Illumination and Reflection

Computer Science 294-13: Advanced Computer Graphics

Lecture 2

2 September 2009

Scribed by Albert Liu

1 The Physical Nature of Light

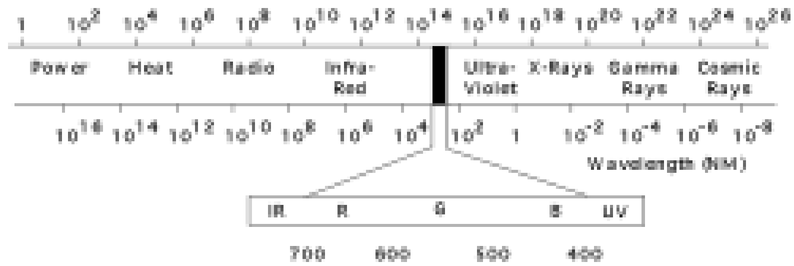


Figure 1: The Electromagnetic Spectrum

Physically speaking, light is visible electromagnetic radiation. While electromagnetic radiation spans from microwaves on the long-wave, low photon energy side of the spectrum to short-wave, high energy gamma rays[1] on the other side, the visible portion of the spectrum spans less than an order of magnitude in wavelength, from about 400 to 700 nanometers. The behavior of real electromagnetic waves has some nuances, including the following:

- **Polarization.** In reality, an electromagnetic wave can be considered as the sum of several polarized components, where the electric and magnetic fields of each component are aligned in a different way. Reflections and certain types of materials may affect the polarization of light and/or may only admit light of a certain polarization.

- **Quantization.** Real electromagnetic waves have a "dual wave-particle" nature, where it behaves like a quantized particle in some respects and like a wave in others. On the particle side, light behaves as a set of packets called photons; as such, light is not a truly continuous stream of energy but is rather a stream of discrete photons.

- **Wave Nature of Light.** On the other hand, electromagnetic waves also have a wave nature (hence their name). A wave passing through an aperture will produce a diffraction pattern instead of carrying energy uniformly. Multiple waves in through the same space will interfere with each other. This occurs despite the quantization of photons–in fact, even a single photon will interfere with itself! More subtly, there is no such thing as a perfectly collimated beam; even laser beams spread out slowly with distance.

In computer graphics, we typically ignore these factors, since they are not usually noticable to the naked eye.

[1]Scribe's note: The diagram in the lecture slides depicts "cosmic rays" at the short-wave end of the electromagnetic spectrum. In reality, what are referred to as cosmic rays are not electromagnetic radiation, but rather high-energy subatomic particles, e.g. alpha particles. Source: Allkofer, Otto Claus. *Introduction to Cosmic Radiation*, Verlag Karl Thiemig, Munchen (1975). Page 39.

Instead, we treat light waves as simple, continuous, independent streams of energy, and rays as perfectly collimated, infinitesimally narrow beams of energy.

2 Mathematical Background

2.1 Solid Angles

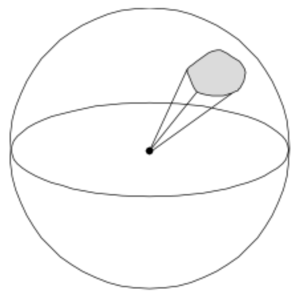


Figure 2: Solid Angles

Before we can have any discussion of radiometry, some mathematical background must be developed. You are probably familar with the concept of angles and arcs. In fact, these two are related by the formula $\theta = l/r$, where θ is the angle, l is the arc length, and r is the radius of the circle. Since a circle has a circumference of $2\pi r$, there are 2π radians in a circle. Technically the radian is dimensionless; however, it is often convenient to carry the label around in order to make sure your answer has the proper dimensions.

In radiometry, we are interested in solid angles. A solid angle is to a "normal" angle as a piece of sky is to a slice of pie; it is like an angle, but one dimension higher. Solid angles are measured in steradians (sr); like radians, the steradian is dimensionless. In analogy to angles, a solid angle is related to the area it occupies on the surface of a sphere by the formula $\Omega = A/r^2$, where Ω is the solid angle, A is the area, and r is the radius of the sphere. Since a sphere has a surface area of $4\pi r^2$, there are 4π steradians in a sphere.

2.2 Spherical Coordinates

Now we can specify how large a solid angle is, but how do we specify *where* a solid angle is? The answer is spherical coordinates. Two angles are sufficient to describe a direction in spherical coordinates. The first is the azimuthal angle, typically denoted by ϕ, describes the direction's angle about the z axis, where the positive x direction is at zero angle and the positive y direction is at an angle of $\pi/2$. In terms of geography, this corresponds to longitude. The second is the angle of elevation, typically denotes by θ; this is the angle between the direction and the positive z direction. This corresponds to 90 degrees minus the latitude, since

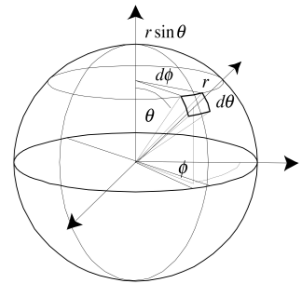


Figure 3: Spherical Coordinates, Differential Solid Angles

the zero is at the north pole and the angle increases toward the south pole, rather than starting at the equator and increasing toward the north pole. If we are also concerned with distance, we use the radius r to describe the distance from the origin.

We can convert between the Cartesian coordinates x, y, z and the spherical coordinates ϕ, θ, r using the following equations:

$$x = r\sin\theta\cos\phi \qquad \phi = \arctan\frac{y}{x} \tag{1}$$

$$y = r\sin\theta\sin\phi \qquad \theta = \arctan\frac{\sqrt{x^2+y^2}}{z} \tag{2}$$

$$z = r\cos\theta \qquad r = \sqrt{x^2+y^2+z^2} \tag{3}$$

$$\tag{4}$$

2.3 Differential Solid Angles

Just as we can perform calculus using angles, we can also perform calculus using solid angles. In computer graphics this will be mostly integration.

Consider a four-sided patch of area on a sphere, where two opposite sides have constant ϕ and the other two have constant θ. In the limit that the patch is a infinitesimally tiny area dA, the patch can be considered flat and its sides straight, so its area is equal to the product of the lengths of two adjacent sides. The sides with constant ϕ have length $r\ d\theta$, being simple arcs on a great circle of radius r. Meanwhile, the sides with constant θ have length $r\sin\theta\ d\phi$, since the circle of which they are arcs of has radius dependent on where it is vertically on the sphere. Multiplying these together, we find that

$$dA = r^2\sin\theta\ d\phi\ d\theta \tag{5}$$

$$= -r^2\ d\phi\ d(\cos\theta) \tag{6}$$

As for limits, to integrate over an entire sphere, we let θ range from 0 to π, and ϕ from 0 to 2π. Integrating dA over a sphere gives a surface area of $4\pi r^2$.

3 Radiometric Quantities

With this mathematical background in place, we can move on to a discussion of radiometry. Radiometry is the physical measurement of electromagnetic energy. We will begin with radiance and irradiance, which describe the flow of electromagnetic energy, then move on to reflection functions, which describe how a surface reflects incident electromagnetic energy.

3.1 Radiance

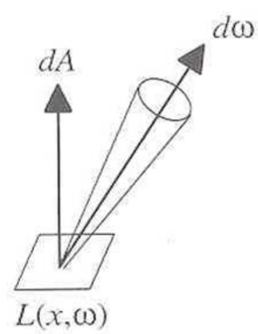


Figure 4: Radiance

Radiance is defined as power per unit projected area perpendicular to the ray per unit solid angle in the direction of the ray. In SI units, radiance is measured in watts per meter squared per steradian ($\mathrm{W/m^2 \cdot sr}$). It is generally denoted by the symbol $L(\vec{x}, \omega)$, being a function of position and solid angle. It is related to the flux (power per unit area) by the equation $d\Phi = L(\vec{x}, \omega) \cos\theta \; d\omega \; dA$. Intuitively, the radiance describes the electromagnetic power coming from a particular direction at a given point.

Since we typically treat rays as perfectly collimated beams of energy, the radiance is constant along a ray.

3.2 Irradiance, Radiant Exitance

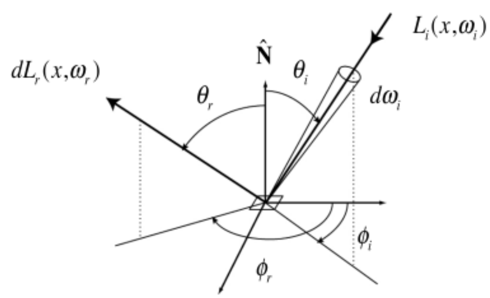
*Ir*radiance is a measure of the total power falling on a unit area of a surface. It is usually denoted by the symbol E. In SI units, irradiance is measured in watts per meter squared per steradian ($\mathrm{W/m^2}$). To determine the irradiance falling on a point on a given surface, we integrate the radiance times the obliquity factor $\cos\theta$ over the visible hemisphere:

$$E = \frac{d\Phi}{dA} = \int_{H^2} L\left(\vec{x}, \omega\right) \cos\theta \; d\omega \tag{7}$$

For uniform illumination of unity radiance, this comes out to an irradiance of π.

A related quantity is radiant exitance; this is exactly like irradiance, except it describes the total power *leaving* the surface.

4 Bidirectional Reflectance Distribution Functions (BRDFs)



$$f_r(\omega_i \to \omega_r) \equiv \frac{dL_r(\omega_i \to \omega_r)}{dE_i} \quad \left[\frac{1}{sr}\right]$$

Figure 5: Bidirectional Reflectance Distribution Functions (BRDFs)

Since we are concerned with the rendering of surfaces, we are concerned with how surfaces reflect light energy. One way of representing this is though a bidirectional reflectance distribution functions (BRDF). The BRDF operates as follows. Given:

- A point on a surface.
- An incident direction of light.
- A reflected direction of interest.

the BRDF describes how much radiance is reflected in the direction of interest per unit of power falling on a unit area (including the obliquity factor) from the incident direction.

The BRDF is denoted in the lecture slides by the symbol $f\left(\omega_i \to \omega_r\right)$, where the ωs indicate solid angles, which can be represented by the angle pairs θ, ϕ. It is defined as the reflected radiance reflected direction divided by the irradiance coming from the incident direction; symbolically, this is

$$f(\omega_i \to \omega_r) = \frac{dL_r(\omega_r)}{L_i(\omega_i)\cos\theta_i \; d\omega_i} \tag{8}$$

Since the projected solid angle $\cos\theta_i \; d\omega_i$ is only applied to the incident term, the BRDF has units of inverse steradians.

4.1 Properties of BRDFs

BRDFs can have several properties, which may reflect the physical properties of light, or may make them easier to store and calculate.

- **Linearity.** There are two properties of linearity here. The first is the property that the light reflected from a surface from two incident light distributions together is the same as the sum of the light reflected from the two incident light distributions separately. This allows us to treat rays as separate and independent, and simply sum the results afterwards.

  The second property of linearity is that a BRDF can be split up into the sum of several components; the total reflected light is the sum of the light reflected by each component. Linearity is implicit in our use of the BRDFs; thus all BRDFs have this property.

- **Reciprocity Principle.** This principle states that the BRDF is invariant with time reversal; that is, the value of f remains the same if ω_i and ω_r are swapped. This is reasonable as light is physically invariant with time reversal; after, it is this property that we are taking advantage of when we trace rays from the eye back to the source instead of from source to eye.

- **Isotropic vs. Anisotropic.** The reflectance of most materials can be considered as isotropic; that is, f's dependence on ϕ only depends on the difference between ϕ_i and ϕ_r. Combined with reciprocity, f's dependence on ϕ only depends on absolute value of this difference. However, not all materials have isotropic reflectance; for example, brushed aluminum's appearance depends on the azimuthal viewing angle. Symbolically, an isotropic BRDF $f(\theta_i, \phi_i, \theta_r, \phi_r)$ can be simplified to $f(\theta_i, \theta_r, |\phi_i - \phi_r|)$.

- **Energy Conservation.** Physically, a surface cannot reflect more light than falls on it; that is, the radiant exitance cannot exceed the irradiance. We can express this mathematically as

$$\frac{E_r}{E_i} = \frac{\int_{H^2} L(\vec{x}, \omega_r)\cos\theta_r \; d\omega_r}{\int_{H^2} L(\vec{x}, \omega_i)\cos\theta_i \; d\omega_i} \tag{9}$$

$$= \frac{\int_{H^2}\int_{H^2} f(\omega_i \to \omega_r) L_i(\omega_i)\cos\theta_i \cos\theta_r \; d\omega_i \; d\omega_r}{\int_{H^2} L(\vec{x}, \omega_i)\cos\theta_i \; d\omega_i} \leq 1 \tag{10}$$

where we have used our BRDF defintion (rearranged)

$$f(\omega_i \to \omega_r) L_i(\omega_i) \cos\theta_i \; d\omega_i = dL_r(\omega_r) \tag{11}$$

and integrated to substitute for $L(\vec{x}, \omega_r)$.

4.2 The Reflection Equation

The linearity of the BRDF allows to express the total light reflected from a surface as the sum (or integral) of the contributions of the reflections of several lights. This gives rise to the reflection equation:

$$L_r(\omega_r) = \sum_i L_i(\omega_i) f(\omega_i \to \omega_r) \cos\theta_i \tag{12}$$

for point lights. For continuous lights, this is

$$L_r(\omega_r) = \int_\Omega L_i(\omega_i) f(\omega_i \to \omega_r) \cos\theta_i \; d\omega_i \tag{13}$$

This forms the core of rendering using BRDFs.

4.3 Reflection Functions

4.3.1 Simple Functions, The Phong Model

Some of the simplest reflection functions are as follows:

- **Mirror Reflection.** Incoming light is simply reflected across the normal. Mathematically,

$$f(\omega_i \to \omega_r) \propto \frac{\delta(\cos\theta_i - \cos\theta_r)}{\cos\theta_r} \delta((\phi_i - \phi_r + \pi) \bmod 2\pi) \tag{14}$$

  where the constant of proportionality is between 0 and 1.

- **Diffuse (Lambertian) Reflection.** Incoming light is reflected uniformly in all directions. Mathematically, this means

$$f(\omega_i \to \omega_r) = \frac{\rho_d}{\pi} \tag{15}$$

  ρ_d is known as the albedo; physically, it must be between 0 and 1.

- **Specular Reflection.** Incoming light is reflected primarily in the mirror direction. Typically the strength of the reflection relative to the mirror direction is equal to the the cosine relative to the mirror direction raised to some power:

$$f\left(\omega_i \to \omega_r\right) \propto \cos^p\left(\omega_{i,\text{reflected}}, \omega_r\right) \tag{16}$$

The Phong model combines the diffuse and specular functions. While not very sophisticated, it is relatively easy to calculate.

4.3.2 Fresnel Reflectance

Physically, the angle at which light strikes a surface influences how much is reflected (and in what polarization, but we are generally not concerned with that). This is especially dramatic for dielectric (non-conducting) materials at shallow angles. While the full Fresnel equations that describe reflectance can become somewhat involved, the Schlick approximation provides a decent estimate:

$$F\left(\theta\right) = F\left(0\right) + \left(1 - F\left(0\right)\right)\left(1 - \cos\theta\right)^5 \tag{17}$$

4.3.3 The Torrance-Sparrow Model

A more physically-based model is the Torrance-Sparrow model. It models a surface as having infinitesimally tiny, perfectly reflecting facets, and considers how light is reflected from it. It considers the fact that light coming from a particular direction may not reach all parts of the surface (shadowing), light reflected off some parts of the surface may not be able to reach the viewer (masking), and light may undergo multiple reflections before reaching the viewer (interreflection). The result is as follows:

$$f = \frac{F\left(\theta_i'\right) G\left(\omega_i, \omega_r\right) D\left(\theta_h\right)}{4\cos\theta_i \cos\theta_r} \tag{18}$$

The terms are as follows:

- $F\left(\theta_i'\right)$ is the Fresnel term, as per the previous heading. θ_i' is the angle between the incident direction and the normal of a facet aligned to reflect the light toward the viewer.

- $G\left(\omega_i, \omega_r\right)$ is the geometric attenuation factor. It models the amount of light that is blocked by shadowing and masking.

- $D\left(\theta_h\right)$ is the distribution function. It models how many of the facets are aligned to reflect light toward the viewer. θ_h is the half-angle between the incident and reflected direction.

- The denominator is simply the obliquity factor.

4.3.4 Other Models

Other models include empirical models, which build a 4-D table from real-world data, and non-photorealistic BRDFs; for example, cartoon shaders.

5 Environment Maps

Figure 6: Environment Maps (Blinn and Newell 1976, Miller and Hoffman, 1984 Later, Greene 86, Cabral et al. 87)

In the limit that the light sources in a scene are infinitely far away (or that one is willing to use this as an approximation), a technique called environment mapping can be used. An environment map is simply a representation of the light coming in at every angle at a particular point in a scene. For a real-world scene, an environment map can be measured by photographing a mirror-reflective sphere. With this data can simply index the mirror angle into the environment map to get the appropriate appearance for a reflective surface. If the surface is not at the same point as where the environment map was constructed, the reflection will not be perfectly accurate, but often the environment map produces a reasonable result without requiring too much computation.