Lecture Notes: Radiosity

Dani Lischinski

March 8, 2004

These notes are, for the most part, a summary of Chapter 3 from the Wallace and Cohen radiosity book.

1 The Rendering Equation

So far, we saw two forms to write down the integral equation describing the radiance $L(x,\omega)$ function at every point and every direction in the scene. The first form expresses reflected light as an integral over all incoming directions:

$$L(x,\omega_r) = L_e(x,\omega_r) + \int_{\omega_i \in \Omega} f_r(x,\omega_i \rightarrow \omega_r)\, L_i(x,\omega_i) \cos\theta_x \, d\omega_i$$

Alternatively, we can write the reflected radiance as an integral over other points in the scene:

$$L(x,\omega_r) = L_e(x,\omega_r) + \int_{y \in \Gamma} f_r(x,\omega_{xy} \rightarrow \omega_r)\, L(y,\omega_{yx})\, G(x,y)\, dy$$

2 Radiosity

Today we'll learn one way to solve the rendering equation. This method is called radiosity, and it is based on a finite-element formulation.

The traditional assumption underlying the radiosity method is that the world is a purely diffuse place. BTW, this assumption has been relaxed, but it makes life much easier when explaining the method (and when implementing it for that matter), so we'll stick with this assumption for a while.

First, we'll introduce a new quantity, called "radiosity" (surprise, surprise...), to make our derivation more consistent with the ones in the literature.

So (going back to physics for a brief while), let's define a new quantity that we'll call radiosity, which is defined as the total (hemispherical) flux density leaving a differential area on a surface.

There is a simple relationship between radiosity and radiance: in order to obtain the radiosity at a point x, we must integrate the radiance leaving the surface in all directions:

$$B(x) = \int_{\Omega} L(x,\omega)\cos\theta\, d\omega$$

The units of radiosity are Watts per square meter, same as irradiance.

Note that if the surface is indeed a Lambertian reflector, than the outgoing radiance does not depend on the direction, so it can be taken outside the integral:

$$B(x) = L(x)\int_{\Omega} \cos\theta\, d\omega = \pi L(x)$$

Now let's rewrite the rendering equation in terms of radiosity:

$$B(x) = E(x) + \rho(x)\int_{y\in S} B(y)\, G(x,y)\, dy$$

where $\rho(x)$ is the ratio of outgoing flux density to incident flux density, which is the same as the ratio of radiosity to irradiance. This is a number between 0 and 1. It can be shown that

$$\rho(x) = \pi f_r(x)$$

It should be easy to convince yourselves that this equation is equivalent to the one in terms of radiance, except that the $G(x,y)$ term now has a π in the denominator:

$$G(x,y) = \frac{\cos\theta_x \cos\theta_y V(x,y)}{\pi \|x-y\|^2}$$

3 Solving the radiosity equation

How do we solve an integral equation like that? There is no closed form solution, except in some very trivial cases, so we must resort to numerical approximation techniques.

Abstract mathematical explanation: the solution is a function in some infinite-dimensional function space (for example, the space of L^2 functions over the appropriate domain). We will project the equation onto a finite-dimensional function space, and we will find a function in that space that "best" approximates the real solution.

Now, in more down-earth-terms:

1. Subdivide the surfaces into n elements;

2. Select locations or *nodes* on the elements;

3. Associate a basis function $N_i(x)$ with each node; Our approximation will be a linear combination of these basis functions, each weighted by the corresponding nodal value B_i:

$$\hat{B}(x) \approx \sum_{i=1}^{n} B_i N_i(x)$$

   How do we know if this approximation is any good? For one thing, we know that if we have a solution, it must satisfy the equation. So, one thing that we can use to indicate the accuracy of the approximation is the *residual function*:

$$r(x) = \hat{B}(x) - E(x) - \rho(x) \int_{y \in S} \hat{B}(y)\, G(x,y)\, dy$$

   So, we'd like find a set of nodal values B_i that minimize the residual function. But the residual is also in a finite-dimensional function space, so it is not clear how to minimize it. Instead of minimizing the residual, the finite-element approach chooses to minimize the projection of the residual in some finite dimensional function space.

4. Choose n weighting functions $W_i(x)$, and project the residual onto each of these n functions. The idea is that each of these n projections is going to provide us with a linear equation in the n unknowns B_i. Using these n simultaneous linear equations, we can find the values of the unknowns for which each of the projections is zero (i.e., the residual is orthogonally projected onto the function space spanned by the weighting functions).

   Ok, but how do we project a function? Just like we do with discrete vectors: we compute the inner product. In the continuous case, the inner product $< r, W_i >$ is an integral instead of a summation:

$$< r, W_i >= \int r(x)\, W_i(x)\, dx$$

5. Compute the coefficients of the following linear system of equations:

$$< r, W_i >= 0 \qquad \text{for } i = 1, \ldots, n$$

   These coefficients are primarily based on the geometric relationships between pairs of basis functions. These relationships are usually referred to as *form factors*.

6. Solve resulting linear system of equations.

7. Now we have an approximation to the radiosity at each point on every surface in the scene, so we can render shaded images from arbitrary viewpoints with little extra work.

4 Point collocation

Let's be more specific now, and show what happens for a particular choice of basis and weighting functions.

The simplest set of weighting functions one can imagine is a set of delta functions: i.e., $W_i(x) = \delta(x - x_i)$, that is $W_i(x) = 0$ unless x happens to coincide with the node x_i.

This choice of basis functions mean that we want the residual to be zero at the nodes that we've chosen. However, note that we can't say much about the behaviour of the residual between the nodes. This technique is referred to as *point collocation*.

In this case, $< r, W_i >= r(x_i)$. So, we get the following set of equations:

$$r(x_i) = 0 \qquad \text{for } i = 1, \ldots, n$$

$$\hat{B}(x_i) - E(x_i) - \rho(x_i) \int_{y \in S} \hat{B}(y)\, G(x_i, y)\, dy = 0$$

Let's replace each occurrence of $\hat{B}$ with it's definition:

$$\sum \hat{B}_j N_j(x_i) - E(x_i) - \rho(x_i) \int_{y \in S} \sum \hat{B}_j N_j(y)\, G(x_i, y)\, dy = 0$$

This can be regrouped as follows:

$$\sum \hat{B}_j \left[N_j(x_i) - \rho(x_i) \int_{y \in S} N_j(y)\, G(x_i, y)\, dy \right] = E(x_i)$$

To simplify things, let's introduce a bit of notation:

$$K_{ij} = \left[N_j(x_i) - \rho(x_i) \int_{y \in S} N_j(y)\, G(x_i, y)\, dy \right]$$

Using this notation we have:

$$\sum K_{ij} B_j = E_i$$

which can also be expressed in matrix notation as $KB = E$.

So far, we haven't even specified what the basis functions are. Well, again, let's pick the simplest choice possible: piecewise constant basis functions. We have n elements, a single node per element (let's put it in the center). Let's define N_i to be 1 over the area of the i-th element, and 0 elsewhere. What are the entries of the matrix K?

They are:

$$K_{ij} = \delta_{ij} - \rho(x_i) \int_{y \in A_j} G(x_i, y)\, dy$$

The integral in this expression is the form factor from point (differential area at) x_i to the finite area element A_j, whose physical interpretation is: "what fraction of the projected hemisphere around x_i is subtended by the element A_j?"

5 Galerkin

In the Galerkin formulation of the weighted residuals approach, the weighting functions W_i are chosen to be equal to the basis functions N_i. Thus,

$$\begin{aligned} < r, W_i > &= < \hat{B}, N_i > - < E, N_i > - \rho(x) \int_{y \in S} \hat{B}(y) < G, N_i > dy \\ &= < \sum B_j N_j, N_i > - < E, N_i > - \rho(x) \int_{y \in S} \sum B_j N_j(y) < G, N_i > dy \\ &= \sum B_j \left[< N_j, N_i > - \rho(x) \int_{y \in S} N_j(y) < G, N_i > dy \right] - < E, N_i > \end{aligned}$$

Again, if we define

$$K_{ij} = < N_j, N_i > - \rho(x) \int_{y \in S} N_j(y) < G, N_i > dy$$

we obtain a linear system of equations

$$KB = E,$$

where $E_i = < E, N_i >$.