Abstract

A Linear Hierarchical Radiosity method using point collocation and triangle meshes is proposed that allows $C^0$ continuity and performs energy exchange at any level both in the shooter and the receiver; this method leads to an exact representation of the Gouraud shading interpolation that will be used for rendering. A new refinement criterion is presented which tries to improve image quality taking into account: the smoothness of the solution based on pixel intensity values instead of energy ones, and visibility changes along the surfaces for high gradient detection (sharp shadows). In order to perform an efficient refinement a data structure is proposed which isolates every shooter contribution over each receiver and allows to only refine high error interactions.

Keywords: hierarchical radiosity, higher order elements, tone operator, refinement criteria.

1 Introduction

Hierarchical Radiosity (HR) [6] is considered as a major step towards realistic lighting simulations systems. At the heart of HR is the oracle function which predicts the error that will be introduced by linking two patches at a given level of the hierarchy. On the basis of the predicted error, either the link is established, or it is refined and a series of links are computed at lower levels of the hierarchy. The acceptability of a predicted error is related to the level of precision desired in the rendering, which, in turn, depends on the quality of the produced images. However, most existing oracle functions are based on the computation of object-space magnitudes, such as form-factors and energy values, and do not take directly into account image-space values such as the intensities of the images.

The view-independency of radiosity is often outlined as a major advantage of the global illumination techniques which compute it. However, as stated in [1] it is not always desirable to generate many different views of a single scene. Based on this assumption, several view-dependent radiosity techniques have been proposed such as the Importance-Driven Radiosity [17] and Two-Pass (Multi-Pass) methods [14, 2].

We herein propose an view dependent Linear Hierarchical Radiosity algorithm using point collocation, based on triangular meshes and a quadtree subdivision scheme. A refinement criterion is introduced that takes into account pixel intensities mapped from the radiosities in order to achieve a given degree of smoothness for a fixed point of view. This criterion along with the use of linear elements allow an efficient representation of the radiosity function and a reduction of the number of links computed for a given image quality.

The paper is divided into 7 sections. In section 2, a summary of the previous related work is presented. Next, the basic linear hierarchical method is described. In section 4, a new refinement criterion is proposed and discussed. In section five interaction graphs are presented. The results of the implementation are discussed in section 6 and the conclusions along with the future direction of research are presented in section 7.
2 Previous work

2.1 View Dependent Radiosity Methods

Although most radiosity methods are inherently view-independent, some techniques incorporate view-dependent information at different radiosity processing stages: Two-Pass [14, 13, 9], Multi-Pass methods [2] and Importance-based radiosity. Since this paper is restricted to diffuse interreflections, the two former ones will not be addressed here.

Importance-based radiosity [17] aims at computing as exactly as possible the radiosities of the elements that will have a quantifiable impact on the final image. Importance is defined as the fraction of radiosity emitted by a patch that ultimately reaches the observer. It is a concept dual to radiosity and it is distributed along with it during the HR process. In [3] a different formulation of importance has been proposed, allowing it to be transported similarly to the radiosity. With this approach, importance and radiosity are transported using the same data structures and algorithms. It should be noted that although Importance Driven Radiosity is view-dependent, it is still based on energy values (object space) rather than in mapped intensities (image space).

2.2 Oracles in Hierarchical Radiosity

Oracle functions in HR were first introduced in [6]. The brightness-weighted refinement criterion is based on an upper-bound approximated value of the form-factors \( F_{ij} \) and \( F_{ji} \) between the involved patches \( i \) and \( j \), and the energy transferred. Both are directly compared with the given thresholds \( \epsilon_F \) and \( \epsilon_{BF} \).

In Importance-Driven Radiosity, the oracle [17] takes into account the importance of an interaction by computing the error as the product of the estimated importance with the estimated radiosity, the reflectance and an error bound \( \Delta F_{ij} \) of the form-factor:

\[
\| \hat{I}_i \rho_i \Delta F_{ij} \hat{B}_j \| \text{, where } \hat{I}_i \text{ is the estimated importance of patch } i \text{, } \rho_i \text{ is the reflectivity of patch } i \text{ and } \hat{B}_j \text{ is the estimated radiosity of patch } j \text{. The error bound } \Delta F_{ij} \text{ is set to the difference between the upper } (\hat{F}^+_{ij}) \text{ and lower bound } (\hat{F}^-_{ij}) \text{ of the form-factor which are computed as the maximum and minimum values respectively of point-to-disk form-factors at given samples on the patches.}
\]

In [10] a new strategy is proposed which refines all links that have a significant effect on the total error bound, independently of the magnitude of their energy transfer. The error is:

\[
\hat{I}_i \Delta K_{ij} \hat{B}_j \text{ where } \hat{I}_i \text{ is the lower bound of the importance, } \Delta K_{ij} \text{ is the difference between the upper and lower bounds of } \rho_i F_{ij} \text{ and } \hat{B}_j \text{ is the upper bound of the radiosity.}
\]

Finally, Sillion and Drettakis [16] propose a feature-based oracle meeting a user-defined criterion. This metric takes into account the image quality by measuring how well illumination features are represented in it.

2.3 Radiosity with Linear Elements

The radiosity equation, which describes the radiosity \( B \) for all points \( x \) in an environment, can be expressed as [7]:

\[
B(x) = E(x) + \rho(x) \int_S B(x')G(x,x')dA'
\]

Being a continuous dimension function, it is generally approximated as a weighted sum of a finite number of simple textibasis functions \( N_i(x) \), defined over a subregion of the function domain:

\[
B(x) \approx \tilde{B}(x) = \sum_{i=1}^{n} B_i N_i(x)
\]

Point collocation and Galerkin Radiosity are the two most referenced methods proposed in order to minimize the error of the approximation of \( B \) [4].

In [12], a linear basis approximation with point collocation requiring a correct energy balance at the vertices of the patches is proposed, leading to the expression of a Vertex-to-Vertex Form-Factor.

3 Linear Hierarchical Radiosity

As stated above, point collocation and Galerkin Radiosity are the two most referenced methods pro-
posed in order to minimize the error of the approximation of the radiosity. Both methods lead to a system of linear equations:

\[ B_i^k = E_i^k + \rho_i^k \sum_j \sum_k E_{ij}^k B_j^l \]

where \( B_i^k, E_i^k, \rho_i^k \) respectively represent radiosity coefficient, emissivity and reflectivity for basis function \( k \) of patch \( i \); \( F_{ij}^k \) is the transport coefficient from basis function \( l \) of patch \( j \) towards basis function \( k \) of patch \( i \); finally \( B_j^l \) is the radiosity coefficient for basis function \( l \) of patch \( j \).

To represent the interaction between two patches nine transport coefficients are needed, the square of the number of basis functions per patch. If we refine the receiver we perform a recursive subdivision following a quadtree scheme (regular subdivision in four subtriangles) and create the new transport coefficients for the new basis functions and destroy the old ones; if we consider one refinement of a receiver, 12 new basis functions are considered (4 patches, 3 basis per patch) and 36 transport coefficients must be computed, thus \( x_i \) becomes \( x_{i0}\ldots x_{i3} \). However, if we look more closely at equation 1 we realize that two basis functions with different domains but with the same node have the same transport coefficients for the same shooter; for example, following Figure 2, basis functions \( N_0^2, N_0^0 \) and \( N_0^0 \) share the same node so \( F_{00,j}^{00} = F_{03,j}^{03} = F_{02,j}^{02} \), as shown by the black dot. Also, transport coefficients from the upper level can be reused; for example, basis functions \( N_0^0 \) and \( N_0^0 \) share the same node and thus \( F_{03,j}^{03} = F_{03,j}^{03} \). Following these rules there are only 9 new transport coefficients to be computed.

However radiosity coefficients cannot be shared because they represent radiosity at different levels and are computed after pushing and pulling radiosity over the hierarchy. In Figure 3 an example of this procedure is shown.

The push-pull procedure is illustrated in Figure 3 for triangle \( P_{01} \). It is performed in three steps after gathering energy through the links:

1. Add the contribution from above. It’s the contribution gathered in higher levels corresponding to the patch \( P_{01} \). This contribution is added to the coefficients \( B_{01}^k \), resulting in the new ones.
and we use linear interpolation to calculate the coefficients to be passed to the children; but it is possible that neighbours have gathered more accurate information. If we consider radiosity from shooter $A$, in Figure 4.c the value passed to node at the middle of the diagonal should be computed by linear interpolation, $(B_{00}^0 + B_{01}^1)/2$, but the exact value has been computed at the neighbours and stored at $B_{10}^0 = B_{12}^0 = B_{13}^0$. So we do not interpolate and use that value to compute the functions to be passed to the children. We will need to isolate each shooter contribution to be able to identify different contributions at different levels and use the right values; we will also need to use a restricted quadtree in order to simplify neighbour identification.

Figure 4: $T$-vertex problem: a) hierarchy of patches, b) contribution of shooter A and B, c) vertex where interpolation is not used

We have described refinement on the receiver and how energy is distributed over the hierarchy, but refinement can also be done at the shooter for accuracy on the transport coefficient. We will call a link the group of three transport coefficients associated to the three basis functions of a single shooter triangle to a certain receiver basis function. If a given refinement criterion decides that this link is not accurate enough then the shooter triangle is subdivided (if necessary) and four new links are established from the children to the receiver basis function. If the shooter was already subdivided then its radiosity coefficient is used directly; if not, the shooter is subdivided and the radiosities of the children are computed as in step 2 of the push-pull procedure.

We have described above a linear hierarchical method for radiosity but a refinement criterion is needed that decides when to perform the refinement of both shooters and receivers. In the next section
we depict this criterion that consists of two oracles: one that tells us if a receiver must be refined, and other one that tells us if a shooter must be refined.

4 Refinement criterion

As said in previous sections, most refinement criteria [6] take into account the energy transported in each interaction. However, the relationship between the energy of a given patch and the intensity of the pixels on which this patch is mapped is not linear [18]. Therefore some refinements in the energy transfer will have a very little noticeable effect on the image. We herein propose a refinement criterion that progressively improves image quality working directly on the Gouraud shaded image to be rendered.

4.1 Receiver oracle

The oracle for receiver refinement takes into account the smoothness of the approximation and the possible visibility artifacts (sharp shadows).

We assume that, at the rendering stage, a hardware implemented Z-Buffer visibility method along with Gouraud shading will be used. In addition, the scene surfaces are meshed in triangular patches as stated before, avoiding Gouraud shading orientation dependence. The intensity value \( E \) at a given pixel \( X \) is therefore obtained by linearly interpolating, at image-precision, the intensity values of the three vertices \( V_i^l \), \( 1 \leq l \leq 3 \) of the patch \( i \) that is visible at the pixel:

\[
E(X) = \text{LinearInterpolation}(E(V_i^0), E(V_i^1), E(V_i^2), X)
\]

The vertices intensities are computed by applying a mapping function to the radiosity values \( B_i^l \) at the vertices:

\[
E(V_i^l) = \text{Map}(B_i^l), \quad 1 \leq l \leq 3
\]

In this paper the mapping function proposed in [18] is used. This mapping uses an empiric model for the human eye response and takes into account the maximum intensity of the image and some parameters related to the display where it will be visualized. However it can be substituted by any other one.

A transition in the Gouraud shading at the shared edge between two neighbouring patches of a same surface is expected to be continuous. In other terms, the change in the intensity along a scan-line across two coplanar adjacent patches is expected to be the same at the right side and the left side of the separating edge. The angle between the two linear intensity distributions at any point on the separating edge is a measure of the shading discontinuity. Figure 5 illustrates this concept.

![Figure 5: Gouraud shading discontinuity; angle \( \alpha \) measures it](image)

Obviously, if the surface is curved, the angle value depends also on the geometrical angle formed by the patches which are not more coplanar.

A discontinuity at the separating edge indicates either an inaccurate radiosity estimation or a poor sampling ratio on the surface [15]. Assuming that the radiosity estimation is acceptable at the vertices, a discontinuity should lead to a subdivision of the two patches and a refinement of their vertices radiosity. However, the discontinuity will be significant only if the vertices radiosities are computed directly and not reconstructed by averaging the radiosities of the neighbouring patches, as with constant basis functions radiosity. With the linear hierarchical radiosity method depicted in previous section we accomplish this goal.

The question that arises is if the inverse statement is valid, i.e. may a continuity in the Gouraud shading across two adjacent patches fail at identifying a poor sampling ratio?

First of all, many different patches acting as shooters contribute to a given radiosity value. These contributions may compensate each other when they are added up. Thus non linear contributions of each patch separately may give a globally linear radiosity distribution. This is illustrated in Figure 6.a, where two shooters 1 and 2 are contributing over ad-
The minimum projection size of the smallest object face but at the receiver this integration is performed least such that the maximum patch size is less than on a single point the initial meshing must thus be all in any scene surface. However, in practice, the initial meshing size can be fixed to a uniform value without loose of accuracy.

Summarizing, for each patch contribution, the proposed oracle function tests Gouraud shading continuity at the edges between adjacent receiving patches and checks the visibility of the edge vertices towards the shooter. The oracle thus decides whether to subdivide or not and, in the former case, new links are established. As the radiosity is linear, the new links are created between the shooter and the new sampling points. The refinement of the new links is exposed below.

In addition, the oracle takes into account importance expressed according to a linear formulation proposed in [11]. Importance is trasmitted through the same links as radiosity and is treated in the same way. Initial importance is computed in terms of the point of view used for the final rendering.

The oracle therefore searches all the edges over each mesh and perform three steps for every shooter contributing at the level of the adjacent patches:

1. Compute the visibility variation between the two vertices $V_i^k$ and $V_j^k$ of the edge corresponding to the shooter $k$:

$$\Delta V_{ij}^k = \| V_i^k - V_j^k \|$$

2. This value is tested against a given threshold in order to decide if visibility change is small enough to consider the change of slope as a valid measure of smoothness:

- If $\Delta V_{ij}^k$ is below a given threshold then we can use the change in the slope as a measure of smoothness. As this oracle is applied many times during the algorithm we will use an easy but non conservative estimation of this change of slope.

Since we are considering image values we must compute the slope change using pixel intensity values instead of radiosity ones. Since this map is not linear we also must take into account the contribution of the other shooters in order to estimate the real change of slope produced by shooter $k$; the greater the energy of the other shooters the smaller the slope change.

Using geometry shown in Figure 7, we compute the slope change as:

![Image](image-url)
\[
\Delta E_{12}^k = \frac{1}{2} \| \text{Map}(B_1 + B_{\text{min}}) + \text{Map}(B_2 + B_{\text{min}}) - \text{Map}(B_3 + B_{\text{min}}) + \text{Map}(B_4 + B_{\text{min}}) \|
\]

where, \( B_{\text{min}} \) is the minimum radiosity of the four vertices, without considering the contribution of the shooting patch \( k \), and \( B_l \), \( 1 \leq l \leq 4 \) are the radiosities at the vertices of the two adjacent patches being \( l = 1 \) and \( l = 2 \) the separating edge vertices.

- If instead \( V_{k,\text{diff}} \) is superior or equal to the threshold, then a more conservative estimation of the slope change is used. We use the difference between the maximum and the minimum pixel intensity values of the four vertices, taking into account again the contribution of other shooters:

\[
\Delta E_{12}^k = \max_{l=1...4} \{ \text{Map}(B_{\text{min}} + B_l) \} - \min_{l=1...4} \{ \text{Map}(B_{\text{min}} + B_l) \}
\]

3. The estimator \( E_{k,\text{diff}} \) is tested against the slope threshold, weighted by the importance of the two adjacent patches:

\[
E_{k,\text{diff}} \cdot (I_{\text{left}} + I_{\text{right}}) < \epsilon_{\text{pixel}}
\]

where \( E_{k,\text{diff}} \) is the estimated intensity variation across the edge, due to the shooting patch \( k \), \( I_{\text{left}} \) and \( I_{\text{right}} \) are the importance of the adjacent patches \( \text{left} \) and \( \text{right} \), and \( \epsilon_{\text{pixel}} \) is the threshold.

If the expression become false then patches \( \text{right} \) and \( \text{left} \) are subdivided and new links created from shooter \( k \).

This formulation has been used assuming that the triangular meshing is regular, that is, the triangles are equilateral. Other expressions could also be used in presence of irregular triangles.

### 4.2 Shooter oracle

The above discussion has described how a receiver has to be subdivided for increasing accuracy of the solution. But the criterion is based on accurate radiosity values at the vertices. If not enough accuracy is achieved at those vertices then the oracle will fail.

When a link is initially established between a shooter and a vertex (associated to a basis function at a given level) we must decide if transferring energy at that level is accurate enough to guarantee a good approximation of the radiosity value at the vertex. If not we destroy that link, subdivide the shooter and create four new links from the children to the vertex, and then test those new links recursively for accuracy.

This lead us to a another oracle that decides if a link at a given level is accurate enough. We use an energy based oracle that compares the energy carried by the link with a given threshold, and if higher refines it. This criterion uses the maximum of the three products of shooter coefficients by the transport coefficients, weighted by the importance of the receiver triangle, and test it against a given threshold.

### 5 Interactions Graphs and Progressive Refinement

As mentioned in the previous section, a requirement of both the hierarchical method and the refinement criterion is to isolate the contribution of each shooting patch to each receiver basis function:

- When performing the push-pull procedure shooter contributions must be kept isolated in order to push the right energies.
- When applying the refinement criteria on the receiver each shooter contribution is tested separately for smoothness.

In order to manage each shooting patch separately, a graph data structure is used. For efficiency
reasons, patches belonging to a same surface are processed altogether in a single mesh; for example, an sphere will be discretized into a set of patches all of them belonging to the same mesh. Thus, for each shooting mesh, a graph is computed for any mesh receiving energy from it. The graph is a representation of the receiver-mesh at the level of refinement in which the energy transport between the shooter and the receiver is computed. The nodes of the graph are vertices of the receiver and the arcs are the edges between them. Figure 8 shows the two different graph representations of a same mesh, for two different shooters 1 and 2. The numbers in the patches of Figure 8.a show at which level each shooter is contributing. Figures 8.b and 8.c show the two graphs for the receiver mesh corresponding to shooters 1 and 2 respectively. Note how the graphs have a restricted quadtree structure.

Figure 8: Two graph corresponding to the same mesh for different shooters 1, 2: a) mesh of triangles showing at which level each shooter contributes, b) graph for shooter 1, c) graph for shooter 2

However, if all meshes are linked against any other one, the memory requirement of $n_{mesh}^2$ graphs may be a serious drawback of the proposed method. Therefore, as proposed in [8], a progressive refinement strategy is used, allowing us to avoid an initial quadratic linking and therefore reducing the number of computed graphs. At the beginning of each iteration, mesh-to-mesh interactions are checked and, if necessary, new interaction graphs are created. The experimental results in section 6 show that the memory requirement overhead due to the interaction graphs has an average value of 30 percent.

6 Implementation and results

In order to evaluate the proposed method we have compared it with the progressive refinement proposed in [8].

Colour plates 1.a and 1.b show the first scene used; it is composed by two rooms with four light sources each one, and some furniture; there are 386 initial surfaces. The second scene in colour plates 2.a and 2.b is composed by four rooms with two light sources each one and one table per room; there are 296 initial surfaces. Colour plates 3.a and 3.b show a detail of the mesh for first and second scene.

Colour plates 1.a and 2.a show the reference images. Colour plates 1.b and 2.b show the images computed with the linear method. Finally, colour plate 3 show a detail of the first scene along with the mesh generated by the linear method; note how the shadows are well recognized and represented.

The scenes have been processed at different levels of image quality, measured in comparison to a reference image with a metric proposed in [5]. This metric is a pixel-by-pixel absolute difference of the approximate image from the reference image; we have used 10% and 15% as differences. The reference image has
been computed with a classical hierarchical method plus a final gathering pass per pixel [3]. The number of links for each image has been measured and Figures 9 and 10 show plots for classical and linear methods.

It can be seen that the proposed method converges more quickly to a good image quality in terms of the ten-per-cent norm. Our method does not perform much better for a small number of links but when detailed images are required classical hierarchical method tends to produce good average images but with no well resolved details like sharp shadows; however the refinement criterion presented here allows to converge to detailed images faster. Other measurements like MSE (Minimum Square Error) or PSNR (Peak Signal to Noise Ratio) do not give much better results, but this a common problem when analyzing image quality. In fact we believe that better measures for image quality would show better results for this method.

In all executions the memory overhead due to the interaction graphs is constant, with an average value of 30 percent.

7 Conclusions and future work

A new linear HR algorithm for triangular meshes using point collocation has been proposed, based on a refinement criterion that takes into account the expected smoothness of Gouraud shading at the separating edges between adjacent patches of a same surface. The method requires the storage of an interaction graph for any important mesh to any shooter that have a significant impact on it. This data structure allows to refine separately the interactions of a patch with each shooter.

The current implementation results show that this criterion reduces significantly the number of links for a given accuracy while keeping a reasonable memory requirement overhead.

We are currently working on the analysis of a better meshing strategy, adapted to the shadow areas. Another future research line is the use of clusters in order to reduce the number of graphs and thus making the progressive linking strategy unnecessary and avoiding energy loses and memory overhead.

Acknowledgments

This work is partially supported by the grant TIC 95-0630-C05-05 of the CICYT. Ignacio Martin has developed this work with a FI (Formacio de Investigadors) grant from the Generalitat de Catalunya. Thanks to George Drettakis and Frederic Perez for his invaluable comments and help.

References


