Improved Sampling and Reconstruction in Radiosity

RUI MANUEL RIBEIRO DE BASTOS[1]
MANUEL MENEZES DE OLIVEIRA NETO[2]

CESUP - National Supercomputing Center
UFRGS - Federal University of Rio Grande do Sul
Av. Osvaldo Aranha, 99 (fundos)
90035-190 - Porto Alegre - RS - BRAZIL
[1]rmb@cesup.ufrgs.br
[2]manuel@cesup.ufrgs.br

Abstract. Radiosity is a sampling and reconstruction method that approximates radiance function of real environments. Bad scene sampling leads to many artifacts in final images. Discontinuity meshing can provide initial meshing that warrants good sampling for regions with discontinuities in the illumination function. Although, that technique is unable to work in regions with no discontinuities. In such situations, adaptive subdivision can improve sampling. Even using adequate sampling, bad results are achieved if inadequate reconstruction techniques are being used. Good results have been obtained by using bicubic reconstruction of illumination functions. This paper analyses different combinations of sampling and reconstruction techniques and their contributions to improve realistic imagery.

1 Introduction

Surface's radiosity describes an arbitrary scalar function whose function space dimension is infinite. This means that solving the radiosity equation for a point x on a surface does not determine the radiosity at an immediately neighboring location [Cohen-Wallace (1993)]. As a result, a full and exact solution to the radiosity equation requires either finding the exact functional form of the radiosity across each surface or computing radiosity values for an infinite number of surface points. Because these are impossible tasks, one must use an approximate solution. The radiosity method computes an approximation of radiance values at discrete locations in the environment which are interpolated during the reconstruction step.

Traditionally, the radiosity method involves four steps (figure 1): environment meshing, form factors calculation, linear system of equations solution and reconstruction.

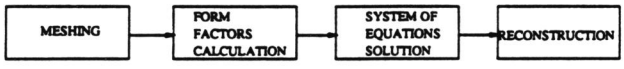

Figure 1: *Usual radiosity pipeline.*

Form factors are geometrical relationships among elements that compose the environment. Their evaluation is the most expensive part of the radiosity method. Although there is no analitycal solution to the general form factor integral when involving complex shapes or occlusions, numerical methods are used to approximate that integral as accurate as desired [Cohen-Greenberg (1985)], [Wallace-Elmquist-Haines (1989)], [Schroder-Hanrahan (1993)], [Malley (1988)].

Linear systems solution and form factors calculation are obtained using numerical methods and the accuracy of the results can be controlled. On the other hand, as the radiosity algorithm is based on sampling and reconstruction of the illumination function, it is very sensitive to sampled points and interpolation techniques used. In this way, impressive improvements to the final images can be obtained by driving effort to meshing and reconstruction steps.

Typical radiosity algorithms approximate illumination functions simulating patches with constant radiosity and use interpolation techniques to smooth results. The real illumination functions of scenes with occlusion are not piecewise smooth nor totally smooth. They are piecewise smooth (C^{∞})[1] within regions bounded by discontinuities of various orders[2].

[1] C^k indicates that a function is continuous in all derivatives up to and including k. C^0 indicates that a function is continuous in value (there are no sudden jumps). C^1 indicates continuity in slope (there are no kinks), and C^{∞} indicates that the function is smooth (continuous in all derivatives) [Cohen-Wallace (1993)].

[2] A function has a D^k discontinuity at x if it is C^{k-1} but not C^k there. D^0 indicates discontinuity in the function itself (value discontinuity). D^1 and D^2 indicate discontinuities in the first and second derivatives of the function, respectively [Cohen-Wallace (1993)].

The discontinuities are caused by transitions in occlusion between sources and receivers of light and can only be resolved at element boundaries, since the polynomial basis functions (interpolation) are continuous on the element interior [Bastos-Sousa-Ferreira (1993)].

Sampling strategies include discontinuity meshing (*a priori* meshing) [Heckbert (1992)], [Lischinski-Tampieri-Greenberg (1992)], [Drettakis (1994)] and adaptive subdivision (*a posteriori* subdivision) [Cohen *et al.* (1986)] techniques. Traditional rendering techniques used for reconstructing the illumination functions are constant (flat), bilinear and quadratic interpolations, but bicubic interpolation has shown to be necessary when continuity of first and second derivatives in the reconstructed function is desired [Bastos-Sousa-Ferreira (1993)].

This paper analyses combined uses of discontinuity meshing and adaptive subdivision either with bilinear or bicubic interpolation of the radiosity function. It shows that the combined use of *a priori* meshing with bicubic reconstruction leads to good compromise between image accuracy and computational costs involved.

2 Radiosity Matrix Solution

Algorithms used for solving radiosity linear systems of equations can be physically interpreted in two basic ways: *gathering* and *shooting* energy methods.

The first approach for solving the systems of equations was the application of algorithms such as Gauss Elimination [Goral *et. al.* (1984)] and Gauss-Seidel [Cohen-Greenberg (1985)]. This approach can be physically understood as gathering for each patch the energy being emitted and/or reflected by all patches in the scene. This means that after each step (each patch processed) the algorithm computes the final radiosity of one patch. For getting an image of the scene with no unprocessed patches (black patches) it is necessary to process every patch in the scene, with a computational cost of $O(n^2)$, where n is the number of patches in the scene.

In 1988, Cohen *et al.* [Cohen *et al.* (1988)] presented a progressive refinement method for solving the systems of equations based on the opposite way of light. At each step of the algorithm the patch with the largest unshot radiosity shoots its energy to all other patches. This means that after each step of the method (with computational cost of $O(n)$) the radiosity of every patch is updated and a new image can be generated each time closer to the final solution.

For patches with constant radiosity the progressive radiosity method follows the pseudocode presented in *Algorithm 1.*

```
1.  for(i = each patch){
2.      RAD_i = EMISSIVITY_i;
3.      unshotRAD_i = EMISSIVITY_i;
4.  }
5.  while(not converged){
6.      i = patch with largest unshotRAD_i * AREA_i;
7.      for(j = each patch){
8.          Δrad = unshotRAD_i * REFL_j * FORM_FACT_ji;
9.          unshotRAD_j = unshotRAD_j + Δrad;
10.         RAD_j = RAD_j + Δrad;
11.     }
12.     unshotRAD_i = 0;
13.     Display new image;
14. }
```

Algorithm 1: Pseudocode for progressive refinement radiosity.

3 Meshing strategies

Scene discretization is a sampling task which can lead to artifacts in the final images. Using large patches in regions of the scene with high radiosity gradients results in inadequate and low precision reconstruction funtions (radiosity) at those regions. The simple use of finer meshes strongly increases the computational cost and can be not enough for resolving some situations as, for instance, bad discontinuity representations in illumination functions [Bastos-Sousa-Ferreira (1993)] [Cohen-Wallace (1993)]. Ideally, the domain must be subdivided more finely only where it will improve the accuracy significantly.

Automatic meshing strategies can or can not use information about the illumination function behavior. If no information is available, only naive uniform or non-uniform subdivision can be applied. Both approaches are inadequate even for simple environments. The use of uniform meshing will frequently distribute the errors in the approximation of the radiosity function in an unevenly way over the mesh. As computational effort is equally distributed over the mesh, it has very unequal contributions to the resulting accuracy. Non-uniform meshing obtained without any knowledge of the radiosity function behavior can not warrant better computational effort distribuition than a uniform one.

Obtaining an optimal mesh normally requires some knowledge about the illumination function. Techniques that use such information can be characterized as either *a priori* or *a posteriori*. *A priori* techniques are used to determine discontinuity edges in the radiosity function before the solution is performed. After solution has been totally or parcially performed *a posteriori* techniques are used to refine

the mesh in regions with high radiosity gradients and no discontinuities.

3.1 *A Priori* Meshing - Discontinuity Meshing

Discontinuity meshing is an *a priori* technique in which the original mesh is improved by generating edges corresponding to boundaries involving shadow regions. These include lit-to-penumbra, penumbra-to-penumbra and penumbra-to-umbra transitions. Such edges are obtained based on purely geometric informations.

Discontinuity edges are used to represent discontinuities in value (C^0), and in first (C^1) and second (C^2) derivatives of the radiosity function that provide clues to three-dimensional shape, proximity and other geometrical relationships. The failure to correctly reproduce discontinuities can degrade image quality dramatically [Cohen-Wallace (1993)].

Discontinuities in value of the radiosity function can occur where one surface touches another. Discontinuities in first and second derivatives of radiosity function can occur at umbra and penumbra boundaries, as well as between different penumbra regions.

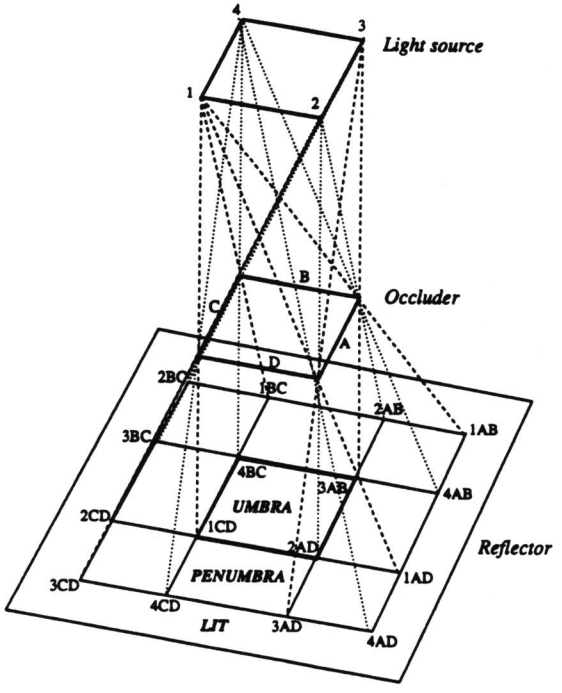

Figure 2: *A priori meshing*

Figure 2 shows the discontinuity edges created by an area light source on a reflector patch shadowed by an occluder. All the patches are aligned rectangles on parallel planes. The discontinuity edges are generated by projecting each edge of the occluder onto the reflector under the point of view of each vertex of the light source. For instance, vertex *1* of the light source projects edge *A* of the occluder onto the reflector from point *1AB* to point *1AD*.

The implementation of discontinuity meshing in *Algorithm 1* can be done before line 1 (before performing the solution of the radiosity matrix). The simplest implementation (brute force) has computational cost of $O(n^3)$, where n is the number of patches in the scene [Lischinski-Tampieri-Greenberg (1992)].

It must be noted that light sources and occluders have to be processed as a whole and not by their compounding patches. That is, if a(n) light souce (occluder) is subdivided into patches, the discontinuity meshing algorithm has to process the vertices (edges) of the whole light source (occluder) without taking into account its subdivision. This is true because across a light source there is no change in emissivity even if it is composed by some patches. In the same way, across an occluder there is no change in visibility between reflector and light source. This observation does not reduce the computational cost order of the discontinuity meshing algorithm, but reduces its final computational cost (reducing the value of n by grouping patches as objects).

3.2 *A Posteriori* Subdivision - Gradients Reduction

A posteriori meshing algorithms refine the mesh after the solution of the radiosity matrix has been at least partially completed. An initial approximation is obtained using a mesh determined *a priori*. That mesh is then refined in regions where the local gradient is high, using information such as the radiosity gradient provided by the initial approximation of the radiosity function. Adaptive subdivision decreases the local error by increasing the mesh density and it has been used in almost all radiosity *a posteriori* meshing implementations.

A posteriori refinement is not very effective for reducing errors in the neighborhood of discontinuites because it generally does not completely eliminate artifacts even when using high density meshes. On the other hand, it can be very effective when high gradients are entirely contained in regions that do not exhibit discontinuities (e.g., completely lit regions). In such situations, the discontinuity meshing algorithm can not act.

The implementation of adaptive subdivision in *Algorithm 1* can be done between lines 12 and 13 (after partially performing the solution of the radiosity matrix). The computational cost of applying adaptive subdivision is $O(n)$ for each step of the progres-

sive refinement algorithm, where n is the number of patches in the scene.

In this work, after processing each shooting patch all patches are processed evaluating vertices radiosity standard deviation for each patch. If the standard deviation is higher than an informed threshold the patch is binary subdivided. Another way of controlling the adaptive subdivision is limiting the maximum depth any quadtree can reach. The data structure used to keep that information is a *quadtree* for each patch. After applying the adaptive subdivision algorithm it is necessary to compute radiosity value at every new (generated) vertex and patch. The traditional approach is to use bilinear interpolation for approximating radiosity at the new vertices and patches. In this work, a gathering approach has been used to evaluate radiosity at those new vertices and patches. Along the progressive refinement algorithm a list of patches that have already shot their energy is kept. After binary subdividing a patch, radiosity is evaluated at the new vertices by gathering the energy from the patches in that list. This approach warrants more accurate radiosity values for the new vertices than the traditional one. Figure 6 illutrates the improvement provided by adaptive subdivision over the uniform mesh of Figure 5.

4 Reconstruction Step

The last step in radiosity pipeline is reconstruction. Images are generated interpolating sampled values of the radiosity function.

Since discontinuities in the illumination functions may be of various orders, interpolation schemes that enforce the appropriate degree of continuity at a particular patch boundary are required. The failure to correctly resolve those discontinuities can result in highly visible artifacts [Cohen-Wallace (1993)].

There are some reconstruction models, and the simplest one is the constant or faceted shading - a single value of the illumination function is used to shade an entire patch. As the name suggests, this model presents images with faceted appearance, caused by value discontinuities in the reconstruction function at the edges of the mesh (C^{-1} - discontinuous in value).

To reduce discontinuity problems a bilinear interpolation model can be used. Bilinear interpolation can ensure continuity of value or magnitude, but cannot eliminate derivative discontinuities (C^0 - continuous in value, discontinuous in derivative).

Discontinuities in value and/or in derivative in places where the illumination function should be smooth may appear as Mach band effects (artifacts) in generated images. To obtain reconstruction functions continuously differentiable higher-order interpolation schemes must be used [Bastos-Sousa-Ferreira (1993)].

Although continuity of value and derivative is desirable for reconstruction functions, care must be taken in cases where the illumination function presents magnitude and/or slope discontinuities. To be faithful to the illumination function a reconstruction function should be continuous where the illumination function is continuous and also preserve its discontinuities.

Quadratic interpolation schemes can avoid Mach band effects in almost every cases, but cannot guarantee derivative continuity between adjacent patches nor correctly handle derivative discontinuities.

A bicubic reconstruction scheme based on parametric bicubic Hermite interpolation has been used to correctly handle all kind of continuity/discontinuities in the reconstruction function up to the second derivative [Bastos-Sousa-Ferreira (1993)].

4.1 Reconstruction Problem

The reconstruction problem corresponds to correctly interpolate radiosity functions based on sampled points warranting that their continuities and discontinuities be perfectly represented.

In radiosity the objects of a scene are tesselated in geometrical patches. For simplicity the bicubic reconstruction function is tesselated in reconstruction patches directly associated to geometrical ones. For each geometrical patch there is a bicubic Hermite one for each color component (Red, Green and Blue).

It was assumed that continuity of the geometrical normal vector implies continuity of the reconstruction function. However, the *a priori* discontinuity meshing algorithm may indicate any kind of discontinuity at any point of the scene. In this way, all the reconstruction patches that share a vertex must share a single tangent plane to the radiosity function (gradient vector) at that point, unless a discontinuity is detected there [Bastos-Sousa-Ferreira (1993)]. If there is a discontinuity at a geometrical point there will be a vertex for each patch sharing the same geometrical properties. These vertices will have different gradient vectors.

The correct reconstruction of discontinuities between patches requires that adjacent elements do not share vertices (nodes) along such boundaries. A vertex is duplicated if there is any situation that causes a discontinuity in the illumination function at that vertex. In the same way, an edge is duplicated if there is any discontinuity, at least, at one of its vertices. A discontinuity between patches at a vertex or edge occurs if:

- the geometrical normal vector is discontinuous;
- the reflectivity is discontinuous;
- the emissivity is discontinuous;
- it is a part of the boundary between umbra and penumbra areas;
- it is a part of the boundary between penumbra areas or
- it is a part of the boundary between penumbra and fully lit areas.

Keeping all these rules in mind a final mesh is created, the radiosity at each vertex computed and an approximation of the gradient vector of the illumination function obtained for each vertex. Using the radiosity value and gradient vector at each vertex of a patch, the final color component of a pixel is evaluated using bicubic Hermite interpolation for any patch in the scene [Bastos-Sousa-Ferreira (1993)].

Special attention must be given to the problem of mapping discontinuity edges when it is necessary to subdivide a reflector introducing discontinuity vertices. T-vertices [Cohen-Wallace (1993)] (see, for instance, figure 2 vertex 3AD) must be avoided in order to warrant good reconstructions. A simple solution able to handle T-vertices is to propagate every new discontinuity edge beyond its discontinuity vertices along the patches and provide that these prolonged edges be treated as continuity ones (figure 3). This can be achieved by duplicating discontinuity vertices as explained above and keeping just one vertex for each continuity geometrical point. The inability to handle such situations introduces some artifacts which can be perceived as Mach band effects (figures 9 and 10).

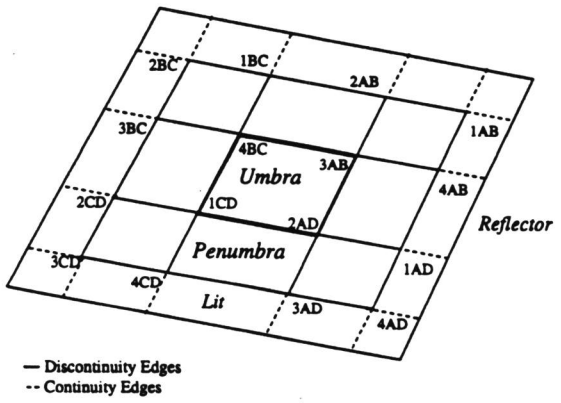

Figure 3: *T-vertices elimination by prolonged edges.*

Other situations that require special attention are those in which no umbra is generated. This can occur when either the relative size of the occluder is small if compared to the light source (figure 4) or when the occluder size is small if compared to its distance to the reflector. In such situations, the use of the simple discontinuity meshing strategy described in section 3.1 is not enough to produce all the actual discontinuity edges (some penumbra-to-lit areas discontinuities are not represented). This problem can be solved by adding to that strategy the projection of the light source edges through the occluder vertices onto the reflector.

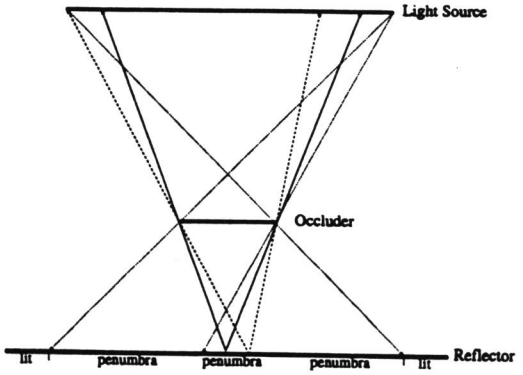

Figure 4: *Shadow with no umbra.*

5 Results

The combined approach of discontinuity meshing and adaptive subdivision techniques, and bilinear and bicubic reconstruction is being implemented on Silicon Graphics Iris Crimson workstations and on a CRAY Y-MP2E in C. It was used a simple geometrical scene with one light source, one occluder and two shadowed patches. Form factors computation was carried out by using ray casting algorithm [Wallace-Elmquist-Haines (1989)] undersampling shooters only at their centers. The results in Figures 5 to 8 are presented with mesh and corresponding bilinear (traditional) reconstruction.

Figure 5 presents the uniform initial mesh (37 vertices) for the scene with its corresponding image. It can be seen that the mesh cannot correctly represent the shadow. Figure 6 presents the mesh (1375 vertices) generated with adaptive subdivision and maximum quadtree depth level equal to four. It can be noted that there is no distinction between umbra and penumbra areas, although directions of these boundaries coincide with binary subdivision directions. Figure 7 presents the mesh (149 vertices) created using the discontinuity meshing algorithm. Although there are fewer patches than with adaptive subdivision, the shadow is perfectly described spliting umbra, penumbra and lit areas. Figure 8 presents the mesh (2217 vertices) created using both dis-

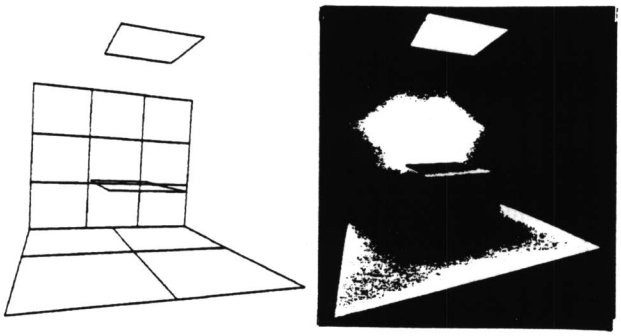

Figure 5: *Initial mesh and image.*

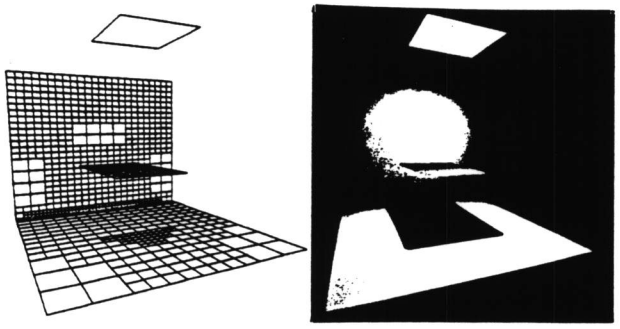

Figure 6: *A posteriori meshing and image.*

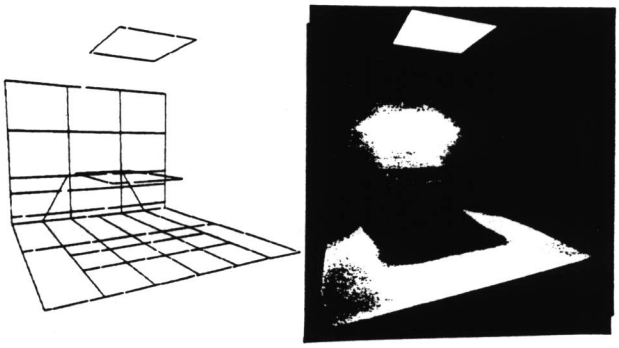

Figure 7: *A priori meshing and image.*

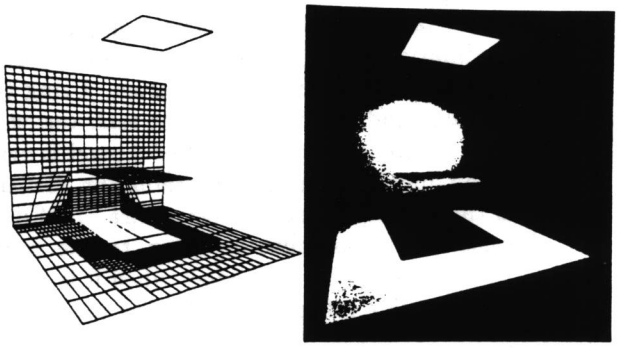

Figure 8: *Final mesh and image.*

continuity meshing and adaptive subdivision algorithms (with maximum quadtree depth level equal to four). The abscence of penumbra is explained by that the shooters were sampled just at their central points for form factors computation.

In order to compare bicubic and bilinear reconstructions it was used one mesh (figure 7) obtained with discontinuity subdivision, since bicubic reconstruction must exhibit continuity in regions with no shadows and preserve discontinuities between discontinuous regions. Figure 9 presents the image generated using bilinear interpolation between vertices of the same patch and figure 10 is the corresponding image for bicubic reconstruction. It can be noted that Mach band effects were avoided in the vertical patch when using bicubic interpolation scheme and that the image looks more realistic. Mach band effects observed on the horizontal reflector patch of figure 10 occur as consequence of not treating prolonged edges (figure 3) as continuity ones.

6 Relative Error Comparison

In order to compare the results accuracy obtained with discontinuity meshing, adaptive subdivision, bilinear and bicubic reconstructions techniques a supersampled reference image of the scene was used. The two shadowed patches were uniformly subdivided into 10,000 polygons each and the light source and the occluder were not subdivided (20,012 vertices). Bilinear interpolation was used as reconstruction technique and form factors were computed using ray casting with 16 sampling points. The relative error was computed by taking the absolute value of the subtraction between the RGB values of the reference image and those of each image to be compared divided by the maximum possible relative error for each pixel. To sample all regions of interest it was chosen a vertical central line. Important transition points are depicted in figure 11 as letters from A to J and refered on figures 12 and 13. These letters correspond to the following points:

A top boundary of the vertical patch;

B sampling edge of the initial mesh;

C top boundary of the occluder;

D bottom boundary of the occluder;

E top boundary of the penumbra region;

F top boundary of the umbra region;

G bottom boundary of the vertical patch;

H bottom boundary of the umbra region;

I bottom boundary of the penumbra region;

J bottom boundary of the horizontal patch.

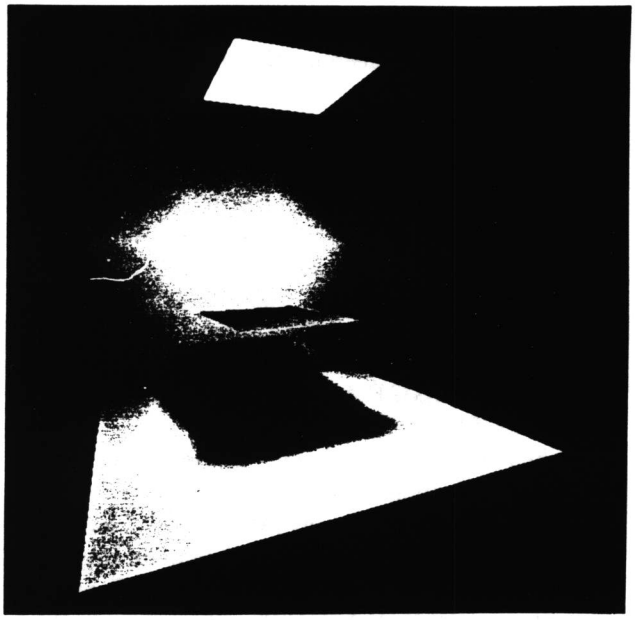

Figure 9: *A priori mesh and bilinear interpolation.*

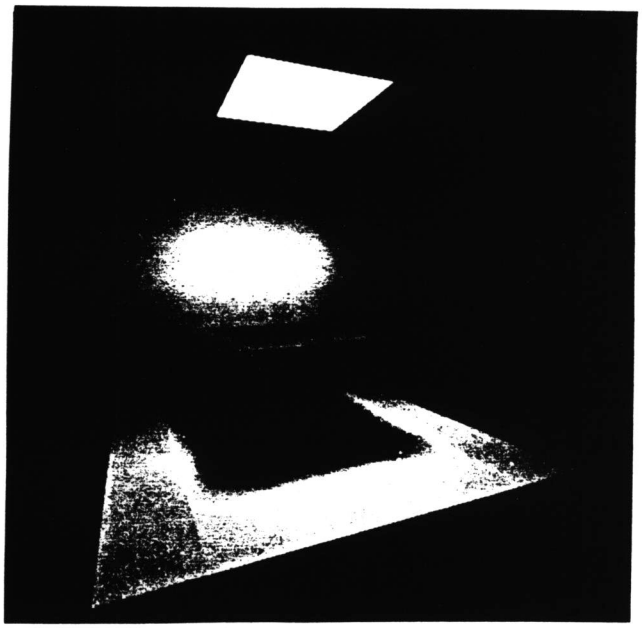

Figure 10: *A priori mesh and bicubic interpolation.*

Figure 12 presents the relative error of the red component of both *a priori* and *a posteriori* meshing techniques combined with bilinear reconstruction. It can be observed that in regions with high gradient in the radiosity function containing no discontinuities (areas from A to C) *a posteriori* subdivision presented more accurate results. On the other hand, *a priori* meshing was superior in regions that present discontinuities (from H to I).

Another measure used was the relative error distribution over the images. It was obtained as the sum of relative errors of all pixels divided by the number of pixels of the image. Using this approach *a priori* strategy shown to be 21.6% more accurate than *a posteriori* strategy.

Figure 13 presents the relative error of the red component in *a priori* and *a posteriori* subdivisions reconstructed with bilinear interpolation compared with *a priori* subdivision reconstructed with bicubic interpolation. It must be noted that in the bilinear case were used 2217 vertices while in the bicubic case only 149 sampling points. It can be observed that in regions from A to E (including high gradient regions) and from H to I (a penumbra region) bicubic interpolation has shown to be more effective on reconstructing the radiosity function. On the other hand, high sampling rates were superior in regions with low gradient (G to H and I to J).

The relative error distribution observed in bicubic and bilinear reconstructions has shown to be 22.7% lesser in bicubic case although it used only 6.7% of the total number of sampling points used in the bilinear case.

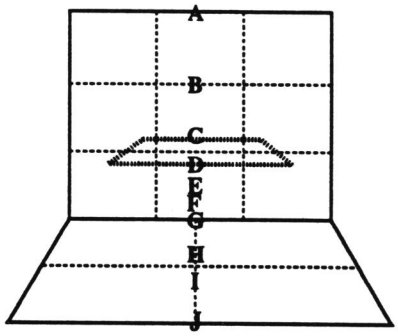

Figure 11: *Reference mesh for relative error analysis.*

7 Conclusion and Further Work

It was presented an analysis of the combined use of discontinuity meshing and adaptive subdivision techniques, and bilinear and bicubic reconstructions. Combination of discontinuity meshing with bicubic reconstruction has proved to be an excellent compromise between image accuracy and computational costs involved, requiring less dense meshes and providing better results than those obtained with finer meshes resonstructed using bilinear interpolation. The additional computation associated to bicubic reconstruction is not significant if compared to the cost of computing radiosity at sampling points.

Aspects of the present technique that suggest further research are:

Reduction of the computational cost of the dis-

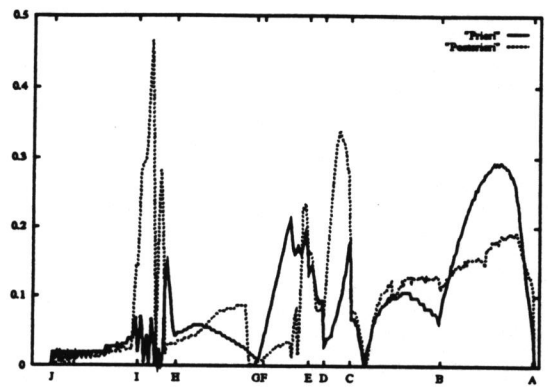

Figure 12: *Comparison of relative error (red component) for images obtained with a priori and a posteriori subdivision techniques.*

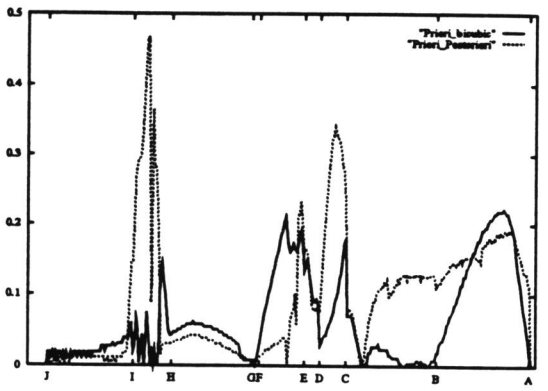

Figure 13: *Comparison of relative error (red component) for images obtained with bilinear and bicubic reconstruction techniques.*

continuity meshing algorithm by using *voxels* (octrees) for selecting patches before applying the algorithm.

For interactive applications the progressive radiosity approach [Cohen et al. (1988)] could be combined with a progressive meshing technique. In such an approach, *a priori* meshing would be applied before each step of the progressive refinement algorithm only for the current shooter if it is an emitter. After each step of the convergence algorithm *a posteriori* meshing is applied for the current mesh. Rapid convergence would be achieved and making as much progress as possible at the beginning of the algorithm would be preserved. It would also eliminate unnecessary wait for a complete traditional *a priori* meshing, allowing the presentation of intermediate results keeping the benefits of *a priori* meshing.

8 References

R. Bastos, A. Sousa, F. Ferreira, Reconstruction of Illumination Functions using Bicubic Hermite Interpolation. In *Fourth Eurographics Workshop on Rendering* (Paris, France, July 1993) 317-326.

M. Cohen, D. Greenberg, The hemi-cube: a radiosity solution for complex environments. *Computer Graphics* (**SIGGRAPH'85** *Proceedings*) **19**:3 (Aug. 1985) 31-40.

M. Cohen *et al.*, An Efficient Radiosity Approach for Realistic Image Synthesis. *IEEE Computer Graphics and Applications* (Mar. 1986) 26-35.

M. Cohen *et al.*, A Progressive refinement approach to fast radiosity image generation. *Computer Graphics* (**SIGGRAPH'88** *Proceedings*) **22**:4 (Aug. 1988) 75-84.

M. Cohen, J. Wallace, Radiosity and Realistic Image Synthesis. Academic Press, 1993, 381.

G. Drettakis, Structured Sampling and Reconstruction of Illumination for Image Synthesis. PhD's thesis, Department of Computer Science, University of Toronto, 1994.

C. Goral *et al.*, Modelling the interaction of light between diffuse surfaces. *Computer Graphics* (**SIGGRAPH'84** *Proceedings*) **18**:3 (Jul. 1984) 212-222.

P. Heckbert, Discontinuity meshing for radiosity. In *Third Eurographics Workshop on Rendering* (Bristol, UK, May 1992) 181-192.

D. Lischinski, F. Tampieri, D. Greenberg, Discontinuity meshing for accurate radiosity. *IEEE Computer Graphics and Applications* **12**:6 (Nov. 1992) 25-39.

T. Malley, A Shading method for computer generated images, Master's thesis, Dept. of Computer Science, University of Utah, (Jun. 1988).

P. Schroder, P. Hanrahan, A Closed form expression for the for factor between two polygons. Tech. Rep. CS-404-93, Dept. Computer Science, Princeton University, (Jan 1993).

J. Wallace, K. Elmquist, E. Haines, A Ray tracing algorithm for progressive radiosity. *Computer Graphics* (**SIGGRAPH'89** *Proceedings*) **23**:4 (July 1989) 315–324.

Figure 9

Figure10

Figuras a cores do artigo ***Improved Sampling and reconstruction in radiosity***..