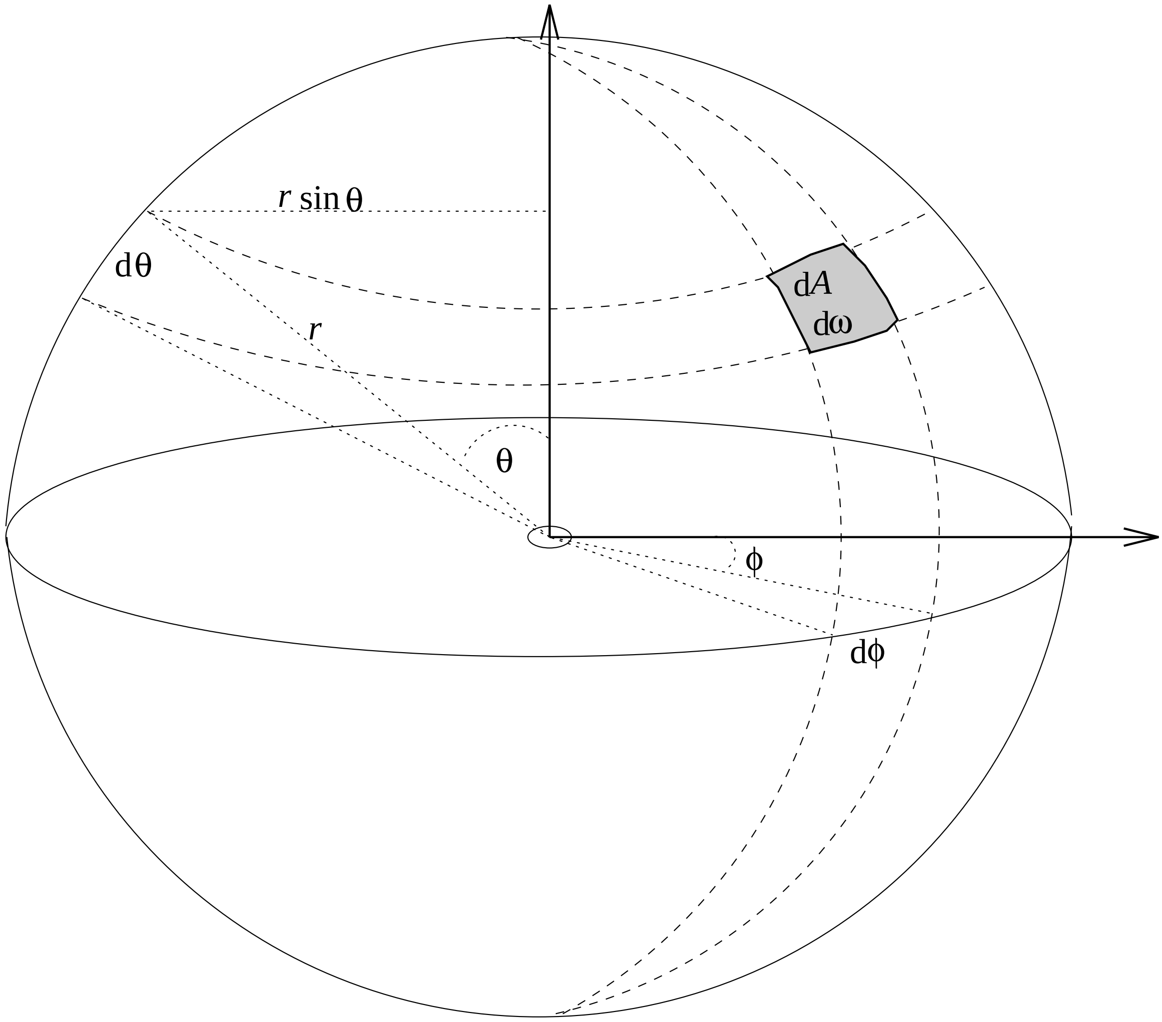
Introduction to Physically-Based Illumination, Radiosity and Shadow Computations for Computer Graphics

Eugene Fiume
Department of Computer Science
University of Toronto

Overview

- physically-based illumination models for computer graphics are maturing.
- tutorial on concepts of physically-based models.
- introduction to light transport (rendering equation).
- overview of radiosity-based approach.

Solid Angle



Differential surface element $\mathrm{d}A$ **on sphere of radius** r**:**

$$\begin{aligned} \mathrm{d}A &= \text{(longitudinal arc length)} \times \text{(latitudinal arc length)} \\ &= (r\,\mathrm{d}\theta)\,(r\,\sin\theta\,\mathrm{d}\phi) \\ &= r^2 \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi. \end{aligned}$$

Differential solid angle $\mathrm{d}\omega$ (on sphere of radius 1):

$$\mathrm{d}\omega = \frac{\mathrm{d}A}{r^2} = \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi.$$

Solid angle (in steradians) requires (messy) contour integral.

I don't think this is very clear

The longitudinal arc is on a circle of radius r that slices through the sphere and includes the origin.

So, what's the arc length of part of a circle? Let's suppose that this circle is parameterised by

$$X(\theta) = r \sin\theta,$$

$$Y(\theta) = r \cos\theta.$$

The arc length of a part of the circle described by angle $\Delta\theta$ is

$$\int_{\theta}^{\theta+\Delta\theta} \sqrt{\dot{X}^2 + \dot{Y}^2}\, d\theta.$$

But look at the integrand:

$$\sqrt{\dot{X}^2 + \dot{Y}^2} = \sqrt{r^2(\sin^2\theta + \cos^2\theta)} = r.$$

So, the length of the longitidinal arc is just

$$r\, d\theta.$$

Remember the parametric definition of a sphere:

$$x(\theta,\phi) = r \sin\theta \sin\phi,$$

$$y(\theta,\phi) = r \sin\theta \cos\phi,$$

$$z(\theta) = r \cos\theta.$$

Holding θ constant gives us a latitudinal circle of radius $r \sin\theta$ in the $z = r \cos\theta$ plane. So the length of the latititudinal arc is

$$r \sin\theta\, d\phi.$$

Solid Angle Properties

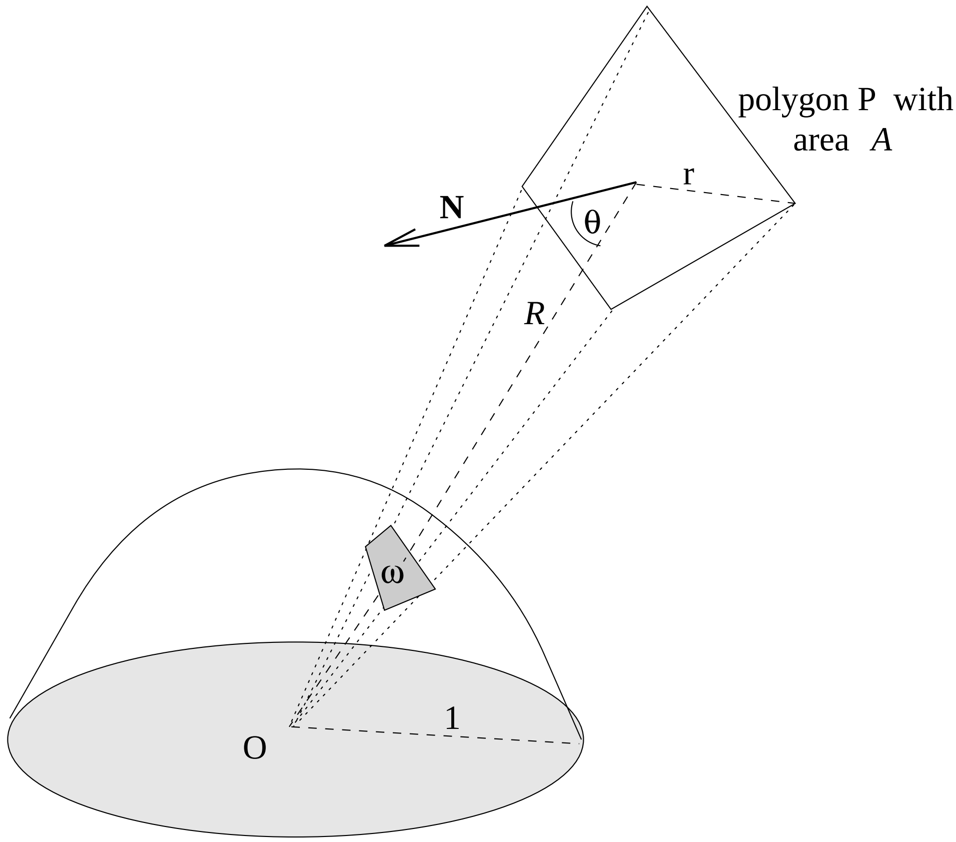
Surface area of unit hemisphere = 2π steradians (sr).
Surface area of unit sphere = 4π steradians (sr).

Just do the integral:

$$\int_0^S \int_0^{2\pi} \sin\phi \, d\phi \, d\theta,$$

where $S = \pi/2$ for the hemisphere and π for the sphere.

Practical Solid Angle Computation



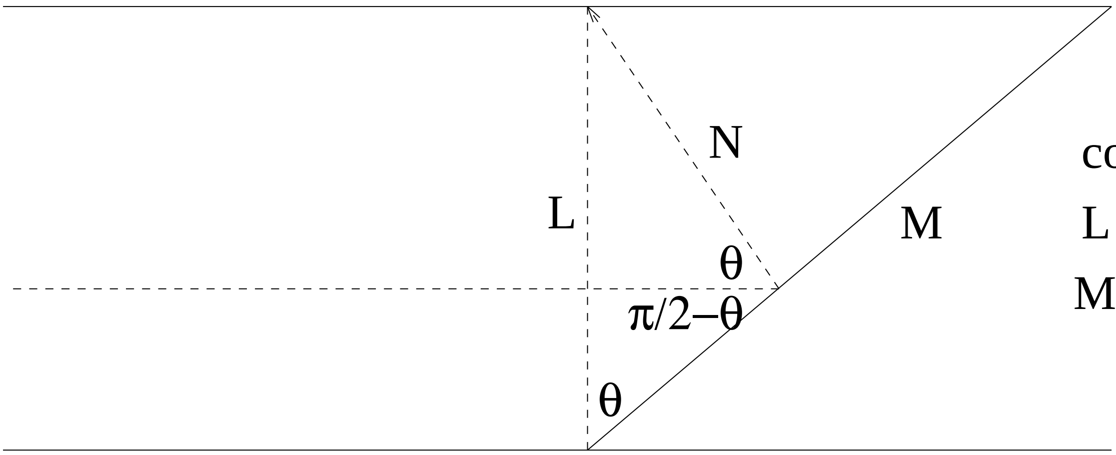
Solid angle of P: projected area of P onto sphere in direction of O.

Approximation: if $r \ll R$ so that projection of P on sphere is almost planar, then approximate solid angle definition by

$$w = \frac{A \cos\theta}{R^2}.$$

Geometry of Basic Projection

Suppose we have a beam of thickness L projecting onto a locally planar region with normal N. Assume the beam make has an incidence angle of θ with the plane.



$$\cos\theta = L / M$$
$$L = M \cos\theta$$
$$M = L / \cos\theta$$

Then we have the above relations.

Physical Lighting Terminology

What is this ‘‘intensity’’ stuff anyway? They are all measures of *power density* w.r.t. solid angles and area (and power is a measure of *energy density* w.r.t. time).

***Radiant power* (Φ):** Rate at which light energy is transmitted (energy/unit time, or Watts, W). Physical term: *flux*.

Radiant Intensity (I): Flux radiated onto a unit solid angle in a given direction (W/sr). E.g., "intensity" of a point light source.

Radiance (L): Radiant intensity per unit projected surface area ($W/(m^2 \cdot sr)$). E.g., "intensity" of reflection of a surface.

Irradiance (E): Incident flux density on a locally planar area, in units of flux per unit surface area (W/m^2). This is a direction-independent quantity.

Radiosity (B): Exitant flux density from a locally planar area, in units of flux per unit surface area (W/m^2).

These are *radiometric* quantities, i.e., physical measurements of electromagnetic energy. There are comparable *photometric* quantities corresponding to psychovisual measurement.

Relationships Between Radiometric Quantities

Radiant power: Φ.

Irradiance/Radiosity:

$$E, B = \frac{d\Phi}{dA},$$

where

$$d\Phi = \left(\int_{\Omega} L \cos\theta \, d\omega \right) dA,$$

Ω is the visible hemisphere, and

L is incoming or outgoing *radiance*.

Radiant Intensity:

$$I(\boldsymbol{\omega}) = \frac{d\Phi}{d\omega}.$$

Radiance: the quantity from which we construct the others. Call it

$$L(\mathbf{x}, \boldsymbol{\omega}),$$

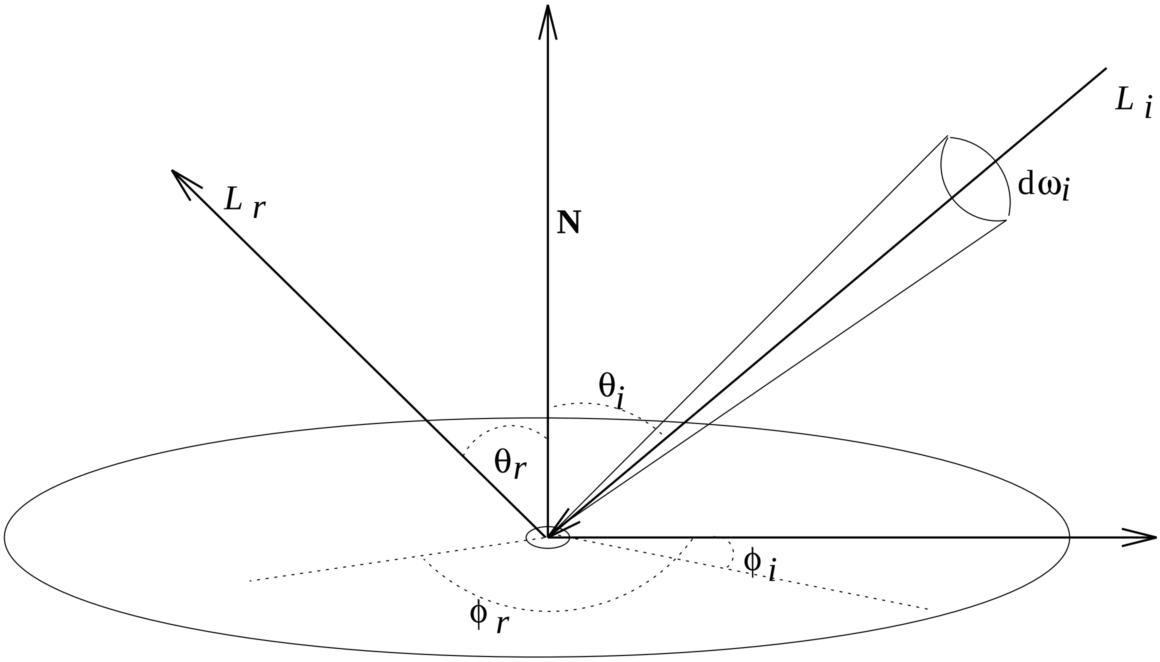
which is power per unit area radiated to point $\mathbf{x}$ along direction $\boldsymbol{\omega}$.

Use *radiance* to denote light carried by a *ray* in a ray tracer. (NO inverse square law in non-participating media.)

Use *radiant intensity* to denote energy distributions of light sources. (Follows inverse square law.)

Incident Irradiance

Suppose incoming radiance from differential solid angle $d\omega_i$ along direction $\boldsymbol{\omega}_i$ is L_i.



Then *incident irradiance* is

$$E_i \quad = \quad L_i \cos\theta_i \, d\omega_i$$

Bidirectional Reflectance Distribution Function

To characterise arbitrary reflection functions, write down an expression that relates incident irradiance to outgoing *radiance* in direction $\boldsymbol{\omega}_r$:

$$f_{i \to r} \quad = \quad \frac{L_r}{E_i}$$

or

$$L_r(\boldsymbol{\omega}_r) \quad = \quad f_{i \to r}(\theta_i, \phi_i, \theta_r, \phi_r) \, E_i,$$

so *bidirectional reflectance distribution function (BRDF)* $f_{i \to r}$ is proportion of incident irradiance that is reflected in direction $\boldsymbol{\omega}_r$.

BRDF and Reflectance

Units of BRDF are sr^{-1}, and gives density of flux per steradian.

BRDF is always positive, and is in fact a true distribution (and really should only be used under an integral).

But it is possible for a BRDF to be an "impulse", meaning perfect specular reflection in one direction.

Reflectivity or the *Reflectance*

Reflectance acts like a BRDF, but is instead a unitless proper fraction in [0,1]. It is defined as

$$\rho \quad = \quad \frac{\text{reflected flux}}{\text{incoming flux}} \quad = \quad \frac{d\Phi_r}{d\Phi_i} \,.$$

The region of integration gives different kinds of canonical reflectances.

Of particular interest to us is this fact: if the surface is *Lambertian* or *diffuse*, meaning that incoming light is equally likely to be scattered in all directions, then the *hemispheric reflectance* is

$$\rho_\mathrm{d} \quad = \quad \pi\, f_{i \to r} \,,$$

or

$$f_{i \to r} \quad = \quad \frac{\rho_\mathrm{d}}{\pi} \,.$$

Another nice fact:

$$\rho_\mathrm{d} \quad = \quad \frac{B}{E} \,.$$

Physically-Plausible Reflectance Models

The use of BRDF allows specification of

- anisotropic reflection models: fixing i and r but rotating the surface will change directional reflectance.
- fully plausible, energy-conserving, empirically validated illumination models.
- models that are really really expensive to compute. Really!

How does this sit with traditional illumination models?
In fact, it's not that hard to make current models more plausible.
Traditionally, local illumination models have been defined as

intensity at a point = ambient + diffuse + + specular.

That is,

$$I_P = k_a I_a + k_\mathrm{d} I_\mathrm{d} + k_s I_s.$$

As a first step, to translate to physical models, use radiance, and define BRDF as

$$\rho = k_\mathrm{d}\rho_\mathrm{d} + k_s\rho_s, \quad k_\mathrm{d} + k_s = 1,$$

where ρ_d is Lambertian reflector and ρ_s is your favourite specular reflection function this week. Leave ambient component *ad hoc.* So,

$$\begin{aligned} L_r &= k_a L_a + (k_\mathrm{d}\rho_\mathrm{d} + k_s\rho_s)E_i \\ &= k_a L_a + (k_\mathrm{d}\rho_\mathrm{d} + k_s\rho_s)\, L_i \cos\theta_i \,\mathrm{d}\omega_i \end{aligned}$$

Sum or integrate as appropriate over all light sources.

Rendering Equation (Kajiya and Hanrahan)

Characterise all interactions of light from any point $\mathbf{x}$ on any surface s with a point $\mathbf{x}'$ on a surface.

The result is an outgoing radiance value.

The *rendering equation* describes a *two-point* (i.e., one-bounce) light transport.

Multi-point transport is got by cascading the equation to get full ray transport problem.

Based on radiance.

$$L(\mathbf{x}, \boldsymbol{\omega}) \quad = \quad L_e(\mathbf{x}, \boldsymbol{\omega}) \;+\; \int_S f(\mathbf{x}'; \mathbf{x})\, L(\mathbf{x}', \boldsymbol{\omega}')\, G(\mathbf{x}, \mathbf{x}')\, V(\mathbf{x}, \mathbf{x}')\, \mathrm{d}A',$$

where

S is all surfaces, and $\mathrm{d}A'$ is a differential element of S ,

$L_e(\mathbf{x}', \boldsymbol{\omega}')$ is self-emission,

$V(\mathbf{x}, \mathbf{x}')$ is 1 if $\mathbf{x}$ is visible to $\mathbf{x}'$, and is 0 otherwise,

$G(\mathbf{x}, \mathbf{x}')$ is a geometric scale factor,

$f(\mathbf{x}'; \mathbf{x})$ is BRDF (actually depends on all angles): light reflected about $\mathbf{x}$ from $\mathbf{x}'$.

Notice that radiance to be computed appears on LHS and under integral.

Kajiya proposed some (theoretically) nice ways to compute the equation (including Monte Carlo path tracing and a series expansion).

Radiosity Equation

Under Lambertian assumption, the BRDF is independent of incoming, outgoing directions, and can be taken out of the integral.

Furthermore, outgoing radiance is direction independent, so we can work with irradiance/radiosity instead.

In fact,

$$L(\mathbf{x}, \boldsymbol{\omega}) \;=\; \frac{B(\mathbf{x})}{\pi}, \qquad (*)$$

so we can rewrite the rendering equation as a *radiosity equation*:

$$B(\mathbf{x}) \;=\; E(\mathbf{x}) \;+\; \frac{\rho_{\mathrm{d}}(\mathbf{x})}{\pi} \int_{S} B(\mathbf{x}')\, G(\mathbf{x}, \mathbf{x}')\, V(\mathbf{x}, \mathbf{x}')\, \mathrm{d}A'$$

Notes:

- This reformulation in terms of radiosity/irradiance can be easily got by multiplying rendering equation by π and changing units.
- Notice the change from BRDF f to reflectance ρ_{d}.

So far, so good, but notice again that radiosity appears on LHS and under integral.

From now on, we will bury the visibility term V and $1/\pi$ into geometric term G. G will resurface as the integrand for a form factor.

Discretisation of the Radiosity Equation

Even the radiosity equation is intractable (let alone the rendering equation), so work with that first.

Break all surfaces up into *patches* or *elements* and make assumptions about the nature of radiosity. Let there be n patches A_i, $i = 1, \cdots, n$. (By convention, areas of the patches will have same name – – ouch!)

The easiest assumption is that radiosity is constant across a surface. If we can solve for radiosities, then, each element i will have constant radiosity B_i.

In that case, we get

$$B_i \quad = \quad E_i \; + \; \rho_i \sum_{j=1}^{n} B_j \, F_{ij}.$$

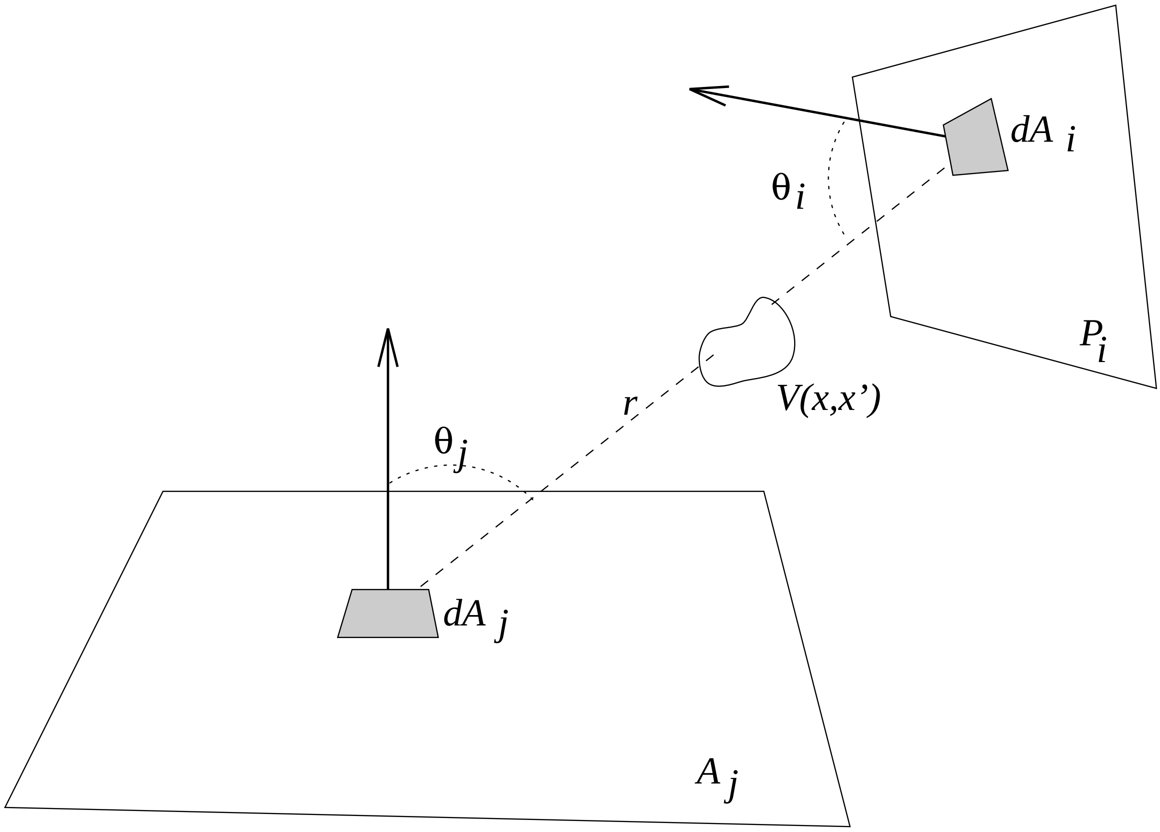
To make the derivation clearer, multiply through by A_i, which is allowed because $A_i > 0$:

$$B_i \, A_i \quad = \quad E_i \, A_i \; + \; \rho_i \sum_{j=1}^{n} B_j \, F_{ij} \, A_i.$$

Physically, what this does is turn an equation in flux density to an equation in flux.

We'll see why this is useful on the next slide, when we talk about F_{ij}, the *form factor*.

Form Factors



The term F_{ij} is a *form factor* and denotes the fraction of energy leaving A_i and arriving at A_j:

$$F_{ij} = \frac{1}{A_i} \int_{A_i} \int_{A_j} \frac{\cos\theta_i \, \cos\theta_j}{\pi r^2} V_{ij} \, \mathrm{d}A_j \, \mathrm{d}A_i .$$

That means that

$$A_i F_{ij} = A_j F_{ji},$$

so that

$$F_{ji} = F_{ij} \frac{A_i}{A_j} .$$

This is called *form-factor* reciprocity.

The Radiosity System of Equations

Folding form factors back into the radiosity equation, we see that since

$$\begin{aligned} B_i\, A_i &= E_i\, A_i \;+\; \rho_i \sum_{j=1}^{n} B_j\, F_{ij}\, A_i . \\ &= E_i\, A_i \;+\; \rho_i \sum_{j=1}^{n} B_j\, F_{ji}\, A_j . \end{aligned}$$

Dividing through by A_i again gives us

$$\begin{aligned} B_i &= E_i \;+\; \rho_i \sum_{j=1}^{n} B_j\, F_{ji}\, \frac{A_j}{A_i} \\ &= E_i \;+\; \rho_i \sum_{j=1}^{n} B_j\, F_{ij} . \end{aligned}$$

This expression states that the radiosity (i.e., outgoing flux density) of element i is an accumulation from the emitted flux density of this element, and of contributions of flux radiated onto element i from all other elements.

Now, the E_i (initial emissions) are the knowns, so rewrite equation in terms of knowns and unknowns in matrix form:

$$\begin{bmatrix} E_1 \\ E_2 \\ \cdots \\ E_n \end{bmatrix} = \begin{bmatrix} 1-\rho_1 F_{11} & \cdot & \cdot & -\rho_1 F_{1n} \\ -\rho_2 F_{21} & \cdot & \cdot & -\rho_2 F_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -\rho_n F_{n1} & \cdot & \cdot & 1-\rho_n F_{nn} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ \cdots \\ B_n \end{bmatrix}$$

More problems than solutions

- matrix is $O(n^2)$ in size, where n, the number of patches, is usually much larger than number of polygons.
- restricted to closed polyhedral environments.
- no refraction, specularity, translucency.
- form-factor computation is very costly, and must be redone whenever a single object moves (but not in principle when the camera moves).
- how are patches chosen?
- how to scale up geometric complexity?
- extension to curved surfaces?
- how to reconstruct continuous tone images from constant-radiosity patches.
- costly visibility computations.
- vast amounts of aliasing.
- fast shadows?

But solution is view independent (?), and independent of initial emissions E_i.

Partial Solutions

Matrix Size

1. Use iterative methods to solve system row-wise (i.e., for each B_i). Corresponds to *gathering* radiosity from other patches.

2. Still costly, so look at system in another way.

   Think instead about distributing radiosity from B_i out to other patches in proportion to form factor.

   In fact, this turns out to solve the system column-wise (incrementally), because for each iteration, i is fixed, corresponding to a column of the radiosity matrix.

   Corresponds to *shooting* radiosity.

   After each iteration, a valid (though incomplete) image can be rendered, so this technique is called *progressive refinement*.

   A full radiosity matrix is not needed (only the current column, assuming form factors are cheap to compute).

3. Hierarchical Approaches. Observe that not all patches (via form factors and radiosity matrix) affect each other at the same scale.

   Create a hierarchy of interactions by grouping together patches with respect to a scale-preserving distance criterion.

   Related to clustering algorithms for n-body systems in applied physics.

   Can have $O(n)$ convergence.

Form Factor Computations

1. Use a projection technique called the *Nusselt analogue*. Use it to discretise the hemisphere into a *hemicube*. Can have substantial approximation error.

2. Analytic form factor techniques for simple polygons. Mathematically very nice, especially for a reference model, but the functions involved are quite expensive to compute (complex dilogarithm).

3. Ray tracing approaches: used differential form factor computations (area/point to point) to approximate area-area form factors.

4. Hierarchical radiosity techniques reduce the number of form-factors to compute at any time.

Optical Phenomena Beyond Diffuse Reflection

1. Some work on participating media: light interacts with environment (e.g., dust, smoke, mist) producing scattering effects.

2. Multipass Algorithms: Do a first-pass radiosity step and use the result like an environment map in a subsequent specular (ray-tracing) pass.

Decomposition into Patches–Meshing

- extremely difficult problem.
- originally, meshing was done by hand and not talked about.
- trade-off between view-dependence and the creation of meshes that were far to fine.
- current approaches now involve tracking *discontinuities* in radiance function due to shadow boundaries and polygon edges. *Very* active research area. (See video).
- adaptive meshing, mesh filtering to reduce mesh size.

Aliasing and Approximation Error

- *constant radiosity syndrome* has now been cured.
- patch elements are now no longer assumed to carry constant radiosity. In finite element terminology, constant radiosity assumption means that order 1 elements are used.
- now, patches are assumed to carry variable radiosity. A polynomial basis (usually degree 1 or 2 lagrange basis) is imposed over patch, permitting subsampling of radiosity over patch.
- improves the approximation, avoids *ad hoc* reconstruction techniques, and avoids some aliasing error.

Shadow Computations for Area Light Sources

- possibly the hardest problem of them all.
- needed for discontinuity meshing.
- extensive geometric computations.
- use of *aspect graphs* to accelerate computation. An aspect graph characterises the parts of a scene that have topologically similar ‘‘view’’ of an area light source.
- can exploit this structure to ‘‘scan convert’’ a penumbral shadow.
- most substantial progress in this area is by U of Toronto!

Most Recent Work at U of Toronto

- accelerated shadow computations (two papers at *SIGGRAPH '94*).
- structured sampling techniques for occluded and unoccluded environments (award winning paper at *Eurographics '93*).
- nonuniform radiant intensity distributions for area light sources (*CVGIP GMIP '93*).
- illumination in the presence of participating media (*SIGGRAPH '93,95*).
- discontinuity meshing (*Graphics Interface '86, TVCG 1997*).
- incremental visibility.

Cast of researchers on illumination at UofT includes:

Eugene Fiume
Sherif Ghali
Marc Ouellette
Michiel van de Panne
James Stewart
Jeff Tupper

References

The last word on heat transfer (including light):

Robert Siegel and John Howell, *Thermal Radiation Heat Transfer*, Third Edition, Hemisphere Publishing Corporation, 1992.

One recent book on this material (despite the backward title):

Michael Cohen and John Wallace, *Radiosity and Realistic Image Synthesis*, Academic Press, 1993.

A more unified and clearer treatment:

Francois Sillion and Claude Puech, *Radiosity and Global Illumination*, Morgan Kaufmann, 1994.

A modern treatment of physically based light transport:

Matt Pharr and Greg Humphreys *Physically Based Rendering: From Theory to Implementation*, Second Edition, Morgan Kaufmann, 2010.

And of course the basic literature.

Discussion/Open Questions

- Is the radiosity-based technique worth the trouble?
- What does a physically-plausible illumination model add to my animated toothpaste commercial?
- Do shadows really need to be computed precisely?