tilde{\lambda}_{l}^{m}(P,\hat{\omega}) Y_{l}^{m}(\hat{\omega})$$
(1)

$$\tilde{\lambda}_{l}^{m}(P,\hat{\omega}) = \frac{n_{\text{texels}}}{\pi \cdot n_{\text{s}}} \sum_{k=1}^{n_{\text{s}}} \frac{\Phi_{k} \cdot V(X_{k}, P_{k}) \cdot Y_{l}^{m}(\hat{\omega}_{k}) \cdot (-\hat{\omega}_{k} \cdot \hat{n}_{k})}{|X_{k} - P_{k}|^{2}}$$
<sup>(2)</sup>

where  $n_{\text{texels}}$  is the total number of texels in the RSM,  $n_s$  is the number of samples drawn form the RSM,  $X_k$ ,  $\hat{n}_k$  and  $\Phi_k$  is the position, surface normal and stored flux at the *k*-th sample, respectively and  $\hat{\omega}_k$  is the direction  $P_k \rightarrow X_k$ .  $V(X_k, P_k)$  is the visibility between  $X_k$  and  $P_k$ .

The parametric-space positions of the RSM samples are chosen based on a precomputed rectangle of low-discrepancy Halton sequence. The total number of samples  $n_s$  can either be split uniformly among the light sources or allocated based on any importance sampling scheme. For each RSM sample  $X_k$  drawn, a corresponding radiance "reception" position  $P_k$  at an arbitrary location inside the cache cell is chosen and incoming radiance is estimated along the path  $X_k \rightarrow P_k$ .

Typically, 50-200 RSM samples are used. The resulting coefficients  $\tilde{\lambda}_l^m(P,\hat{\omega})$  are stored in in the RH cell render target channels and this process is repeated for all cache volume cells.

**Secondary Inter-reflections**. For each cache point  $P_c$  to be updated,  $n_{sec}$  sampling directions  $\hat{\omega}_s$  are uniformly chosen around  $P_c$  and the contribution of an arbitrarily chosen point  $P_s$  in this direction is sampled. The radiance field is incrementally updated by accumulating the gathered radiance spherical harmonics coefficients at  $P_c$ .

Let  $L(P,\hat{\omega})^{(i)}$  be the *i*-th light bounce, where  $L(P,\hat{\omega})^{(0)} = L(P,\hat{\omega})$  from Eq 1. Using a uniform spherical distribution of *n* sampling directions,  $L(P,\hat{\omega})^{(i)}$  can also be approximated by a truncated spherical harmonics series, in the spirit of Eq 1, as:

$$L(P,\hat{\omega})^{(i)} \approx L(P,\hat{\omega})^{(i-1)} + \sum_{l=0}^{n} \sum_{m=-l}^{l} \tilde{\lambda}_{l}^{m}(P,\hat{\omega})^{(i)} Y_{l}^{m}(\hat{\omega})$$
(3)  
$$\tilde{\lambda}_{l}^{m}(P,\hat{\omega})^{(i)} = \frac{4\pi}{n} \sum_{k=1}^{n} L_{i}(P,\hat{\omega}_{k}) Y_{l}^{m}(\hat{\omega}_{k}) = \frac{4\pi}{n} \sum_{k=1}^{n} L_{0}(X_{k},-\hat{\omega}_{k}) Y_{l}^{m}(\hat{\omega}_{k}) =$$
(4)  
$$\frac{4\rho_{ave}}{n} \sum_{k=1}^{n} E(X_{k},-\hat{\omega}_{k}) Y_{l}^{m}(\hat{\omega}_{k})$$

where  $\hat{\omega}_k = P \rightarrow X_k$ .  $E(X_k, -\hat{\omega}_k)$  is the estimated irradiance on a virtual reflective surface positioned at the sampled location  $X_k$ , which is oriented towards P, i.e.  $-\hat{\omega}_k$ . Note that in the above formulation contributing points are not necessarily the nearest ones to P, nor is occlusion taken into account. Having expressed the radiance field as spherical harmonics,  $E(X, \hat{\omega})$  can be efficiently reconstructed using a closed form solution as proposed Ramamoorthi et al. [2001]. To compensate for the unknown albedo at  $X_k$ , a fixed average scene albedo  $\rho_{ave}$  is used for the secondary inter-reflections.

**Irradiance reconstruction**. The surface irradiance at the visible fragments is reconstructed as follows: A number of samples are selected on a normal-aligned rotating hemisphere above the shaded point. The fragment's final irradiance value is then estimated by the dot product between the interpolated radiance SH coefficients and the SH coefficients of the oriented hemisphere.

# 3. Method Overview

In CRC, a uniform volume grid for maintaining the radiance cache data is created, similar to the Radiance Hints algorithm, with a pre-defined maximum resolution r corresponding to the longest side of the radiance field bounding box. The CRC volume may occupy the entire scene, a part of it or be centered at the user and translated in voxel-sized increments. To support more than one bounce a second volume is created, for interleaved radiance field update.

The steps required for determining the position of the occupied cache cells in the current frame, the population of the cache, the compressed encoding of the radiance field in them, the secondary inter-reflections and the final irradiance reconstruction are outlined in the following paragraphs and summarized in Figure 2.



**Figure 2**. The radiance field generation and caching pipeline. The rectangles correspond to separate shader passes.

**Cache points determination**. The cache point occupancy information is generated per frame based on a 2D binary voxelization of the scene at a higher resolution (quadruple resolution in each dimension), the *geometry* volume, which is also maintained and used for visibility tests. Then, the *occupancy volume* is generated through a special mipmapping procedure and is also stored as a 2D bit-mask texture. For single-bounce indirect illumination, only cells visible to the camera need to be enabled, so a *depth occupancy volume* is created instead. This can reduce significantly the number of occupied cache points, without any degradation or view-dependent behaviour. In a single pass, the appropriate occupancy volume is *dilated* to avoid interpolation errors during irradiance reconstruction. The process is described in more detail in Section 5.

**First-bounce radiance field estimation and compression**. The radiance field of the first light bounce is estimated similarly to the Radiance Hints method, but the computations are restricted to the occupied cache cells rather than the entire volume. Also, indirect visibility is estimated by ray-marching through the geometry volume,

rather than using the depth buffer, thus dispensing with any view dependencies. The first-bounce radiance caching requires one pass and one drawing call for the entire volume per light source. The CRC volume cell fragments are generated with layered rendering and geometry instancing of a volume-sweeping quad. For each light source, the occupancy of all radiance cache volume cells is checked and for occupied cells only, the RSM is sampled and each sample is checked for visibility. The incoming linear RGB radiance from visible samples is transformed to the YCoCg space and compressed so that the luminance is always projected to a 3rd order SH basis and the chrominance values are encoded in 1st, 2nd or 3rd order spherical harmonics, depending on the quality setting of the method. The resulting coefficients are finally accumulated in the CRC cell c data. See Section 4 for more details the compression method.

Secondary diffuse inter-reflections. For the energy exchange in secondary light bounces, two copies of the CRC volume are maintained. In contrast to the Radiance Hints method, where radiance field was evaluated in void space, in CRC we only consider energy exchange among mutually visible, occupied cells. For each occupied cache cell c,  $n_{sec}$  rays are marched around the location of c on the geometry volume. Once an intersection is found, the radiance from the hit cache point in the input CRC volume is sampled and accumulated along with the radiance field of the previous bounce in the output volume. For subsequent bounces, the roles of the input and output cache volumes are interchanged and the process is repeated.

**Temporal blending**. For dynamic scenes, where the light sources and the geometry can change arbitrarily, flickering of the reconstructed irradiance may occur, which is a defect common to both radiance caching and instant radiosity approaches. To remedy this, a third volume is used, which, similar to the method by Kaplanyan et al. [2011], contains the exponential moving average of the SH coefficients between the previous frame and the current one.

**Irradiance reconstruction and decompression**. The surface irradiance at the visible fragments is reconstructed similarly to the Radiance Hints method. However, since the radiance field is stored in YCoCg space, a final step is required to convert the irradiance back to the linear RGB space. To avoid interference of cache locations behind the shaded point for thin structures, all samples are shifted by half a voxel in the direction of the surface normal. This, along with the requirement that interpolated radiance coefficients must not be affected by unoccupied cells, are the reasons for the dilation of the occupancy volume by one voxel.

#### 4. Radiance Field Compression

An RGB color value can be decomposed into a luminance component and two chrominance (orange and green) offsets using the YCoCg transform:

$$Y = 0.25 \cdot R + 0.5 \cdot G + 0.25 \cdot B$$
  

$$Co = 0.5 \cdot R - 0.5 \cdot B$$
  

$$Cg = -0.25 \cdot R + 0.5 \cdot G - 0.25 \cdot B$$
(5)

The lower sensitivity of the human visual system to tonal variations as compared to the luminance gradients is exploited in many compression algorithms for spatiallyvarying color data.

We claim that for typical environments spatial variations of luminance are dominant over chrominance ones. We exploit the fact that the projection of spatial light variations (RSM samples) on the unit sphere remain coherent. This means that chrominance at cache points can be sufficiently approximated by a truncated series of spherical harmonics or other spherical domain basis functions of *lower order than the luminance*. As demonstrated by our diverse examples shown in all figures, this assumption holds for most typical environments and lighting conditions (see discussion about quality in Section 7).

To this end, in CRC, before accumulating the incoming radiance at a cache point, the radiometric RGB values are converted to the YCoCg space. Then, the luminance and chrominance components are projected to the spherical harmonics basis using different orders. Due to the linear form of the transformation, all radiance accumulation and interpolation calculations still hold in the YCoCg space. During the irradiance reconstruction step, the irradiance is converted back to to linear RGB values and used for any shading calculations.

Converting the radiance values to YCoCg space and encoding the color offset components in a lower order can drop the number of coefficients down to 17 or even 11 for order l = 2, with negligible or acceptable visual quality loss, respectively. We project the luminance in the 3rd order spherical harmonics basis and the two chrominance offsets in either 1st, 2nd or 3rd order spherical harmonics, depending on a quality setting (see Figure 3). The reduction of SH coefficients directly affects the number of volume textures to store them. Compressing the chrominance to spherical harmonics with l = 1 and l = 0 reduces the number of volume textures (rendering targets) to 5 and 3 respectively and has a direct performance benefit.

Specular events require much higher order spherical or hemi-spherical harmonics to be approximated in a sufficient manner. Therefore, we were not able to accommodate the required coefficients in the graphics pipeline implementation of our radiance caching scheme.

A handy consequence of using radiance expressed in the YCoCg space is the direct artistic control of GI "presence" (Y) and color bleeding (CoCg) by linear scaling





**Figure 3**. The effect of radiance field directional compression in our real-time diffuse-only GI method; luminance transitions are maintained, while suppressing the chrominance detail.

of the respective coefficients. Presence and color bleeding can be adjusted either in the reconstruction stage (scale factor per polygon or global) or per cache point (via volume- / area-based factor selection).

# 5. Geometry and Cache Occupancy Data

Our approach stores the radiance field only near locations, where the irradiance is going to be evaluated, by keeping track of occupancy information for each cell. This process can reduce significantly the number of occupied cache points and improve the rendering times, especially in open environments. For single bounce indirect illumination, the cache points are further reduced to only those that are visible to the camera. Also, having this occupancy information available, allows us to reuse it for view-independent indirect visibility tests for an arbitrary number of bounces, highly improving the visual quality of the result.

Initially, the scene is voxelized in a higher resolution volume, the *geometry* volume. This resolution is four-times the CRC resolution. In a second step, the geometry volume is downsampled to create the *occupancy* volume, which determines the *active* cache points, using the same resolution as the CRC grid. To avoid any interpolation



**Figure 4**. Radiance field compression. An extreme case of a red and green spotlight illuminating the scene in opposite directions. From left to right, top to bottom: linear RGB space (reference), YCoCg space with no SH order reduction, 2nd and 1st order (constant) SH chrominance. The subscripts denote the SH order.

artifacts in the irradiance reconstruction step due to incorrect sampling of unoccupied cells, the set of occupied cells is further dilated by one voxel.

**Geometry Volume**. The geometry volume is generated by a three-way variation of the binary voxelization proposed by Eisemann et al. [2006]; the geometry is orthographically rasterized in a single pass along the three world-space axes of the CRC volume bounding box and each view is encoded separately as a bit-plane image layer. A geometry shader selectively routes each polygon to the dominant rasterization axis layer according to its normal vector. The partial and potentially sparsely-sampled results are subsequently merged to form a single complete volume.

Our current implementation uses a single 128bit buffer, allowing a maximum resolution of 128<sup>3</sup> voxels. Higher resolutions are of course achievable with a multiple rendering target binary voxelization.

**Volume-based Occupancy**. The occupancy volume, i.e. the mask of occupied CRC cells, is bit-encoded in a two-dimensional texture, similar to the geometry volume. It is simply a down-scaled version of the geometry volume and is produced by generating 2 additional mipmap levels of the latter, using the maximum (OR) operator in



**Figure 5**. Depth Occupancy volume behaviour on the canyon scene. As the user moves from p1 towards p5, a gradually smaller portion of the scene is visible. This reduces the number of occupied voxels, resulting in a smaller occupancy ratio and computation time.

a low-cost additional pass. Since the maximum resolution of the geometry texture is 128 in the current implementation, the maximum CRC volume resolution is restricted to 32.

During CRC occupancy checks in the radiance caching and secondary bounce steps, the occupancy volume is considered *dilated* by one voxel. For this reason, after the initial occupancy volume is generated, a fast dilation pass also marks a cell as occupied, if any of its 26 neighbours is also occupied. Since the volume is encoded as a binary mask image, only 9 coherent texture fetches per cell are required in total.

**Depth-based Occupancy**. When only one bounce of indirect illumination is estimated, there is no need to cache the radiance field near invisible geometry, since this is never going to be exploited in any irradiance reconstruction calculation. Therefore, we can further optimize the occupancy of the cache points by marking as occupied only those CRC cells that are visible to the camera.

To achieve this, instead of extracting the occupancy from the geometry volume, we inject the camera depth buffer fragments in the occupancy volume using a fixed horizontal and vertical stride *s*. This separate bit mask texture, the *depth occupancy volume*, is used instead of the full occupancy volume to determine which cache points must be evaluated.

The occupied cache points in the depth occupancy volume can be significantly fewer than those of the geometry-based one (typical cache point reduction up to 90%), resulting in a sizeable speed-up of the radiance cache population.

An example of this is illustrated in Figure 5. The depth occupancy is tracked along a certain path as the user navigates from a high-coverage part of the scene (bottom left - similar view to the canyon scene in Figure 7), marked as p1, towards the robot model at the top right, marked as p5. As the view closes in on p5, the number of occupied



Figure 6. Illustration of the indirect occlusion in 1- and 2 bounce indirect illumination.

cells is reduced, with a respective drop of the cache generation time. In the current scene, the caching requires 0.54 ms at the start of the path with 40% occupancy and it gradually drops to 0.37 ms for a 5% occupancy ratio. It should be noted that 0.3 ms is a lower limit for this step in our implementation due to the rasterization overhead and the required state changes.

Note that using the camera-based injected points introduces no view-dependent artifacts, since the state of marginally invisible voxels or even voxels missed by the injection is also flagged as "occupied" due to the occupancy dilation.

#### 6. Visible Sample Determination

In the RSM sampling pass (first light bounce), visibility  $V(X_k, P_k)$  is determined between the RSM sample  $X_k$  and the CRC cell sample  $P_k$  by ray marching in the geometry volume along the segment  $P_k \rightarrow X_k$ . We offset the start and end position to avoid incorrect self-occlusion checks, and sample the segment sparsely, using a finite number of jittered locations.

In contrast to the Radiance Hints method, for the secondary light bounces, ray marching from the gathering location is used to determine the nearest visible point. Since the contribution of secondary light bounces is less perceptually significant compared to the first bounce, a smaller number of per ray samples is devoted to them. The first bounce uses 8-12 ray samples, while for the secondary bounces, 4-6 sam-

ples suffice. A downside of using sparse ray marching is that intersections may be missed. However, the impact of occasionally missed secondary occlusion in typical environments is hardly objectionable and does not justify the computational penalty introduced by increasing the (fixed) number of samples or dynamically determining the samples per ray.

Since diffuse indirect illumination is mainly of low frequency and accurate visibility calculations are not necessary [Yu et al. 2009], the granularity offered by the geometry volume is usually sufficient (see Figure 6). For (near-field) contact shadows, we employ ambient occlusion in some of the example scenes.

We also experimented with hierarchical empty space skipping using the already available geometry volume mipmaps. However, since ray marching always begins at an occupied high-level cell, there is no safe way to make the first leap without losing intersections. Furthermore, any hierarchical traversal would lead to unbalanced thread execution.

### 7. Results and Discussion

**Test cases**. Our method was tested on an Nvidia GTX 670 GPU on various scenes of varying geometric complexity, volume coverage and dimensions. Apart from example scenes with full CRC volume coverage, we have also performed testing for expansive environments, where building an all-encompassing CRC volume is impractical. Instead, GI is restricted to an axis-aligned moving volume centered at the user. The extents of the volume are updated in voxel-sized increments as the user explores the environment. At the extents of the boundaries of the volume, indirect lighting is blended with a constant ambient color. Alternatively, a cascaded approach can be utilized.

**Quality**. We evaluate the quality of the compressed-chrominance radiance field by both visual inspection and two metrics: the RMS error in the CIE Lab space and the PerceptualDiff metric [Yee and Newman 2004] with respect to the uncompressed linear space RGB values. PerceptualDiff measures the percentage of different pixels from the reference image. Both measurements are shown in the "RMSE" and the "pdiff" rows in Table 1.

Order 2 spherical harmonics for the CoCg channel works remarkably well even in extreme lighting conditions, such as the one presented in Figure 4. Keeping only the constant term for the CoCg (l = 0) results in no noticeable color shifting of the GI in the majority of the scenes and lighting conditions (see Figure 7), while the rendering times are decreased by 28-44%.

Corner cases, where the tonal variations in the original signal are of higher frequency, may arise. In these situations, chrominance cannot be efficiently reconstructed especially in the Y3Co1Cg1 compression, where the chromicance compo-



**Figure 7**. Screenshots on several scenes. Left: (1 bounce) Canyon, Ruins, Barn. Right: (2 bounces) Factory, Level1, Temples (moving volume). The radiance field is compressed using the  $Y_3Co_1Cg_1$  setting for all scenes.

nents are represented by a constant color. These are manifested either as desaturation of the light field at the problematic cache cells or the predominance of the strongest chrominance value of the incident radiance. The former is shown in Figure 4, where two lights sources illuminating the scene in opposite directions. Chrominance is neutralized at cache points in the middle of the room as they gather incident radiance of a different color from opposite directions. The color predominance can be observed in Figure 8. In the bottom row, where the light field consists of orthogonal hues from different directions, notice the prevailing orange tint on the nearest columns and arches (indicated by the red arrows) and the color desaturation on the ceiling in the Y3Co1Cg1 inset. Conversely, in the top row, where the reflected light from the floor is consistent with the rest of the environment, no visible change can be detected. A similar case can be observed on the right inset of Figure 4 along the protrusions of the side walls where the directionality of the green and red indirect diffuse color is lost. A higher volume resolution would mitigate the issue, but not remedy it.

For the same memory (and bandwidth) budget, we favor an unbalanced luminance/chrominance representation (Y3Co1Cg1 - 11 coefficients) over a balanced one (Y2Co2Cg2 or R2G2B2 - 12 coefficients). In most scenes, such as the one in Figure 9, the error introduced by the 2nd order spherical harmonics in the luminance transitions exceeds the over/under saturation caused by the constant chrominance term of



Vol. 3, No. 4, 2014 http://jcgt.org



**Figure 8**. Perceptual evaluation of chrominance directionality in variable GI conditions. Percentages indicate the RMS error compared to the uncompressed lightfield. Light is reflected off the floor (top: grey, bottom: sienna) and enters the building as well, reflected off grass and stone. The red and orange arrows denote the observed orange color predominance due to radiance averaging and color desaturation respectively.



**Figure 9**. Comparison of similar budget radiance field encodings indicates more severe errors when luminance and chrominance are equally but inadequately represented.

the Y3Co1Cg1 encoding. Furthermore, the desaturation effect is not usually objectionable or even noticeable, while the over-saturation can be easily treated by scaling down the CoCg coefficients during irradiance reconstruction. Most importantly, on the other hand, problems in excessive smoothing of illumination or leaking caused by



**Figure 10**. Cumulative GI timings with 2 light bounces for the hillside scene (see teaser) with respect to the CRC volume size. 100 RSM samples were used for one light source.

the 2nd order luminance SH representation cannot be remedied.

**Performance**. Cumulative timings of the pipeline stages are presented in Figure 10 for the "hillside" scene of Figure 1 (right), a moderately populated outdoor environment. The maximum coverage of the scene by cache points is also shown above the timings.

The factors that influence the performance of all stages except from the irradiance reconstruction, are the total number of RSM samples, the CRC volume resolution (which consequently affects the number of occupied cells) and the number of secondary inter-reflections. The irradiance reconstruction is only affected by the framebuffer size. All measurements regarding this stage are per MPixel and the rendering time scales linearly with the GI buffer size.

The cost of the occupancy and blending stages depends on the total number of voxels, but is significantly lower than that of the other stages (approx. 0.1 ms each). The mipmap generation for the Occupancy Volume and volume injection for the Depth Occupancy volume have a very small cost. The first depends on the CRC resolution, but is relatively constant (~0.11 ms) and the second depends on the size of the framebuffer (usually between 0.47-0.85 ms for stride s = 10 and a frame buffer of 1MPixel).

It is important to note that the performance benefit of the occupancy optimization has a non-linear relationship to the actual occupancy. This is an expected behavior, since during the caching stage, the unoccupied voxels are skipped using a simple discard statement. If one of the voxels that belong to the same warp is flagged as occupied, the performance increase will be zero. As a result, the occupancy optimization does not perform as well for densely occupied volumes. In our tests, the highest performance gain was 100% in the caching stage. This was reported for occupancy ratios lower than 30%. On the other hand, only a 2 - 10% gain was measured for occupancy ratios higher that 50%.

The performance of our method for some of the scenes of Figure 7 is listed in Table 1. The voxelization times are given separately, since alternative methods can be used instead. The unoptimized GI time corresponds to no occupancy volume gen-

Scene (indirect bounces)	Canyon (1)	Barn (1)	Level1 (2)	Temples (2)	
Polygon / Voxelization (ms)	412k / 1.44	230k / 0.87	50k / 0.26	1.56M / 4.96	
Res / Lights / RSM Samples	16 / 1 / 100	24 / 2 / 200	24 / 2 / 150	32/2/100	
Volume Generation (ms)	0.73	0.47	0.12	0.12	
Occupancy Ratio	24.13%	42.72%	49.65%	72.67%	
Unoptimized GI time (ms)					
(occupancy & compression	2.24	3.3	4.33	2.88	
disabled)					
Occupancy Enabled (ms)	1.65	3.12	4.25	2.65	
$Y_3Co_2Cg_2$ compressed (ms)	1.52	2.62	3.61	2.10	
$Y_3Co_1Cg_1$ compressed (ms)	1.42	2.61	3.10	1.97	
Total speed improvement	57 750	26 1101	20 6901	46 1007	
(over unoptimized time)	51.15%	20.44%	39.00%	40.19%	
RMSE	0 4507	0.2507	0 4507	0.5407	
$(Y_3Co_3Cg_3 \text{ vs } Y_3Co_1Cg_1)$	0.45%	0.25%	0.45%	0.54%	
pdiff	0.030/-	0.220%	0.07%	0.80%	
$(Y_3Co_3Cg_3 \text{ vs } Y_3Co_1Cg_1)$	0.05%	0.52%	0.07%	0.80%	

**Table 1.** Performance results on some scenes of Figure 7. Global illumination framebuffer size is 1MP and the timings are measured on an Nvidia GTX 670. The scenes with one indirect bounce use the Depth Occupancy Volume, so the occupancy ratio (occupied voxels / total voxels) represents the occupied voxels from the current view instead of the whole volume. Note: The unoptimized time represents the performance of the original RH method under the current implementation

eration (all cells are considered occupied by default) and no radiance field compression. It essentially represents the performance of the original RH method under the current implementation. The other two timings show the rendering time as the occupancy optimization and radiance field compression are enabled and stacked. The Occupancy volume used for 1-bounce indirect illumination is the Depth Occupancy volume, which means that the timings represent the exact viewing position of Figure 7. In the case of 1-bounce indirect illumination, the overall speed improvement ranges from 29% to 58%, depending on the visible geometry in the camera frustum. In the case of multi-bounce illumination the overall gain is smaller, due to the fact that CRC cells are flagged as occupied regardless of their visibility.

Please note that we stress our tests by completely rebuilding the volumes on each frame. A more realistic scenario would include action-triggered updates or amortized voxelization (3-way) and occupancy data generation and dilation, spreading the cost of these passes to 5 or more frames.

# 8. Conclusion

We have presented a real-time radiance caching scheme for diffuse global illumination based on the idea of chrominance compression in the spherical domain. The proposed reduction in SH coefficients and the respective number of texture fetches are further complemented by an optimization of the cache point generation in dynamic scenes. Our method is fast, can handle fully dynamic and large environments, requires no precomputations, has no view dependencies and is thus aimed at real-time applications. Finally, it is worth noting that the proposed radiance field compression is a generic strategy that can be applied to any other interactive or offline technique and other spherical basis functions.

### Acknowledgements

This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: ARISTEIA II-GLIDE (grant no.3712). The decimated Stanford Lucy model was downloaded from http://animium.com. All versions of the Sponza Atrium model are based on the Crytek Sponza mode which was downloaded from http://www.crytek.com/cryengine/cryengine3/downloads.

# References

- CRASSIN, C., NEYRET, F., SAINZ, M., GREEN, S., AND EISEMANN, E. 2011. Interactive indirect illumination using voxel cone tracing. *Computer Graphics Forum (Proceedings of Pacific Graphics 2011) 30*, 7 (sep). URL: http://maverick.inria.fr/ Publications/2011/CNSGE11b. 114
- DACHSBACHER, C., AND STAMMINGER, M. 2005. Reflective shadow maps. In *Proceedings of the 2005 Symposium on Interactive 3D Graphics and Games*, ACM, New York, NY, USA, I3D '05, 203–231. URL: http://doi.acm.org/10.1145/1053427.1053460, doi:10.1145/1053427.1053460. 113
- EISEMANN, E., AND DÉCORET, X. 2006. Fast scene voxelization and applications. In *Proceedings of the 2006 Symposium on Interactive 3D Graphics and Games*, ACM, New York, NY, USA, I3D '06, 71–78. URL: http://doi.acm.org/10.1145/1111411. 1111424, doi:10.1145/1111411.1111424. 121
- GREGER, G., SHIRLEY, P., HUBBARD, P. M., AND GREENBERG, D. P. 1998. The irradiance volume. *IEEE Comput. Graph. Appl. 18*, 2 (Mar.), 32–43. doi:10.1109/38.656788. 113
- KAPLANYAN, A., ENGEL, W., AND DACHSBACHER, C. 2011. Diffuse global illumination with temporally coherent light propagation volumes. In *GPU Pro 2*, W. Engel, Ed. A.K. Peters, Feb., 185–203. 114, 118

- KELLER, A. 1997. Instant radiosity. In Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, SIGGRAPH '97, 49–56. doi:10.1145/258734.258769. 113
- KŘIVÁNEK, J., GAUTRON, P., PATTANAIK, S., AND BOUATOUCH, K. 2008. Radiance caching for efficient global illumination computation. In ACM SIGGRAPH 2008 Classes, ACM, New York, NY, USA, SIGGRAPH '08, 75:1–75:19. URL: http://doi.acm. org/10.1145/1401132.1401228, doi:10.1145/1401132.1401228. 114
- MALVAR, H., AND SULLIVAN, G. 2003. Ycocg-r: A color space with rgb reversibility and low dynamic range. Tech. Rep. JVTI014r3. 21, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, July. URL: http://ftp3.itu.ch/av-arch/ jvt-site/2003\_09\_SanDiego/JVT-I014r3.doc. 114
- MAVRIDIS, P., AND PAPAIOANNOU, G. 2012. The compact YCoCg frame buffer. *Journal of Computer Graphics Techniques (JCGT) 1*, 1 (Sept), 19–35. URL: http://jcgt.org/ published/0001/01/02/. 112, 114
- NIJASURE, M., PATTANAIK, S., AND GOEL, V. 2005. Real-time global illumination on gpus. *Journal of Graphics, GPU, & Game Tools 10*, 2, 55–71. 114
- PAPAIOANNOU, G. 2011. Real-time diffuse global illumination using radiance hints. In Proceedings of the ACM SIGGRAPH Symposium on High Performance Graphics, ACM, New York, NY, USA, HPG '11, 15–24. URL: http://doi.acm.org/10.1145/ 2018323.2018326, doi:10.1145/2018323.2018326. 112, 114
- RAMAMOORTHI, R., AND HANRAHAN, P. 2001. An efficient representation for irradiance environment maps. In SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, ACM, New York, NY, USA, 497–500. 116
- 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. ACM Trans. Graph. 27, 5 (Dec.), 129:1–129:8. URL: http://doi.acm.org/10.1145/1409060.1409082, doi:10.1145/1409060.1409082. 113
- RITSCHEL, T., DACHSBACHER, C., GROSCH, T., AND KAUTZ, J. 2012. The state of the art in interactive global illumination. *Comput. Graph. Forum 31*, 1 (Feb.), 160–188. URL: http://dx.doi.org/10.1111/j.1467-8659.2012.02093.x. 113
- YEE, Y. H., AND NEWMAN, A. 2004. A perceptual metric for production testing. In ACM SIGGRAPH 2004 Sketches, ACM, New York, NY, USA, SIG-GRAPH '04, 121-. URL: http://doi.acm.org/10.1145/1186223.1186374, doi:10.1145/1186223.1186374. 124
- YU, I., COX, A., KIM, M. H., RITSCHEL, T., GROSCH, T., DACHSBACHER, C., AND KAUTZ, J. 2009. Perceptual influence of approximate visibility in indirect illumination. *ACM Trans. Appl. Percept.* 6, 4 (Oct.), 24:1–24:14. URL: http://doi.acm.org/10. 1145/1609967.1609971, doi:10.1145/1609967.1609971. 124

# **Author Contact Information**

Kostas Vardis	Georgios Papaioannou
kvardis@aueb.gr	gepap@aueb.gr

Anastasios Gkaravelis agkar@aueb.gr

Department of Informatics Athens University of Economics & Business 76 Patission Str Athens, 10434 Greece

Vardis, Papaioannou, Gkaravelis, Real-time Radiance Caching using Chrominance Compression, *Journal of Computer Graphics Techniques (JCGT)*, vol. 3, no. 4, 111–131, 2014 http://jcgt.org/published/0003/04/06/

Received:	2014-05-27		
Recommended:	2014-08-30	Corresponding Editor:	Marc Stamminger
Published:	2014-12-16	Editor-in-Chief:	Morgan McGuire

© 2014 Vardis, Papaioannou, Gkaravelis (the Authors).

The Authors provide this document (the Work) under the Creative Commons CC BY-ND 3.0 license available online at http://creativecommons.org/licenses/by-nd/3.0/. The Authors further grant permission reuse of images and text from the first page of the Work, provided that the reuse is for the purpose of promoting and/or summarizing the Work in scholarly venues and that any reuse is accompanied by a scientific citation to the Work.

