IMPROVING HIERARCHICAL MONTE CARLO RADIOSITY ALGORITHMS

Jackson Pope and Alan Chalmers

Department of Computer Science
University of Bristol, Bristol, UK
[pope|alan]@cs.bris.ac.uk

ABSTRACT

Hierarchical subdivision techniques remove the need for a-priori meshing of surfaces when approximating global illumination. In addition they allow progressive refinement of the solution. However, when subdivision is based upon Monte Carlo methods, due to the stochastic nature of such techniques, subdivision decisions cannot be made unless a sufficiently large number of samples have been considered. Shadow boundaries are one of the main features such subdivision algorithms are designed to detect, but mesh elements that are in shadow receive less light, and hence are slower to subdivide. In this paper we investigate methods for modifying the Monte Carlo hierarchical subdivision algorithm to improve the detection of shadow boundaries and caustics.

Keywords: radiosity, stochastic, Monte Carlo, hierarchical

1 INTRODUCTION

Global illumination is a task that requires a large number of computations to calculate. Monte Carlo based techniques have previously been introduced as methods to model complex properties of the global illumination solution [Kajiy86a]. Pattanaik and Mudur presented particle tracing, a Monte Carlo method modelling the propagation of 'photons' throughout the environment [Patta92a]. Pattanaik's method relied on the surfaces within the scene being discretised into a-priori elements before the rendering stage. This restriction leads to major problems. The scene modeller rarely knows what resolution to mesh the surfaces within the scene, as lighting variations across surfaces are expected to be resolved by the render. Both over- and under-estimation have significant drawbacks: under-estimating the surface gridding can cause the surfaces to appear discretised—with very noticable artifacts; over-estimation will take far longer for the approximation to converge, with typical Monte Carlo noise clearly visible if the approximation has not yet converged.

Hierarchical subdivision techniques have been introduced in other global illumination techniques as a method to adapt the mesh automatically to reduce the error. Hanrahan et al. developed an algorithm which subdivided mesh elements based upon the illumination gradient within that element[Hanra91a]. Heckbert used hierarchical techniques to store adaptive radiosity textures on surfaces when generating images using bidirectional ray tracing [Heckb90a]. Hierarchical subdivision was later incorporated into Monte Carlo radiosity techniques by Tobler et al. and Bekaert et al. [Toble97a, Bekae98a].

Due to the stochastic nature of Monte Carlo methods the radiosity approximation within a mesh element cannot be considered stable—locally converged enough for extra samples to make a negligable difference to the radiosity within that element—until a statistically significant number of samples have been collected. Until a mesh element is stable it cannot be considered for subdivision, as any imbalance in illumination across its surface could be due to stochastic noise. Current depth-first Hierarchical Monte Carlo Radiosity algorithms work well for areas of high illumination—where a statistically significant number of samples will arrive quickly. However, due to the partial occlusion of elements containing shadow boundaries those elements will be slower to subdivide than their unoccluded neighbours. This leads to scenes where a good approximation has been reached for most of the scene, but the shadow boundaries appear jagged due to the lower subdivision of the elements containing them. This paper proposes two techniques to improve the detection of illumination differences when subdividing in order to reduce the number of computations needed to reach an accurate enough solution.

The next section summarises the algorithm proposed by Tobler et al. for adding hierarchical subdivision to a Monte Carlo method for solving global illumination. In section 3 an improved subdivision metric is discussed and its behaviour is compared with the existing metric for two different scenes. Section 4 presents and evaluates a new heuristic for improving the subdivision along shadow boundaries where previous methods have been slow to subdivide. Section 5 describes how the resulting meshes may be triangulated to avoid T-vertex discontinuities, and shows the results for the method proposed in section 4 compared to the original algorithm using bilinear interpolation between mesh elements. The paper finishes with conclusions about the two methods and future improvements.

2 PREVIOUS WORK

This paper is based upon the particle tracing technique developed by Pattanaik [Patta93a] modified to include the hierarchical subdivision algorithm published by Tobler et al. [Toble97a]. However, the techniques presented in this paper are equally applicable to any method of generating a hierarchical subdivision of surface radiosity where a depth-first stochastic technique is used to model the propagated light [Heckb90a]. The particle tracing algorithm models global illumination by emitting 'photons' from the light sources within the scene and tracking them as they interact with the surfaces within the scene. When a photon is reflected by a surface, the radiosity within the element on that surface which the photon hit is increased by the energy the particle was carrying. The particle then loses energy according to the reflectance of the surface hit and is reflected. The particles are then absorbed when their energy drops below some energy threshold. To compensate for the bias this culling of particles introduces we use Russian roulette. The hierarchical algorithm, rather than depending upon pre-meshing of the surfaces, maintains a hierarchy of meshes in a quad-tree structure.

When a particle hits a surface, its intersection point is passed to the top level of a quad-tree representing piece-wise approximations to the radiosity across the surface at different resolutions. The particle hit is registered in the appropriate element at each level in the quad-tree. The levels within the quad-tree maintain counters indicating the number of photons that have been received by that mesh element, and the numbers that have propagated to each one of the elements on the level below. At any given time there is a set of nodes within the tree that represents the piece-wise functional representation of the radiosity over the surface. These nodes are the 'current' layer and need not all be on the same level of the quad-tree. One or more levels below these current elements are 'preview' elements which are used to gauge the accuracy of the functional representation at the current level.

Once a statistically significant number of particles have hit a current element, that element is considered for subdivision. If the difference in illumination between that element and the subelements on the preview layer is significantly large, the element is subdivided. The level below the current level becomes current and a new level is created below each of the preview subpatches. Throughout this paper we have used a constant representation of radiosity within an element—that the whole patch has radiosity equal to the average for that patch, however the techniques presented are equally valid for higher order representations.

3 MULTI-CHANNEL SUBDIVISION

Tobler et al. have determined whether a patch needs to be subdivided or not by comparing the number of incident photons on each of its subnodes in the preview layer. If the difference between nodes is sufficiently high then the patch is subdivided. This method relies on two assumptions: that the majority of incident photons within a scene carry direct illuminiation and hence have similar energy or 'weight'; and that there are no areas where the number of incident photons is constant but the radiosity changes. Let us consider these two assumptions. The particle tracing algorithm developed by Pattanaik either reflected photons at a surface (with probability equal to the surface reflectance p) or absorbed them (with probability $1 - p$) [Patta93a]. We choose instead to suppress the absorption of particles by decreasing the weight of particles according to the surface reflectance. Using this method only those photons carrying direct illumination are guaranteed to have maximum weight. The percentage of photons carrying maximum weight will vary depending upon the average reflectance of the scene and the threshold weight below which particles are subject to Russian roulette. Tobler's second assumption can be broken if the scene contains lights with different colours or brightly coloured texture maps.

3.1 Method

Since we cannot assume that the majority of incident photons have similar weights, it is necessary to test a subdivision criterion in which no assumption about incident particle weights is made. In addition to storing the number of incident particles—which is still used in this method as a node stability metric—we also store the total weight of all incident particles received and propagated by each node in the quad-tree in each channel of light being modelled. Light is modelled in the standard RGB triplet form, so each node contains both total received and total propagated weights of red, green and blue light. However, this method will extend to the spectral treatment of light.

The decision as to whether or not an element requires subdivision is now based upon the differences in received light at each patch for each channel. If any one channel has a difference large enough to warrant splitting, the element is subdivided.

3.2 Results

The results presented in this section compare the multi-channel subdivision metric described in the previous section with the method used by Tobler et al. which used the number of incident photons as a simple approximation to the average radiosity within a patch.

Two different test scenes are used: the first is a simple scene with a single light source, grey walls, a white floor, yellow ceiling, an occluder and a caustic generating transmissive sphere; the second is a similar sized scene but with two different coloured light sources occupying symmetrical positions on the ceiling. In the second scene the occluder and the transmissive sphere have been removed from the scene to minimise the differences in the number of particle hits received by mesh elements on the floor and walls. The two scenes are shown in Figures 1 and 2.

We would expect the multi-channel metric to perform similarly to the existing one on the first

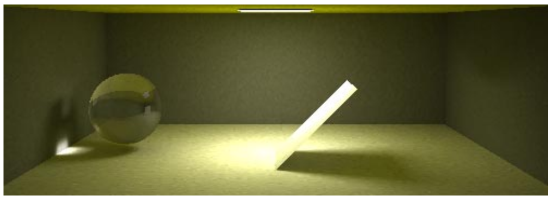

Figure 1: Scene 1 full solution.

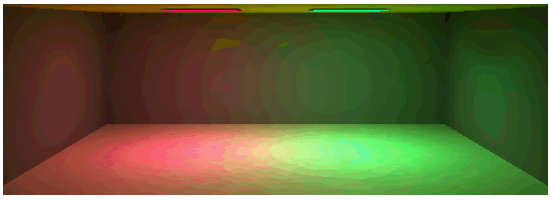

Figure 2: Scene 2 full solution.

scene as there are no areas where the radiosity changes without a similar difference in incident particles. However the second scene with the two light sources should display the multi-channel method to its full advantage. The different coloured light sources should lead to areas where there is a difference in radiosity without there being a difference in the number of particle hits. In addition, the way in which particles have their weight decreased as they interact with the surfaces of the scene should disadvantage the original method which relies on only small differences in particles weights.

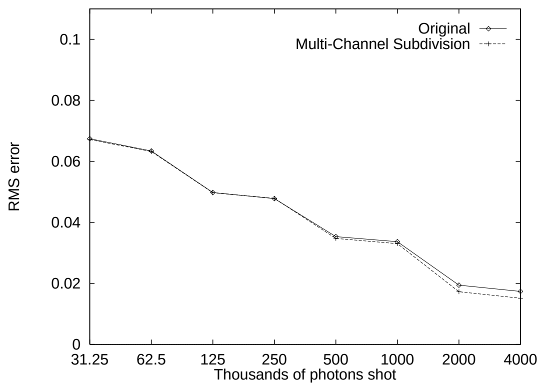

Figure 3: Convergence for original method and multi-channel method for scene 1.

Figure 3 illustrates that both methods perform very similarly for test scene 1. It is not until a large number of photons have been shot that the advantages of the multi-channel metric become apparent, providing better convergence to the solution. However, in Figure 4 where the two methods are tested on the scene containing coloured lights the multi-channel method performs significantly better when a large number of photons have been traced.

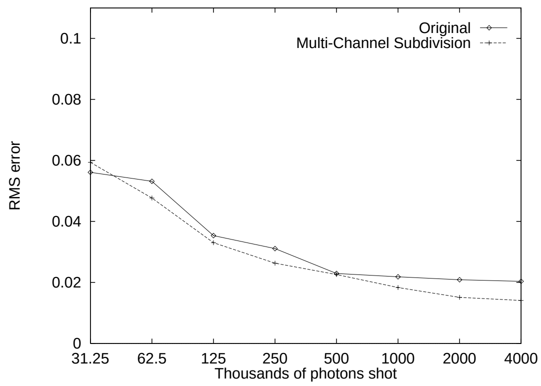

Figure 4: Convergence for original method and multi-channel method for scene 2.

From these results it is clear that the simple approximation proposed by Tobler et al. is sufficiently good to be used as a subdivision metric when rendering approximate images. However, if the full solution is required, the multi-channel metric will converge quicker to the solution. In the following sections we will use the simple subdivision metric to enable a performance comparison between the metric presented in this section and the enhancement presented in the next sections.

4 MULTI-STAGE SUBDIVISION

This section presents an extension to existing subdivision techniques to aid the detection of shadow boundaries and caustics. By exploiting the highly uneven distribution of particle hits in the elements which contain shadow boundaries we should be able to enhance the subdivision of these—typically slow to subdivide—elements.

Due to the stochastic nature of Monte Carlo techniques the radiosity of an element cannot be considered stable until a suitably large number of photons have hit it. This premise causes problems when considering shadow boundaries and caustics. Those elements that contain large radiosity gradients are the elements that we want to subdivide as soon as possible as it is within these patches that any aliasing artifacts due to an inadequate mesh will be most noticable. However, if a patch is partially occluded—and hence constitutes a shadow boundary—then it will obviously receive less light, and hence less particles. Therefore, existing hierarchical subdivision techniques divide the elements that cause the largest errors slower than they subdivide fully lit elements. In the next section we present a method that enables quicker subdivision of high-gradient patches which should reach a more accurate solution in less time.

4.1 Method

Our method is based upon the observation that we expect large differences in the number of hits between subnodes on the preview layer if a node contains a shadow boundary or caustic. Previously we have been unable to make a decision about whether or not to subdivide an element until a large number of particles have hit that element. Monte Carlo theory states that the variance in the illumination of a node's subnodes decreases with the number of particles that have hit that node in accordance with Shirley's results[Shirl91a]. Chebyshev's Inequality [Rice95a] states that for a random variable X, with mean μ and variance σ^2, and for any $t > 0$:

$$P\left(|X-\mu| > t\right) \leq \frac{\sigma^2}{t^2} \quad (1)$$

In this case, X is the fraction of current layer particle hits that have hit a particular subnode on the preview layer. Assuming that there is no difference in illumination across the element, the fraction of hits each element on the preview layer will receive can be extrapolated from the current level. The variance, σ^2 is controlled by the number of particles that have hit the element, and t is a threshold value which is used to determine whether or not an element needs to be subdivided. If we establish a confidence value, c, which is the probability of subdividing an element incorrectly, i.e. the probability of $|X-\mu|$ being greater than t, when μ is extrapolated from the current layer, then from Equation 1 we have:

$$c = \frac{\sigma^2}{t^2} \quad (2)$$

We note that, for elements containing high illumination gradients, at least one of the subnodes on the layer below will have large $|X-\mu|$. This can be detected by setting a higher threshold, t. From Equation 2, we note that to maintain the same level of confidence in our decision to subdivide the element, if t increases, then we can increase the variance, σ^2. Our confidence measure states the probability of incorrectly subdividing a patch—leading to an increase in Monte Carlo noise. If we can maintain the same level of confidence we will not increase the noise in our radiosity approximation. Since we know that the variance decreases as the number of particle hits increases, if using a larger threshold, t, then we can reach a decision with fewer particle hits.

In the previous sections we required a large number of photons before the illumination within an element could be considered stable. We determined this number by observing how many photons must hit an element before extra photons will only negligably change the illumination. Thus we set 4800 photon hits on an element in the current layer as our single stage. Rather than waiting until a suitably large number of photons have hit a patch before testing for subdivision against a small threshold, we can now perform multi-stage subdivision. In this section we introduce a number of new stages, each requiring fewer photons hits to perform the subdivision algorithm. As the number of photons hits required decreases we increase the threshold difference—between the current and preview layers—required for subdivision. In this manner we can detect high illumination gradients such as caustics and shadow boundaries quicker, and subdivide them accordingly. This leads to a technique which should perform much better in those areas where subdivision is critical to the perceived realism of scenes.

4.2 Results

The results presented in this section compare the multi-staged subdivision technique described in the section above to the original method propsed by Tobler et al. used in section 3. The results we calculated using the first test scene shown in Figure 1. In addition to the original subdivision criterion of a difference of 6% between the constant representations at the current and preview layers after 4800 photon hits at the current level, extra stages of subdivision were included, as shown in Table 1.

In Figure 5 the performance of the new technique is compared with the original method. It is clear that the multi-stage method converges quicker to the solution. The convergence in the early stages of the rendering is much better due to the multi-stage method detecting the caustic caused by the transmissive sphere with less than half the

Number of hits	Difference
300	33%
600	22%
1200	16%
2400	11%

Table 1: Subdivision criterion used in multi-staged subdivision technique.

number of traced photons than the conventional method. Figures 6 and 7 show the images generated by tracing 125,000 photons with and without using multi-stage subdivision. These images demonstrate the higher level of subdivision along the shadow boundary and within the caustic.

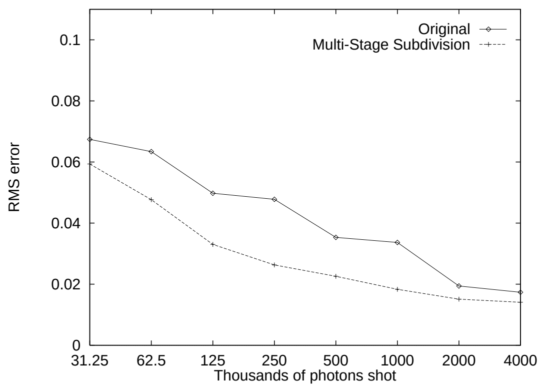

Figure 5: Convergence for original method and multi-stage method for scene 1.

5 TRIANGULATION

The results presented in previous sections have all been using a constant function to approximate the radiosity within a surface patch. Using a constant function enables easier visualisation of the patch size—as the patches appear discrete within the image—which was important for determining how quickly various elements subdivided. However, when rendering it is important that the disadvantages associated with constant approximation of illumination across a discretised surface such as Mach banding are minimised. In this section we test our multi-stage subdivision technique to ensure that the benefits provided by our method in the previous sections are also present when we interpolate the radiosity between the mesh elements.

First we describe the method we have used to determine how to interpolate between the cur-

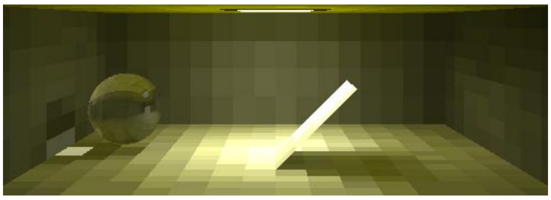

Figure 6: 125,000 photons without multi-stage subdivsion.

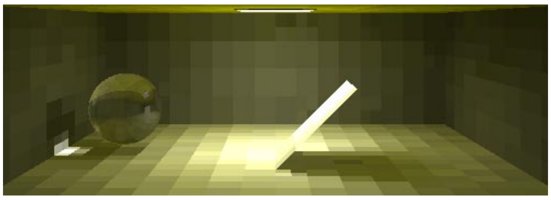

Figure 7: 125,000 photons with multi-stage subdivsion.

rent elements which can lie of different layers of the quad-tree. One possible solution is to perform some bilinear interpolation of the illumination between average radiosity values at the patch centres, using a technique such as Gouraud shading [Goura71a]. However, as we have placed no restrictions on the depth of subdivision it is possible for adjacent current elements to be on different layers. This causes problems with T-vertices, which are corner vertices of a mesh element, but fall along the edge of a neighbour which lies on a higher layer. These T-vertices can lead to very noticeable discontinuities in the illumination across a surface. Techniques have been developed to handle T-vertices [Silli94a], but many of these rely on a mesh being restricted—e.g. not allowing more than one level of subdivision difference between adjacent elements.

5.1 Method

As a post-processing stage, once the particle tracing stage is complete we compute a triangulation of each surface within the scene. For our research it is important that the triangulation can be used both for swift rendering of an image, and for interactive exploration of the particle-traced scene. For the first use, it is important that we do not just generate a list of triangles, as it should be possible to quickly ascertain which triangle an eye ray from the image-rendering stage has intersected. At the end of the particle-tracing stage we have the average radiosity values for each mesh element, as shown in Figure 8.

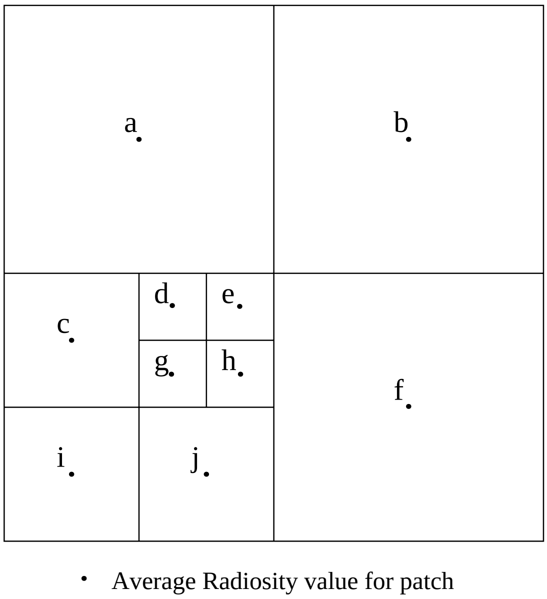

Figure 8: Example final mesh.

The next stage is to generate an ordered list of all adjacent elements centre points, and all those corner vertices of the element that lie on the surface edge. These centre points are then used to generate triangles, each consecutive pair of points forming a triangle with the centre point of the element. However, this method leads to mesh inconsistencies—where two mesh elements will cover the same area with different triangles. These mesh inconsistencies are undesirable as they can lead to visible discontinuities in the interpolated radiosity. To avoid these discontinuities we need to remove some of the triangles we generate from our triangulation to ensure that no two different triangles cover the same area. The simplest way to achieve this is to not include some of the centre points in our ordered list. We thus remove the centre points of any adjacent elements that lie diagonally above and left of the element, or diagonally below and right. For example in Figure 8 the naive method would generate the following list of centre points for the patch marked d: a, e, h, g, c—each of pair of which would form a triangle with the centre point of d. However, the generated list for g: d, e, h, j, i, c would then describe a different way to grid the square $d - e - h - g$ leading to a mesh inconsistency. Hence, as described above we cull the centre point diagonally below and right from d from its list (h), yielding the triangulation shown in Figure 9.

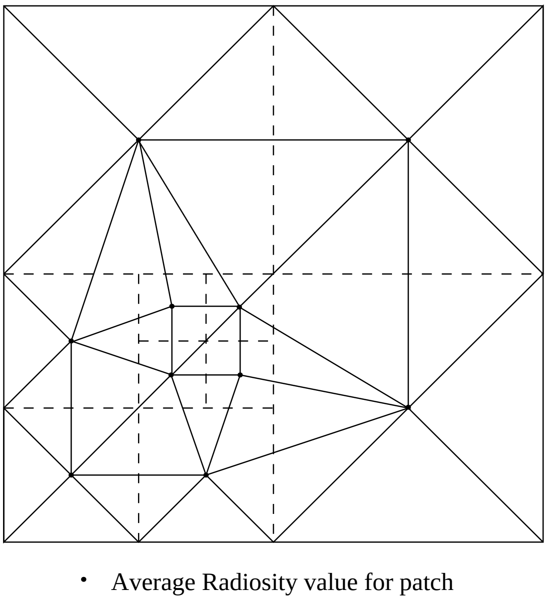

Figure 9: Example mesh after triangulation.

5.2 Results

The results presented in the previous sections of this paper have all been using a constant approximation of the radiosity within a mesh element. We chose to use a constant approximation as it would best display the subdivision of the surfaces to enable conclusions to be drawn about the subdivision method. In this section we will present results in which radiosity values have been bilinearly interpolated between the current layer mesh elements—as described in the last section—to converge quicker to the full solution. Figure 10 shows the results of the multi-stage subdivision method presented in section 4, compared with the original method. Results are shown both with and without the interpolation. From this figure it is clear that the multi-stage subdivision technique performs equally well for both interpolated and constant approximations. In both cases the multi-stage subdivision technique produces images of similar quality to the original method using half the number of photons—and hence half the rendering time. In addition the interpolated results—for both original and multi-stage methods—also produce the same error using half the number of photons required for the constant approximation to reach that error. These two improvements are cumulative with each other. In Table 2 we show mean rendering times and errors for the original method and the multi-stage subdivision method. These means were averaged over fifteen runs. The renderings are of the scene shown in Figure 1, with bilinear interpolation. They were rendered on a 333MHz Sun Ultra 10, with 128MB of RAM.

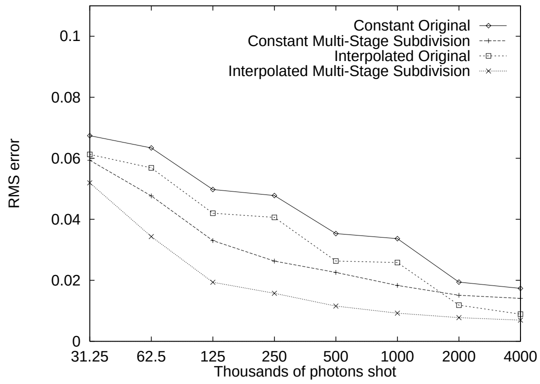

Figure 10: Convergence for original method and multi-stage method for scene 1.

Photons shot	Using M-S subdivision	Mean time	Mean error
125000	Yes	2mins 06.37	0.01947
500000	No	8mins 03.19	0.02615
250000	Yes	4mins 02.57	0.01599
1000000	No	16mins 07.40	0.02501

Table 2: Timing results for the multi-stage subdivision technique.

Using the multi-stage subdivision technique we can generate an image with significantly smaller error, using a quarter of the number of photons. This enables scenes to be generated with the better convergence in a quarter of the time using the multi-stage subdivision technique, as shown in Table 2. Our results also showed that the multi-stage subdivision technique gives more reliable results. For the sets of fifteen runs used to generate the results in Table 2, the multi-stage subdivision technique runs had time and error variances half those of the standard method.

6 CONCLUSIONS

In this paper we have presented two modifications to hierarchical subdivision algorithms for Monte Carlo radiosity methods. These modifications have been implemented using the hierarchical stochastic radiosity technique proposed by Tobler et al.—but are equally applicable to other depth-first methods [Heckb90a]. The first, a more accurate subdivision heuristic based upon the incident radiosity in each channel rather than the number of particle hits proved to offer some small improvement in results over the original method, including better convergence to the solution in the later stages of rendering. This was shown to be the case both for scenes in which the original method was expected to do well and scenes in which the new method was expected to prove more accurate.

The second technique presented used statistical theory to suggest a technique for more effective subdivision along shadow boundaries and caustics where subdivision is usually slower. By staging the subdivision and detecting large differences in the illumination across an element's preview subnodes after fewer particle hits we can detect high illumination gradients with the same confidence that we have not subdivided erroneously. This allows us to render images to a similar quality in a quarter of the time.

To ensure the robustness of our multi-stage subdivision technique we repeated our experiments using a triangulation technique to enable bilinear interpolation of radiosity values between mesh elements. The results showed a similar error between our method, using a quarter of the number of photons, and the original. This result was consistent for both the constant and interpolated radiosity values.

There are several improvements that could be made to this work. We have for the moment used only a constant representation of the radiosity over the surface elements for the subdivision metric. This work could be expanded to consider using the linear approximation used by Tobler et al.. The use of a linear approximation of the radiosity within a patch would decrease memory usage by limiting excessive subdivision in areas of constant radiosity gradient. Linear approximation would require up to ten times as many photon hits before a node could be considered stable. Using the multi-stage subdivision technique, that number could be significantly decreased in areas where there are large gradients.

REFERENCES

[Bekae98a] Bekaert, P., Neumann, L., Neumann, A., Sbert, M., Willems, Y.:*Hierarchical Monte Carlo Radiosity,* In Proceedings of the Ninth Eurographics Workshop on Rendering, pp. 259–268, Vienna, Austria, June 1998.

[Bekae99a] Bekaert, P.:*Hierarchical and Stochastic Algorithms for Radiosity,* PhD Thesis, Katholieke Universiteit Leuven, Heverlee, Belgium, 1999.

[Goura71a] Gouraud, H.: *Continuous Shading of Curved Surfaces,* IEEE Transactions on Computers, Vol. 20, No. 6, pp. 623–629, June 1971.

[Hanra91a] Hanrahan, P., Salzman, D. and Aupperle, L.:*A Rapid Hierarchical Radiosity Algorithm,* Computer Graphics (SIGGRAPH '91 Proceedings), Vol. 25, No. 4, pp. 197–206, 1991.

[Heckb90a] Heckbert, P.:*Adaptive Radiosity Textures for Bidirectional Ray Tracing,* Computer Graphics (SIGGRAPH '90 Proceedings), Vol. 24, No. 4, pp. 145–154, 1990.

[Kajiy86a] Kajiya, James T.:*The Rendering Equation,* Computer Graphics (SIGGRAPH '86 Proceedings), Vol. 20, No. 4, pp. 143–150, 1986.

[Patta92a] Pattanaik, S. N. and Mudur, S. P.:*Computation of Global Illumination by Monte Carlo Simulation of the Particle Model of Light,* In Proceedings of the Third Eurographics Workshop on Rendering, pp. 71–83, Bristol, UK, May 1992.

[Patta93a] Pattanaik, S. N.:*Computational methods for Global Illumination and Visualisation of Complex 3D Environments,* PhD Thesis, Birla Institute of Technology and Science, Pilami, India, 1993.

[Rice95a] Rice, J. A.:*Mathematical Statistics and Data Analysis*, Duxbury Press, Belmont, CA, 1995.

[Shirl91a] Shirley, P.:*Time Complexity of Monte Carlo Radiosity,* In Werner Purgathofer, editor, Eurographics '91, pp.459–465, Holland, September, 1991.

[Silli94a] Sillion, F. X., Puech, C.:*Radiosity & Global Illumination,* Morgan Kaufmann, San Francisco, CA, 1994.

[Toble97a] Tobler, R. F., Wilkie, A., Feda, M. and Purgathofer, W.:*A Hierarchical Subdivision Algorithm for Stochastic Radiosity Methods,* In Proceedings of the Eighth Eurographics Workshop on Rendering, pp. 193–203, St. Etienne, France, June 1997.