Contents

Course Notes:

The following papers (in order) are also included:

Papers are included with permission of the authors.

Acknowledgements.

Notes on Basic Radiometry and BRDF Models

Peter Shirley

1   Introduction

In these notes we discuss the practical issues of measuring light, an endeavor usually called *radiometry*, and the basics of BRDF models. These notes are derived from the draft for *Computer Graphics* under prparation with AK Peters.

The terms that arise in radiometry may at first seem strange and have terminology and notation that may be hard to keep straight. However, because radiometry is so fundamental to computer graphics it is worth studying radiometry until it sinks in. This chapter also covers *photometry*, which takes radiometric quantities and scales them to estimate how much "useful" light is present. For example, a green light may seem twice as bright as a blue light of the same power because the eye is more sensitive to green light. Photometry attempts to quantify such distinctions.

The reflective properties of a surface can be summarized using the BRDF. We then discuss some of the most visually important aspects of material properties, and a few models that are useful in capturing these properties while remaining fairly simple.

2   Radiometry

Although we can define radiometric units in many systems, we use we use the *SI* units. Familiar SI units include the familiar metric units of meter (m) and gram (g). Light is fundamentally a propagating form of energy, so it is useful to define the SI unit of energy which is the Joule (J).

2.1   Photons

To aid intuition, we will describe radiometry in terms of collections of large numbers of *photons*, and this section establishes what is meant by a photon on this context. For the purposes of this chapter, a photon is a quantum of light that has a position, direction of propagation, and a wavelength λ. Somewhat strangely, the SI unit used for wavelength is *nanometer* (nm). This is mainly for historical reasons, and $1nm = 10^{-9}m$. Another unit, the *angstrom* is sometimes used, and one nanometer is ten angstroms. A photon also has a speed c that depends only on the refractive index n of the medium it propagates through. Sometimes the frequency $f = c/\lambda$ is also used for light. This is convenient because unlike λ and c, f does not change when the photon refracts into a medium with a new refractive index. Another invariant is the amount of energy q carried by a photon, which is given by the following relationship:

$$q = hf = \frac{hc}{\lambda}, \tag{1}$$

where $h = 6.63 \times 10^{-34} Js^{-1}$ is Plank's Constant. Although these quantities can be measured any many unit systems, we use *SI* units whenever possible.

2.2 Spectral Energy

If we have a large collection of photons, their total energy Q can be computed by summing the energy q_i of each photon. A reasonable question to ask is "how is the energy distributed across wavelengths?" An easy way to answer this is to partition the photons into bins, essentially histogramming them. You would then have an energy associated with an interval. For example, you might count all the energy between $\lambda = 500nm$ and $\lambda = 600nm$, and have it turn out to be $10.2J$ and this might be denoted $q[500, 600] = 10.2$. If we divided the wavelength interval into two $50nm$ intervals, we might find that $q[500, 550] = 5.2$ and $q[550, 600] = 5.0$. This tells us there was a little more energy in the short wavelength half of the interval $[500, 600]$. If we divide into $25nm$ bins we might find $q[500, 525] = 2.5$ and so on. The nice thing about the system is that it is straightforward. The bad thing about it is that the choice of the interval size determines the number.

A more commonly used system is to divide the energy by the size of the interval. So instead of $q[500, 600] = 10.2$ we would have:

$$Q_\lambda[500, 600] = \frac{10.2}{100} = 0.12 J(nm)^{-1}$$

This is nice because the size of the interval has much less impact on the overall size of the numbers. An immediate idea would be to drive the interval size $\Delta\lambda$ to zero. This would be awkward because for a sufficiently small $\Delta\lambda$, Q_λ will either be zero or huge depending on whether there is a single photon or no photon in the interval. There are two schools of thought to solve that dilemma. The first is to assume that $\Delta\lambda$ is small, but not so small that the quantum nature of light comes into play. The second is to assume that the light is a continuum rather than individual photons, so a true derivative $dQ/d\lambda$ is appropriate. Both ways of thinking about it are appropriate and lead to the same computational machinery. In practice it seems that most people who measure light prefer small but finite intervals because that is what they can measure in the lab. Most people that do theory or computation prefer infinitesimal intervals because that makes the machinery of calculus available.

The quantity Q_λ is called *spectral energy* and is an *intensive* quantity. This is opposed to an *extensive* quantity such as energy, length, or mass. Intensive quantities can be thought of as density functions that tell you the density of an extensive quantity at an infinitesimal point. For example, the energy Q at a specific wavelength is probably zero, but the spectral energy (energy density) Q_λ is a meaningful quantity. A probably more familiar example is that the population of a country may be 25 million, but the population at a point in that country is meaningless. However, the population *density* measured in people per square meter is meaningful provided it is measured over large enough areas. Much like with photons, population density works best if we pretend that we can view population as a continuum where population density never becomes granular even when the area is small.

We will follow the convention of graphics where spectral energy is almost always used, and energy is rarely used. This results in a proliferation of λ subscripts if "proper" notation is used. Instead we will drop the subscript and use Q to denote spectral energy. This can result in some confusion when people outside of graphics read graphics papers, so be aware of this standards issue. Your intuition about spectral power might be aided

by imagining a measurement device with an energy sensor that measures light energy q. If you place a colored filter in front of the sensor that allows only light in the interval $[\lambda - \Delta\lambda/2, \lambda + \Delta\lambda/2]$, then the spectral power at λ would be $Q = \Delta q/\Delta\lambda$.

2.3   Power

Its useful to estimate a rate of energy production for light sources. This rate is called *power* and is measured in *Watts W* which is another name for *Joules per second*. This is easiest to understand in a *steady state*, but because power is an intensive quantity (a density over time) it is well defined even when energy production is varying over time. The units of power may be more familiar because of lights, e.g. a 100 watt light bulb. Such bulbs draw approximately $100J$ of energy each second. The power of the light produced will be lower than $100W$ because of heat loss, etc., but that can still be used to help intuition. For example, we can get a feel for how many photons are produced in second by a $100W$ light. Suppose the average photon produced has the energy of a $\lambda = 500nm$ photon. The frequency of such a photon is:

$$f = \frac{c}{\lambda} = \frac{3 \times 10^8 ms^{-1}}{500 \times 10^{-9}m} = 6 \times 10^{14}s^{-1}.$$

The energy of that photon is $hf \approx 4 \times 10^{-19}J$. That means a staggering 10^{20} photons are produced each second even if the bulb is not very efficient. This explains why simulating a camera with a fast shutter speed and directly simulated photons is an inefficient choice for producing images.

Like with energy, we are really interested in *spectral power* measured in $W(nm)^{-1}$. Again, although the formal standard symbol for spectral power is Φ_λ, we will use Φ with no subscript for convenience and consistency with most of the graphics literature. One thing to note is that the spectral power for a light source is usually a smaller number than the power. For example, if a light emits a power of $100W$ evenly distributed over wavelengths $400nm$ to $800nm$, then the spectral power will be $100W/400nm = .25W(nm)^{-1}$. This is something to keep in mind if you set the spectral power of light sources by hand for debugging purposes.

The measurement device for spectral energy in the last section could be modified by taking a reading with a shutter that is open for a time interval Δt centered at time t. The spectral power would then be $\Delta Q/(\Delta t \Delta\lambda)$.

2.4   Irradiance

The quantity *irradiance* arises naturally if you ask the question "how much light hits this point?". Of course the answer is "none" and again we must use a density function. If the point is on a surface, it is natural to use area to define our density function. We modify the device from the last section to have a finite ΔA area sensor that is smaller than the light field being measured. The spectral irradiance H would just be the power per unit area $\Delta\Phi/\Delta A$. Fully expanded this is:

$$\Phi = \frac{\Delta q}{\Delta A \, \Delta t \Delta\lambda} \tag{2}$$

Thus the full units of irradiance are $Jm^{-2}s^{-1}(nm)^{-1}$. Note that it is *not* $Jm^{-3}s$. This is a perversity introduced by using square meters for area and nanometers for wavelength, but it is standard radiometry.
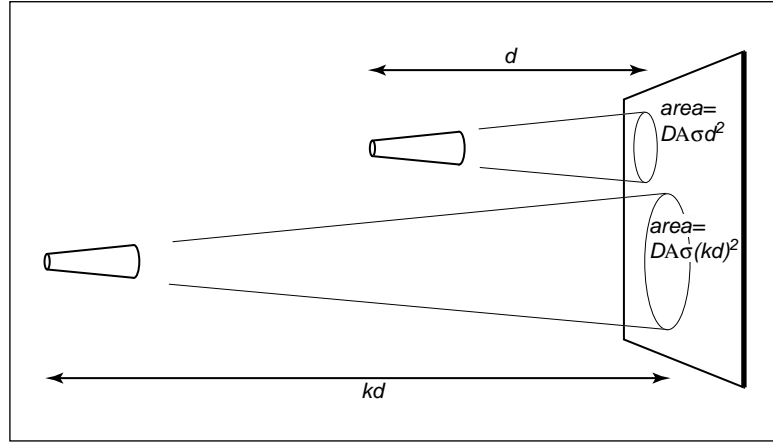
Figure 2 The signal a radiance detector receives does not depend on distance to the surface being measured. This figure assumes the detectors are pointing at areas on the surface that are emitting light in the same way.

When the light is leaving a surface, e.g., when it is reflected, the same quantity as irradiance is called *radiant exitance E*. It is useful to have different words for incident and exitant light because the same point has potentially different irradiance and radiant exitance.
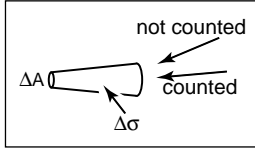
2.5 Radiance

Figure 1 By adding a blinder that shows only a small solid angle $\Delta\sigma$ to the irradiance detector, we measure radiance.

Although radiant exitance tells us how much light is arriving at a point, it tells us little about the direction that light comes from. To measure something analogous to what we see with our eyes, we need to be able to associate "how much light" with a specific direction. We can imagine a simple device to measure such a quantity (Figure 1). We use a small irradiance meter and add on conical "baffler" which limits light hitting the counter to a range of angles with solid angle $\Delta\sigma$. The response of the detector is thus:

$$\text{response} = \frac{\Delta H}{\Delta\sigma}$$
$$= \frac{\Delta q}{\Delta A \; \Delta\sigma \; \Delta t \; \Delta\lambda}$$

This is the spectral *radiance* of light travelling in space. Again, we will drop the "spectral" in our discussion and assume that it is implicit.

Radiance is what we are usually computing in graphics programs. A wonderful property of radiance is that it does not vary along a line in space. To see why this is true, examine the two radiance both looking at a surface as shown in Figure 2. Assume the lines the detectors are looking along are close enough together that the surface is emitting/reflecting light "the same" in both of the areas being measured. Because the area of the surface being sampled is proportional to squared distance, and because the light reaching the detector is *inversely* proportional to squared distance, the two detectors should have the same reading.

It is useful to measure the radiance hitting a surface. We can think of placing the cone baffler from the radiance detector at a point on the surface and measuring the irradiance H on the surface originating from directions in the cone (Figure 3). Note

4

that the surface "detector" is not aligned with the cone. For this reason we need to add a cosine correction term to our definition of radiance:

$$\text{response} = \frac{\Delta H}{\Delta \sigma \cos \theta}$$
$$= \frac{\Delta q}{\Delta A \cos \theta \; \Delta \sigma \; \Delta t \; \Delta \lambda}$$

As with irradiance and radiant exitance, it is useful to distinguish between radiance incident at a point on a surface and exitant from that point. Terms for these concepts sometimes used in the graphics literature are *surface radiance* L_s for the radiance of (leaving) a surface, and *field radiance* for the radiance incident at a surface. Both have the cosine term because they both correspond to the configuration in Figure 3:

$$L_s = \frac{\Delta E}{\Delta \sigma \cos \theta}$$
$$L_f = \frac{\Delta H}{\Delta \sigma \cos \theta}$$

2.5.1 Radiance and Other Radiometric Quantities

If we have a surface whose field radiance is L_f, then we can derive all of the other radiometric quantities from it. This is one reason radiance is considered the "fundamental" radiometric quantity. For example, the irradiance is:

$$H = \int_{\text{all } \omega} L_f(\omega) \; \cos \theta \; d\sigma$$

This formula has several notational conventions that are common in graphics that make such formulae opaque to readers not familiar with them (Figure 4). First, ω is an incident direction, and can be thought of as a unit vector, a direction, or a (θ, ϕ) pair in spherical coordinates with respect to the surface normal. The direction has a differential solid angle $d\omega$ associated with it. The field radiance is potentially different for every direction, so we write it as a function $L(\omega)$.

As an example, we can compute the irradiance H at a surface that has constant field radiance L_f in all directions. To integrate, we use a classic spherical coordinate system, and recall that the differential solid angle is

$$d\sigma \equiv \sin \theta \; d\theta \; d\phi$$

so the irradiance is:

$$H = \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\frac{\pi}{2}} L_f \; \sin \theta \; d\theta \; d\phi$$
$$= \pi L_f$$

This relation shows us our first occurrence of a potentially surprising constant π. These factors of π occur frequently in radiometry and are an artifact of how we chose to measure solid angles, i.e., the area of a unit sphere is a multiple of π rather than a multiple of one.

Similarly, we can find the power hitting a surface by integrating the irradiance across the surface area:

$$\Phi = \int_{\text{all } \mathbf{x}} H(\mathbf{x}) dA,$$
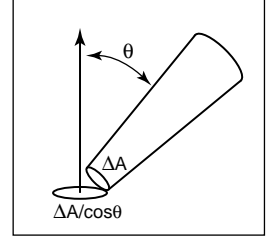
Figure 3 The irradiance at the surface as masked by the cone is smaller than that measured at the detector by a cosine factor.
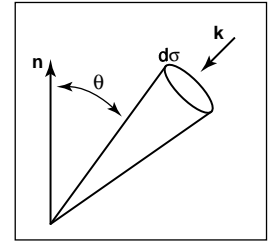
Figure 4 The direction ω has a differential solid angle $d\sigma$ associated with it.

Figure 5 A simple measurement device for directional reflectance. The positions of light and detector are moved to each possible pair of directions.

where $\mathbf{x}$ is a point on the surface, and dA is the differential area associated with that point. Note that we don't have special terms of symbols for incoming versus outgoing power. That distinction does not seem to come up enough to have encouraged the distinction.

2.6 BRDF

Because we are interested in surface appearance, we would like to characterize how a surface reflects light. At an intuitive level, for any incident light coming from direction $\mathbf{k}_i$, there is some fraction scattered in a small solid angle near outgoing direction $\mathbf{k}_o$. There are many ways we could formalize such a concept, and not surprisingly the standard way to do so is inspired by building a simple measurement device. Such a device is shown in Figure 5, where a small light source is positioned in direction $\mathbf{k}_i$ as seen from a point on a surface, and a detector is placed in direction $\mathbf{k}_o$. For every directional pair $(\mathbf{k}_i, \mathbf{k}_o)$ we take a reading with the detector.

Now we just have to decide how to measure the strength of the light source and make our reflection function independent of this strength. For example, if we replaced the light with a brighter light, we would not want to think of the surface as reflecting light differently. We could place a radiance meter at the point being illuminated to measure the light. However, for this to get an accurate reading that would not depend on the $\Delta\sigma$ of the detector, we would need the light to subtend a solid angle bigger than $\Delta\sigma$. Unfortunately, the measurement taken by our roving radiance detector in direction $\mathbf{k}_o$ will also count light that comes from points outside the new detector's cone. So this does not seem like a practical solution.

Alternatively, we can place an irradiance meter at the point on the surface being measured. This will take a reading that does not depend strongly on subtleties of the light source geometry. This suggests characterizing reflectance as a ratio:

$$\rho = \frac{L_s}{H}$$

where this fraction ρ will vary with incident and exitant directions $\mathbf{k}_i$ and $\mathbf{k}_o$, H is the irradiance for light position $\mathbf{k}_i$ and L_s is the surface radiance measured in direction $\mathbf{k}_o$. If we take such a measurement for all direction pairs, we end up with a 4D function $\rho(\mathbf{k}_i, \mathbf{k}_o)$. This function is called the *bidirectional reflectance distribution function*

(BRDF). The BRDF is all we need to know to characterize the directional properties of how a surface reflects light.

2.6.1 Directional Hemispherical Reflectance

Given a BRDF it is straightforward to ask "what fraction of incident light is reflected?" However, the answer is not so easy; the fraction reflected depends on the directional distribution of incoming light. For this reason we typically only set a fraction reflected for a fixed incident direction $\mathbf{k}_i$. This fraction is called the *directional hemispherical reflectance*. This fraction, $R(\mathbf{k}_i)$ is defined:

$$R(\mathbf{k}_i) = \frac{\text{power in all outgoing directions } \mathbf{k}_o}{\text{power in a beam from direction } \mathbf{k}_i}.$$

Note that this quantity is between zero and one for reasons of energy conservation. If we allow the incident Φ_i power to hit on a small area ΔA, then the irradiance is $\Phi_i / \Delta A$. Also, the ratio of the incoming power is just the ratio of the radiance exitance to irradiance:

$$R(\mathbf{k}_i) = \frac{E}{H}$$

The radiance in a particular direction resulting from this power is by the definition of BRDF:

$$L(\mathbf{k}_o) = H\rho(\mathbf{k}_i, \mathbf{k}_o)$$
$$= \frac{\Phi_i}{\Delta A}$$

And from the definition of radiance we also have:

$$L(\mathbf{k}_o) = \frac{\Delta E}{\Delta\sigma_o \cos\theta_o}$$

Where E is the radiant exitance of the small patch going in direction $\mathbf{k}_o$. Using these two definitions for radiance we get:

$$H\rho(\mathbf{k}_i, \mathbf{k}_o) = \frac{\Delta E}{\Delta\sigma_o \cos\theta_o}$$

Rearranging terms we get:

$$\frac{\Delta E}{H} = \rho(\mathbf{k}_i, \mathbf{k}_o)\Delta\sigma_o \cos\theta_o$$

This is just the small contribution to E/H that is reflected near the particular $\mathbf{k}_o$. To find the total $R(\mathbf{k}_i)$ we sum over all outgoing $\mathbf{k}_o$. In integral form this is:

$$R(\mathbf{k}_i) = \int_{\text{all } \mathbf{k}_o} \rho(\mathbf{k}_i, \mathbf{k}_o) \cos\theta_o \, d\sigma_o$$

2.6.2 Ideal diffuse BRDF

An idealized diffuse surface is called *Lambertian*. Such surfaces are impossible in nature for thermodynamic reasons, but they do mathematically conserve energy. The

lambertian BRDF has ρ equal to a constant for all angles. This means the surface will have the same radiance for all viewing angles, and this radiance will be proportional to the irradiance.

If we compute $R(\mathbf{k}_i))$ for a a Lambertian surface with $\rho = C$ we get:

$$
\begin{aligned}
R(\omega_i) &= \int_{\text{all } \mathbf{k}_o} C \cos \theta_o \ d\sigma_o \\
&= \int_{\phi_o=0}^{2\pi} \int_{\theta_o=0}^{\pi} k \cos \theta_o \sin \theta_o \ d\theta_o \ d\phi_o \\
&= \pi C
\end{aligned}
$$

Thus for a perfectly reflecting Lambertian surface ($R = 1$) we have $\rho = 1/\pi$ and for a Lambertian surface whose $R(\mathbf{k}_i) = r$ we have:

$$
\rho(\mathbf{k}_i, \mathbf{k}_o) = \frac{r}{\pi}
$$

This is another example where the use of steradians for solid angle determines the normalizing constant and thus introduces factors of π.

3 Transport Equation

With the definition of BRDF, we can describe the radiance of a surface in terms of the incoming radiance from all different directions. Because in computer graphics we can use idealized mathematics that might be impractical to instantiate in the lab, we can also write the BRDF in terms of radiance only. If we take a small part of the light with solid angle $\Delta\sigma_i$ with radiance L_i, and "measure" the refected radiance in direction $\mathbf{k}_o$ due to this small piece of the light, we can compute a BRDF. The irradiance due to the small piece of light is $H = L_i \cos \theta_i \Delta\sigma_i$. Thus the BRDF is:

$$
\rho = \frac{L_o}{L_i \cos \theta_i \Delta\sigma_i}.
$$

That form can be useful in some situations. Rearranging terms we can write down the part of the radiance that is due to light coming from direction $\mathbf{k}_i$:

$$
\Delta L_o = \rho(\mathbf{k}_i, \mathbf{k}_o) L_i \cos \theta_i \Delta\sigma_i
$$

If there is light coming from many directions $L_i(\mathbf{k}_i)$ we can sum all of them. In integral form, with notation for surface and field radiance this is:

$$
L_s(\mathbf{k}_o) = \int_{\text{all } \mathbf{k}_i} \rho(\mathbf{k}_i, \mathbf{k}_o) L_f(\mathbf{k}_i) \cos \theta_i d\sigma_i
$$

This equation is often called the *rendering equation* in computer graphics.

Sometimes it is useful to write the transport equation in terms of surface radiances only. Note that in a closed environment the field radiance $L_f(\mathbf{k}_i)$ comes from some surface with surface radiance $L_s(-\mathbf{k}_i) = L_f(\mathbf{k}_i)$ (Figure 6). The solid angle subtended by the point $\mathbf{x}'$ in the Figure is given by:

$$
\Delta\sigma_i = \frac{\Delta A' \cos \theta'}{\|\mathbf{x} - \mathbf{x}'\|^2}
$$

8

where $\Delta A'$ the the area we associate with $\mathbf{x}'$. Substituting for $\Delta\sigma_i$ in terms of $\Delta A'$ suggests the following transport equation:

$$L_s(\mathbf{x},\mathbf{k}_o) = \int_{\text{all } \mathbf{x}' \text{ visible to } \mathbf{x}} \frac{\rho(\mathbf{k}_i,\mathbf{k}_o)L_s(\mathbf{x}',\mathbf{x}-\mathbf{x}')\cos\theta_i\cos\theta'}{\|\mathbf{x}-\mathbf{x}'\|^2}$$

Note that we are using a non-normalized vector $\mathbf{x}-\mathbf{x}'$ to indicate the direction from $\mathbf{x}'$ to $\mathbf{x}$. Also note that we are writing L_s as a function of position and direction.

The only problem with this new transport equation is that the domain of integration is awkward. If we introduce a visibility function we can trade off complexity in the domain with complexity in the integrand:

$$L_s(\mathbf{x},\mathbf{k}_o) = \int_{\text{all } \mathbf{x}'} \frac{\rho(\mathbf{k}_i,\mathbf{k}_o)L_s(\mathbf{x}',\mathbf{x}-\mathbf{x}')v(\mathbf{x},\mathbf{x}')\cos\theta_i\cos\theta'}{\|\mathbf{x}-\mathbf{x}'\|^2}$$

where

$$v(\mathbf{x},\mathbf{x}') = \begin{cases} 1 & \text{if } \mathbf{x} \text{ and } \mathbf{x}' \text{ are mutually visible} \\ 0 & \text{otherwise} \end{cases}$$

Figure 6 The light coming into one point comes from another point.

4 Photometry

For every spectral radiometric quantity there is a related *photometric quantity* that measures how much of that quantity is "useful" to a human observer. Given a spectral radiometric quantity $f_r(\lambda)$, the related photometric quantity f_p is:

$$f_p = 683\frac{lm/W}{\int} \int_{\lambda=380nm}^{800nm} \bar{y}(\lambda)f_r(\lambda)\,d\lambda,$$

where $\bar{y}$ is the *luminous efficiency function* of the human visual system. This function is zero outside the limits of integration above, so the limits could be 0 and ∞ and f_p would not change. The leading constant is to make the definition consistent with historical absolute photometric quantities.

The luminous efficiency function quantifies the human visual system is not equally sensitive to all wavelengths (Figure 7). For wavelengths below 380nm (the *ultraviolet range*) the light is not visible to humans and thus has a $\bar{y}$ value of zero. From 380nm it gradually increases until $\lambda = 555nm$ where it peaks. This is a pure green light. Then it gradually decreases until it reaches the boundary of the infrared region at 800nm.

The photometric quantity that is most commonly used in graphics is *luminance*, the photometric analog of radiance:

Figure 7 The luminous efficiency function versus wavelength (nm).

$$Y = 683\frac{lm/W}{\int} \int_{\lambda=380nm}^{800nm} \bar{y}(\lambda)L(\lambda)\,d\lambda.$$

Using the symbol Y for luminance is from colorimetry. Most fields use the symbol L, but we will not follow that convention because it is to confusing to use L for both luminance and spectral radiance. Luminance gives one a general idea of how "bright" something is independent of the adaptation of the viewer. Note that the black paper under noonday is subjectively darker than the lower luminance white paper under moonlight, so reading too much into luminance is dangerous, but it is a very useful quantity

for getting a quantitative feel for relative perceivable light output. The units lm stands for lumens. Note that most light bulbs are rated in terms of the power they consume in Watts, and the useful light they produce in lumens. More efficient bulbs produce more of their light where $\bar{y}$ is large and thus produce more lumens per Watt. A "perfect" light would convert all power into $555nm$ light, and would produce 683 lumens per Watt. The units of luminance are thus $(lm/W)(W/(m^2 sr)) = lm/(m^2 sr)$. The quantity one lumen per steradian is defined to be one *candela* (cd), so luminance is usually described in units cd/m^2.

5 Real-world Materials
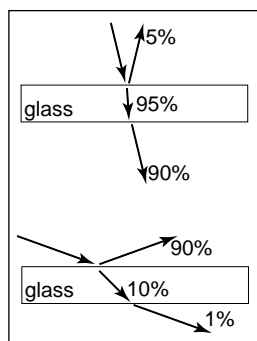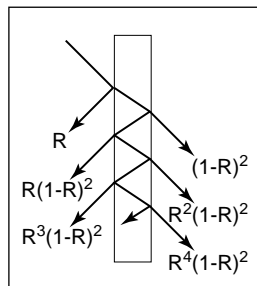
Many real materials have a visible structure at normal viewing distances. For example, most carpets have easily visible pile that contributes to appearance. For our purposes, such structure are not part of the material property but are instead part of the geometric model. Structure whose details are invisible at normal viewing distances that do determine macroscopic material appearance are part of the material property. For example, the fibers in paper have a complex appearance under magnification, but they are blurred together into an homogeneous appearance when viewed at arm's length. This distinction between microstructure that is folded into BRDF is somewhat arbitrary and depends on what one defines as "normal" viewing distance and visual acuity, but the distinction has proven quite useful in practice.

In this section we define some categories of materials. Later in the chapter we present reflection models that target each type of material. In the notes at the end of the chapter some models that account for more exotic materials are also discussed.

5.1 Smooth Dielectrics and Metals

Dielectrics are clear materials that refract light. Metals reflect and refract light much like dielectrics, but they absorb light very very quickly. Thus only very thin metal sheets are transparent at all, e.g., the thin gold plating on some glass objects. For a smooth material there are only two important properties:

1. How much light is reflected at each incident angle and wavelength,

2. What fraction of light is absorbed as it travels through the material for a given distance and wavelength.

The amount of light transmitted is whatever is not reflected because of energy conservation. For a metal, in practice, we can assume all the light is immediately absorbed. For a dielectric, the fraction is determined by the constant used in Beer's Law The amount of light reflected is determined by the *Fresnel Equations*. These equations are straightforward but cumbersome. We will show an approximation later in the chapter that is enough for almost all graphics applications. The main effect of the Fresnel Equations is to increase the reflectance as the incident angle increases, particularly near grazing angles as shown in Figure 8. This effect works for transmitted light as well. This is shown diagrammatically in Figure 9. Note that the light is repeatedly reflected and refracted as shown in Figure 10. Usually only one or two of the reflected images is easily visible.

Figure 9 The amount of light reflected and transmitted by glass varies with angle.

Figure 10 Light is repeatedly reflected and refracted by glass, with the fractions of energy shown.
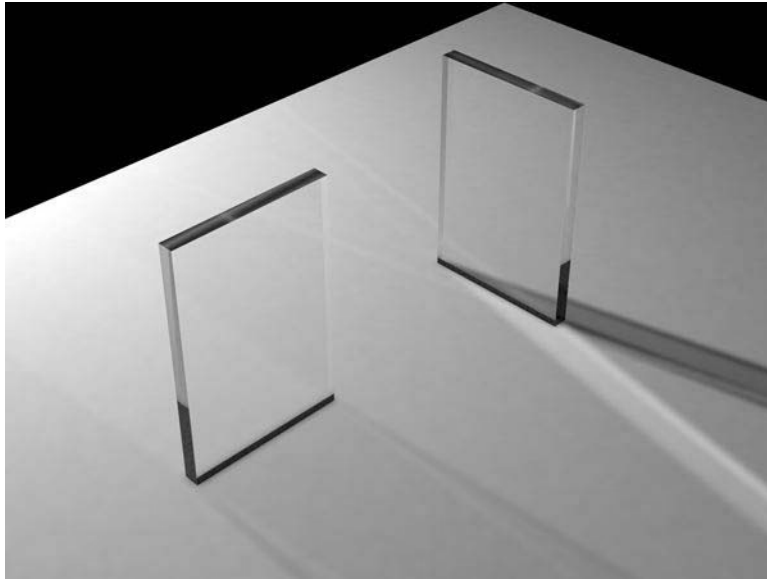
Figure 8 The way the glass reflects and transmits light is shown by the light patterns on the base plane. On the left the light hits near normal incidence and most light is transmitted, so the shadow is light. On the right the angle is more acute so the reflected light paints the base plane. Because this light is reflected, there is less light transmitted and the shadow is dark.

5.2   Rough Surfaces

If a metal or dielectric is roughened at a small scale, but not so small that diffraction occurs, then we can think of it as a surface with *microfacets*. Such surfaces behave specularly at a small scale, but viewed at a distance seem to spread the light out in a distribution. For a metal this could be brushed steel, or the "cloudy" side of most aluminum foil.

For dielectrics, such as a sheet of glass, scratches or other irregular surface features makes the glass blur the reflected and transmitted images we can usually see clearly. If the surface if heavily scratched we call it *translucent* rather than transparent. This is a somewhat arbitrary distinction, but it is usually clear whether we would consider it translucent or transparent.

5.3   Diffuse Materials

A material is *diffuse* if it is matte, i.e., not shiny. Many surfaces we see are diffuse, such as most stones, paper, and unfinished wood. To a first approximation, diffuse surfaces can be approximated with a Lambertian (constant) BRDF. Real diffuse materials usually become somewhat specular for grazing angles. This is a subtle effect, but can be important for realism.

5.4   Translucent Materials

Many thin objects such as leaves and paper both transmit and reflect light diffusely. For all practical purposes no clear image is transmitted by these objects. These surfaces can add a hue shift to the transmitted light. For example, red paper is red because it filters

11

Figure 12 Layered surfaces for two viewing angles. Note that the specular coefficient increases for acute viewing angles. Also note that the diffuse component fades in tandem.
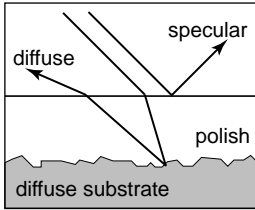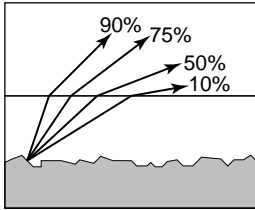
out non-red light for light that penetrates a short distance into the paper and then scatters back out. The paper also transmits light with a red hue because the same mechanisms apply, but the transmitted light makes it all the way through the paper. One implication of this is that the transmitted coefficient should be the same in both directions.

5.5 Layered Materials

Many surfaces are composed of "layers" or are dielectrics with embedded particles that give it a diffuse property. The surface of such materials reflects specularly as shown in Figure 11 and thus obeys the Fresnel equations.

An example of a layered surface are the glazed ceramic tiles shown in Figure 12. Note that the diffuse component also is attenuated with angle because the Fresnel equations make reflection back into the surface as the angle increases as shown in Figure 13.

6 Implementing Reflection Models

When we want to do a more physically-based rendering than is done with point light sources and Phong-like models, our basic goal is to implement a BRDF model as described in Section 2.6. Unfortunately, real BRDFs are typically quite complicated and cannot be deduced from first principles. Instead they must either be measured and used directly approximated from raw data, or they must be crudely approximated in an empirical fashion. The latter empirical strategy is what is usually done, and the development of approximate models is still an area of research. This section discusses several desirable properties of such empirical models.

First, there are two properties of a BRDF model that physical constraints imply. First is energy conservation.

$$\text{for all } \mathbf{k}_i, R(\mathbf{k}_i) = \int_{\text{all } \mathbf{k}_o} \rho(\mathbf{k}_i, \mathbf{k}_o) \cos\theta_o \, d\sigma_o \leq 1.$$

This just says that if you send a beam of light at a surface from any direction $\mathbf{k}_i$ then the total amount of light reflected over all directions will be at most the incident amount. The second physical property we expect all BRDFs to have is reciprocity:

$$\text{for all } \mathbf{k}_i, \mathbf{k}_o, \rho(\mathbf{k}_i, \mathbf{k}_o) = \rho(\mathbf{k}_o, \mathbf{k}_i)$$

Figure 11 Light hitting a layered surface and be reflected specularly or it can be transmitted and then scatter diffusely off the substrate.

Figure 13 The light scattered by the substrate is less and less likely to make it out of the surface as the angle increases.

Second, we want a clear separation between diffuse and specular components. The reason for this is that although there is a mathematically clean delta function formulation for ideal specular components, delta functions must be implemented as special cases in practice. Such special cases are only practical if the BRDF model clearly indicates what is specular and what is diffuse.

Third, we would like intuitive parameters. For example, one reason the Phong model has enjoyed such longevity is that its diffuse constant and exponent are both clearly related to the intuitive properties of the surface, namely surface color and highlight size.

Finally, we would like the BRDF function to be amenable to Monte Carlo sampling. Recall that an integral can be sampled by N random points $x_i \sim p$ where p is defined with the same measure as the integral:

$$\int f(x)d\mu \approx \frac{1}{N} \sum_{j=1}^{N} \frac{f(x_j)}{p(x_j)}.$$

Recall from Section 3 that the surface radiance in direction $\mathbf{k}_o$ in a direction is given by a transport equation:

$$L_s(\mathbf{k}_o) = \int_{\text{all } \mathbf{k}_i} \rho(\mathbf{k}_i, \mathbf{k}_o)L_f(\mathbf{k}_i)\cos\theta_i d\sigma_i$$

If we sample directions with pdf $p(\mathbf{k}_i)$ then we can approximate the surface radiance with samples:

$$L_s(\mathbf{k}_o) \approx \frac{1}{N} \sum_{j=1}^{N} \frac{\rho(\mathbf{k}_j, \mathbf{k}_o)L_f(\mathbf{k}_j)\cos\theta_j}{p(\mathbf{k}_j)}.$$

This approximation will converge for any p that is non-zero wherever the integrand is non-zero. However, it will only converge well if there is nowhere where the integrand is very large relative to p. Ideally, $p(\mathbf{k})$ should be approximately shaped like the integrand $\rho(\mathbf{k}_j, \mathbf{k}_o)L_f(\mathbf{k}_j)\cos\theta_j$. In practice L_f is complicated and the best we can accomplish is to have $p(\mathbf{k})$ shaped somewhat like $\rho(\mathbf{k}, \mathbf{k}_o)L_f(\mathbf{k})\cos\theta$.

For example, if the BRDF is Lambertian, then it is constant and the "ideal" $p(\mathbf{k})$ is proportional to $\cos\theta$. Because the integral of p must be one, we can deduce the leading constant:

$$\int_{\text{all } \mathbf{k} \text{ with } \theta < \pi/2} C\cos\theta d\sigma = 1$$

This implies that $C = 1/\pi$ so we have:

$$p(\mathbf{k}) = \frac{1}{\pi}\cos\theta.$$

However, an acceptably efficient implementation would result as long as p doesn't get too small where the integrand is non-zero. Thus the constant pdf would also suffice:

$$p(\mathbf{k}) = \frac{1}{2\pi}.$$

This emphasizes that many pdf's may be acceptable for a given BRDF model.

7 Specular Reflection Models

For a metal, we typically specify the reflectance at normal incidence $R_0(\lambda)$. The reflectance should vary according to the Fresnel Equations, and a good approximation developed by Christophe Schlick to this is given by:

$$R(\theta, \lambda) = R_0(\lambda) + (1 - R_0(\lambda))(1 - cos\theta)^5$$

This allows one to just set the normal reflectance of the metal either from data or by eye.

For a dielectric, the same formula works for reflectance. However, we can set $R_0(\lambda)$ in terms of the refractive index $n(\lambda)$:

$$R_0(\lambda) = \left(\frac{n(\lambda) - 1}{n(\lambda) + 1}\right)^2$$

Typically, we don't have n vary with wavelength, but for applications where dispersion is important it can vary. The refractive indices that are often useful include water ($n = 1.33$), glass ($n = 1.4$ to $n = 1.7$), and diamond ($n = 2.4$).

8 Smooth Layered Model

Matte/specular materials such as plastics or polished woods have reflection governed by Fresnel equations at the surface, and scattering within the subsurface. An example of this reflection can be seen in the tiles in the photographs in Figure 12. Also note that the blurring in the specular reflection is mostly vertical. This is due to the compression of apparent bump spacing in the view direction. This effect causes the vertically-streaked reflections seen on lakes on windy days, and can either be modeled using explicit micro-geometry and a simple smooth-surface reflection model or by a more general model that accounts for this asymmetry.

We could use the traditional Lambertian-specular model for the tiles, which uses two constants to modulate a constant and specular component of the BRDF. In standard radiometric terms, this idea is expressed as:

$$\rho(\theta, \phi, \theta', \phi' \lambda) = \frac{R_d(\lambda)}{\pi} + R_s \rho_s(\theta, \phi, \theta', \phi'),$$

where $R_d(\lambda)$ is the hemispherical reflectance of the matte term, R_s is the specular reflectance, and ρ_s is the normalized specular BRDF (a weighted Dirac delta function on the sphere). This equation is a simplified version of the BRDF where R_s is independent of wavelength. This independence causes a highlight that is the color of the luminaire, so a polished rather than a metal appearance will be achieved. Ward suggests that in order to conserve energy, $R_d(\lambda) + R_s \leq 1$. However, such models with constant R_s fail to show the increase in specularity for steep viewing angles. This is the key point: in the real world the relative proportions of matte and specular appearance change with viewing angle.

He et al. suggest using the Fresnel equation for the coefficient of the specular term, but do not address the subsurface term's angular behavior because this model is intended primarily to simulate surface physics. Since the Fresnel term of the He model goes to one for $\theta = 90°$, the Lambertian term would have to be set to zero to enforce

energy conservation for all incident (θ, ϕ). Because in the case of smooth polished surfaces energy conservation is important to us, and the explicit spread reflection is not (we assume that we will model it with micro-geometry for this discussion), the He model is not appropriate for our purposes.

Shirley attempted to simulate the change in the matte appearance with angle by explicitly dampening $R_d(\lambda)$ as R_s increases:

$$
\rho(\theta, \phi, \theta', \phi', \lambda) = R_f(\theta)\rho_s(\theta, \phi, \theta', \phi') + \frac{R_d(\lambda)(1 - R_f(\theta))}{\pi},
$$

where $R_f(\theta)$ is the Fresnel reflectance for a polish-air interface. The problem with this equation is that it is not reciprocal, as can been seen by exchanging θ and θ' which changes the value of the matte dampening factor because of the multiplication by $(1 - R_f(\theta))$. The specular term, a scaled Dirac delta function, is reciprocal, but this does not make up for the non-reciprocity of the matte term. Because Shirley's BRDF is not physically plausible, it will cause some rendering methods to have ill-defined solutions.

Schlick proposed a general reflectance model tuned for efficiency. In his model a matte/specular surface could have constant Lambertian and specular coefficents, or the Fresnel reflectance could be used. In the latter case, which he calls a *double* surface the BRDF becomes:

$$
\rho(\theta, \phi, \theta', \phi', \lambda) = R_f(\theta)\rho_s(\theta, \phi, \theta', \phi') + \frac{R_d(\lambda)(1 - R_f(\alpha))}{\pi},
$$

where $R_d(\lambda)$ is a matte coefficient and α is half the angle between incident and outgoing directions. However, this form does not conserve energy for all incident angles: for example, at $\theta = 90°$ the specular reflectivity goes to one, and the fraction of the hemispherical reflectance is still above zero (e.g,. plug in $\theta' = 0$). So that part of the Schlick model is not appropriate for our purposes.

In review of our attempt to model a smooth matte-specular surface, none of the commonly used general models is appropriate for our purposes. The Lambertian-specular and Ward models do not have the appropriate angular trade-off between the matte and specular term. For smooth surfaces, the He model has a constant subsurface term that must be set to zero if energy is to be conserved. The Schlick model either defaults to the Lambertian-specular model, or it accounts for the Fresnel equation effects but does not conserve energy. One reason these models fail for this case is that they are all intended to model spread reflection for a variety of material types. In our case-study we do not need this generality, so we can develop a simple model that is customized for this narrow class of materials and captures the angular-dependent relationship of the matte and specular coefficients. This model uses a physically-based specular coefficient derived from the Fresnel equations, and a heuristic matte component of the BRDF. To our knowledge, it is the first model that produces the matte/specular tradeoff while remaining reciprocal and energy conserving. Because the key feature of the new model is that it couples the matte and specular scaling coefficients, we will hereafter refer to it as the *coupled* model.

Surfaces which have a glossy appearance are often a clear dielectric, such as polyurethane or oil, with some subsurface structure. The specular (mirror-like) component of the reflection is caused by the smooth dielectric surface and is independent of the structure

below this surface. The magnitude of this specular term is governed by the Fresnel equations.

The light that is not reflected specularly at the surface is transmitted through the surface. There either it is absorbed by the subsurface or it is reflected from a pigment or a subsurface and transmitted back through the surface of the polish. This transmitted light forms the matte component of reflection. Since the matte component can only consist of as much light as is transmitted, it will naturally decrease in total magnitude for increasing angle.

To avoid choosing between physically plausible models and models with good qualitative behavior over a range of incident angles, we note that the Fresnel equations that account for the specular term, $R_f(\theta)$, are derived directly from the physics of the dielectric-air interface. Therefore the problem must lie in the matte term. We could use a full-blown simulation of subsurface scattering as implemented by Hanrahan and Krueger , but this technique is both costly and requires detailed knowledge of subsurface structure, which is usually neither known nor easily measurable. Instead, we can modify the matte term to be a simple approximation that captures the important qualitative angular behavior shown in Figure 12. Let us assume that the matte term is not Lambertian, but instead is some other function that depends only on θ, θ' and λ: $\rho_m(\theta, \theta', \lambda)$. We discard behavior that depends on ϕ or ϕ' in the interest of simplicity. We try to keep the formulas reasonably simple because the physics of the matte term is complicated and sometimes requires unknown parameters. We expect the matte term to be close to constant, and roughly rotationally symmetric, as is argued in He's dissertation.

An obvious candidate for the matte component $\rho_m(\theta, \theta', \lambda)$ that will be reciprocal is the *separable* form $kR_m(\lambda)f(\theta)f(\theta')$ for some constant k and matte reflectance parameter $R_m(\lambda)$. We could merge k and $R_m(\lambda)$ into a single term, but we choose to keep them separated because this makes it more intuitive to set $R_m(\lambda)$ which must be between 0 and 1 for all wavelengths. Separable BRDFs have been shown to have several computational advantages, which suggests the separable model:

$$
\begin{aligned}
\rho(\theta, \phi, \theta', \phi', \lambda) \quad = \quad & R_f(\theta)\rho_s(\theta, \phi, \theta', \phi') + \\
& kR_m(\lambda)f(\theta)f(\theta').
\end{aligned}
\tag{3}
$$

We know that the matte component can only contain energy not reflected in the surface (specular) component. This means that for $R_m(\lambda) = 1$, the incident and reflected energy are the same, which suggests the following constraint on the BRDF for each incident θ and λ:

$$
R_f(\theta) + 2\pi k f(\theta) \int_0^{\frac{\pi}{2}} f(\theta') \cos \theta' \sin \theta' d\theta' = 1.
\tag{4}
$$

We can see that $f(\theta)$ must be proportional to $(1 - R_f(\theta))$. If we assume matte components that absorb some energy have the same directional pattern as this ideal, we get a BRDF of the form:

$$
\begin{aligned}
\rho(\theta, \phi, \theta', \phi', \lambda) \quad = \quad & R_f(\theta)\rho_s(\theta, \phi, \theta', \phi') + \\
& kR_m(\lambda)[1 - R_f(\theta)][1 - R_f(\theta')].
\end{aligned}
$$

This is similar to a BRDF model used in the sensor community, although the constants used in that model do not have the normalization properties we desire. We could now

16
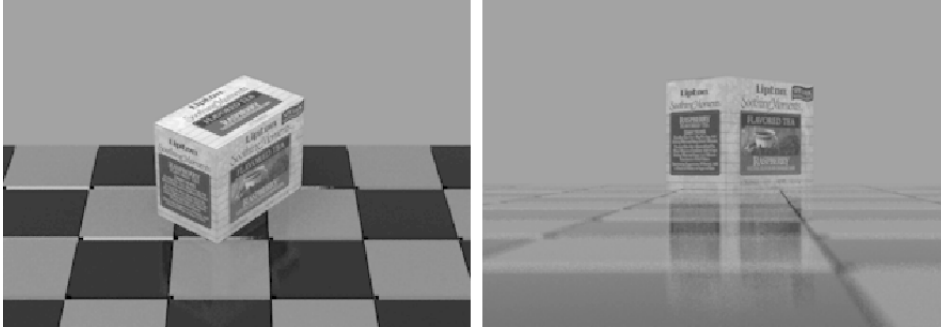
Figure 14 Renderings of polished tiles using coupled model.

insert the full form of the Fresnel equations to get $R_f(\theta)$ and then use energy conservation to solve for constraints on k. Instead we will use the approximation discussed in Section 5.1 This implies that

$$f(\theta) \propto (1 - (1 - \cos\theta)^5).$$

Applying Equation 4 gives:

$$k = \frac{21}{20\pi(1 - R_0)}. \tag{5}$$

The full coupled BRDF is then:

$$\begin{aligned}
\rho(\theta, \phi, \theta', \phi', \lambda) = \\
\left[R_0 + (1 - \cos\theta)^5(1 - R_0) \right] \rho_s(\theta, \phi, \theta', \phi') + \\
k R_m(\lambda) \left[1 - (1 - \cos\theta)^5 \right] \left[1 - (1 - \cos\theta')^5 \right].
\end{aligned} \tag{6}$$

To test this new coupled model we created an approximate geometric model of the scene in Figure 12, where each tile has a displacement map that roughly corresponds to the subtle roughness of the real tiles. We took two photographs of the test scenes from different view angles, resulting in two different values of θ.

The results of running the coupled model is shown in Figure 14. Note that for the high viewpoint the the specular reflection is almost invisible, but is clearly visible in the low-angle photograph image, while the matte behavior is less obvious.

For reasonable values of refractive indices, the R_0 is limited to approximately the range 0.03 to 0.06 (the value $R_0 = 0.05$ was used for the figures). The value of R_s in a traditional Phong model is harder to choose because it must typically be tuned for viewpoint in static images, and tuned for a particular camera sequence for animations. Thus, the coupled model is easier to use in a "hands-off" mode.

We did not attempt to mimic all subtleties of geometry exactly, so the reader should concentrate on the gross appearance features in the rendered and photographic images of the model. These images were produced using a Monte Carlo path tracer. The sampling distribution for the diffuse term is $\cos\theta/\pi$.

9   Rough Layered Model

The previous model is fine if the surface is smooth. However, if the surface is not ideal, some spread is needed in the specular component is needed. An extension of the

17

coupled model to this case is *An Anisotropic Phong Reflection Model* by Ashikhmin and Shirley, included in these notes.

Notes

A common radiometric quantity not described in the chapter is *radiant intensity* (I), which is the spectral power per steradian emitted from an infinitesimal point source. It should usually be avoided in graphics programs because point sources cause implementational problems. A more rigorous treatment of radiometry can be found in James Arvo's dissertation (Yale University, 1995).

There are many BRDF models described in the literature, and only a few of them have been described here. Many of them, as well as more on the physics of surface reflection, are surveyed in Glassner's two-volume set *Principles of Digital Image Synthesis* (Morgan Kaufmann, 1994).

SIGGRAPH 2001 Course: State of the Art in Modeling
and Measuring of Surface Reflection

BRDF Measurement

Stephen R. Marschner

April 25, 2001

1   Measuring Reflectance

Before I discuss measuring the complete BRDF, I'll first talk about some of
the simpler reflectance measurements that are commonly made. Because the
BRDF completely describes what goes on at an opaque surface, we can think
of all reflectance measurements as special cases of BRDF measurement.

What does it mean to make a measurement of the BRDF? The BRDF is a
function of three things (six variables): incident direction, exitant direction,
and position on the surface. Since we need a finite (i.e. not infinitesimal)
amount of energy before we can measure it, we have to integrate over some
area in all the dimensions of the BRDF—incident solid angle, outgoing solid
angle, and surface area. The way to describe any reflectance measurement
is to say what region of the BRDF you are integrating to get the number
that you write down.

1.1   Single-sample diffuse and specular measurements

The easiest way to get a single number for the reflectance of a surface (or
a set of numbers to describe its color) is to illuminate the surface from a
small solid angle somewhere on the hemisphere (integrating over incident
solid angle) and measure the reflected radiance with a detector that sees a
small area of the surface (integrating over surface area) and collects the light
that goes into a small solid angle somewhere on the hemisphere (integrating
over outgoing solid angle). Since the region of the BRDF's domain being
measured is small, this kind of measurement approximates measuring the
BRDF's value at a single point.
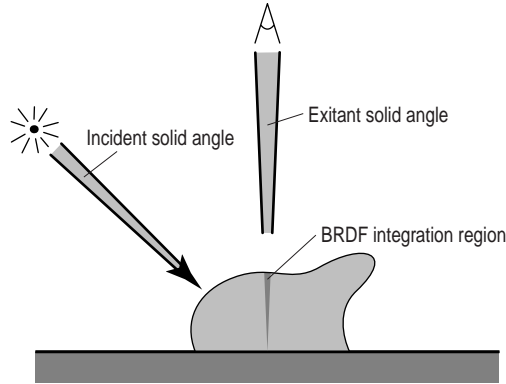
Figure 1: Measuring diffuse reflectance in a 45°/0°configuration.

This last statement is a little misleading, though, because of how BRDFs behave. To say that we are approximating the BRDF's value, we'd like the details of the measurement not to matter too much. In particular, we'd like to report the reflectance measurement without having to say what solid angles we used. But the measurements will depend differently on the solid angles depending on which part of the BRDF we are looking at.

Away from the specular direction, where the BRDF is a continuous function, the number from the detector will be proportional to sizes of the solid angles (for fixed light source radiance), so we'll divide by the product of the two solid angles to get the number we report. This is a "diffuse" measurement and is generally used to describe the color of the diffuse component of a surface's BRDF (Figure 1). Often the configuration that is measured has the source at 45 degrees incidence and the detector normal to the surface ("45°/0°") or vice versa ("0°/45°"). Sometimes it's not desirable to have the specular reflection of the source end up back in the source, or conversely to have the detector looking at its own specular reflection, in which case a slightly off-normal configuration like 8°/45° or 45°/8° can be used. Usually these measurements are referenced to a measurement of a known surface (for example a diffuse 100% white reflector) and expressed as a percentage diffuse reflectance.

But if we're looking at a specular reflection, things are more complicated—in that case our goal is to ensure that all the specularly reflected light ends up being captured by the detector. Once this condition is satisfied, the measurement is proportional only to the solid angle of the source (again, with constant source radiance). This type of measurement can be referenced to a measurement from a known specular reflector (a first-surface mirror,
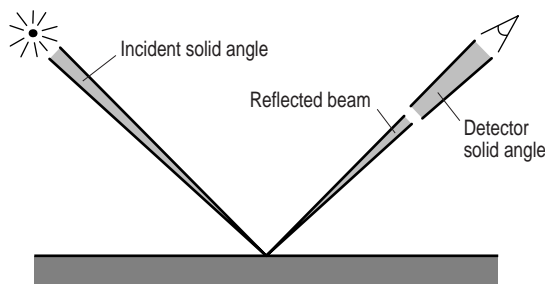
Figure 2: Measuring specular reflectance at 45°.

for example) or to a measurement of light going directly from the source to the detector without reflecting, and it is reported as a percentage specular reflectance.

1.2 Hemispherical measurements

A carefully done measurement of specular reflectance says what there is to say about the surface's ideal specular reflection at a particular angle, but a single point measurement of the diffuse component is only as meaningful as the surface is diffuse. For repeatability, the configuration used in the measurement needs to be carried around with the measurement itself as part of the description of what the number means.

A more well-defined measurement is one that captures all the diffusely reflected light. This can be done using an integrating sphere, which is a hollow sphere painted white on the inside. Any light that enters such a sphere is reflected many times and distributed uniformly around its interior. By using an integrating sphere to collect and sum up all the light that leaves a surface, one can make a "directional-hemispherical" measurement (Figure 3).

You can measure the diffuse and specular parts of the BRDF together by using a complete integrating sphere; or, if you only want the diffuse component (as you do if you're making a color measurement) you can put a small hole in the sphere at the specular direction to let the specularly reflected light escape and just add up all the diffusely reflected light.

Careful measurements of hemispherical reflectance involve making multiple measurements to correct for the fact that introducing the sample makes the sphere not quite an integrating sphere any more.

Of course, you still need to report the incident angle: "0°/d", "8°/d". The reciprocal measurements—d/0°, etc.—are equivalent and are sometimes
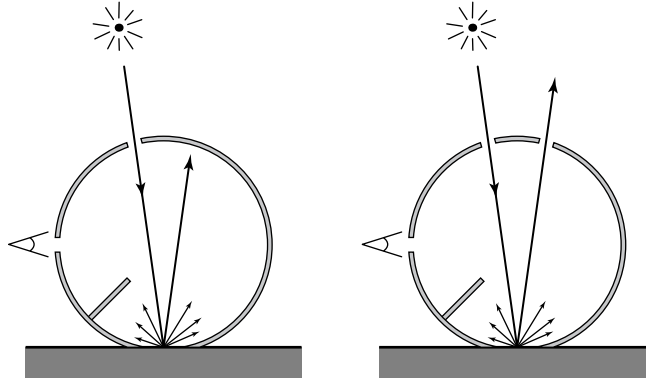
Figure 3: Measuring directional-hemispherical reflectance, including (left) and excluding (right) the specular component.

more convenient to make.

1.3   Gloss measurements

For highly polished surfaces, in which you see a perfect mirror reflection, just a single measurement of percentage reflection at a particular angle says what there is to say about the specular reflection. However, many surfaces are not perfectly smooth, but are still glossy. Quantifying the differences in appearance between these surfaces is important, particularly if you are manufacturing such surfaces.

Gloss is measured by comparing a measurement at the specular direction with another measurement at an angle α from the specular direction (Figure 4). Although gloss measurements go by many names, the general idea is that you use a value of α that is appropriate for the breadth of the specular lobe you are measuring. If the surface is very nearly a mirror, you set α to a fraction of a degree, and the resulting ratio is "distinctness-of-image gloss." If your surface is slightly less perfect α is 2 to 6 degrees and you are measuring "haze" or "bloom." If the surface is not really that shiny, you put the second measurement far away and you are measuring what amounts to a specular/diffuse ratio, called "contrast gloss."

2   Traditional BRDF Measurement

All these single-number measurements are fine if you are dealing with surfaces that are near the ideal diffuse-plus-mirror behavior, or if you just need

4
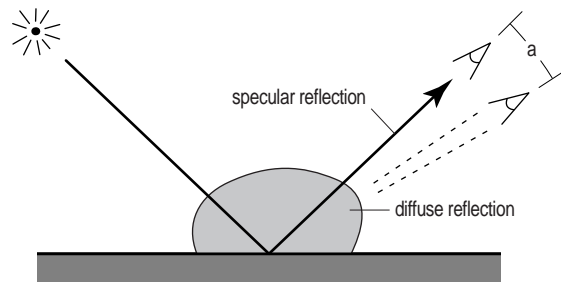
Figure 4: Measuring gloss by comparing a specular measurement with a nearby diffuse measurement.

specific, repeatable measurements, for instance for quality control. But for rendering in computer graphics we want to deal with more complex materials, and we want to know all about how they look, since in our models we have to be able to illuminate and view surfaces from arbitrary directions.

For these reasons we need to do full BRDF measurements, for which an appreciable number of individual samples of the BRDF are taken systematically over the whole domain. Devices for doing this are called gonioreflectometers.

2.1   Designs for gonioreflectometers

The simplest gonioreflectometers measure only incidence-plane BRDF, which means the viewing direction always lies in the plane defined by the incident direction and the surface normal (Figure 5). For getting insight into the behavior of a material this type of measurement is helpful, because it can observe the most frequently occurring reflectance phenomena: diffuse reflection, specular and near-specular reflection, and retroreflection.

To get the complete picture we need for rendering, however, we have to make measurements of the whole BRDF, not just a slice through the interesting part. This means we have to build a three- or four-axis device to get to all the configurations we have to measure. Four axes are required if we need to measure anisotropic BRDFs in full generality; the extra symmetry in isotropic materials means three will suffice in that case (Figure 6)

An issue that always arises in BRDF measurement, and is very important in gonioreflectometers, is what defines the surface area we are integrating over. There are basically three ways for this to be set up, and each requires a different normalization on the measurements to get BRDF values out.

The arrangement that is most familiar from rendering is to have the

5
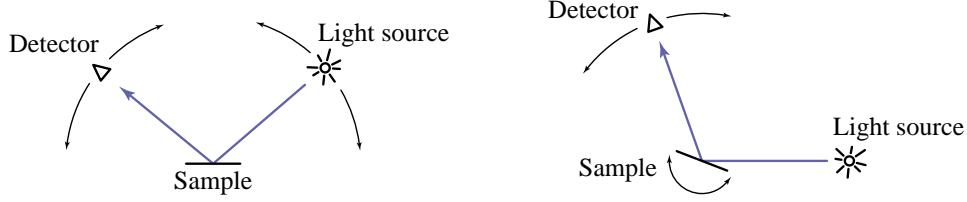
Figure 5: Two devices for measuring incidence plane (2D) BRDFs.

source uniformly illuminate a large area of surface, but have the detector focused to measure only a small area within the illumination spot. This case is like a pixel in a camera—the radiance integrated over that area will be proportional to $\rho(\omega_i, \omega_e) \cos \theta_i$, and we have to divide by $\cos \theta_i$ to get ρ.

However, equally valid is the dual case: the detector measures everything coming from near the sample, but the light source is focused down so that it illuminates only a small area. In this case the key surface area gets larger in proportion to $1/\cos \theta_i$ and the measurement is proportional to $\rho(\omega_i, \omega_e) \cos \theta_e$ and needs to be divided by $\cos \theta_e$.

In both of the preceding cases it's important to make sure that as the surface area expands toward grazing angles it does not fall off the edge of the sample. The third approach is to make sure the measured area always falls of the edge—that is, the source illuminates a large area and the detector measures a large area, so that the edges of the sample itself define the surface area. In this case the measurements are proportional to $\rho(\omega_i, \omega_e) \cos \theta_i \cos \theta_e$, so the normalization factor is $1/(\cos \theta_i \cos \theta_e)$.

Since the normalization factors for these three cases are different, it's obvious that any gonioreflectometer needs to be built to stay in one of these cases all the time. If it wanders from one case to another as the angles change, it will be impossible to normalize the results into meaningful BRDF values. This requirement is often what limits the maximum incident and exitant angles that can be measured in a particular gonioreflectometer.

2.2 Suitability for graphics

Traditional gonioreflectometers work well, and they can produce very precise and accurate results, even for difficult problems like measuring the tiny bit of diffuse scattering from the surface of a mirror or the very dark reflection from a black coating for optical components. However, these devices, since they measure one sample at a time, take hours or even days to make densely-sampled BRDF measurements.

6
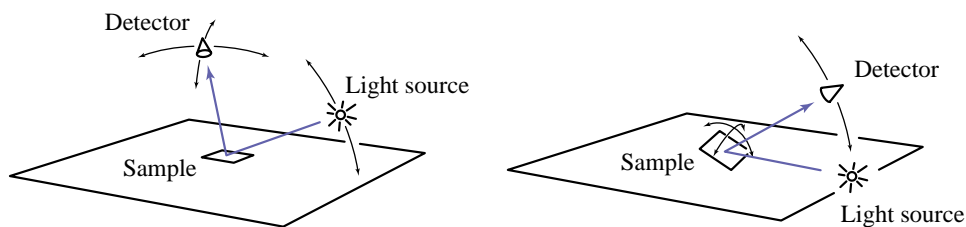
Figure 6: Two devices for measuring full isotropic (3D) BRDFs.

For graphics our accuracy requirements are not so high, and we are interested mainly in everyday materials and phenomena that are visible to the naked eye. Some important materials don't come in small, flat pieces that can be put in a gonioreflectometer. Also, complex environments contain large numbers of materials, so we are willing to make some sacrifices for speed and generality.

3   Image-based BRDF Measurement

With digital image sensors, such as CCD arrays, we can accelerate the process of BRDF measurement by measuring many samples in parallel. The idea is to take the three (or four) dimensions we need to measure, and instead of handling all of them with mechanical degrees of freedom, take two of them and assign them to the axes of an image sensor. With the right geometric arrangement, we can measure thousands of BRDF samples with a single brief exposure.

3.1   Ward's mirror-based device

Ward [5] describes a BRDF measurement system that sets the incident direction in the usual way, by moving a source along an arc, but uses a spherical reflector and a fisheye lens to capture all the reflected directions in a single exposure (Figure 7). (The paper describing this device is reprinted in these course notes.)

Anisotropic surfaces can be handled by repeating the BRDF measurement with the sample in several different orientations. This device is simpler than a four-axis gonioreflectometer, with many fewer moving parts: only the source is required to move. It does, however, require an unusual component: the hemispherical (or preferably hemiellipsoidal) part-silvered mirror.
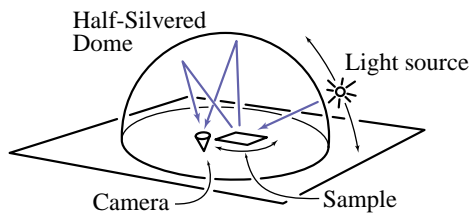
Figure 7: Ward's image-based BRDF measurement device.

3.2 BRDF measurements using curved surfaces

Another, simpler, approach to using a camera to accelerate BRDF measurement is to use standard camera optics with a curved sample. This approach, used by Lu et al. [1] in the incidence plane and by Marschner et al. [3, 2] to measure full BRDFs, is attractive because it requires no special equipment and few moving parts, and because it can measure materials that cannot be obtained in the flat, uniform samples normally required for BRDF measurement (Figure 8). (The first paper [3] is reprinted in these notes.)

In this approach, every pixel of a digital camera is used as a detector for BRDF measurement. With the camera focused on the surface and the surface illuminated by a small source, each pixel returns a measurement of the BRDF averaged over a small spot on the surface (the image of the pixel) and over solid angles defined by the camera aperture and the light source size. Since the surface is curved, you get a measurement of a different configuration at each pixel, and if the BRDF is uniform across the surface you can consider all the pixels to be measurements of a single BRDF. This gives you a very dense collection of samples of the BRDF, scattered through the domain in a semi-arbitrary way that depends on the shape of the surface and on the camera and light source positions. The coverage will depend on the range of surface normals presented by the visible parts of the surface—a sphere, for instance, will give you the broadest possible coverage because it presents all surface normals. If you take a series of photographs and choose the right camera and light source positions, you can densely sample very nearly the entire domain of an isotropic BRDF.

An interesting connection can be made between this photograph-based measurement and the traditional gonioreflectometer design shown on the right of Figure 6. One way to think of the photographic BRDF setup is that in one image it measures all the same configurations that you would make with this gonioreflectometer using a single detector position and all sample orientations. But it makes all the measurements in parallel, in one exposure,
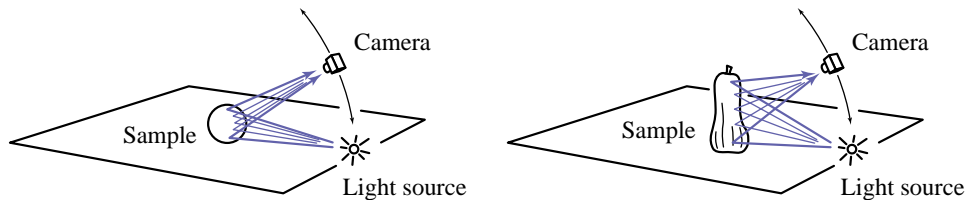
Figure 8: Image-based BRDF measurement using a camera and a curved surface with regular (left) and irregular (right) geometry.

and it avoids the need to build a multi-axis rotation device to position the sample.

Because the material sample being measured need not—in fact, must not—be flat, this approach can handle a range of important materials that cannot be measured any other way. In particular, the BRDF of human skin can be measured *in vivo* with this technique [3].

4   Dropping the BRDF assumption

The assumption that we made when we first started talking about measuring and modeling the BRDF was that surfaces are opaque, so that the BRDF is an accurate model for light reflection. If you examine the materials around you, you will realize (especially if you are armed with a laser pointer) that many materials—skin, leaves, milk, fruit, marble, and more—don't conform to this assumption. In fact, when Nicodemus originally introduced the BRDF [4] it was as a special case of the more general BSSRDF, the bidirectional scattering-surface reflectance distribution function. The BSSRDF drops the assumption that reflection occurs at each surface point in isolation, and lets the reflectance depend on the points where the incident light arrives at and the reflected light exits from the surface (Figure 9):

$$\rho(\mathbf{x}, \omega_i, \omega_e) \rightarrow S(\mathbf{x}_i, \mathbf{x}_e, \omega_i, \omega_e)$$

For homogeneous surfaces, all that matters is the distance r between $\mathbf{x}_i$ and $\mathbf{x}_e$.

The BSSRDF can be measured using techniques similar to BRDF measurement techniques, but this time the source and the detector both must be focused to small spots, and the distance between the spots has to be able to vary. Again image-based measurement can speed things up: by focusing the light to a small area and measuring the reflected light with a camera, the reflectance for all values of r can be measured at once.

9
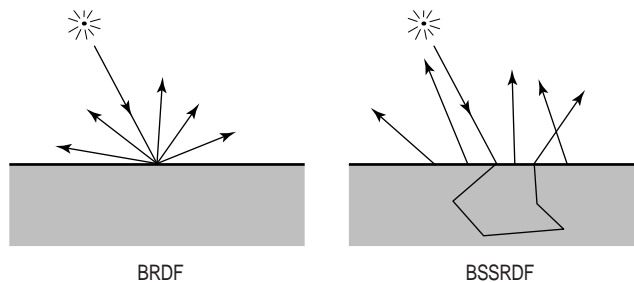
Figure 9: The BSSRDF drops the assumption of single-point reflection, which is inherent to the BRDF.

For more about measuring and modeling the BSSRDF, see the paper, "A practical model for subsurface scattering" in this SIGGRAPH's papers program.

References

[1] Rong Lu, Jan J. Koenderink, and Astrid M. L. Kappers. Optical properties (bidirectional reflection distribution functions) of velvet. *Applied Optics*, 37(25):5974–5984, 1998.

[2] Stephen R. Marschner, Stephen H. Westin, Eric P. F. Lafortune, and Kenneth E. Torrance. Image-based measurement of the bidirectional reflectance distribution function. *Applied Optics*, 39(16), 2000.

[3] Stephen R. Marschner, Stephen H. Westin, Eric P. F. Lafortune, Kenneth E. Torrance, and Donald P. Greenberg. Image-based BRDF measurement including human skin. In *Proceedings of 10th Eurographics Workshop on Rendering*, pages 139–152, June 1999.

[4] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis. Geometric considerations and nomenclature for reflectance. Monograph 161, National Bureau of Standards (US), October 1977.

[5] Gregory J. Ward. Measuring and modeling anisotropic reflection. In *Computer Graphics (SIGGRAPH '92 Proceedings)*, pages 265–272, July 1992.

Diffraction Shaders

From microsurface models to reflection models

Jos Stam
Alias | wavefront
Seattle, WA USA

Shaders

Model reflection from surfaces
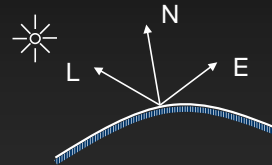
Shaders

Perfectly smooth surface
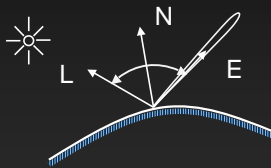
Not very interesting

Shaders
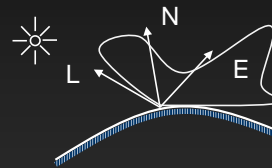
Rough surfaces

More interesting

Shaders
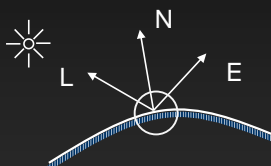
Rough surfaces

Zoom in on microstructure

Shaders

Rough surfaces
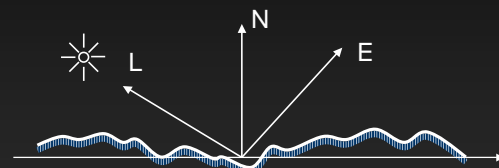
Reflection depends on the microsurface

Microsurface Models

Model surface as a two-dimensional (random) height field
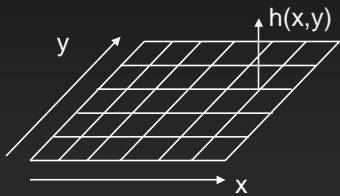
Microsurface Models
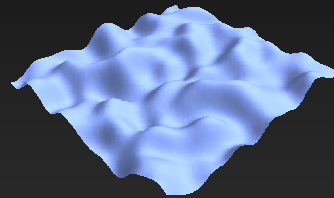
Isotropic Gaussian (smooth)

Microsurface Models

Isotropic Gaussian (smooth)
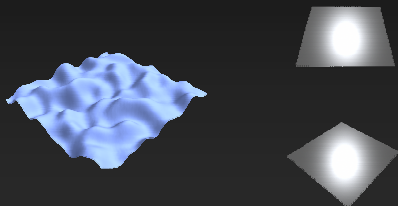
Microsurface Models

Anisotropic Gaussian (brushed metal)
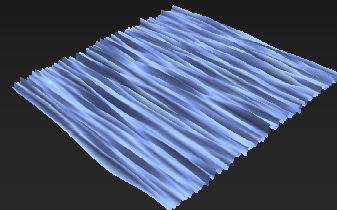
Microsurface Models

Anisotropic Gaussian (brushed metal)
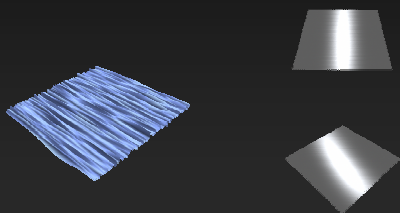
Microsurface Models
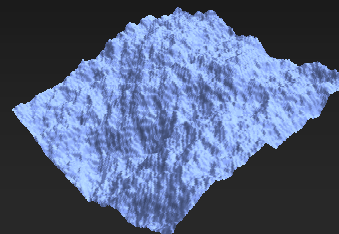
Fractal

Microsurface Models

Fractal
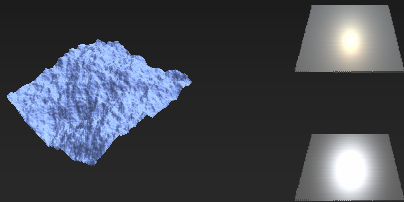
Microsurface Models

Periodic (compact disk)
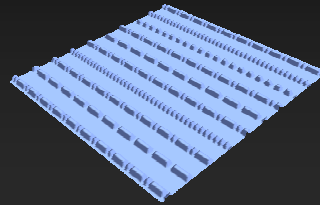
Microsurface Models

Periodic (compact disk)
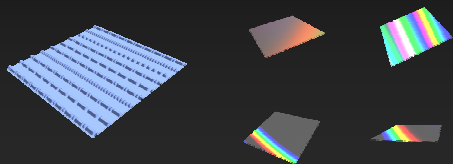
Our Approach

Use waves to model both

• the propagation of light

• the microsurface (Fourier Analysis)

Generalization of previous models

Wave Theory of Light

700nm

400nm

wavelength

λ $k = \dfrac{2\pi}{\lambda}$

Wave Theory of Light

L_1

$e^{ikL_1} = \cos(kL_1) + i\sin(kL_1)$ $k = \dfrac{2\pi}{\lambda}$

Wave Theory of Light

$$e^{ik(L_1+L_2)} = e^{ikL_1} \, e^{ikL_2}$$
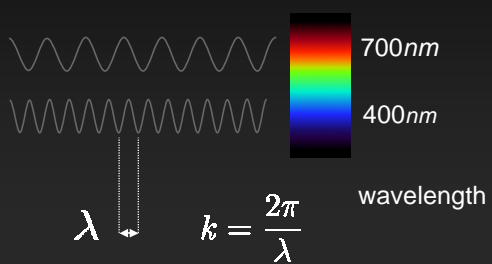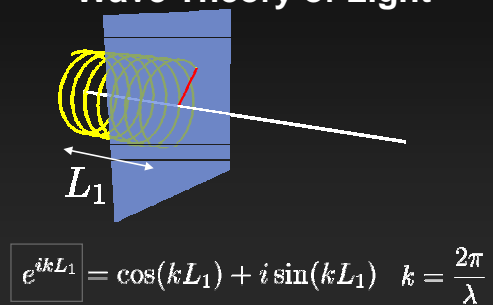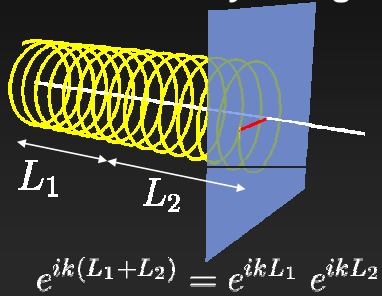
Intensity of a Wave
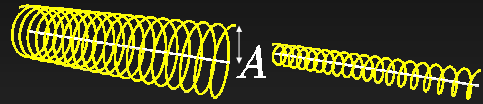
$$I = \left| A \, e^{ikL} \right|^2 = A^2$$

Diffraction

$$\left| A_1 \, e^{ikL_1} + A_2 \, e^{ikL_2} \right|^2 \neq A_1^2 + A_2^2$$

Intensities don't add up in general

Interaction: Light-Surface
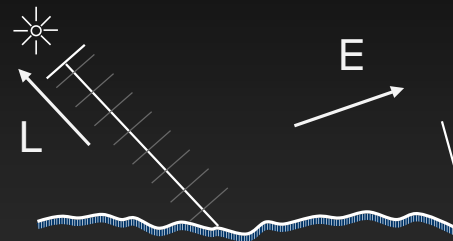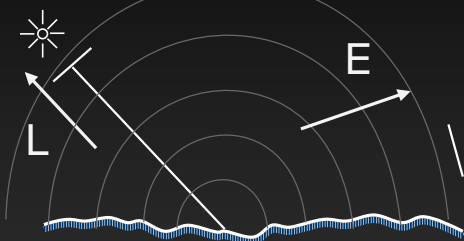
E

L

Interaction: Light-Surface

E

L

Interaction: Light-Surface

E

L

Interaction: Light-Surface

$$e^{ik(L_1+L_2)} = e^{ikL_1}\, e^{ikL_2}$$

Interaction: Light-Surface

Source and observer "far away"
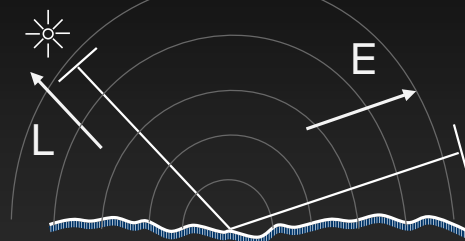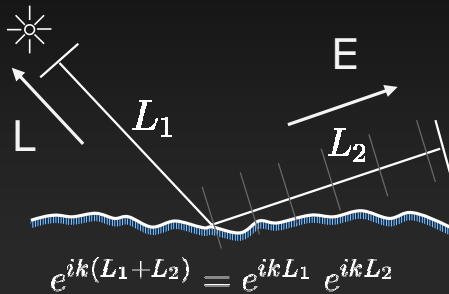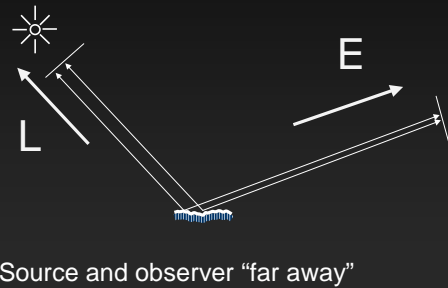
Interaction: Light-Surface

Add up contributions

$$\psi = e^{ikL(x_1)} + e^{ikL(x_2)} + e^{ikL(x_3)} + \cdots$$
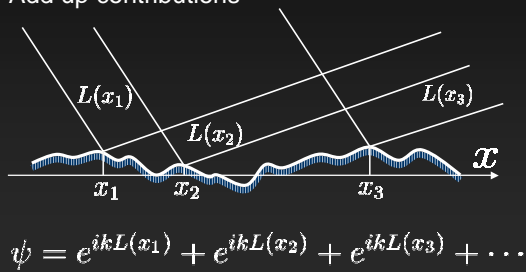
Interaction: Light-Surface

Integral in the limit
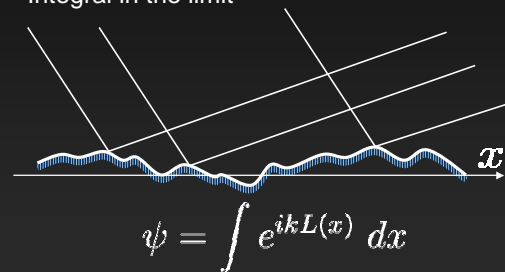
$$\psi = \int e^{ikL(x)}\, dx$$

Compute Path Length

Distance from source

L_1

$(x, h(x))$

Compute Path Length

R_1

d_1

L_1

$$L_1 = R_1 - d_1$$

5

Compute Path Length

$$L_1 = R_1 + x\sin\theta_1 - h(x)\cos\theta_1$$

Compute Path Length

Distance to the eye

$$(x, h(x))$$

Compute Path Length

$$L_2 = R_2 + d_2$$

Compute Path Length

$$L_2 = R_2 - x\sin\theta_2 - h(x)\cos\theta_2$$

Compute Path Length

Putting the two together:

$$L_1 = R_1 + x\sin\theta_1 - h(x)\cos\theta_1$$
$$L_2 = R_2 - x\sin\theta_2 - h(x)\cos\theta_2$$

Compute Path Length

Putting the two together:

$$L_1 = R_1 + x\sin\theta_1 - h(x)\cos\theta_1$$
$$+\ L_2 = R_2 - x\sin\theta_2 - h(x)\cos\theta_2$$

$$L(x) = R_1 + R_2 + ux + wh(x)$$

$$u = \sin\theta_1 - \sin\theta_2$$
$$w = -\cos\theta_1 - \cos\theta_2$$
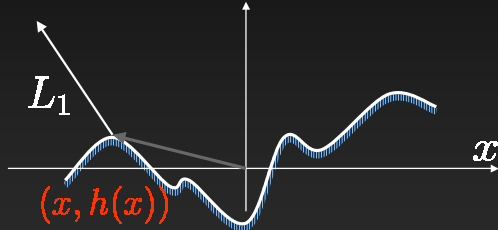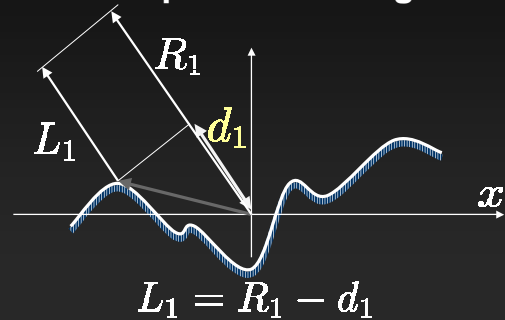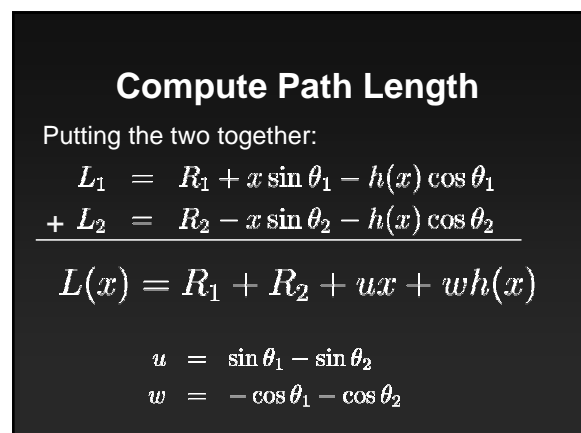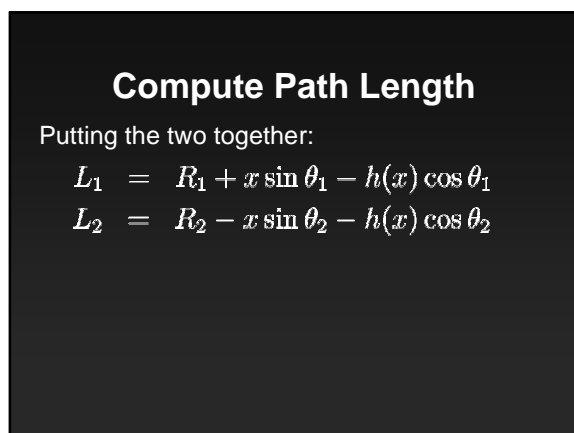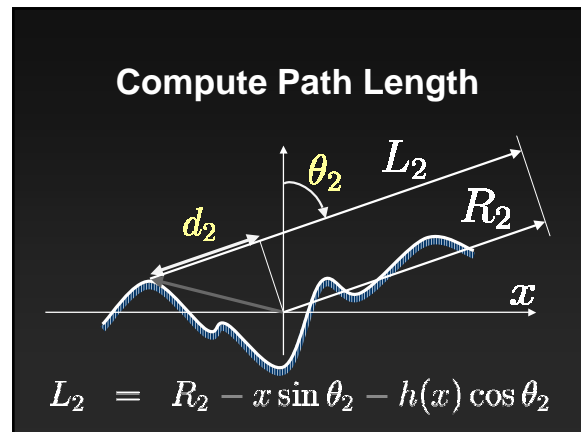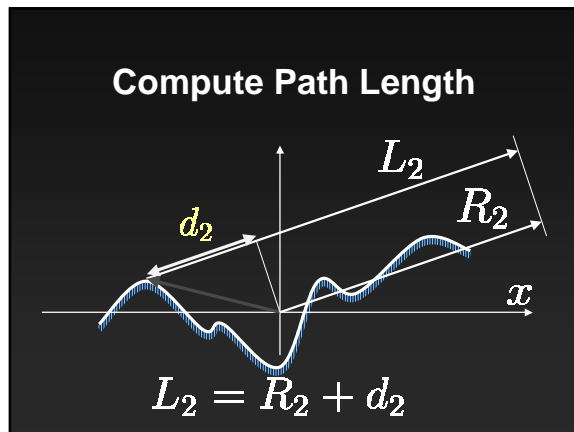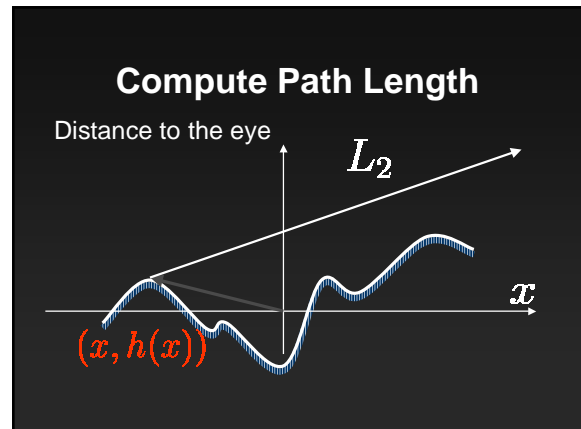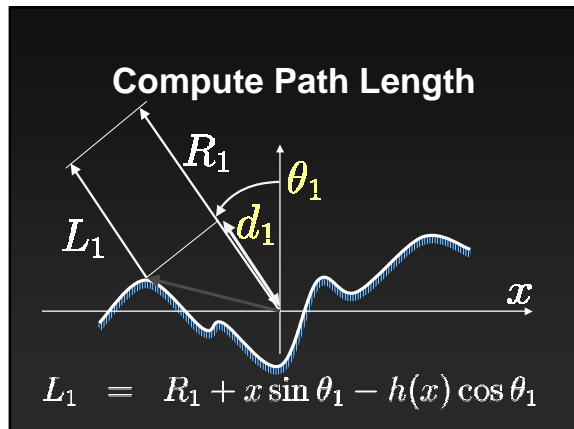
Reflected Wave

Now integrate over the surface:

$$L(x) = \underbrace{R_1 + R_2 + ux + wh(x)}$$

$$\psi = \int e^{ikL(x)}\, dx$$

Reflected Wave

$$L(x) = \underbrace{R_1 + R_2} + ux + wh(x)$$

$$\psi = C \int e^{ikwh(x)}\, e^{ikux}\, dx$$

Fourier Transform

$$\boxed{\psi = C \int p(x)\, e^{ikux}\, dx}$$

$$p(x) = e^{ikwh(x)}$$

Key insight

Fourier Transform

$$\boxed{\psi = C\, P(ku)}$$

$$p(x) = e^{ikwh(x)}$$

Key insight

Fourier Transform

Simple relationship:

$$\boxed{I = |P(ku)|^2}$$

$$|C| = 1$$

Example

Smooth surface:

$$h(x) = 0$$

$$p(x) = e^{ikwh(x)} = e^0 = 1$$

$$P(ku) = \delta(ku)$$

$$u = \sin\theta_1 - \sin\theta_2$$

Example

Almost smooth surface:

$$p(x) = e^{ikwh(x)} \approx 1 + ikwh(x)$$

$$P(ku) = \delta(ku) + ikwH(ku)$$

$$|P(ku)|^2 = \delta(ku) + k^2w^2|H(ku)|^2$$

Two dimensions

Previous derivation extends to 2D
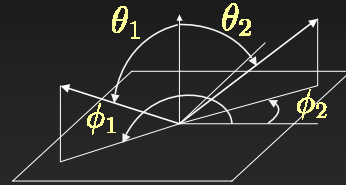
Two dimensions

$$I = |P(ku, kv)|^2$$

$$p(x, y) = e^{ikwh(x,y)}$$

$$k = \frac{2\pi}{\lambda}$$

$$u = -\cos\phi_1\sin\theta_1 - \cos\phi_2\sin\theta_2$$
$$v = \sin\phi_1\sin\theta_1 - \sin\phi_2\sin\theta_2$$
$$w = -\cos\theta_1 - \cos\theta_2$$

Computing Shaders

Shader = computing Fourier transforms

I have done this for:

- Gaussian random surfaces
- Fractal random surfaces
- Periodic surfaces

Details in   http://reality.sgi.com/jstam_sea/Research/ps/diff.ps.gz

Implementation

Implemented as MAYA plugin

$$I = |P(ku, kv)|^2$$

Straightforward

Implementation

Assign frame to the surface
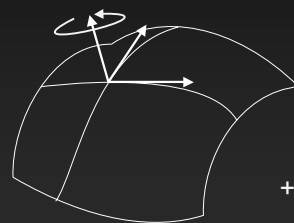
+ twist angle

Brushed Metal
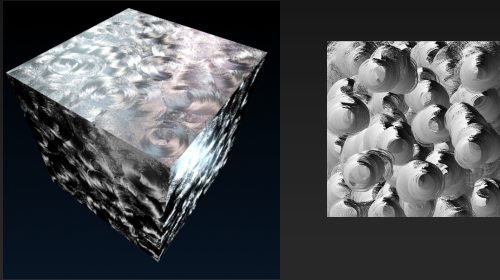
Diffraction
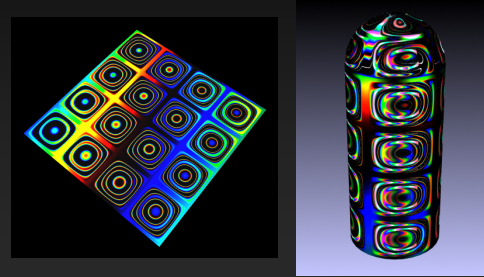
Compact Disk

Used Physical Dimensions:

- bump height : 150 *nm*
- bump width : 500 *nm*
- separation between tracks : 2500 *nm*

Results

Animations Rendered in MAYA 2.0

Conclusion

Fourier transform very powerful tool
Most general Illumination model in CG
He-Torrance special case

Experimental validation (?)

Future Work

Our Model

Wave Theory
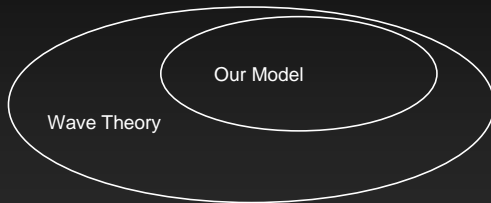
Multiple scattering, varying Fresnel coefficient, any distances, polarization, non-height field surfaces, subsurface scattering, etc.

Microfacet-based BRDFs

Michael Ashikhmin

University of Utah

1 Introduction

In this section of the notes we provide a brief overview of popular reflection models which are based on the principles of geometric optics. An equivalent term we use is "microfacet-based". This deviates slightly from traditional use of this last term which usually includes only analytical models and excludes models based on explicit simulation of light scattering from a surface. We do not see a need for such separation since in both cases the same model of the surface-light interaction (to be described later) is used. The emphasis of presentation will be put on practical aspects of the most popular general models (or very recent models which have a chance to become popular). No attempt is made to mention every published work. In particular, models which consider subsurface scattering are not covered at all. For a more complete survey one should consult [25]. Relatively few models in the domain we consider have been published since this review and we mention many of them. Reference lists of other papers included with these notes can also serve as a valuable resource if completeness is desired. We also will not discuss in detail computational efficiency of different models - this important topic is beyond the scope of this course. For software implementation, efficiency of the model can be approximately judged by its complexity (this is especially true for analytical models). Readers interested in hardware-assisted rendering with arbitrary BRDFs can consult [15] and references therein to get them started.

It is common to describe reflection properties of a surface by factoring BRDF into at least two parts: diffuse and specular reflectances (In addition, a mirror reflectance term is sometimes included). Most of the discussion in this section of the notes refers to the specular part of the BRDF. Unless specified overwise, the diffuse part is modeled by a Lambertian (constant) BRDF.

The initial title of this document was "Microfacet-based and empirical reflection models". After some consideration, however, it became clear that no general-purpose model currently used is truly empirical. (A major exception of the original Phong model and some of its derivatives will be discussed later.)

This statement needs some clarification. It is quite possible that multiple special-purpose (i.e. describing one particular material) and maybe even more general empirical reflection models are being created every year in universities and industry research labs. Unfortunately, these rarely stand a chance of being published in visible open literature. The main reason for this is the general climate in computer graphics research community which seems to effectively *require* a physical justification for the model, often leaving its practical usefulness as a secondary criteria. To the best of our knowledge, there has been no narrowly specialized empirical model published in recent years and all more general models do provide rather extensive physical justification. (It is interesting to contrast this with recent publication of several special-purpose wave optics based BRDFs [9, 28].)

We start this paper from a brief description of relevant physics of light-surface interaction, section 2. The primary goal of this section is to understand the range of applicability and physical limitations of any geometric optics based model. We then proceed with discussion of important classes of BRDFs in the remainder of the paper.
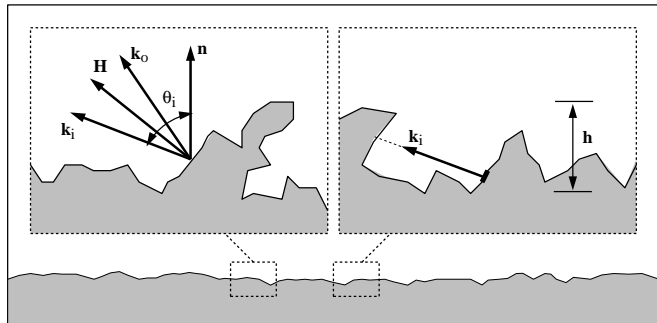
Figure 1: *Geometry of reflection. N is geometric surface normal, k_i is the direction to the light, k_o is the direction to the observer and H is half-vector between them. These three vectors share a plane which might not contain n. h is the characteristic height of surface microstructure. An example of shading is shown on the right.*

2 Physical foundations

The reason for non-trivial (not perfect mirror) reflection from a surface is its microstructure. Mirror reflection takes place if $h \cos \theta_i << \lambda$ where h is characteristic height of the surface microstructure features, θ_i is the incident angle and λ is light wavelength. By microstructure here we mean the fine structure of a surface on the scale smaller than anything we are interested in representing explicitly for the rendering of a given image. In other words, it's the geometry which is too small to be visible. To avoid any confusion, we explicitly state that geometry represented by bump maps does not constitute microgeometry for our purposes. The only effect the microstructure has is a change of the way surface reflects light. Note that this definition depends on particular viewing conditions and, in fact, the same geometry can be represented differently for rendering different frames or even in the same frame [4].

In this part of the notes will concentrate on models which do not take into account the wave nature of light and use only geometric optics. Ignoring wave effects immediately excludes materials which exhibit noticeable diffraction and interference effects. One could argue that such materials are in weak minority in everyday life. This is not the case, however: *every* material in theory exhibits wave effects under certain conditions, as we will show below.

Physically, geometric optics condition restricts the feature size of the surface microstructure from below to at least several wavelengths. A useful rule of thumb to estimate the strength of wave effects in a general setting is to remember that because of the diffraction on an object of size d light (a plane wave) deviates from its initial propagation direction by an angle of the order λ/d (small deviation angle is assumed for this approximation). Applied to our problem of reflection, one should keep in mind that close to grazing incidence the effective size of the surface microstructure will decrease ($d \approx h \cos \theta_i$). This is what makes it possible to observe interference effects on very coarse structures at near-grazing angles. Note that the condition to reliably ignore wave effects which can be written as $\lambda/h \cos \theta_i << 1$ (very small deviation angle)

is exactly the opposite to the mirror reflection condition. These two inequalities together give a range where non-trivial wave optics based models should be used. Accidentally, they also show that the most trivial reflection (perfect mirror) is a strictly wave optics effect, even though in practice it is always handled with geometric optics techniques. Another immediate result is that strictly speaking, one *never* can describe reflection by only geometric optics: even if under normal incidence geometric optics is adequate, close to grazing angle first wave effects and then mirror reflection prevail. Although we observed the full transition (geometric to wave to mirror) for a few surfaces, it is really quite unusual effect to worry much in practice. This is due to the fact that for most surfaces to which geometric optics can be applied under normal incidence, wave optics effects start to appear so extremely close to the grazing incidence that observing them would require very special conditions for these effects not to be completely obscured by other factors. An example of such special circumstances is looking at a semi-infinite plane (or extremely large polygon) from its edge. For the sake of completeness, we should also mentioned that if one wants to use a coherent light source (a laser) to light a scene, wave optics should be used regardless of the characteristic size of surface microstructure - even overwise "large" bumps can produce well known speckle patterns in laser reflections.

The choice of geometric optics as the tool for describing light-surface interaction completely describes the presentation of light as a flux of classical particles (commonly misnamed as "photons") propagating along straight paths (rays). It also specifies the way the direction of a single ray changes when it hits a flat surface. This is done through the well known geometric laws of reflection and refraction. One still needs to specify the surface description. Microfacet models assume that surface consists of a large number of small flat "micromirrors" (facets) each of which reflect light only in the specular direction. An extension of this view is considering microfacets to have some other reflection properties (not that of a perfect mirror). By computing the number of visible facets at appropriate orientations to transport light from the source to the viewer, one can determine the BRDF. Multiple analytical derivations of the final result of microfacet models are available [2, 29] and we do not repeat this process here. We will discuss this result in section 4 but will start from approaches based on direct simulation of light scattering by a surface done within the boundaries of geometric optics.

3   Non-analytical models

One very general approach to considering surface reflection is to split the problem into two parts. First, one explicitly models the microgeometry of the surface which reflection properties are of interest. With such geometric model it is possible to use common techniques such as raytracing to simulate the process of light scattering by the surface. Effectively, at this phase the microgeometry is "promoted" to the rank of "normal" geometry. This representation is only used during this preprocessing step and is never included as a part of actual scene to be rendered. The result of the simulation process is a numerical representation of the complete BRDF. During the second stage this result is usually presented in a way convenient to use during rendering. Some representations which have been explored include spherical harmonics [31], and, more recently, multiple cosine lobes [12], spherical wavelets [26] and neural networks [8]. A representative paper by Westin at al. [31] giving the details of the technique is included with the notes.

The main advantages of this approach over other methods are its unsurpassed generality and accuracy (within the limits of geometric optics). As long as surface microstructure can be modeled, the process will create a valid BRDF. No approximations are fundamentally necessary to succeed: the only penalty is potential run-time increase of the simulation step. It's completely up to the user whether to include multiple bounces of rays withing the microstructure, whether the elements of the surface are just small mirrors (as we specified them before) or they have more complex behavior (diffuse, Phong-like, transparent or any combination of these), the flexibility analytical models usually lack. If this elementary BRDF of a single facet is reciprocal and conserves energy, the result will automatically have these properties too. Depending on what effects are included in the simulation, the result can represent only specular part of BRDF (if, for example, the elements of the surface are treated as mirrors and only single ray bounces are considered) or the full BRDF.

There are two main disadvantages of non-analytic methods. The first one is that the surface has to be modeled. This requires direct human intervention, a good procedural technique or some measurement of the surface microstructure, which is not easy to obtain. The second disadvantage is the fact that the true result of the method is a numerical representation of a four dimensional function (BRDF). Direct use of these numbers is cumbersome and alternative more convenient representations can introduce artifacts, diminishing the accuracy. All representations are also much less compact than analytical models. Several methods to factor reflection models into lower dimensional functions have been proposed [10, 15, 22], but they are often designed with specific goals in mind (such as hardware rendering) and might not be as useful in a general setting while adding an extra preprocessing step to the technique.

4   Analytical Models

Analytical models are perceived as more convenient by most people since they do not require (or require very moderate) preprocessing and are usually easier to implement and use. The main result of microfacet theory can be loosely written in its most general form as

$$BRDF(k_i, k_o) = p(H)S(k_i, k_o)F((kH)) \qquad (1)$$

where k_i and k_o are light's in and out directions (vectors), H is the half-vector between the two and (kH) is a dot product. All vectors are assumed to be normalized. Different terms of this equation specify, in turn, the distribution of microfacet orientations (p), shadowing and masking effects (S, which we will simply call "shadowing term" and will include all other normalization functions in it) and Fresnel effects (F). Differences among available analytical microfacet models are primarily in how they treat the shadowing term, what kinds of distributions are allowed and whether the Fresnel term is included.

Almost every model mentioned in this section can be written in this form, regardless of whether it was originally developed through the apparatus of microfacet theory or even with microfacets in mind. (One exception is described in section 4.4). For some of these models one or even two of the last terms are simply set to one, which means that particular phenomenon is simply ignored by the model.

One should realize that the convenience of having a closed form solution comes for a price. To obtain such a solution, *every* model makes additional assumptions about the surface and its interaction with light. This always leads to restricting the class of surfaces handled by a given model and can produce non-physical artifacts such as lack of energy conservation. Most common assumptions are:

- only single bounce of light from microfacets is included in the analytical solution. To the best of our knowledge, this condition is employed by every model except [19]
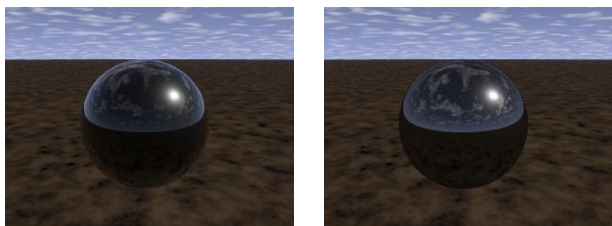
Figure 2: *Two dielectric spheres, rendered with (left) and without (right) Fresnel term. Images are courtesy of Stephen Westin.*

- the effect of multiple bounces is approximated by an additional Lambertian diffuse term. More recent models make an attempt to design a somewhat better diffuse function, but complex multiple bounce behavior (such as that of colored velvet) is still not well handled.

- specific conditions are imposed on the distribution of microfacets. These can range from very mild [2] to as severe as considering only one particular distribution and even some specific parameter range for it [30].

- as mentioned, one or two of the terms in equation 1 might be completely ignored and any or all of the three can be simplified to obtain analytic solution. The exact effect of this depends on the particular model and the degree of deviation from the reality.

We will now discuss the relative importance of different terms in equation 1.

4.1 Relative importance of different factors

By far the most important of the three factors in equation 1 is the microfacet distribution, $p(H)$. It gives the shape of the specular highlight and is at the core of any model. One should note that the exact physical interpretation of $p(H)$ can differ from model to model. For example, [19] uses slope-area probability distribution while [2] uses the distribution of the number of microfacets with given normal in its derivation. This subtle question is, of course, only relevant for models which explicitly use microfacet theory. For more simple models we just identify the dominant term in BRDF expression as the one corresponding to $p(H)$.

We argue that the presence or absence of the Fresnel term is the second most important factor for the visual appearance of the model. Many surfaces do exhibit noticeable Fresnel effects and ignoring them can introduce substantial deviations from expected appearance. There are two main Fresnel-term related effects: I) the difference in the amount of light being reflected by the surface for different incident angles (roughly, more reflection as one gets close to grazing incidence) and II) shifted (off-specular for single Phong-style highlight) position of the highlight maximum. For metals, the first effect is usually minor since they already reflect 80 percent of light or more at normal incidence. For dielectrics the effect is quite noticeable, see Figure 2. More examples are available in [3, 12]. Off-specular position of reflection maximum is really a result of the interplay between Fresnel term and most commonly used shapes of $p(H)$. It can be important if compared directly with measured (photographed) data but lack of this effect is usually not disturbing visually.

Inclusion of the full Fresnel expression found in physics books can lead to a substantial complication of the model and sizable computational penalty. However, it seems to be the case that fine details of the Fresnel term do not make much visual difference as long as the term itself is present. This observation has lead Schlick to

the development of an approximation [23, 24] the use of which we highly recommend. Approximating Fresnel term looks even more viable if one remembers that exact formulation depends on light polarization - something which is almost never considered in graphics. While one can use polarization- averaged Fresnel expression for single reflection, multiple reflections are very common in graphics and polarization information would be necessary to adequately deal with them if exact Fresnel term is used. With Schlick's formula, it should be possible to augment in an inexpensive way models lacking Fresnel term in their original formulation.

Finally, the term which accounts for shadowing effects, S, is visually probably the least important of the three (see also Figure 3 and discussion related to it below). For simple models, this term is simply "everything which is not covered by p and F". Its details are essential, however, to keep energy conserved by the model. This can be important in some rendering algorithms, such as some variations of raytracing which may rely on energy conservation to be able to terminate the rays. Other than that, the issue of good energy behavior is largely a matter of intellectual satisfaction and many popular models do not conserve energy under certain conditions. By "good behavior" we mean both not reflecting more energy than is incident and not uncontrollably loosing too much energy. In particular, most models have problems maintaining good energy behavior properties near the grazing angle. Some possible solutions to this problem for popular models are proposed in [13, 14]. More principle solution is provided by the model described in [2]. One should remember, however, that no geometric optics based model is adequate very close to grazing angle anyway (section 2), so this problem does not have a really satisfying solution within our domain.

The issue of having more (or less) energy than necessary reflected by the model is usually handled in practice by ad hoc adjustments to the total reflectance of the surface. Moreover, even systems credited for some of the most realistic images ever produced (such as Pixar's RenderMan) often do so by using non-physical adjustments to their lighting and reflection models [1]. Another factor which lessens the importance of having physically justified energy behavior is that most specular highlights, if computed in physically meaningful units, are out of range of common display devices and some type of tone mapping process is necessary. This is usually a nonlinear process which, while mostly preserving the shape of highlight, can change the visual perception of the total amount of reflected energy dramatically.

Everything which is said here about the relative importance of the terms in equation 1 is based on the author's personal experience only. To the best of my knowledge, no formal study has been done to objectively evaluate the importance for a human observer of different reflection effects included to or excluded from BRDF in a setting of general computer graphics imagery. "A sphere under a single light" images widespread in BRDF literature are designed to show reflection model differences and hardly qualify for such a study. The author is willing to speculate that under more general lighting conditions and with complex scene geometry minor differences between models (for example, models described in section 4.2) will be very hard to observe and the images will be almost identical visually.

It is also assumed in the above discussion that the terms are designed within certain reasonable limits. For example, Fresnel effects are not inverted (more reflection near normal incidence) or shadowing term does not make the energy reflected by a surface infinite (this actually can happen even for a very useful model near grazing angle [30], but it is probably not a good idea to tolerate such behavior in any other case).

4.2 Phong-style single highlight

Most common effect modeled by microfacet-based analytical models is a single specular highlight. The original Phong model is probably the only truly empirical model currently in wide use. In its original formulation it is not even a BRDF, just a shading function and some effort is needed to convert it to a BRDF [20]. Several modifications to this model have been proposed to improve its properties [11, 13]. The most popular modification of Phong shading term is due to Blinn and uses $p(H) \sim (Hn)^m$ where m is the parameter controlling the sharpness of specular highlight. With this modification, Phong shading function can be considered as a microfacet model where both shading and Fresnel effects are ignored.

It is interesting to note that Blinn's paper [6] actually introduced microfacet models originally developed in physics [29] to computer graphics and used three different distributions for $p(H)$, Phong's being just one of the three and not the one recommended. The paper did not contain any discussion of color and did not demonstrate very significant differences from Phong model, which is probably the reason why Cook and Torrance [7] are often given the credit for the first practical microfacet-based model in computer graphics. They use gaussian distribution of microfacets. Both papers used the original shadowing term from [29]. This shadowing expression is still the one most commonly used in graphics even though others are available (see [2] and references in it). One should remember that it was derived using geometrically inconsistent assumption of a surface consisting exclusively from randomly oriented V-grooves with length much greater than width. This inconsistency is the main reason for this shadowing term often not doing its main job, i.e. not providing full energy conservation [13, 14]. Another reason for this problem is that true gaussian distribution of microfacets has to contain additional angle-dependent normalization factors, which are almost never included, see Beckman's book [5] for details.

Ward [30] introduced a model designed without explicit use of microfacet theory but rather to fit experimental data. This model, however, can be represented by equation 1 with Fresnel term omited, anisotropic gaussian $p(H)$ and inverse square roots playing the role of the shadowing term. This is one of the two currently available simple models which include anisotropy and are reasonably Monte-Carlo friendly [3, 30]. One should keep in mind though that expressions for importance sampling presented Ward's paper only approximate the needed distribution. This model has become a very popular one, to the degree that original warning by its author about applicable parameter range for the gaussian distribution is now often ignored. Such carelessness can produce noticeable artifacts near grazing angles since the model's total reflected energy is infinite there. The reader is encouraged to look at the Ward's paper which is included with the notes for more details. Ward also advocated reverse-engineering of a reflection model by obsedvation of the scpecular highlight and even designed a simple device that was similar to a cup with a hole in the base and a light emitting diode inside it. If one puts the cup on a surface and looks trhough the hole, the reflection of the diode will have a certain shape one can match to a chart for Ward's model. (Greg Ward, Peter Shirley, personal communication, 2001). We describe below a similar in spirit approach with respect to the model presented in [2].

Another model for anisotropic reflection is that of Poulin and Fournier [21] which has been developed somewhat earlier. It models a surface as consisting of cylindrical bumps or grooves and proceeds first to derive expressions for shadowing and then the complete BRDF based on this particular microgeometry. This model is much more complicated than Ward's (the paper does not even provide a single formula for the BRDF).

Schlick investigated the problem of making single highlight models more efficient computationally [23, 24]. He starts from models derived using full microfacet theory and explicitly containing all three terms of equation 1. The result is several approximate

formulations for different terms of equation 1 all of which, especially the already mentioned approximation to Fresnel expression, can be used in other models as well.

More recently, Neumann et al. have developed several variations of single highlight analytic models [16–18]. They modify the analog of shadowing term to adjust the behavior of Phong model. An interesting example of the process of reflection model design is presented in [18] where one first ensures the needed properties and then try to preserve them while building up the model. Some of the results of Neumann et al. [16] were extended in [3] to obtain a model which includes anisotropy and Fresnel effects in a very simple formulation. A technical report which gives an extended version of this paper is included with the notes. This is the simplest model known to us which includes all major effects which are of interest to single highlight models: anisotropy, Fresnel effects, good energy behavior, Monte Carlo friendliness. It also provides a non-Lambertian diffuse term which can be used independently with any other model which includes Fresnel effects it its specular part.

In our experience, different models which include the same effects often produce visually extremely similar images for general scenes. As a result, many building blocks of different models can be used interchangeably with each other. An example is given on Figure 3 where the result of Neumann's metallic model [16] is first augmented with Fresnel term and this result is then compared with a newly assembled model in which we replaced $max((k_in), (k_on))$ with Ward's square roots. Then compared with the original Phong model, however, one can notice a difference due to the absence of Fresnel term (look at the ground plane), but even this difference is minor since all objects are made from metal (see section 4.1). Under special viewing/lighting conditions the differences might be more noticeable but these images support our speculation expressed near the end of section 4.1.

4.3 More general models

There is only a handful of models which attempt to describe analytically reflection which is more general than a single Phong-Ward style highlight. The main reason for this is that for physical reasons this simple behavior describes a great majority of the materials we see around us.

Lafortune et. al. [12] presented a generalization of single lobe models described in the previous section (the paper is included with the notes). They allowed multiple lobes to be fit to some external data. This work can also be used as a stand alone model if the lobes and their weightings are adjusted as free parameters and not given by numerical fits. The authors can also be credited for calling the attention of the graphics research community to the need for a non-Lambertian diffuse term. Such terms were later included with models described in [2, 3, 27]. The author of this section of the notes has no personal experience with this model. Lafortune provided the following comment when he was asked for insights about his work which are not widely known: "The most difficult aspect of this model appears to be the non-linear fitting to measurement data. Non-linear optimization often is something of an art. I haven't found any magical recipe so far; I use standard Levenberg-Marquardt, as described in the paper, and try to follow up on the convergence of the process." (Eric Lafortune, Personal communication, 2001). Note also that visually similar results can sometimes be obtain with a simpler model [3].

The most general analytic model developed so far is presented in [2]. A copy of this paper is included with the notes. The result follows quation 1 with only mild constrains on the type of microfacet distribution and a special shadowing term which enforces energy conservation. If strict energy conservation is not of primary importance, we noticed that, in agreement with our earlier observations in sections 4.1 and 4.2, we can interchange the shad-
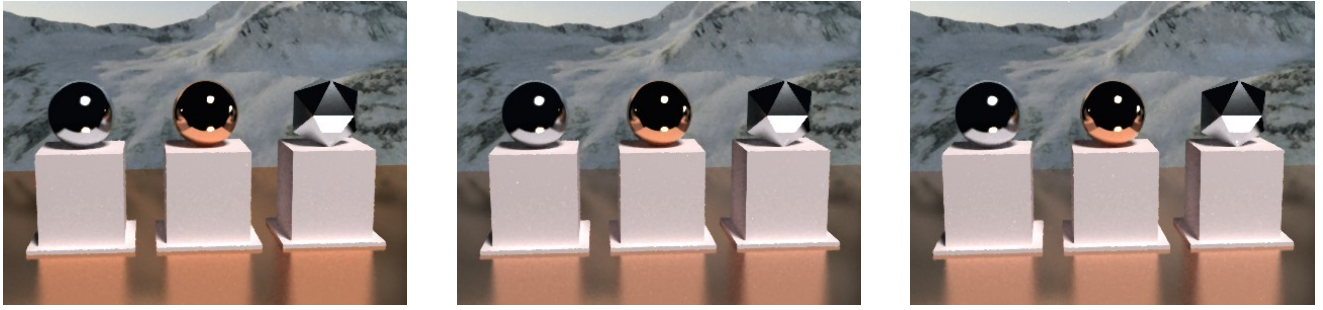
Figure 3: *Metallic objects in these images have the following reflection models: left - original Phong, center - Neumann augmented with Fresnel term, right - same as center, with $max((k_i n), (k_o n))$ replaced by $\sqrt{((k_i n)(k_o n))}$. Images are courtesy of Peter Shirley.*

owing function g (paper notations) computed for one distribution with that computed for another. For example, g corresponding to a usual gaussian distribution (which width should be approximately the same as that of the real distribution $p(H)$ used) could have been used to create images on figures 1 in the paper (and even 11, left) with very minor, if any, visual differences. This observation can improve practical utility of the model significantly by simplifying precomputation step to creating a library of a few g functions and using the one corresponding to the distribution closest to the one currently needed.

Another probable misconception about this model is that an analysis of the details of surface microgeometry is necessary to come up with analytic distributions of the needed form. This is only one way to go, however. Often, it is much easier to reverse-engineer the distribution and its parameters from the observed shape of the specular highlight. This is what has been done for cloth distributions and their parameters, as illustrated by Figure 4. A piece of cloth was wrapped around a cylinder and the highlights this configuration produced were studied. The double structure of satin highlight and slant of the velvet distributions are quite obvious. Rotation of the cloth by 90 degrees reveals anisotropic character of the reflection. In the paper, only major characteristics of the reflection were modeled. If desired, secondary effects (such as the dim highlight for satin visible on the bottom photograph) can be included in a similar way. Parameters of distributions are effectively read off from the positions of the highlights. Although this was done manually for the paper, one can imagine a simple automatic measurement system which does this job. Another interesting possibility is interactive design of a reflection model. Since our paper uses only two-dimensional function $p(H)$ as the input and the result usually conforms to intuitive expectations, a human user can design this function which would have been excessively difficult to do for a full three or four dimensional BRDF. Figure 1 in the paper scratches the surface of possible effects.

4.4   Generalized Lambertian model

A model developed by Oren and Nayar [19] has a special place among microfacet models. First of all, the model concentrates on modifying the diffuse reflection of the surface rather than specular behavior. It also includes interreflections (multiple ray bounces) among the microfacets. It assumes gaussian distribution of microfacets and applies a variation of the V-grove shadowing term. Since the model assumes diffuse microfacets, there is no Fresnel term. The result of all these differences is that equation 1 is not followed. The (rather complicated) final formula of the model is then approximated by a more manageable expression. Visual results of the model are close to observed appearance of clay and dust but overall are not very far from the Lambertian reflectance.
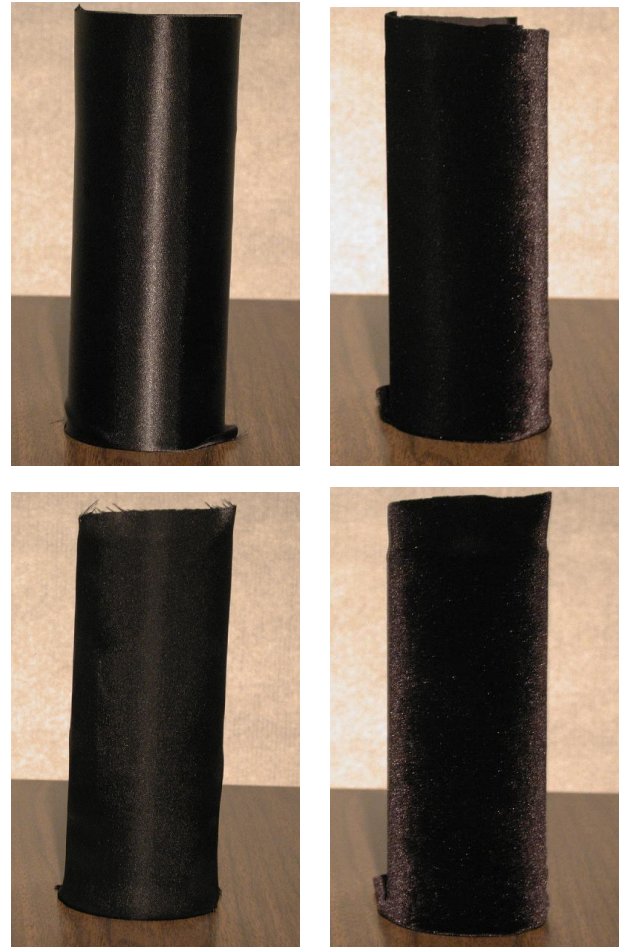
Figure 4:   *Photographs of cloth-wrapped cylinders. Left: satin Right: velvet. Below: same pieces of cloth after 90 degrees rotation. The light comes from the front and above (notice no shadows of the cylinder on the table). Lighting conditions, including intensity, are identical for all images.*

5 Conclusion and acknowledgements

In this paper we presented an overview of popular geometric optics based reflection models. An attempt was made to concentrate on practical issues which are not always covered in the literature. The author greatly appreciates the help of everyone who suggested and contributed papers for all parts of these notes, allowed their images to be used and provided comments about their work. In connection with this section of the notes in particular, I would like to thank Eric Lafortune, Peter Shirley, Greg Ward and Stephen Westin. However, much of the material in this paper is based on the author's personal experiences and is, therefore, subjective. The author is the only one responsible for the opinions expressed in this paper which might not coincide with those of other course participants or people mentioned in this section.

References

[1] APODACA, A., AND GRITZ, L. *Advanced RenderMan*. Morgan Kaufman, San Francisco, 2000.

[2] ASHIKHMIN, M., PREMOZE, S., AND SHIRLEY, P. A microfacet-based brdf generator. *Proceedings of SIGGRAPH 2000* (July 2000), 65–74.

[3] ASHIKHMIN, M., AND SHIRLEY, P. An anisotropic phong brdf model. *Journal of Graphics Tools 5*, 2 (2000), 25–32.

[4] BECKER, B. G., AND MAX, N. L. Smooth transitions between bump rendering algorithms. *SIGGRAPH 93 Proceedings* (August 1993), 183–190.

[5] BECKMANN, P., AND SPIZZICHINO, A. *The Scattering of Electromagnetic Waves from Rough Surfaces*. MacMillan, 1963.

[6] BLINN, J. F. Models of light reflection for computer synthesized pictures. *Computer Graphics (Proceedings of SIGGRAPH 77) 11*, 2 (July 1977), 192–198.

[7] COOK, R. L., AND TORRANCE, K. E. A reflectance model for computer graphics. *SIGGRAPH 81 Proceedings* (August 1981), 307–316.

[8] GARGAN, D., AND NEELAMKAVIL, F. Approximating reflectance functions using neural networks. In *Proceedings of the 9th Eurographics Workshop on Rendering* (1998), pp. 23–34.

[9] ICART, I., AND ARQUES, D. A physically-based brdf model for multilayer systems with uncorrelated rough boundaries. In *Proceedings of the Eleventh Eurographics Workshop on Rendering* (June 2000), pp. 353–365.

[10] KAUTZ, J., AND MCCOOL, M. Interactive rendering with arbitrary brdfs using separable approximations. In *Proceedings of the 10th Eurographics Workshop on Rendering* (1999), pp. 247–260.

[11] LAFORTUNE, E. P., AND WILLEMS, Y. D. Using the modified phong BRDF for physically based rendering. Tech. Rep. CW197, Computer Science Department, K.U.Leuven, November 1994.

[12] LAFORTUNE, E. P. F., FOO, S.-C., TORRANCE, K. E., AND GREENBERG, D. P. Non-linear approximation of reflectance functions. *Proceedings of SIGGRAPH 97* (August 1997), 117–126.

[13] LEWIS, R. Making shaders more physically plausible. In *Fourth Eurographics Workshop on Rendering* (1993), M. F. Cohen, C. Puech, and F. Sillion, Eds., pp. 47–62. Paris, France, June 1993.

[14] LEWIS, R. Making shaders more physically plausible. *Computer Graphics Forum 13*, 2 (June 1994), 109–120.

[15] MCCOOL, M., ANG, J., AND AHMAD, A. Homomorphic factorization of brdfs for high-performance rendering. *SIGGRAPH 01 proceedings* (August 2001).

[16] NEUMANN, L., NEUMANN, A., AND SZIRNAY-KALOS, L. Compact metallic reflectance models. *Computer Graphics Forum 18*, 3 (1999), 161–171.

[17] NEUMANN, L., NEUMANN, A., AND SZIRNAY-KALOS, L. Reflectance models by pumping up the albedo function. *Machine Graphics and Vision 8*, 1 (1999), 3–17.

[18] NEUMANN, L., NEUMANN, A., AND SZIRNAY-KALOS, L. Reflectance models with fast importance sampling. *Computer Graphics Forum 18*, 4 (1999), 249–270.

[19] OREN, M., AND NAYAR, S. K. Generalization of lambert's reflectance model. In *Proceedings of SIGGRAPH '94* (July 1994), A. Glassner, Ed., pp. 239–246.

[20] PHONG, B.-T. Illumination for computer generated images. *Communications of the ACM 18*, 6 (June 1975), 311–317.

[21] POULIN, P., AND FOURNIER, A. A model for anisotropic reflection. *SIGGRAPH 90 Proceedings* (August 1990), 267–282.

[22] RUSINKIEWICZ, S. A new change of variables for efficient brdf representation. In *Proceedings of the 9th Eurographics Workshop on Rendering* (1998), pp. 11–22.

[23] SCHLICK, C. A customizable reflectance model for everyday rendering. In *Fourth Eurographics Workshop on Rendering* (June 1993), pp. 73–84.

[24] SCHLICK, C. An inexpensive BRDF model for physically-based rendering. *Computer Graphics Forum 13*, 3 (1994), 233—246.

[25] SCHLICK, C. A survey of shading and reflectance models for computer graphics. *Computer Graphics Forum 13*, 2 (June 1994), 121–132. Paris, France, June 1993.

[26] SCHRODER, P., AND SWELDENS, W. Spherical wavelets: efficiently representing functions on the sphere. *Proceedings of SIGGRAPH 95* (1995), 161–172.

[27] SHIRLEY, P., HU, H., SMITS, B., AND LAFORTUNE, E. A practitioners' assessment of light reflection models. In *Pacific Graphics* (October 1997), pp. 40–49.

[28] SUN, Y., FRACCHIA, F., DREW, M., AND CALVERT, T. Rendering iridescent colors of optical disks. In *Proceedings of the Eleventh Eurographics Workshop on Rendering* (June 2000), pp. 341–353.

[29] TORRANCE, K. E., AND SPARROW, E. M. Theory for off-specular reflection from roughened surfaces. *Journal of Optical Society of America 57*, 9 (1967), 1105–1114.

[30] WARD, G. J. Measuring and modeling anisotropic reflection. *SIGGRAPH 92 proceedings* (July 1992), 265–272.

[31] WESTIN, S. H., ARVO, J. R., AND TORRANCE, K. E. Predicting reflectance functions from complex surfaces. *SIGGRAPH 92 proceedings* (July 1992), 255–264.

Illumination for Computer Generated Pictures

Bui Tuong Phong
University of Utah

The quality of computer generated images of three-dimensional scenes depends on the shading technique used to paint the objects on the cathode-ray tube screen. The shading algorithm itself depends in part on the method for modeling the object, which also determines the hidden surface algorithm. The various methods of object modeling, shading, and hidden surface removal are thus strongly interconnected. Several shading techniques corresponding to different methods of object modeling and the related hidden surface algorithms are presented here. Human visual perception and the fundamental laws of optics are considered in the development of a shading rule that provides better quality and increased realism in generated images.

Key Words and Phrases: computer graphics, graphic display, shading, hidden surface removal.

CR Categories: 3.26, 3.41, 8.2

Introduction

This paper describes several approaches to the production of shaded pictures of solid objects. In the past decade, we have witnessed the development of a number of systems for the rendering of solid objects by computer. The two principal problems encountered in the design of these systems are the elimination of the hidden

parts and the shading of the objects. Until now, most effort has been spent in the search for fast hidden surface removal algorithms. With the development of these algorithms, the programs that produce pictures are becoming remarkably fast, and we may now turn to the search for algorithms to enhance the quality of these pictures.

In trying to improve the quality of the synthetic images, we do not expect to be able to display the object exactly as it would appear in reality, with texture, overcast shadows, etc. We hope only to display an image that approximates the real object closely enough to provide a certain degree of realism. This involves some understanding of the fundamental properties of the human visual system. Unlike a photograph of a real world scene, a computer generated shaded picture is made from a numerical model, which is stored in the computer as an objective description. When an image is then generated from this model, the human visual system makes the final subjective analysis. Obtaining a close image correspondence to the eye's subjective interpretation of the real object is then the goal. The computer system can be compared to an artist who paints an object from its description and not from direct observation of the object. But unlike the artist, who can correct the painting if it does not look right to him, the computer that generates the picture does not receive feedback about the quality of the synthetic images, because the human visual system is the final receptor.

This is a subjective domain. We must at the outset define the degree of realism we wish to attain, and fix certain goals to be accomplished. Among these goals are:
1. "Real time" display of dynamic color pictures of three-dimensional objects. A real time display system is one capable of generating pictures at the rate of at least 30 frames a second.
2. Representation of objects made of smooth curved surfaces.
3. Elimination or attenuation of the effects of digital sampling techniques.
The most important consideration in trying to attain these goals is the object modeling technique.

Existing Shading Techniques

Methods of Object Modeling

Image quality depends directly on the effectiveness of the shading algorithm, which in turn depends on the method of modeling the object. Two principal methods of object description are commonly used:
1. Surface definition using mathematical equations.
2. Surface approximation by planar polygonal mosaic.

Several systems have been implemented to remove hidden parts for mathematically defined curved surfaces [1, 2, 3, 4, 5]. With these systems, exact information at each point of the surface can be obtained, and the result-

ing computer generated pictures are most realistic. The class of possible surfaces is restricted, however, and the computation time needed to remove the hidden parts and to perform shading is very large. Up to the present time, these systems have usually considered the class of surfaces represented by quadric patches. Although higher degree surfaces are desirable and are sometimes necessary to model an object, they have not been taken into consideration due to an increase in computation time to remove hidden surfaces and to perform shading computations. Even when only quadric surfaces are considered, the implementation of a real time display system using this type of model is too expensive and complex.

A simple method of representing curved surfaces and objects of arbitrary shape is to approximate the surfaces with small planar polygons; for example, a cone might be represented as shown in Figure 1. This type of representation has the advantage that it avoids the problem, posed by mathematically curved surface approaches, of solving higher order equations.

Planar approximation also offers the only means of reducing hidden surface computation to within reasonable bounds, without restricting the class of surfaces that can be represented. For this reason, all recent attempts to devise fast hidden surface algorithms have been based on the use of this approximation for curved surfaces; these algorithms have been summarized and classified by Sutherland et al. [6]. The next section discusses their influence on the way shading is computed.

While planar approximation greatly simplifies hidden surface removal, it introduces several major problems in the generation of a realistic displayed image. One of these is the *contour edge* problem: the outline or silhouette of a polygonally approximated object is itself a polygon, not a smooth curve. The other problem is that of shading the polygons in a realistic manner. This paper is concerned with the shading problem; the contour edge problem is discussed by the author and F.C. Crow in [7].

Influence of Hidden Surface Algorithms

The order in which a hidden surface algorithm computes visible information has a decided influence on the way shading is performed. For example Warnock, who developed one of the first such algorithms [8], computed display data by a binary subdivision process: this meant that the order of generating display data was largely independent both of the order of scanning the display and of the order of the polygons in memory. This made it difficult to perform effective shading on curved objects.

The two major advances in the development of fast hidden surface algorithms have been made by Watkins [9] and by Newell, Newell, and Sancha [10]. Watkins generates the displayed picture scan line by scan line. On each scan line he computes which polygons intersect the scan line, and then computes the visible *segment* of each polygon, where this segment is the visible strip of
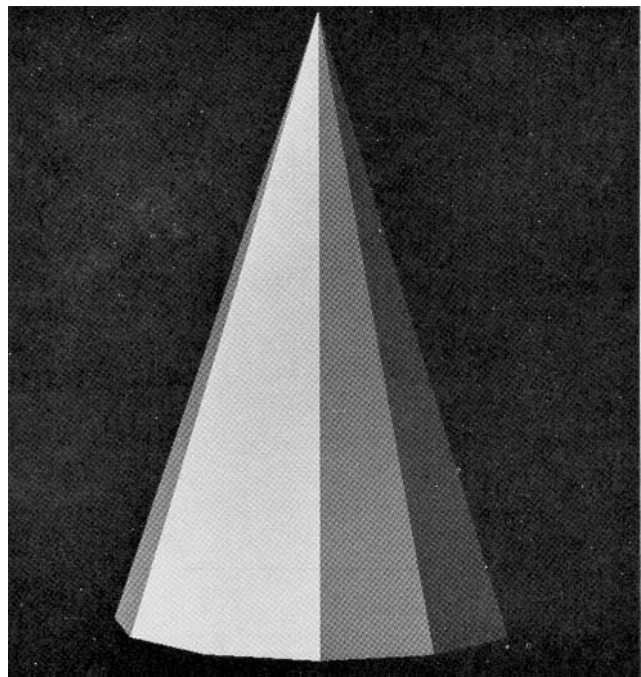
the polygon, one screen resolution unit in height, that lies on the scan line.

Newell, Newell, and Sancha adopt a different approach, using a *frame buffer* into which the object is painted, face by face. The hidden surface problem is solved by painting the farthest face first, and the nearest last. Each face is painted scan line by scan line, starting at the top of the face.

From the shading aspect, the important attribute of these algorithms is that they both generate information scan line by scan line in order to display the faces of an object. This information is in the form of segments, one screen resolution unit high, on which the shading computation may then be performed. The main differences between the algorithms, from the point of view of shading, are (a) the order in which the segments are generated, and (b) the fact that Watkins generates each screen dot only once, whereas the Newell-Sancha algorithm may overwrite the same dot several times.

Shading with the Polyhedral Model

When planar polygons are used to model an object, it is customary to shade the object by using the *normal vectors* to the polygons. The shading of each point on a polygon is then the product of a shading coefficient for the polygon and the cosine of the angle between the polygon normal and the direction of incident light. This cosine relationship is known in optics as the "cosine law," and allows us to compute the shading S_p for a polygon p as

$$S_p = C_p \cos(i), \tag{1}$$

where C_p is the reflection coefficient of the material of p relative to the incident wavelength, and i is the incident angle.

Fig. 2. An example of the use of Newell, Newell, and Sancha's shading technique, showing transparency and highlight effects.
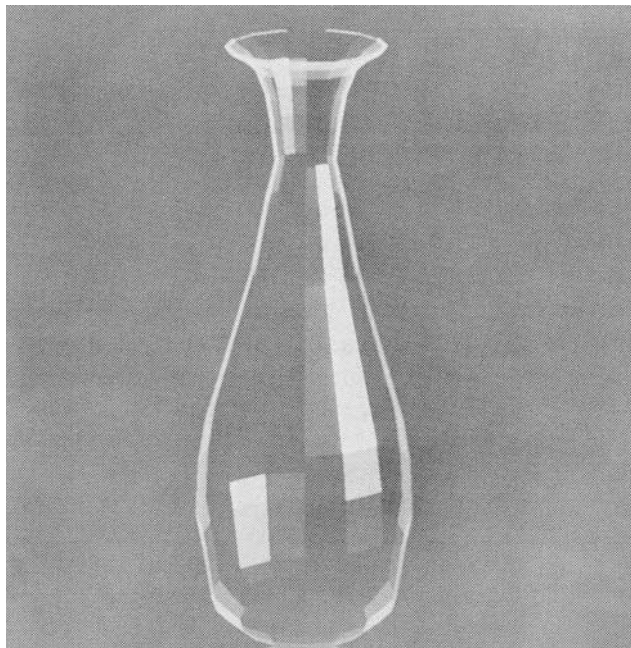
Fig. 3. Computation of the shading at point R using the Gouraud method. There are two successive linear interpolations: (1) across polygon edges, i.e. P between A and B, Q between A and D; and (2) along the scan line, i.e. R between P and Q.
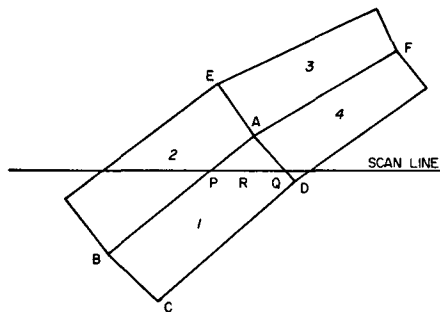
Fig. 4. Gouraud shading, applied to approximated cone of Fig. 1.
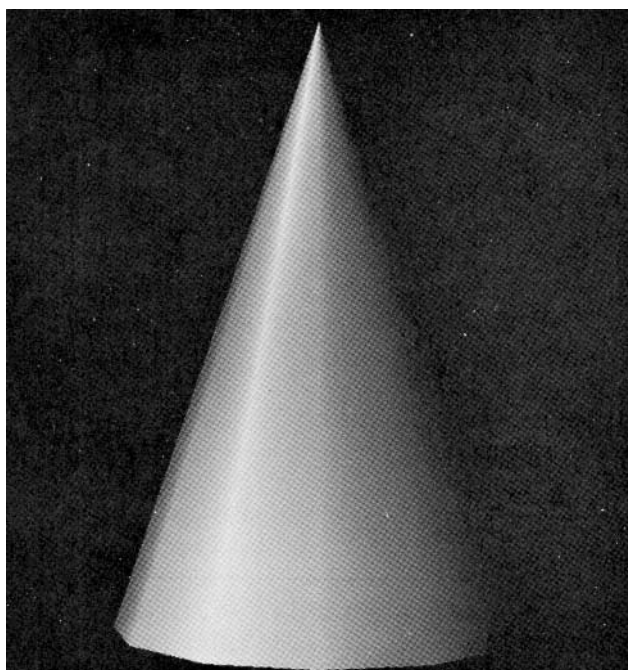
This shading offers only a very rough approximation of the true physical effect. It does not allow for any of the *specular* properties of the material, i.e. the ability of the material to generate highlights by reflection from its outer surface, and the position of the observer, which is ignored. A more serious drawback to this method, however, is the poor effect when using it to display smooth curved surfaces. The cosine law rule is appropriate for objects that are properly modeled with planar surfaces, such as boxes, buildings, etc., but it is inappropriate for smoothly curved surfaces such as automobile bodies. This does not mean, however, that we should abandon the use of such a polygon-oriented shading rule and search for a different rule for curved surfaces. Recent research in shading techniques demonstrates that significant results can be achieved by using the basic shading rule of eq. (1) and modifying the results to reduce the discontinuities in shading between adjacent polygons.

1. Warnock's shading. As three-dimensional objects are projected onto the cathode-ray tube screen, the depth sensation is lost, and the images of those objects appear flat. In order to restore the depth sensation, two effects were simulated by Warnock:

1. Decreasing intensity of the reflected light from the object with the distance between the light source and the object.

2. Highlights created by specular reflection.

Warnock placed the light source and the eye at the same position, so that the shading function was the sum of two terms, one for the normal "cosine" law, and the other term for the specularly reflected light. The resulting pictures have several desirable attributes; for example, identical parallel faces, located differently in space, will be shaded at different intensities, and facets which face directly toward the light source are brighter than adjacent facets facing slightly away from the incident light. However, the polygonal model gives a discontinuity in shading between faces of an approximated curved surface. When a curved surface is displayed, the smoothness of the curved surface is destroyed by this discontinuity. This is clearly visible in Figure 1.

2. Newell, Newell, and Sancha's shading. Newell, Newell, and Sancha presented some ideas on creating transparency and highlights. From observations in the real world, they found that highlights are created not only by the incident light source but also by the reflection of light from other objects in the scene; this is especially true in the case of objects made of highly reflective or transparent materials. In the Newell-Sancha model, curved surfaces are approximated with planar polygons. Unfortunately, the ability to generate highlights is severely limited due to the inability to vary light intensity over the surface of any single polygon. This problem is apparent in Figure 2.

3. Gouraud's shading. While working on a technique to represent curved objects made of "Coons surfaces"

313

Communications
of
the ACM

June 1975
Volume 18
Number 6

or "Bezier patches," Gouraud [11] developed an algorithm to shade curved surfaces. With his algorithm, a surface represented by a patch is approximated by polygonal planar facets. Gouraud computes information about the curvature of the surface at each vertex of each of these facets. From the curvature, a shade intensity is computed and retained. For example, the shade intensity may be computed for each vertex using eq. (1), with i as the angle between the incident light and the normal to the surface at this vertex. When the surface is displayed, this shade intensity is linearly interpolated along the edge between adjacent pairs of vertices of the object. The shade at a point on the surface is also a linear interpolation of the shade along a scan line between intersections of the edges with a plane passing through the scan line (Figure 3). This very simple method gives a continuous gradation of shade over the entire surface, which in most cases restores the smooth appearance. An example of Gouraud's shading is shown in Figure 4.

With the introduction of the Gouraud smooth shading technique, the quality of computer-generated images improved sufficiently to allow representation of a large variety of objects with great realism. Problems still exist, however, one of which is the apparent discontinuity across polygon edges. On surfaces with a high component of specular reflection, highlights are often inappropriately shaped, since they depend upon the disposition and shape of the polygons used to approximate a curved surface and not upon the curvature of the object surface itself. The shading of a surface in motion (in a computer generated film) has annoying frame to frame discontinuities due to the changing orientation of the polygons describing the surface. Also the shading algorithms are not invariant under rotation.

Frame-to-frame discontinuities of shade in a computer generated film are illustrated in the following situation. A curved surface is approximated with planar facets. When this surface is in motion, all the facets which are perpendicular to the direction of the light take on a uniform shade. In the next frame the motion of the object brings these facets into a different orientation toward the light, and the intensity of the shade across their surfaces varies continuously from one end to the other. Thus the surface appears to change from one with highlights to one of uniform shade. Moreover, the position of these highlights is not steady from frame to frame as the object rotates.

Mach Band Effect

Many of the shading problems associated with planar approximation of curved surfaces are the result of the discontinuities at polygon boundaries. One might expect that these problems could be avoided by reducing the size of the polygons. This would be undesirable, of course, since it would increase the number of polygons and hence would increase both the memory requirements for storing the model and the time for hidden surface removal.

Fig. 5. Normal at a point along an edge.   Fig. 6. Shading at a point.
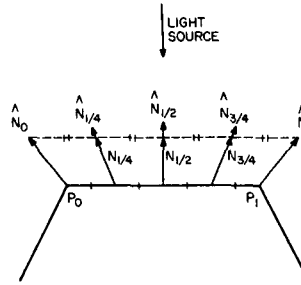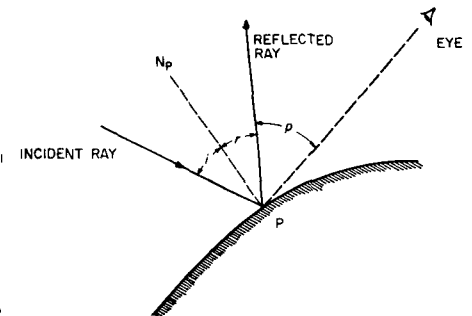
Unfortunately, because of visual perception effects, the reduction of polygon size is not as beneficial as might be expected. The particular effect responsible is the *Mach Band* effect. Mach established the following principle:

Wherever the light-intensity curve of an illuminated surface (the light intensity of which varies in only one direction) has a concave or convex flection with respect to the axis of the abscissa, that particular place appears brighter or darker, respectively, than its surroundings [E. Mach, 1865].

Whenever the slope of the light intensity curve changes, this effect appears. The extent to which it is noticeable depends upon the magnitude of the curvature change, but the effect itself is always present.

Without the Mach Band effect, one might hope to achieve accurate shading by reducing the size of polygons. Unfortunately the eye enhances the discontinuities over polygon edges, creating undesired areas of apparent brightness along the edges. Therefore unless the size of the displayed facets is shrunk to a resolution point, increasing the number of facets does not solve the problem. Using the Gouraud method to interpolate the shade linearly between vertices, the discontinuities of the shading function disappear, but the Mach Band effect is visible where the slope of the shading function changes. This can be seen in Figure 4. The subjective discontinuity of shade at the edges due to the Mach Band effect then destroys the smooth appearance of the curved surface.

A better shading rule is therefore proposed for displaying curved surfaces described by planar polygons. This new technique requires the computation of the normal to the displayed surface at each point. It is therefore more expensive in computation than Gouraud's technique; but the quality of the resulting picture, and the accuracy of the displayed highlights, is much improved.

Using a Physical Model

Specular Reflection

If the goal in shading a computer-synthesized image is to simulate a real physical object, then the shading model should in some way imitate real physical shading situations. Clearly the model of eq. (1) does not accomplish this. As mentioned before, it completely

314

Fig. 7(a). Determination of the reflected light.
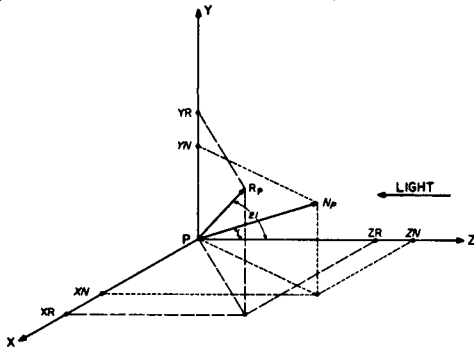
Fig. 7(b). Projections of the reflected light.
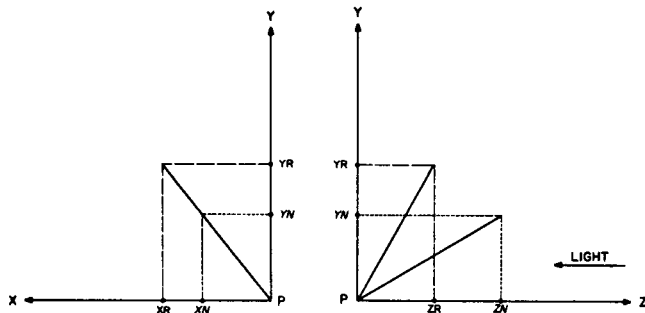
ignores both the position of the observer and the specular properties of the object. Even with the improvements introduced by Gouraud, which provide remarkably better shading, these properties are still ignored.

The first step in accounting for the specular properties of objects and the position of the observer is to determine the normal to the surface at each point to be shaded, i.e. at each point where a picture element of the raster display projects onto the surface. It is only with this knowledge that information about the direction of reflected rays can be acquired, and only with this information can we model the specular properties of objects. It is evident from the preceding discussion, however, that our polyhedral model provides information about normals only at the vertices of polygons. Thus the first step in improving our shading model is to devise a way to obtain the normal to the surface for each raster unit.

Computation of the Normal at a Point on the Surface

The normal at each vertex can be approximated by either one of the methods described by Gouraud [10]. It is now necessary to define the normal to the surface along the edges and at a point on the surface of a polygon.

The normal to the surface at a point along the edge of a polygonal model is the result of a linear interpolation to the normals at the two vertices of that edge. An example is given in Figure 5: the normal N_t to the surface at a point between the two vertices P_0 and P_1 is computed as follows:

$$N_t = tN_1 + (1-t)N_0, \qquad (2)$$

where $t = 0$ at N_0 and $t = 1$ at N_1.

The determination of the normal at a point on the

surface of a polygon is achieved in the same way as the computation of the shading at that point with the Gouraud technique. The normal to the visible surface at a point located between two edges is the linear interpolation of the normals at the intersections of these two edges with a scan plane passing through the point under consideration. Note that the general surface normal is quadratically related to the vertex normal.

From the approximated normal at a point, a shading function determines the shading value at that point.

The Shading Function Model

In computer graphics, a shading function is defined as a function which yields the intensity value of each point on the body of an object from the characteristics of the light source, the object, and the position of the observer.

Taking into consideration that the light received by the eye is provided one part by the diffuse reflection and one part by the specular reflection of the incident light, the shading at point P (Figure 6) on an object can be computed as:

$$S_p = C_p[\cos(i)(1-d)+d] + W(i)[\cos(s)]^n, \qquad (3)$$

where:

C_p is the reflection coefficient of the object at point P for a certain wavelength.

i is the incident angle.

d is the environmental diffuse reflection coefficient.

$W(i)$ is a function which gives the ratio of the specular reflected light and the incident light as a function of the incident angle i.

s is the angle between the direction of the reflected light and the line of sight.

n is a power which models the specular reflected light for each material.

The function $W(i)$ and the power n express the specular reflection characteristics of a material. For a highly reflective material, the values of both $W(i)$ and n are large. The range of $W(i)$ is between 10 and 80 percent, and n varies from 1 to 10. These numbers are empirically adjusted for the picture, and no physical justifications are made. In order to simplify the model, and thereby the computation of the terms $\cos(i)$ and $\cos(s)$ of formula (3), it is assumed that:

1. The light source is located at infinity; that is, the light rays are parallel.

2. The eye is also removed to infinity.

With these two considerations, the values of $\cos(i)$ and $\cos(s)$ of the shading function in (3) can be rewritten as: $\cos(i) = kN_p / |N_p|$ and $\cos(s) = uR_p / |R_p|$ where k and u are respectively the unit vectors in the direction of the light and the line of sight, N_p is the normal vector at P, and R_p is the reflected light vector at P.

The quantity $kN_p / |N_p|$ can be referred to as the projection of a normalized vector N_p on an axis parallel to the direction of the light. If $|N_p|$ is unity, the previous

315

Communications   June 1975
of   Volume 18
the ACM   Number 6

Fig. 8. Improved shading, applied to approximated cone of Fig. 1.

Fig. 9. Improved shading, applied to the example of Figure 2.
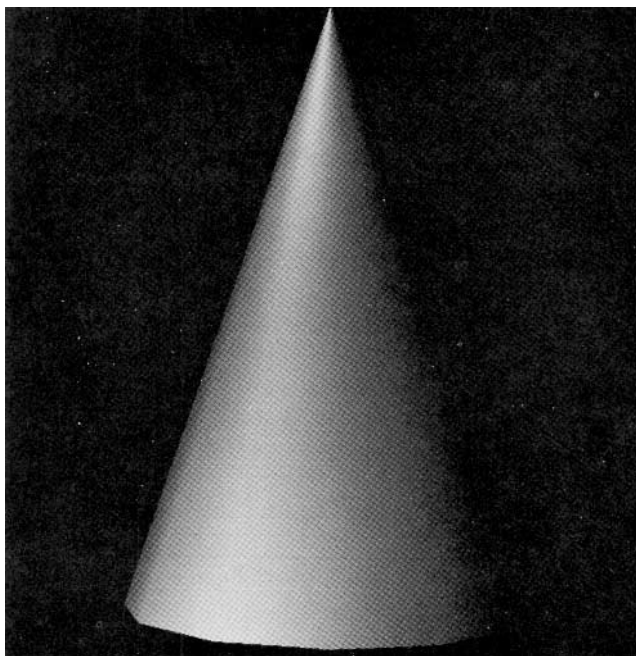
quantity is one component of the vector N_p in a coordinate system where the direction of light is parallel to one axis. In this case, the quantity $uR_p / |R_p|$ can be obtained directly from the vector N_p in the following way.

Let us consider a Cartesian coordinate system having the origin located at point P and having the z axis parallel to the light but opposite in direction (Fig. 7(a)).

We have the following assumptions about the model:

1. The normalized vector N_p makes an angle i with the z axis, and the reflected light vector R_p makes an angle $2i$ with the same axis.

2. Only incident angles less than or equal to 90 degrees are considered in the shading computation. For a greater angle, this means that the light source is behind the front surface. In the case where a view of the back surface is desired when it is visible, it can be assumed that the normal will always point toward the light source.

3. If k is the unit vector along the PZ axis, then by simple geometry, it may be shown that the three vectors k, N_p, and R_p are coplanar.

4. The two vectors N_p and R_p are of unit length.

From assumption (3), the projections of the vectors N_p and R_p onto the plane defined by (PX, PY) are merged into a line segment (Figure 7(b)). Therefore,

$$X_r/Y_r = X_n/Y_n, \qquad (4)$$

where X_r, X_n, Y_r, and Y_n are respectively the components of R_p and N_p in the x and y directions.

From assumptions (1) and (2), the component Z_n of N_p is:

$$Z_n = \cos(i), \qquad (5)$$

where $0 \leq i \leq 90$ degrees.

By simple trigonometry, we obtain the following expressions:

$$Z_r = \cos(2i) = 2[\cos(i)]^2 - 1 = 2Z_n^2 - 1, \qquad (6)$$

$$X_r^2 + Y_r^2 = [\sin(2i)]^2 = 1 - [\cos(2i)]^2. \qquad (7)$$

From (4) and (7), we obtain:

$$X_r = 2Z_n X_n, \quad Y_r = 2Z_n Y_n, \quad 0 \leq Z_n \leq 1.$$

The three components of R_p are then known in the light source coordinate system. The projection of the vector R_p onto the z-axis of the eye coordinate system may be found by a simple dot product of the reflected vector with this z-axis. The component of R_p on an axis parallel to the line of sight is the value of the cosine of the angle between the reflected light and the line of sight. The value of this cosine will be used in the simulation of the specular reflection.

This method of calculating the direction of the reflected light for each point from the orientation of the normal is preferred over the computation of the reflected light vector at vertices and the subsequent interpolation of them in the same way as the normal. It is faster and it requires less storage space than the interpolation scheme.

With the described method, the shading of a point is computed from the orientation of the approximated normal; it is not a linear interpolation of the shading values at the vertices. Therefore, a better approximation of the curvature of the surface is obtained, and highlights due to the simulation of specular reflection are properly rendered. Examples of application of the shading technique are shown in Figures 8 and 9. Figure 10 compares a display generated by this technique with a photograph of a real object.

316

Fig. 10(a). A sphere displayed with the improved shading.
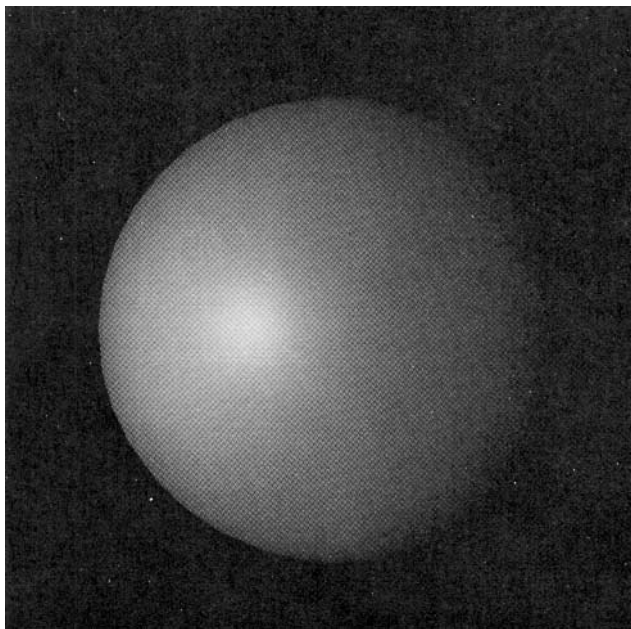
Fig. 10(b). A photograph of a real sphere.
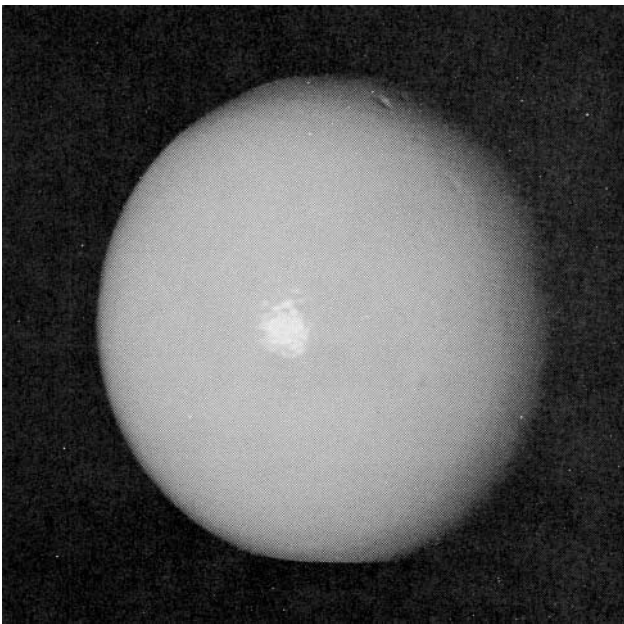
Conclusion

The linear interpolation scheme used here to approximate the orientation of the normal does not guarantee a continuous first derivative of the shading function across an edge of a polygonal model. In extreme cases where there is an abrupt change in the orientation of two adjacent polygons along a common edge, the subjective brightness due to the Mach Band effect will be visible along this edge. However, this effect is much less visible in the described model than in the Gouraud smooth shading model. Also, an interesting fact discussed previously on Mach Band effect shows

that this effect is visible whenever there is a great change in the slope of the intensity distribution curve, even if the curve has a continuous first derivative. When a higher degree interpolation curve is used, it will make the presence of the edges unnoticeable, although it will still give some Mach Band effect.

When a comparison was made of pictures of the same object generated with different shading techniques, it was found that little difference existed between pictures generated with the new shading and the ones created with a cubic interpolant curve for the shading computation. Furthermore, as time is the critical factor in a real time dynamic picture display system, the use of a high degree interpolation curve does not seem to be possible at the moment with the current techniques to compute the coefficients of such a function.

A hardware implementation of this shading model would of course require more hardware than the simpler Gouraud method. The Gouraud model needs one interpolator for the shading function. It must compute a new shading value for each raster unit, and hence must be very high speed to drive a real time display. The model proposed here requires three of these interpolators operating in parallel. In addition, since the results of the interpolation do not yield a unit vector, and since eqs. (6), (7), and (8) require a unit normal vector, some extra hardware is necessary to "normalize" the outputs of the interpolators. This requires a very fast mechanism for obtaining square roots. None of these problems is too difficult to solve; and judging from the improvements in image quality obtained using the new model, it may well be worth the extra expense to provide such hardware in applications for which real time display is important.

References
1. MAGI, Mathematical Applications Group Inc. 3-D simulated graphics. *Datamation 14* (Feb. 1968), 69.
2. Comba, P.G. A procedure of detecting intersections of three-dimensional objects. Rep. 39,020, IBM New York Scientific Center, Jan. 1967.
3. Weiss, R.A. BE VISION, a package of IBM 7090 FORTRAN programs to draw orthographic views of combinations of plane and quadric surfaces. *J. ACM 13*, 2 (Apr. 1966), 194–204.
4. Mahl, R. Visible surface algorithm for quadric patches. *IEEE Trans. C-21*, (Jan. 1972), 1–4.
5. Catmull, E.E. A subdivision algorithm for computer display of curved surfaces. Ph.D th., Dep. of Comput. Sci., U. of Utah.
6. Sutherland, I.E., Sproull, R.F., and Schumacker, R.A. A characterization of ten-hidden surface algorithms. *Computing Surveys 6* (Mar. 1974), 1–56.
7. Bui Tuong Phong and Crow, F.C. Improved rendition of polygonal models of curved surfaces. To be presented at the joint USA-Japan Computer Conference.
8. Warnock, J.E. A hidden-line algorithm for halftone picture representation. Dep. of Comput. Sci., U. of Utah, TR 4-15, 1969.
9. Watkins, G.S. A real-time visible surface algorithm. Dep. of Comput. Sci., U. of Utah, UTEC-CSc-70-101, June 1970.
10. Newell, M.E., Newell, R.G., and Sancha, T.L. A new approach to the shaded picture problem. Proc. ACM 1973 Nat. Conf.
11. Gouraud, H. Computer display of curved surfaces. Dep. of Comput. Sci., U. of Utah, UTEC-CSc-71-113, June 1971. Also in *IEEE Trans. C-20* (June 1971), 623–629.

A Comprehensive Physical Model
for Light Reflection

Xiao D. He
Kenneth E. Torrance
François X. Sillion
Donald P. Greenberg

Program of Computer Graphics
Cornell University
Ithaca, NY 14853

Abstract

A new general reflectance model for computer graphics is presented. The model is based on physical optics and describes specular, directional diffuse, and uniform diffuse reflection by a surface. The reflected light pattern depends on wavelength, incidence angle, two surface roughness parameters, and surface refractive index. The formulation is self consistent in terms of polarization, surface roughness, masking/shadowing, and energy. The model applies to a wide range of materials and surface finishes and provides a smooth transition from diffuse-like to specular reflection as the wavelength and incidence angle are increased or the surface roughness is decreased. The model is analytic and suitable for Computer Graphics applications. Predicted reflectance distributions compare favorably with experiment. The model is applied to metallic, nonmetallic, and plastic materials, with smooth and rough surfaces.

CR Categories and Subject Descriptors: I.3.7—[Computer Graphics]: Three-Dimensional Graphics and Realism; I.3.3—[Computer Graphics]: Picture/Image Generation; J.2—[Physical Sciences and Engineering]: Physics.

Additional Key Words and Phrases: reflectance model, specular and diffuse reflection, comparison with experiment.

1 Introduction

Photorealistic image generation is an active research area in Computer Graphics. Ray-tracing and Radiosity have been developed to obtain realistic images for specular and diffuse environments, respectively. However, applications of these methods to general environments have been hindered by the lack of a broadly-applicable local light reflection model. To obtain a true global illumination solution of a general environment, a physically based reflection model of general applicability is needed.

A comprehensive light reflection model is presented in this paper. The model compares favorably with experiment and describes specular, directional diffuse, uniform diffuse and combined types of

reflection behavior. The model is analytic and provides a smooth transition from specular to diffuse-like behavior as a function of wavelength, incidence angle and surface roughness.

As illustrated in Figure 1, we classify the reflection process from
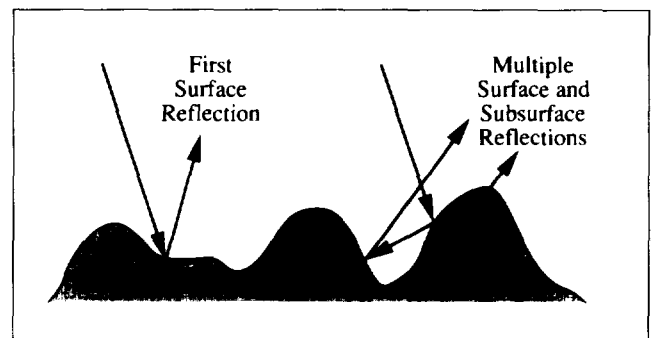
Figure 1: Reflection processes at a surface.

an arbitrary surface as consisting of first-surface reflections and multiple surface and/or subsurface reflections. The first-surface reflection process is described by physical optics and is strongly directional. As the surface becomes smooth this part evolves toward specular or mirror-like behavior. As the surface becomes rough, a diffuse-like behavior due to diffraction and interference effects becomes more important and, at larger roughnesses, it controls the directional distribution of the first-surface reflected light. The model partitions energy into specular and diffuse-like components according to the roughness of the surface. The multiple surface and subsurface reflections sketched in Figure 1 are geometrically complex, but may be expected to be less strongly directional than the first-surface reflected light. Hence, they are approximated as uniform diffuse. Our model leads to analytic expressions suitable for the full range of surface roughnesses and thus is useful for implementation in computer graphics.

The present model builds on, and extends, existing models from optics [3] [5]. It allows for polarization and masking/shadowing effects. The model extends the geometric optics model of Cook [8] to the physical optics region, and correctly includes specular reflection as the surface roughness is decreased. The model is physically based in contrast to empirical approaches [13].

The following sections provide a conceptual introduction, the model, a comparison with physical experiments, and example implementations. The mathematical derivation of the model appears in Appendix A. For unpolarized incident light, the reflectance model is summarized in Appendix B.

A	=	projected area of the surface (Figure 5)
$BRDF$	=	bidirectional reflectance distribution function
$C(r)$	=	correlation coefficient, equation (48)
c	=	complex coefficient of polarization state
D	=	distribution function, equation (78)
$E_s, \vec{E}_s$	=	scalar and vector electric fields
F	=	Fresnel reflection coefficient, equation (44)
$|F|^2$	=	Fresnel reflectivity
$\bar{F}$	=	Fresnel matrix, equation (44)
G	=	geometrical factor, equation (76)
G'	=	Green's function, equation (2)
g	=	surface roughness function, equation (9)
I	=	intensity
$\bar{I}$	=	unit tensor
$i \quad .$	=	unit imaginary number, i.e., $i = \sqrt{-1}$
k	=	wave number, i.e., $k = 2\pi/\lambda$
$\vec{k}$	=	wave vector
$\hat{k}$	=	unit vector in wave direction
L	=	length
L_x, L_y	=	length dimensions of the surface
m	=	summation index
$\bar{n}$	=	refractive index
$\hat{n}$	=	local surface normal, unit vector
$\hat{n}_b$	=	bisecting unit vector, equation (51)
p	=	incident polarization state vector, equation (34)
$p(z)$	=	Gaussian distribution function, equation (3)
R	=	distance from origin to field point
$\vec{R}$	=	positional vector to field point
$\vec{r}$	=	positional vector of a surface point
S	=	shadowing function, equation (23)
$\hat{s}, \hat{p}$	=	s and p polarization unit vectors
$\bar{T}$	=	transformation matrix, equation (39)
$\vec{v}$	=	wave vector change, equation (20)

v_{xy}	=	$\sqrt{v_x^2 + v_y^2}$
$\hat{x}, \hat{y}, \hat{z}$	=	unit vectors in Cartesian coordinates
z	=	surface height
Γ	=	area of bounding surface, Figure 2
Δ	=	delta function
$\vec{\eta}$	=	horizontal distance vector, equation (28)
θ, ϕ	=	polar and azimuthal angles (Figure 5)
λ	=	wavelength
$\xi(x, y)$	=	Gaussian distributed random function
ρ_{bd}	=	bidirectional reflectivity, equation (4)
ρ_{dh}	=	directional-hemispherical reflectivity
ρ_{hd}	=	hemispherical-directional reflectivity
σ^2	=	apparent variance of $z = \xi(x, y)$
σ_0^2	=	variance of $z = \xi(x, y)$
τ	=	autocorrelation length, equation (48)
ω	=	solid angle

Subscripts

a	=	ambient
b	=	bisecting
bd	=	bidirectional
dd	=	directional-diffuse
i	=	incident
p	=	p polarization
r	=	reflected
s	=	s polarization
sp	=	specular
ud	=	uniform-diffuse
x, y, z	=	Cartesian coordinates
$1, 2$	=	surface points

Superscripts

n	=	local plane
$*$	=	complex conjugate

Table 1: Nomenclature

2 Theory of light reflection

This section introduces the principal techniques often used to analyze the reflection of an electromagnetic wave by a general surface [3] [5]. The improved model presented later in this paper uses all of these techniques.

2.1 Kirchhoff theory

Consider the geometry sketched in Figure 2. According to classical electromagnetic theory, the scalar electromagnetic field $E(\vec{R})$ at an arbitrary point in space can be expressed as a function of the scalar field E_s and its normal derivative $\partial E_s/\partial n$ on any enclosing surface Γ. The governing equation is [5]

$$E(\vec{R}) = \frac{1}{4\pi} \int_\Gamma \left(E_s(\vec{r}) \frac{\partial G'(\vec{R}, \vec{r})}{\partial n} - G'(\vec{R}, \vec{r}) \frac{\partial E_s(\vec{r})}{\partial n} \right) d\Gamma \quad (1)$$

where G' is the free space Green's function given by [12]

$$G'(\vec{R}, \vec{r}) = \frac{e^{ik|\vec{R} - \vec{r}|}}{|\vec{R} - \vec{r}|} \quad (2)$$

Equation (1) is an integral representation of the wave equation and is known as the Kirchhoff integral of scalar diffraction theory.

For a single reflecting surface, the domain of integration Γ reduces to the area of the reflecting surface. This has allowed a class

of surface reflection models, known as "physical or wave optics" models, to be derived [5]. "Physical optics" uses a complete physical or wave description of the reflection process, thus allowing for diffraction and interference effects. Wave effects must be included if a reflection model is to describe both specular and diffuse-like reflection from a surface.
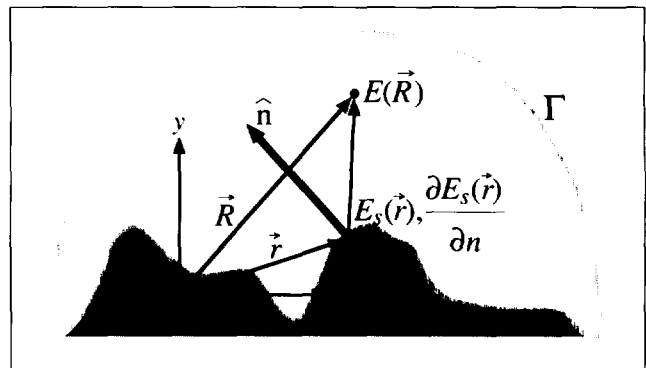
Figure 2: Geometry for application of the Kirchhoff integral. $\hat{n}$ is the local surface normal.

2.2 Tangent plane approximation

For reflection processes, the Kirchhoff formulation reduces the general problem of computing the field everywhere in space to the simpler one of determining the field on the reflecting surface. However, even this is a complex task, and the so-called "tangent plane approximation" is often used. This is done by setting the value of the field at a given point on the surface to be the value that would exist if the surface were replaced by its local tangent plane. This is sketched in Figure 3 where E_i and E_s are the incident and scattered fields, respectively, and $F(\theta)$ is the local Fresnel (electric field) reflection coefficient. The approximation is valid when the local radius of curvature of the surface is large compared to the wavelength. The reflected field depends on the Fresnel reflection coefficients for horizontal and vertical polarizations, as well as on the local slope and position of the reflecting point.
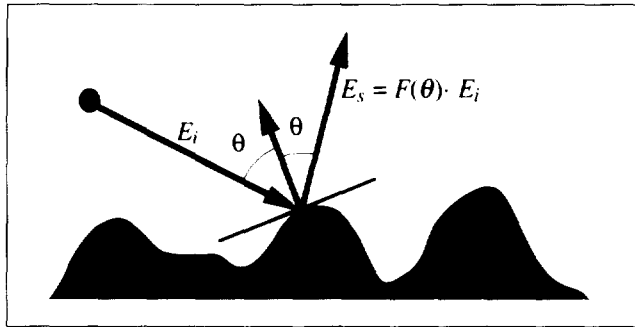
Figure 3: Tangent plane approximation for a reflecting surface. The statistical parameters σ and τ for the surface are indicated schematically.

2.3 Statistical surfaces

The complete geometrical specification of a reflecting surface is rarely known, but information at length scales comparable to the radiation wavelength is required when the Kirchhoff theory is used. However, small scale variations of the electromagnetic field on the surface are averaged out when viewed from a distance. This averaging over points on a surface is statistically equivalent to averaging over an entire class of surfaces with the same statistical description. Interesting quantities, such as the reflected intensity in a given direction, can then be obtained by a weighted average of the Kirchhoff integral.

Frequently, the height distribution on a surface (Figure 3) is assumed to be Gaussian and spatially isotropic. Under such conditions, the probability that a surface point falls in the height range z to $z + dz$ is given by $p(z)dz$, with a probability distribution

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-(z^2/2\sigma_0^2)} \qquad (3)$$

A mean value of $z = 0$ is assumed and σ_0 is the *rms roughness* of the surface. To fully specify an isotropic surface a horizontal length measure is also needed. One such measure is the *autocorrelation length* τ (defined in equation (48)), which is a measure of the spacing between surface peaks. The *rms slope* of the surface is proportional to σ_0/τ.

2.4 Shadowing and masking

The effect of self-shadowing and self-masking by a rough surface (Figure 4) was introduced in computer graphics by Blinn [6]

and Cook [8]. This effect manifests itself at large angles of incidence or reflection, where parts of the surface are shadowed and/or masked by other parts, reducing the amount of reflection. Beckmann [4] argued that to first order, the effect of shadow-
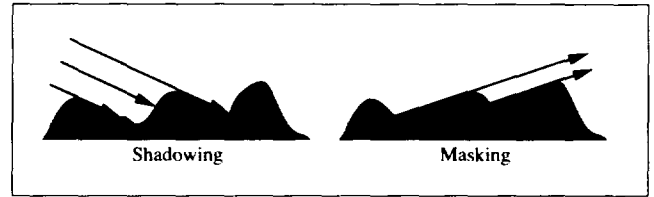
Figure 4: Shadowing and masking.

ing/masking can be obtained by using a multiplicative factor which accounts for the fraction of the surface that is visible both to the source and the receiver. Such a concept was used by both Blinn and Cook in their geometrical optics approaches, but the V-groove shadowing/masking factor they used [20] is first-derivative discontinuous. Many other shadowing/masking factors have appeared in the literature. Of these, the one due to Smith [16] is continuous in all derivatives and has been found to agree with statistical numerical simulations of a Gaussian rough surface [7].

2.5 Discussion

An early comprehensive model of light reflection from a rough surface, using physical optics, was introduced by Beckmann [5]. Beckmann applied the scalar form of the Kirchhoff theory, used the tangent plane approximation, and performed a statistical average over the distribution of heights to get the reflected intensity. The Beckmann distribution function was used by Blinn and Cook for their computer graphics applications.

Stogryn applied a more general, vector form of the Kirchhoff theory, thus taking polarization effects and the correct dependency of the Fresnel reflectivity into account [18]. Furthermore, he used a more complete statistical averaging scheme that averages over both height and slope. However, shadowing/masking was not considered, and the derivation of the reflected intensity was limited to special cases of incident polarization. A more general model, which accounts for polarization, Fresnel, and shadowing/masking effects, has been described by Bahar [1] [2]. However, it is difficult to implement because it relies on the solution of a set of coupled integro-differential equations.

Finally, it should be noted that these models were very rarely compared with experimental results.

3 An improved model

This section presents an improved light reflection model of broad applicability. Section 3.1 summarizes the techniques and key assumptions; Section 3.2 presents the improved model. Details of the mathematical derivation appear in Appendix A and a full set of equations for unpolarized incident light in Appendix B.

3.1 Techniques and key assumptions

To develop a general reflection model which avoids many of the limitations of previous models, the overall formulation of Beckmann was used, but with the following improvements:

- The vector form of the Kirchhoff diffraction theory is used. This allows, for the first time, a complete treatment of polarization and directional Fresnel effects to be included. Such

effects are required for a comprehensive formulation. The model permits abitrary incident polarization states (e.g., plane, circular, unpolarized, partially polarized, etc.) and includes effects like depolarization and cross-polarization.

- The surface averaging scheme of Stogryn [18] is employed with its improved representation of the effects of surface height and slope. Averaging of the Kirchhoff integral is over a four-fold joint probability function (i.e., height, slope, and two spatial points).

- The scheme of Stogryn [18] is extended to average only over the *illuminated* (unshadowed/unmasked) parts of the surface. This requires a modified probability function with an *effective roughness*, σ, given by equation (53). When roughness valleys are shadowed/masked (Figure 4), the effective surface roughness can be significantly smaller than the rms roughness, σ_0, especially at grazing angles of incidence or reflection. For the first time, the concept of an effective roughness, which depends on the angles of illumination and reflection, is applied.

- The geometrical shadowing/masking factor of Smith [16] is introduced as a multiplicative factor. The function has appropriate smoothness and symmetry.

With the above, the model leads to a fairly-complex integral formulation. Simplifications result by making the local tangent-plane approximation and assuming gentle roughness slopes. These assumptions should be realistic for many surfaces over a wide range of radiation wavelengths. Significantly, the assumptions lead to an analytical form for the light reflection model.

3.2 The improved light-reflection model

The light reflection model is presented in terms of the bidirectional reflectivity ρ_{bd}, also called the bidirectional reflectance distribution function (BRDF). The coordinates are shown in Figure 5, together with the propagation unit vectors $(\hat{k}_i, \hat{k}_r)$ and the polarization unit vectors $(\hat{s}, \hat{p})$ for the polarization components perpendicular $(\hat{s})$ and parallel $(\hat{p})$ to the incident and reflecting planes (i.e., the $(\hat{k}, \hat{z})$ planes). The total BRDF is defined as the ratio of the total reflected
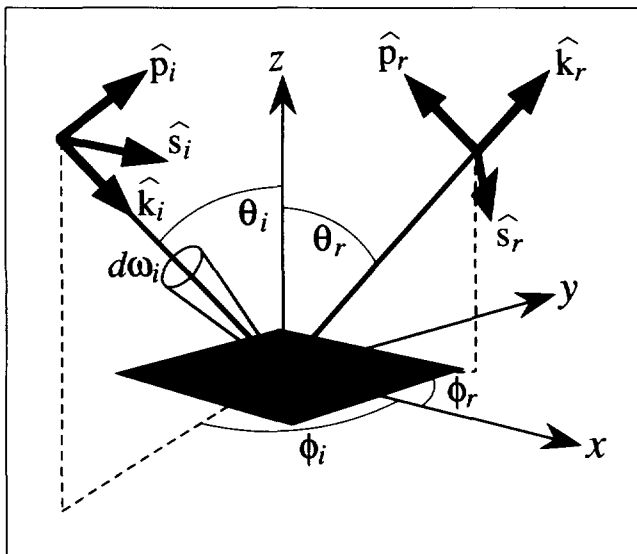
Figure 5: Coordinates of illumination and reflection.

intensity (i.e., the sum of reflected s and p intensities) in the direction (θ_r, ϕ_r) to the energy incident per unit time and per unit area onto the surface from the direction (θ_i, ϕ_i) [14]. The incident energy flux may be expressed in terms of the incident intensity I_i and the incident solid angle $d\omega_i$:

$$\rho_{bd}(\theta_r, \phi_r, \theta_i, \phi_i) = \frac{dI_r(\theta_r, \phi_r; \theta_i, \phi_i)}{I_i(\theta_i, \phi_i)\cos\theta_i d\omega_i} \qquad (4)$$

The BRDF may also be defined for each polarization component of the reflected intensity (see Appendix A). Equation (4) gives the frequently-used total BRDF.

We propose a bidirectional reflectivity consisting of three components:

$$\rho_{bd} = \rho_{bd,sp} + \rho_{bd,dd} + \rho_{bd,ud} \qquad (5)$$

The additional subscripts correspond to specular (sp), directional-diffuse (dd), and uniform-diffuse (ud) reflection. The first two components in (5) result from the first-surface reflection process (see Figure 1) and are respectively due to specular reflection by the mean surface and diffraction scattering by the surface roughness. The third component, taken as uniform diffuse, is attributed to multiple surface and/or subsurface reflections.

An example of a light intensity distribution corresponding to equation (5) is shown in Figure 6. A general reflecting surface is
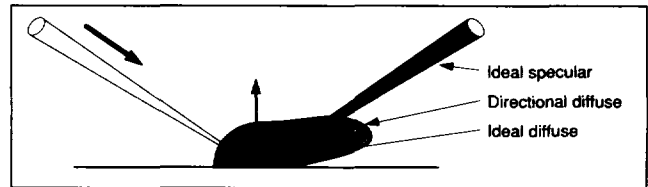
Figure 6: Example of a light intensity distribution.

assumed, with some specular reflection, some diffraction scattering due to roughness, and some multiple or subsurface scattering. The specularly-reflected part is contained within the specular cone of reflection. The diffraction-scattered part shows a directional distribution which is far from ideal diffuse. The last part is uniform diffuse (Lambertian).

An analytic form for the first two terms in (5) is derived in Appendix A. With the local-tangent-plane and gentle-slope assumptions for the first-surface reflection process, and for arbitrary incident polarization, we have:

$$\rho_{bd,sp} = \frac{\rho_s}{\cos\theta_i d\omega_i} \cdot \Delta = \frac{|F|^2 \cdot e^{-g} \cdot S}{\cos\theta_i d\omega_i} \cdot \Delta \qquad (6)$$

$$\rho_{bd,dd} = \frac{\mathcal{F}(\hat{n}_b, \hat{n}_b, \mathbf{p}) \cdot S}{\cos\theta_i \cdot \cos\theta_r} \cdot \frac{\tau^2}{16\pi} \cdot \sum_{m=1}^{\infty} \frac{g^m e^{-g}}{m! \cdot m} \cdot \exp(-\frac{v_{xy}^2 \tau^2}{4m}) \qquad (7)$$

$$\rho_{bd,ud} = a(\lambda) \qquad (8)$$

where ρ_s is the specular reflectivity of the surface, Δ is a delta function which is unity in the specular cone of reflection and zero otherwise, $|F|^2$ is the Fresnel reflectivity which depends on the index of refraction ($\bar{n}(\lambda)$) of the surface material [14, p.100], g is a function of the effective surface roughness given by

$$g = [(2\pi\sigma/\lambda)(\cos\theta_i + \cos\theta_r)]^2, \qquad (9)$$

S is the shadowing function (see equation (23)), $\mathcal{F}$ is a function involving the Fresnel reflection coefficients (see equations (68) and

(59), (60)), p is the polarization state vector of the incident light (see equation (34)), v_{xy} is a function which depends on the illumination and reflection angles (see equation (20)), and $a(\lambda)$ is a parameter to be discussed later.

For convenience and for the special case of incident unpolarized light, the governing equations are gathered together and presented in Appendix B. The directional-diffuse term in this appendix (equation (71)) uses nomenclature to permit comparison with the geometric optics model of Cook-Torrance [8].

The physical basis of the three reflection components in (5) is discussed in the following subsections. Before proceeding, we note that the dependence of the specular component on $d\omega_i$ drops out if equation (5) is converted to an intensity basis by multiplying by $I_i \cos\theta_i d\omega_i$. From (6), the specular term becomes $\rho_s I_i \Delta$, which is the well known form used in Ray-tracing. The specular intensity is then independent of $d\omega_i$, but the directional-diffuse and uniform-diffuse intensities are proportional to $d\omega_i$.

3.2.1 Specular contribution: $\rho_{bd,sp}$

The specular term accounts for mirror-like reflection from the *mean* plane of the reflecting surface. The term is proportional to the Fresnel or mirror reflectivity, $|F|^2$. For rough surfaces, the specular term is reduced by the roughness and shadowing factors e^{-g} and S, respectively.

For a smooth surface, as the wavelength of the incident light becomes large relative to the projected surface roughness, i.e., $\lambda \gg \sigma\cos\theta_i$, the specular term is not attenuated since $g \rightarrow 0$ and $S \rightarrow 1$. Also in this limit, the specular component dominates the first-surface reflection process, since the contribution from equation (7) diminishes as $g \rightarrow 0$. For smooth surfaces, equation (6) reduces to

$$|F|^2/\cos\theta_i d\omega_i, \qquad (10)$$

which is the usual form of the bidirectional reflectivity for a specular surface.

3.2.2 Directional diffuse contribution: $\rho_{bd,dd}$

When the wavelength of the incident light is comparable to or smaller than the projected size of surface roughness elements (i.e., $\lambda \sim \sigma\cos\theta_i$), the first-surface reflection process introduces diffraction and interference effects. The reflected field is spread out to the hemisphere above the reflecting surface. We call this *directional diffuse*, to indicate that the field is diffused to the hemisphere but may have a directional, nonuniform character.

The reflected light pattern given by equation (7) depends on surface statistics through the effective roughness σ and the autocorrelation length τ. For smooth surfaces, as σ/λ or g approach zero, the bidirectional reflectivity given by equation (7) diminishes to zero. For rough surfaces, with σ/λ or g large, equation (7) describes the directional distribution of the first-surface reflected light. The reflected pattern can be complex with maximal values in the specular direction for slightly rough surfaces, at off-specular angles for intermediate roughnesses, or at grazing reflection angles for very rough surfaces.

3.2.3 Uniform diffuse contribution: $\rho_{bd,ud}$

The light reflected by multiple surface reflections or by subsurface reflections is generally more difficult to describe analytically than light reflected by the first-surface reflection process. This contribution is small for metallic (opaque) surfaces with shallow roughness slopes. However, the contribution can be important for surfaces with large slopes, or for nonmetals if significant radiation crosses the

first surface and is reflected by subsurface scattering centers (e.g., paints, ceramics, plastics).

Estimates of the multiple-reflection process within surface V-grooves, based on geometrical optics, have been carried out [10] [17]. Also, estimates of the subsurface scattering are available [14]. The analytical results often suggest that the reflected field due to these two processes may be approximated as nearly directionally uniform. Therefore, the multiply-reflected and/or subsurface scattered light is approximated as *uniform-diffuse* (i.e., Lambertian), and we denote it by $a(\lambda)$.

The coefficient $a(\lambda)$ can be estimated theoretically if the V-groove geometry is applicable, or if the subsurface scattering parameters are known. Alternatively, $a(\lambda)$ can be estimated experimentally if equation (5) is integrated over the reflecting hemisphere, and the results are compared with measured values of the directional-hemispherical reflectivity, ρ_{dh}. This reflectivity is equal to the hemispherical-directional reflectivity ρ_{hd} (for the case of uniform incident intensity [14]), and which can be easily measured using an integrating sphere reflectometer. For the present paper, in the absence of additional surface or subsurface scattering parameters, or experimental measurements, we will treat $a(\lambda)$ as a constrained, but otherwise free, parameter. The constraint is based on energy conservation and gives an upper bound for $a(\lambda)$.

3.3 Discussion

The theoretical model described by equation (5) allows specular, directional-diffuse, and uniform-diffuse reflection behavior as sketched in Figure 6. The governing equations in general form are given in equations (5) to (8) and Appendix A, or for unpolarized incident light in Appendix B. The actual reflection patterns depend on wavelength, incidence angle, surface roughness and subsurface parameters, and index of refraction. The model provides a unified approach for a wide range of materials and surface finishes, and is in a form suitable for use in computer graphics.

4 Comparison with experiments

In this section we compare the reflection model with experimental measurements. Appropriate comparison experiments appear only infrequently in the literature, since well-characterized surfaces as well as good wavelength and directional resolution are required. The measurements selected for comparison consist of BRDF's for roughened aluminum [19], roughened magnesium oxide ceramic [19], sandpaper [9], and smooth plastic [11]. The comparisons cover a wide range of materials (metallic, nonmetallic) and reflection behavior (specular, directional diffuse, uniform diffuse).

Polar comparisons are presented in Figures 7 to 10. Results are shown in the plane of incidence; the polar angle is θ_r and the curve parameter is the angle of incidence θ_i. Theoretical predictions are shown with solid lines and experimental measurements with dashed lines. The polar radius is the BRDF normalized with respect to the specular reflecting ray direction, i.e.,

$$\frac{\rho_{bd}(\theta_i, 0; \theta_r, \phi_r)}{\rho_{bd}(\theta_i, 0; \theta_i, 0)} \qquad (11)$$

Results for an aluminum surface (very pure; measured roughness: $\sigma_0 = 0.28\mu m$) are shown in Figures 7 and 8, respectively, for wavelengths of $\lambda = 2.0\mu m$ and $0.5\mu m$. These figures illustrate the effects of wavelength and incidence angle. The autocorrelation length and measured hemispherical reflectances were not reported. Therefore, values of $\tau = 1.77\mu m$ and $a(\lambda) = 0$ were selected as best fits at both wavelengths. Several points can be noted.
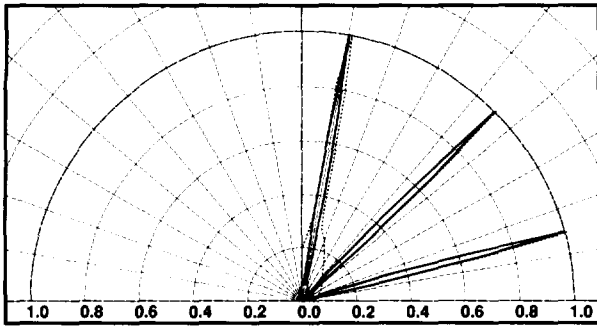
Figure 7: Normalized BRDF's of roughened aluminum as obtained from theory (solid lines) and experiment (dashed lines) for incidence angles of $\theta_i = 10°, 45°$, and $75°$. $\lambda = 2.0\mu m$. This is the same surface as in Figure 8. The surface shows strong specular reflection at this wavelength.

Figure 8: Normalized BRDF's of roughened aluminum as obtained from theory (solid lines) and experiment (dashed lines) for incidence angles of $\theta_i = 10°, 30°, 45°, 60°$, and $75°$. $\lambda = 0.5\mu m$. This is the same surface as in Figure 7. The surface shows strong directional diffuse and emerging specular reflection at this wavelength.
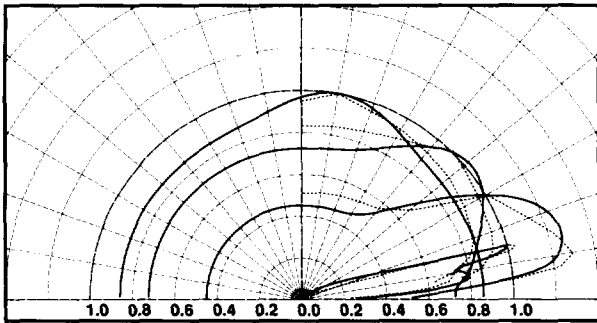
Figure 9: Normalized BRDF's of roughened magnesium oxide ceramic as obtained from theory (solid lines) and experiment (dashed lines) for incidence angles of $\theta_i = 10°, 45°, 60°$, and $75°$. $\lambda = 0.5\mu m$. The surface shows strong uniform diffuse and emerging specular reflection.
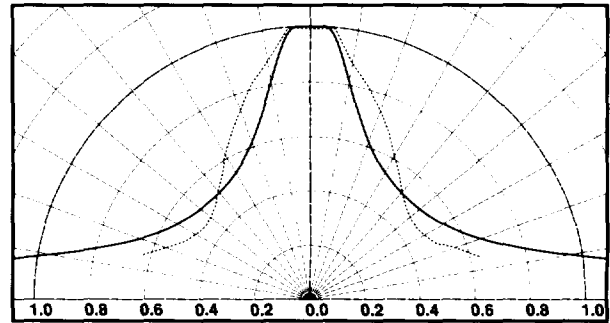
Figure 10: Normalized BRDF's of sandpaper as obtained from theory (solid lines) and experiment (dashed lines) for normal incidence, $\theta_i = 0°$. $\lambda = 0.5\mu m$. The surface shows a large reflectance at grazing reflection angles.

When σ_0 is small compared to λ, as in Figure 7, strong specular reflection occurs. The angular width of the measured specular peak is determined by the solid angles of incident and received light in the experiments ($d\omega_i = d\omega_r = \pi/1024$). To allow comparisons, the theoretical peaks have been averaged over the same solid angles. For incidence at $\theta_i = 10°$, the reflected pattern displays both specular and directional diffuse components. In Figure 8, when the roughness is more comparable to the wavelength, a strong directional diffuse pattern appears, and for $\theta_i = 10°, 30°, 45°$, and $60°$, the reflected intensity is maximal at larger-than-specular angles. For $\theta_i = 75°$, a specular peak emerges as the surface appears somewhat smoother to the incident radiation.

A comparison with a magnesium oxide ceramic (very pure; measured roughness: $\sigma_0 = 1.90\mu m$, but model best fit $\sigma_0 = 1.45\mu m$) at $\lambda = 0.5\mu m$ is displayed in Figure 9. This surface shows nearly uniform diffuse behavior at $\theta_i = 10°$ and an emerging specular peak for larger values of θ_i. The model employed best-fit parameters of $\tau = 13.2\mu m$ and $a(\lambda) = 0.9$, the latter expressing the relatively stronger role of subsurface scattering as compared to the aluminum surface. Significantly, the experimental and theoretical trends in Figures 7 to 9 for both the metal and the nonmetal are in qualitative accord. Importantly, both materials display an emerging specular peak as the angle of incidence is increased, and, for the metal, as the wavelength is increased. Further, the metal shows a strong directional diffuse pattern, and the nonmetal a strong uniform diffuse

pattern, both of which are in accord with the model.

A dramatically different reflection pattern is displayed in Figure 10, corresponding to 220 grit sandpaper at $\theta_i = 0°$ and $\lambda = 0.55\mu m$. Parameters used for the comparison are $\sigma_0/\tau = 4.4$ and $a(\lambda) = 0$. For very rough surfaces, only the ratio σ_0/τ is required, not σ_0 and τ separately [5]. Although the large ratio of σ_0/τ challenges the gentle slope assumption of the model, the agreement between experiment and theory is striking as both display large reflected intensities at grazing angles of reflection.

A comparison of experiment and theory in terms of absolute BRDF's is shown in semilog form in Figure 11 for a smooth blue plastic at $\lambda = 0.46\mu m$. The shape of the specular spikes is determined by the geometry of the incident and receiving optical systems. The distributions for four incidence angles reveal a linear combination of specular and uniform diffuse behavior. This is consistent with the model (equations (5) to (8)). For a smooth surface with $\sigma_0 = 0$, the directional-diffuse term drops out and the specular term reduces to equation (10). The directional-hemispherical reflectivity at $\theta_i = 0°$ and $\lambda = 0.46\mu m$ was measured ($\rho_{dh} = 0.195$) and yields the value $a(\lambda) = 0.15$ used for the uniform diffuse term in the model. The agreement between experiment and theory in Figure 11 in terms of shape and absolute magnitude is encouraging.

In conclusion, the experimentally-measured directional distributions in Figures 7 to 11 show a wide range of behavior and complexity. The present model describes the major features of the dis-
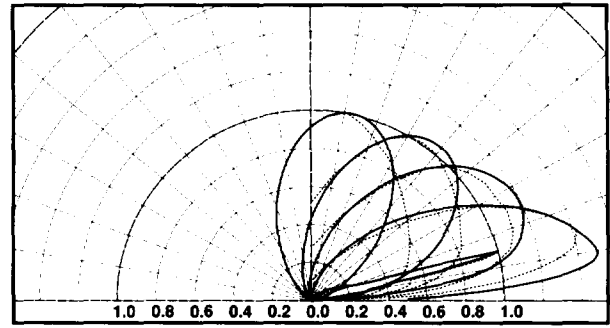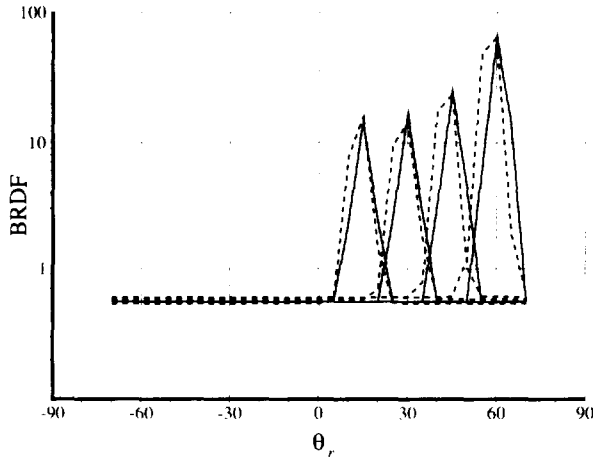
Figure 11: Absolute BRDF's for smooth blue plastic as obtained from theory (solid lines) and experiment (dashed lines) for incidence angles of θ_i = 15°, 30°, 45°, and 60°. λ = 0.46μm. This surface shows a typical smooth plastic reflection pattern with combined specular and uniform diffuse behavior.

tributions.

5 Example scenes

The reflection model described by equations (5) to (8) can be incorporated in ray-tracing or extended radiosity [15] methods. We have employed ray tracing. A single reflected ray is used together with ambient and point source illumination. The reflected intensity is given by

$$I_r(\lambda) = \sum_{i=1}^{N_i} \left\{ |F(\theta_i)|^2 \cdot e^{-g_i} \cdot S \cdot \Delta + \left[(\rho_{bd,dd})_i + a(\lambda)\right] \right. $$
$$\left. \cdot \cos\theta_i \cdot d\omega_i \right\} \cdot I_i(\lambda) + \rho_{hd}(\lambda) \cdot I_a(\lambda)$$

(12)

where N_i is the number of light sources, subscript i denotes the ith light source, the terms inside the braces respectively correspond to the three terms in equation (5), $\rho_{hd}(\lambda)$ is the hemispherical-directional reflectivity of the surface (taken as a function of λ only, and found from experiment or by integrating (5) over the incident hemisphere), and I_a is the uniform ambient illumination. The directional-diffuse term is included only for light sources. To include a directional-diffuse term from the environment, a distributed ray-tracer or an extended radiosity method [15] must be employed.

Figure 12 displays six aluminum cylinders in front of a brick wall. Each cylinder is rendered in isolation. Cylinders (a) to (f) are in order of increasing surface roughness. Other parameters are τ = 3.0μm for cylinders (a) to (e) and τ = 16.0μm for cylinder (f), and $a(\lambda)$ = 0. Note that the sharp specular image in the top faces of the cylinders diminishes, but is not blurred, with increasing surface roughness, and the image of the light source on the front vertical face spreads out. These are characteristics, respectively, of the specular and directional diffuse terms in the reflection model that are derived from physical optics. Note also that the apparent roughness of a given cylinder varies with viewing angle. The top and lateral edges can appear specular or nearly specular at grazing angles, even when the vertical face on the front side appears to be rough. A slight color shift is also apparent for a given rough surface (i.e., as λ in σ_0/λ varies). For visible light, this is most apparent in the blue shift on the front faces of the cylinders. The enhanced red shift

of the specular images is not so apparent. Clearly, the specular and directional diffuse terms of the model vary with wavelength, incidence angle, and roughness, and are responsible for the realism of the cylinders in Figure 12.

The aluminum cylinders (a) to (c) in Figure 13 illustrate limiting cases of each of the three terms in the reflection model. Cylinder (a) in Figure 13 is the same as cylinder (f) in Figure 12. Cylinder (b) is a smooth cylinder described by the specular term, in which the reflectance is a function of incidence angle according to the Fresnel reflectivity. Specular images are apparent on the top and lateral edges. (To emphasize the specular images, we have set the ambient illumination term to zero in rendering cylinder (b).) Cylinder (a) represents the directional diffuse term in the limit of $\sigma_0/\lambda \rightarrow \infty$ with σ_0/τ fixed at 0.16 (i.e., a limiting form for very rough surfaces). Cylinder (c) is ideal diffuse and is described by the uniform diffuse term. Note the striking differences between the three cylinders.

Figure 14 illustrates a scene consisting of a rough aluminum cylinder (σ_0 = 0.18μm, τ = 3.0μm, $a(\lambda)$ = 0), a rough copper sphere (σ_0 = 0.13μm, τ = 1.2μm, $a(\lambda)$ = 0), and a smooth plastic cube (σ_0 = 0, τ = 2.0μm, $a(\lambda)$ = 0.55μm) = 0.28), all resting on a rough plastic table (σ_0 = 0.20μm, τ = 2.0μm, $a(\lambda)$ = 0.55μm) = 0.28). The cube and table have the same Fresnel reflectivity.

Several effects can be noted in Figure 14. On the faces of the cube, the specular image varies with reflection angle, an effect caused solely by the Fresnel reflectivity $|F|^2$ in equation (6). The specular images on the table top also vary with reflection angle (and disappear), but this is caused mainly by roughness effects (i.e., e^{-g}) in equation (6). The cylinder in Figure 14 corresponds to cylinder (a) in Figure 12 and displays some of the specular and directional diffuse characteristics of that image.

Figure 14 gives a hint of the comprehensiveness of the light reflection model derived in this paper. Several materials of different roughnesses appear. A given surface can display specular or diffuse-like behavior depending on reflection angles and surface properties. Specular images appear or disappear based on correct physical principles. The high level of realism in Figure 14 is due to a physically-correct treatment of specular, directional diffuse, and uniform diffuse effects by the reflection model.

6 Conclusions

1. The general reflection model given by equations (5) to (8), in a single formulation, describes specular, directional diffuse, and uniform diffuse behavior. For unpolarized incident light, the model reduces to the form given in Appendix B. All of the parameters of the model are physically based.

2. The model compares favorably with experimental measurements of reflected radiation for metals, nonmetals, and plastics, with smooth and rough surfaces.

3. The model accurately predicts the emergence of specular reflection with increasing wavelength or angle of incidence, or decreasing surface roughness.

4. The model predicts a directional-diffuse pattern which can have maximal values at specular, off-specular, or grazing angles, depending on surface roughness.

5. The model is in analytical form and can improve the realism of synthetic images.

6. The model can be employed for ray-tracing or extended radiosity [15] methods.

(a) $\sigma_0 = 0.18$ (b) $\sigma_0 = 0.28$ (c) $\sigma_0 = 0.38$ (d) $\sigma_0 = 0.48$ (e) $\sigma_0 = 0.58$ (f) $\sigma_0 = 2.50$
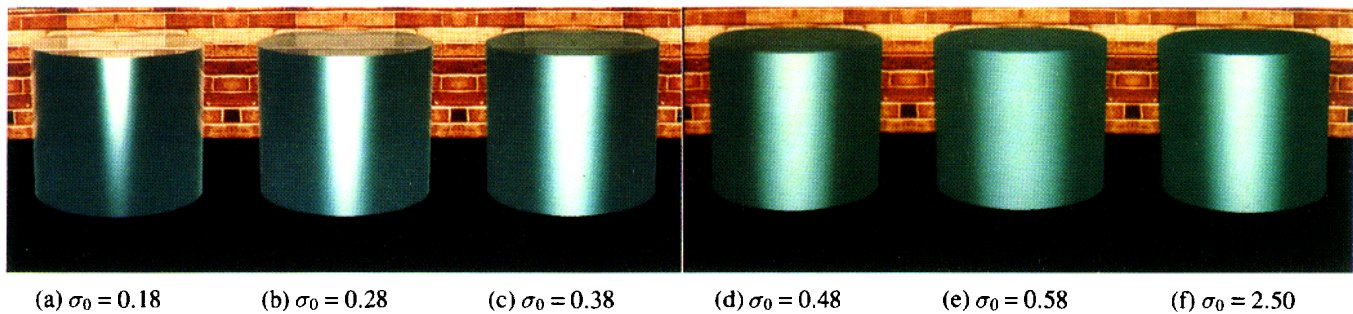
Figure 12: Aluminum cylinders with different surface roughnesses. σ_0 is in μm. $\tau = 3.0\mu m$ for cylinders (a) to (e) and $\tau = 16.0\mu m$ for cylinder (f). Note that the specular and directional-diffuse reflection characteristics vary with reflection angle and roughness.

7. The model highlights the need for tabulated databases of parameterized bidirectional reflectivities. The parameters include two surface roughness parameters (σ_0, τ), the index of refraction (as a function of wavelength), and the constrained parameter $a(\lambda)$. The latter can be inferred from measured hemispherical reflectivities.

In conclusion, the reflection model is comprehensive, physically-based, and provides an accurate transition from specular to diffuse-like reflection. Further, the model is computable and thus useful for graphics applications.

Figure 14: A general scene with metallic and plastic objects in the foreground, with smooth and rough surfaces. The specular images in the smooth plastic box vary with incidence angle due to the Fresnel effect. In the table top, the decay of the specular images with reflection angle is due to roughness. In the rough metallic surfaces, the glossy highlights result from directional diffuse reflection.

(a) $\sigma_0 = 2.5$ (b) $\sigma_0 = 0.0$ (c) diffuse

Figure 13: Aluminum cylinders in extreme limiting cases. Each cylinder corresponds to one of the three terms in the reflection model. σ_0 is in μm. (a) Directional diffuse reflection; (b) Ideal specular reflection; (c) Uniform diffuse (Lambertian) reflection.

discussions, and to Emil Ghinger for photographing the raster images. We thank the reviewers for their extensive and constructive comments which have helped to clarify a difficult paper.

Acknowledgments

References

[1] Bahar, E. and S. Chakrabarti. "Full wave theory applied to computer-aided graphics for 3-D objects," *IEEE Computer Graphics and Applications*, 7(7),1987, pages 46–60.

[2] Bahar, Ezekiel. "Review of the full wave solutions for rough surface scattering and depolarization," *Journal of Geophysical Research*, May 1987, pages 5209–5227.

[3] Bass, F.G. and I.M. Fuks. *Wave Scattering from Statistically Rough Surfaces*, Pergamon Press, 1979.

[4] Beckmann, Petr. "Shadowing of Random Rough Surfaces," *IEEE Transactions on Antennas and Propagation*, May 1965, pages 384–388.

[5] Beckmann, Petr and André Spizzichino. *The Scattering of Electromagnetic Waves from Rough Surfaces*, Pergamon Press, 1963.

[6] Blinn, James F. "Models of Light Reflection for Computer Synthesized Pictures," *Computer Graphics*, 11, 1977, pages 192–198. (Proceedings SIGGRAPH '77.)

[7] Brockelman, R. A. and T. Hagfors. "Note on the Effect of Shadowing on the Backscattering of Waves from a Random Rough Surfaces," *IEEE Transactions on Antennas and Propagation*, AP-14(5), September 1966, pages 621–629.

[8] Cook, Robert L. and Kenneth E. Torrance. "A Reflectance Model for Computer Graphics," *ACM Transactions on Graphics*, 1, 1982, pages 7–24.

[9] Feng, Xiaofen. *Comparison of methods for generation of absolute reflectance factors for BRDF studies*, Master's thesis, Rochester Institute of Technology, 1990.

[10] Hering, R.G. and T.F. Smith. "Apparent radiation properties of a rough surface," *Application to Thermal Design of Spacecraft*, 23, 1970, pages 337–361.

[11] Himlan, Theodore H., Michael C. Monks, Stephan H. Westin, Donald P. Greenberg, and Kenneth E. Torrance. "Physical measurement Techniques for Improving and Evaluating Computer Graphic Simulations.," 1991. (To be published.)

[12] Jackson, John D. *Classical Electrodynamics*, John Wiley & Son Inc., 1975.

[13] Phong, Bui Tuong. "Illumination for Computer Generated Pictures," *Communications of the ACM*, 18(6), June 1975, pages 311–317.

[14] Siegel, Robert and John R. Howell. *Thermal Radiation Heat Transfer*, McGraw-Hill book Company, 2nd edition, 1981.

[15] Sillion, François, James Arvo, Stephen Westin, and Donald P. Greenberg. "A Global Illumination Solution for General Reflectance Distributions," *Computer Graphics*, 25(4), August 1991. (Proceedings SIGGRAPH '91 in Las Vegas.)

[16] Smith, Bruce G. "Geometrical Shadowing of a Random Rough Surface," *IEEE Transactions on Antennas and Propagation*, AP-15(5), September 1967, pages 668–671.

[17] Smith, T.F. and K.E. Nichols. "Effects of polarization on bidirectional reflectance of a one-dimensional randomly rough surface," *Spacecraft Radiative Transfer and Temperature Control*, 83, 1981, pages 3–21.

[18] Stogryn, Alex. "Electromagnetic Scattering From Rough, Finitely Conducting Surfaces," *Radio Science*, 2(4), 1967, pages 415–428.

[19] Torrance, K.E. and E.M. Sparrow. "Off-Specular Peaks in the Directional Distribution of Reflected Thermal Radiation," *Journal of Heat Transfer – Transactions of the ASME*, May 1966, pages 223–230.

[20] Torrance, K.E. and E.M. Sparrow. "Theory for Off-Specular Reflection from Roughened Surfaces," *Journal of the Optical Society of America*, 57(9), September 1967, pages 1105–1114.

A Appendix: Derivations

A.1 Reflected intensities

The reflected intensities for the s and p components of polarizations are given by [14][18]

$$dI_r(\theta_r, \phi_r; \theta_i, \phi_i)_s = \frac{R^2}{A \cdot \cos\theta_r} < |\hat{s}_r \cdot \vec{E}_r(\vec{R})|^2 >$$

$$dI_r(\theta_r, \phi_r; \theta_i, \phi_i)_p = \frac{R^2}{A \cdot \cos\theta_r} < |\hat{p}_r \cdot \vec{E}_r(\vec{R})|^2 > \quad (13)$$

where the coordinates are as shown in Figure 5, $\vec{E}_r(\vec{R})$ is the reflected field in vector form, R is the distance from the origin to an arbitrary point in space, A is the area of the reflecting surface projected on the x-y plane, and $\hat{s}_r, \hat{p}_r$ are unit polarization vectors, given by

$$\hat{s}_r = \frac{\hat{k}_r \times \hat{z}}{|\hat{k}_r \times \hat{z}|}$$

$$\hat{p}_r = \hat{s}_r \times \hat{k}_r \quad (14)$$

which are normal and parallel, respectively, to the plane formed by the viewing direction and the mean surface normal. The symbol $<>$ denotes an average over the joint probability distribution function of the random rough surface characterized by

$$z = \xi(x, y). \quad (15)$$

The reflected field can be expressed in terms of the scattered field on the surface by using the vector form of the Kirchhoff diffraction theory [12]:

$$\vec{E}_r(\vec{R}) = \frac{e^{ikR}}{4\pi R} \left(\bar{I} - \hat{k}_r \hat{k}_r \right) \cdot$$
$$\int_\Gamma e^{-i\vec{k}_r \cdot \vec{r}} \left\{ -i\vec{k}_r \times (\vec{E}_s \times \hat{n}) - (\nabla \times \vec{E}_s) \times \hat{n} \right\} d\Gamma$$

$$(16)$$

where $\vec{k}_i, \vec{k}_r$ are wave vectors in the incident and reflection directions, $|k| = 2\pi/\lambda$ is the wave number, $\vec{r}$ is the position vector for a point on the surface, and the tensor $\bar{I} - \hat{k}_r \hat{k}_r = \hat{s}_r \hat{s}_r + \hat{p}_r \hat{p}_r$ is introduced to to make the reflected field transverse.

Substituting (16) into (13), we have

$$dI_s = \frac{1}{A \cos\theta_r (4\pi)^2} \cdot < | \int_\Gamma e^{-i\vec{k}_r \cdot \vec{r}} \cdot$$
$$\left\{ ik\hat{p}_r \cdot (\vec{E}_s \times \hat{n}) + \hat{s}_r \cdot [(\nabla \times \vec{E}_s) \times \hat{n}] \right\} d\Gamma |^2 >$$

$$dI_p = \frac{1}{A \cos\theta_r (4\pi)^2} \cdot < | \int_\Gamma e^{-i\vec{k}_r \cdot \vec{r}} \cdot$$
$$\left\{ ik\hat{s}_r \cdot (\vec{E}_s \times \hat{n}) - \hat{p}_r \cdot [(\nabla \times \vec{E}_s) \times \hat{n}] \right\} d\Gamma |^2 >$$

$$(17)$$

To evaluate the right side of (17), the surface element $d\Gamma$ is expressed in terms of the planar surface area $dA = dx \cdot dy$ by

$$d\Gamma = dA/(\hat{n} \cdot \hat{z}) \quad (18)$$

Further, the squares of the absolute values of the integrals in (17) can be expanded in terms of double surface integrals. We find

$$< | \int_\Gamma e^{-i\vec{k}_r \cdot \vec{r}} \{\} d\Gamma |^2 >=$$

$$< \int_A dA_1 \int_A dA_2 e^{-i\vec{v}\cdot(\vec{r}_1-\vec{r}_2)} \cdot$$
$$\left(e^{-i\vec{k}_i\cdot\vec{r}_1}\{\}_1 \right)\left(e^{-i\vec{k}_i\cdot\vec{r}_2}\{\}_2 \right)^* / (\hat{n}_1\cdot\hat{z})(\hat{n}_2\cdot\hat{z}) >$$

$$(19)$$

where v is the wave vector change

$$\vec{v} = k(\hat{k}_r - \hat{k}_i), \tag{20}$$

$*$ denotes a complex conjugate, $\{\}$ refers to the terms in braces in (17), and the subscripts refer to points on area elements dA_1 and dA_2.

The $<>$ in (19) commutes with the surface integral and a term of the form

$$< e^{-i\vec{v}\cdot\hat{z}(\xi_1-\xi_2)}\{\}_1\{\}_2^* / (\hat{n}_1\cdot\hat{z})(\hat{n}_2\cdot\hat{z}) > \tag{21}$$

results. Since the surface is assumed to be isotropic and stationary, (21) is a function only of $x_1 - x_2$ and $y_1 - y_2$. Thus, by making the change of variables

$$x' = x_1 - x_2 \qquad x'' = x_2$$
$$y' = y_1 - y_2 \qquad y'' = y_2 \tag{22}$$

the integrals over x'' and y'' may be carried out separately to give a factor $S \cdot A$, where S is the fraction of the surface that is both illuminated and viewed and represents the shadowing function given by [16]:

$$S = S_i(\theta_i) \cdot S_r(\theta_r) \tag{23}$$

where

$$S_i(\theta_i) = (1 - \frac{1}{2}\text{erfc}(\frac{\tau\cot\theta_i}{2\sigma_0}))/(\Lambda(\cot\theta_i)+1)$$
$$S_r(\theta_r) = (1 - \frac{1}{2}\text{erfc}(\frac{\tau\cot\theta_r}{2\sigma_0}))/(\Lambda(\cot\theta_r)+1)$$

$$(24)$$

and

$$\Lambda(\cot\theta) = \frac{1}{2}\left(\frac{2}{\pi^{1/2}}\cdot\frac{\sigma_0}{\tau\cot\theta} - \text{erfc}(\frac{\tau\cot\theta}{2\sigma_0}) \right) \tag{25}$$

Hence, the reflected intensities in (17) are

$$dI_s = \frac{S}{\cos\theta_r(4\pi)^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dx'dy'\, e^{-i\vec{v}\cdot\vec{\eta}}B_s \quad (26)$$

$$dI_p = \frac{S}{\cos\theta_r(4\pi)^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dx'dy'\, e^{-i\vec{v}\cdot\vec{\eta}}B_p \quad (27)$$

where

$$\vec{\eta} = x'\hat{x} + y'\hat{y} \tag{28}$$

and

$$B_s = < e^{-i\vec{v}\cdot\hat{z}(\xi_1-\xi_2)}\mathcal{F}(\hat{n}_1,\hat{n}_2)_s >$$
$$B_p = < e^{-i\vec{v}\cdot\hat{z}(\xi_1-\xi_2)}\mathcal{F}(\hat{n}_1,\hat{n}_2)_p > \tag{29}$$

where

$$\mathcal{F}(\hat{n}_1,\hat{n}_2)_s = e^{-i\vec{k}_i\cdot(\vec{r}_1-\vec{r}_2)}/(\hat{n}_1\cdot\hat{z})(\hat{n}_2\cdot\hat{z})$$
$$\cdot \left(\{ik\hat{p}_r\cdot(\vec{E}_s\times\hat{n}) + \hat{s}_r\cdot[(\nabla\times\vec{E}_s)\times\hat{n}]\} \right)_1$$
$$\cdot \left(\{ik\hat{p}_r\cdot(\vec{E}_s\times\hat{n}) + \hat{s}_r\cdot[(\nabla\times\vec{E}_s)\times\hat{n}]\} \right)_2^*$$

$$(30)$$

$$\mathcal{F}(\hat{n}_1,\hat{n}_2)_p = e^{-i\vec{k}_i\cdot(\vec{r}_1-\vec{r}_2)}/(\hat{n}_1\cdot\hat{z})(\hat{n}_2\cdot\hat{z})$$
$$\cdot \left(\{ik\hat{s}_r\cdot(\vec{E}_s\times\hat{n}) - \hat{p}_r\cdot[(\nabla\times\vec{E}_s)\times\hat{n}]\} \right)_1$$
$$\cdot \left(\{ik\hat{s}_r\cdot(\vec{E}_s\times\hat{n}) - \hat{p}_r\cdot[(\nabla\times\vec{E}_s)\times\hat{n}]\} \right)_2^*$$

$$(31)$$

The functions B_s and B_p in (29) depend only on x' and y'. Notice that dI_s and dI_p are the s and p polarized reflected intensities, respectively. The total reflected intensity, as used in equation (4), is given by

$$dI_r = dI_s + dI_p \tag{32}$$

A.2 Tangent plane approximation

The reflected intensities in (26) and (27) are expressed in terms of the scattered field $\vec{E}_s$ on the surface. In turn, $\vec{E}_s$ depends on the incident field, and may be related to the incident field by using the local tangent plane approximation.

For the case of a unidirectional incident field, we have

$$\vec{E}_i = E_0 e^{i\vec{k}_i\cdot\vec{r}}\mathbf{p} \tag{33}$$

$$\mathbf{p} = c_s\hat{s}_i + c_p\hat{p}_i \tag{34}$$

where E_0 is the wave amplitude, $\mathbf{p}$ is the polarization state vector of the incident radiation, c_s, c_p are called the polarization coefficients[1], and $\hat{s}_i$, $\hat{p}_i$ are unit polarization vectors with respect to the plane of incidence $(\hat{k}_i, \hat{z})$. The unit vectors are given by

$$\hat{s}_i = \frac{\hat{k}_i\times\hat{z}}{|\hat{k}_i\times\hat{z}|}$$
$$\hat{p}_i = \hat{s}_i\times\hat{k}_i \tag{35}$$

Equation (33) can be written in the more compact matrix form

$$\vec{E}_i = E_0 e^{i\vec{k}_i\cdot\vec{r}}\left(c_s, c_p\right)\cdot\left(\begin{array}{c}\hat{s}_i\\\hat{p}_i\end{array}\right) \tag{36}$$

$\hat{s}_i, \hat{p}_i$ decompose into incident local polarization unit vectors $\hat{s}_i^n$, $\hat{p}_i^n$ with respect to the local incident plane $(\hat{k}_i, \hat{n})$, given by

$$\hat{s}_i^n = \frac{\hat{k}_i\times\hat{n}}{|\hat{k}_i\times\hat{n}|}$$
$$\hat{p}_i^n = \hat{s}_i^n\times\hat{k}_i \tag{37}$$

Therefore,

$$\left(\begin{array}{c}\hat{s}_i\\\hat{p}_i\end{array}\right) = T_{in}\cdot\left(\begin{array}{c}\hat{s}_i^n\\\hat{p}_i^n\end{array}\right) \tag{38}$$

where T_{in} is the transformation matrix from incident coordinates to local coordinates

$$\bar{T}_{in} = \left(\begin{array}{cc}\hat{s}_i\cdot\hat{s}_i^n & \hat{s}_i\cdot\hat{p}_i^n\\\hat{p}_i\cdot\hat{s}_i^n & \hat{p}_i\cdot\hat{p}_i^n\end{array}\right) \tag{39}$$

Substituting (38) into (36), we have the incident field in terms of $\hat{s}_i^n, \hat{p}_i^n$ as

$$\vec{E}_i = E_0 e^{i\vec{k}_i\cdot\vec{r}}\left(c_s, c_p\right)\cdot\bar{T}_{in}\left(\begin{array}{c}\hat{s}_i^n\\\hat{p}_i^n\end{array}\right) \tag{40}$$

Reflections of the $\hat{s}_i^n, \hat{p}_i^n$ fields are found from the local Fresnel reflection coefficients for each component of polarization, i.e.,

$$\hat{s}_i^n \longrightarrow F_s\cdot\hat{s}_r^n$$
$$\hat{p}_i^n \longrightarrow F_p\cdot\hat{p}_r^n \tag{41}$$

---

[1]For example: for s polarization, $c_s = 1.0$, $c_p = 0$; for p polarization, $c_s = 0, c_p = 1$.

where F_s and F_p are the Fresnel reflection coefficients for s and p polarizations, respectively [14, p.100]. The unit vectors $\hat{s}_r^n, \hat{p}_r^n$ are the local polarization unit vectors for reflection from the tangent plane:

$$\hat{s}_r^n = \frac{\hat{k}_r \times \hat{n}}{|\hat{k}_r \times \hat{n}|} \qquad (42)$$

$$\hat{p}_r^n = \hat{s}_r^n \times \hat{k}_r \qquad$$

where $\hat{k}_r$ is the unit vector in the specular direction from the tangent plane, given by

$$\hat{k}_r = \hat{k}_i - 2(\hat{k}_i \cdot \hat{n})\hat{n} \qquad (43)$$

Using the Fresnel matrix

$$\bar{F} = \begin{pmatrix} F_s & 0 \\ 0 & F_p \end{pmatrix} \qquad (44)$$

we have in more compact form

$$\begin{pmatrix} \hat{s}_i^n \\ \hat{p}_i^n \end{pmatrix} \longrightarrow \bar{F} \begin{pmatrix} \hat{s}_r^n \\ \hat{p}_r^n \end{pmatrix} \qquad (45)$$

From equations (40) and (45), the scattered field on the surface can be expressed as a linear combination of the Fresnel reflection coefficients

$$\vec{E}_s = E_0 e^{i\hat{k}_i \cdot \vec{r}} (c_s, c_p) \cdot \begin{pmatrix} \hat{s}_i \cdot \hat{s}_i^n & \hat{s}_i \cdot \hat{p}_i^n \\ \hat{p}_i \cdot \hat{s}_i^n & \hat{p}_i \cdot \hat{p}_i^n \end{pmatrix} \cdot$$

$$\begin{pmatrix} F_s & 0 \\ 0 & F_p \end{pmatrix} \cdot \begin{pmatrix} \hat{s}_r^n \\ \hat{p}_r^n \end{pmatrix} \qquad (46)$$

The scattered field is a function of the incident polarization state, the local surface normal $\hat{n}$, the Fresnel reflection coefficients F_s and F_p of the surface, and the incident and reflection directions $\hat{k}_i, \hat{k}_r$.

A.3 Representation of the surface

Specification of the surface topography is required to carry out the surface integrals and surface averages appearing in equations (26), (27) and (29). Without losing generality, we assume the surface to be Gaussian distributed [5], i.e., we assume the surface height in (15) to be a stationary normally distributed random process whose mean value is zero. In addition we assume the surface to be directionally isotropic. An appropriate two-point joint probability function is given by

$$P(z_1, z_2) = \frac{exp\left[-(z_1^2 + z_2^2 - 2C(r)z_1 z_2)/2\sigma_0^2(1 - C(r)^2)\right]}{2\pi\sigma_0^2\sqrt{1 - C(r)^2}} \qquad (47)$$

where $r^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2$, σ_0^2 is the variance of $z_1 = \xi(x_1, y_1)$ and $z_2 = \xi(x_2, y_2)$, and $C(r)$ is the correlation coefficient, which is assumed to be [5]

$$C(r) = e^{-\frac{r^2}{\tau^2}} \qquad (48)$$

where τ is the autocorrelation length.

The parameters σ_0 and τ are the only two surface parameters required for the surface integrations.

A.4 Analytic evaluation of the integrals

Substituting (46) into (29) to (31), B_s and B_p are expressed in terms of known quantities and depend on the surface only through the normals n_1 and n_2 at two surface points. Further, the integrals in equations (26) and (27) can be written as:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-i\vec{v}\cdot\vec{\eta}} < e^{-i\vec{v}\cdot\hat{z}(\xi_1 - \xi_2)} \mathcal{F}(n_1, n_2, \mathbf{p}) > dxdy \qquad (49)$$

Stogryn [18] has shown that an integral and average of the form in (49) can be approximatly evaluated under either of the following two conditions:

- the surface is very rough (i.e., $(v_z\sigma)^2 \gg 1$)

- the surface has gentle slopes (i.e. $(\frac{\sigma}{\tau}) \ll 1$)

As a result, (49) reduces to

$$\mathcal{F}(\hat{n}_b, \hat{n}_b, \mathbf{p}) \cdot \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-i\vec{v}\cdot\vec{\eta}} < e^{-i\vec{v}\cdot\hat{z}(\xi_1 - \xi_2)} > dxdy \qquad (50)$$

where $\mathcal{F}$ is evaluated at $\hat{n}_b$, which is the unit vector bisecting $\hat{k}_i$ and $\hat{k}_r$, given by

$$\hat{n}_b = \frac{\hat{k}_r - \hat{k}_i}{|\hat{k}_r - \hat{k}_i|} \qquad (51)$$

Furthermore, the $< >$ in (50) can be shown to be [5]:

$$< e^{-i\vec{v}\cdot\hat{z}(\xi_1 - \xi_2)} > = e^{-(v_z\sigma)^2[1 - C(\eta)]} \qquad (52)$$

where $C(\eta)$ is given by (48).

Note that σ in (52) is the effective surface roughness, not σ_0. This is because the surface averaging is carried over illuminated and visible parts only. σ is given by [4]:

$$\sigma = \frac{\sigma_0}{\sqrt{1 + \frac{z_0^2}{\sigma_0^2}}} \qquad (53)$$

where z_0 depends on θ_i and θ_r and is the root of the following equation

$$\sqrt{\frac{\pi}{2}} z = \sigma_0 K \cdot \exp(-\frac{z^2}{2\sigma_0^2}) \qquad (54)$$

and

$$K = K_i + K_r$$

$$K_i = \frac{1}{4} \tan\theta_i \cdot erfc(\frac{\tau}{2\sigma_0} \cot\theta_i)$$

$$K_r = \frac{1}{4} \tan\theta_r \cdot erfc(\frac{\tau}{2\sigma_0} \cot\theta_r) \qquad (55)$$

The double integral in (50) can be evaluated analytically [5]:

$$N = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-i\vec{v}\cdot\vec{\eta}} < e^{-i\vec{v}\cdot\hat{z}(\xi_1 - \xi_2)} > dxdy$$

$$= e^{-g} \cdot A \cdot sinc^2(v_x L_x) sinc^2(v_y L_y) +$$

$$\pi\tau^2 \sum_{m=1}^{\infty} \frac{g^m e^{-g}}{m! \cdot m} \cdot \exp(-v_{xy}^2 \tau^2/4m) \qquad (56)$$

where L_x, L_y are the dimensions of the reflecting surface. Since we are only interested in cases when $L_x, L_y \gg \lambda$, the first term is nonzero only in the specular direction and zero otherwise. For the case of unidirectional incidence with solid angle $d\omega_i$ and $L_x, L_y \gg \lambda$, the averaged form of the first term in (56) is

$$A \cdot sinc^2(v_x L_x) sinc^2(v_y L_y) \rightarrow (2\pi\lambda)^2 \cdot \Delta/(d\omega_i \cdot \cos\theta_r) \quad (57)$$

Hence, (56) becomes

$$N = e^{-g} \cdot (2\pi\lambda)^2 \cdot \Delta/(d\omega_i \cdot \cos\theta_r) +$$
$$\pi\tau^2 \sum_{m=1}^{\infty} \frac{g^m e^{-g}}{m! \cdot m} \cdot \exp(-v_{xy}^2 \tau^2/4m) \quad (58)$$

Next, $\mathcal{F}_s$ and $\mathcal{F}_p$ in (30) and (31) are evaluated. First, $\hat{n}_1, \hat{n}_2$ are replaced by $\hat{n}_b$ defined in (51). Then they are substituted into (30) and (31). After lengthy vector manipulations, we find

$$\mathcal{F}(\hat{n}_b, \hat{n}_b, \mathbf{p})_s = \delta \cdot |c_s M_{ss} + c_p M_{sp}|^2 \quad (59)$$
$$\mathcal{F}(\hat{n}_b, \hat{n}_b, \mathbf{p})_p = \delta \cdot |c_s M_{ps} + c_p M_{pp}|^2 \quad (60)$$

where

$$M_{ss} = \left(F_s(\hat{p}_i \cdot \hat{k}_r)(\hat{p}_r \cdot \hat{k}_i) + F_p(\hat{s}_i \cdot \hat{k}_r)(\hat{s}_r \cdot \hat{k}_i)\right) \quad (61)$$
$$M_{sp} = -\left(F_s(\hat{s}_i \cdot \hat{k}_r)(\hat{p}_r \cdot \hat{k}_i) - F_p(\hat{p}_i \cdot \hat{k}_r)(\hat{s}_r \cdot \hat{k}_i)\right) \quad (62)$$
$$M_{pp} = \left(F_s(\hat{s}_i \cdot \hat{k}_r)(\hat{s}_r \cdot \hat{k}_i) + F_p(\hat{p}_i \cdot \hat{k}_r)(\hat{p}_r \cdot \hat{k}_i)\right) \quad (63)$$
$$M_{ps} = \left(F_s(\hat{p}_i \cdot \hat{k}_r)(\hat{s}_r \cdot \hat{k}_i) - F_p(\hat{s}_i \cdot \hat{k}_r)(\hat{p}_r \cdot \hat{k}_i)\right) \quad (64)$$
$$\delta = (\frac{2\pi}{\lambda})^2 \cdot \frac{|\hat{k}_r - \hat{k}_i|^4}{|\hat{k}_r \times \hat{k}_i|^4 \left(\hat{z} \cdot (\hat{k}_r - \hat{k}_i)\right)^2} \quad (65)$$

The Fresnel reflection coefficients F_s and F_p in (61) to (64) are evaluated at the bisecting angle given by $\cos^{-1}(|\hat{k}_r - \hat{k}_i|/2)$.

Using (59)-(65) and (58) in (26) and (27), we find an analytical expression for the reflected intensity

$$dI_s = \frac{|E_0|^2}{\cos\theta_r (4\pi)^2} \mathcal{F}(\hat{n}_b, \hat{n}_b, \mathbf{p})_s \cdot N$$
$$dI_p = \frac{|E_0|^2}{\cos\theta_r (4\pi)^2} \mathcal{F}(\hat{n}_b, \hat{n}_b, \mathbf{p})_p \cdot N \quad (66)$$

where the square of the absolute value of the incident field amplitude, $|E_0|^2$, is related to the incident intensity I_i by

$$|E_0|^2 = I_i d\omega_i \quad (67)$$

Note that the right side of (66) has the correct dimensions of intensity since N has dimension $[L^2]$ whereas the $\mathcal{F}$'s have dimension $[L^{-2}]$.

Finally, substituting (67) into (66) and using (4) and (32), we get exactly the first two terms in (5), given that

$$\mathcal{F}(\hat{k}_i, \hat{k}_r, \mathbf{p}) = \mathcal{F}(\hat{n}_b, \hat{n}_b, \mathbf{p})_s + \mathcal{F}(\hat{n}_b, \hat{n}_b, \mathbf{p})_p \quad (68)$$

since the BRDF defined in (4) is the total BRDF, which is the sum of the BRDF's for the reflected s and p components.

B Appendix: Governing equations of the reflectance model for unpolarized incident light

Equations (5) to (8) together with the defining equations for all the symbols in (5) to (8) completely define the general BRDF for arbitrarily-polarized incident light. In most applications, however, we are only interested in the BRDF for unpolarized incident light. The expressions for the BRDF are greatly simplified for this special but useful case. For convenience, the BRDF equations for unpolarized incident light are presented in this appendix. The reader should refer to Figure 5 and the nomenclature list in Table 1 for the angular coordinates and other physical parameters that appear in the reflectance model:

$$\rho_{bd} = \rho_{bd}(\lambda, \sigma_0, \tau, \bar{n}(\lambda), a(\lambda))$$
$$= \rho_{bd,sp} + \rho_{bd,dd} + \rho_{bd,ud} \quad (69)$$
$$\rho_{bd,sp} = \frac{\rho_s}{\cos\theta_i d\omega_i} \cdot \Delta \quad (70)$$
$$\rho_{bd,dd} = \frac{|F|^2}{\pi} \cdot \frac{G \cdot S \cdot D}{\cos\theta_i \cos\theta_r} \quad (71)$$
$$\rho_{bd,ud} = a(\lambda) \quad (72)$$
$$\rho_s = |F|^2 \cdot e^{-g} \cdot S \quad (73)$$
$$\Delta = \begin{cases} 1 & \text{if in specular cone} \\ 0 & \text{otherwise} \end{cases} \quad (74)$$
$$|F|^2 = \frac{1}{2}(F_s^2 + F_p^2) = f(\theta_i, \theta_r, \bar{n}(\lambda)) \quad (75)$$
$$G = \left(\frac{\vec{v} \cdot \vec{v}}{v_z}\right)^2 \cdot \frac{1}{|\hat{k}_r \times \hat{k}_i|^4} \cdot [(\hat{s}_r \cdot \hat{k}_i)^2 + (\hat{p}_r \cdot \hat{k}_i)^2] \cdot$$
$$[(\hat{s}_i \cdot \hat{k}_r)^2 + (\hat{p}_i \cdot \hat{k}_r)^2] \quad (76)$$
$$S = S(\theta_i, \theta_r, \sigma_0/\tau) \quad (77)$$
$$D = \frac{\pi^2 \tau^2}{4\lambda^2} \cdot \sum_{m=1}^{\infty} \frac{g^m e^{-g}}{m! \cdot m} \cdot \exp(-v_{xy}^2 \tau^2/4m) \quad (78)$$
$$g = [(2\pi\sigma/\lambda)(\cos\theta_i + \cos\theta_r)]^2 \quad (79)$$
$$\sigma = \sigma_0 \cdot [1 + (\frac{z_0}{\sigma_0})^2]^{-1/2} \quad (80)$$
$$\sqrt{\frac{\pi}{2}} z_0 = \frac{\sigma_0}{4}(K_i + K_r) \cdot \exp(-\frac{z_0^2}{2\sigma_0^2}) \quad (81)$$
$$K_i = \tan\theta_i \cdot erfc(\frac{\tau}{2\sigma_0} \cot\theta_i) \quad (82)$$
$$K_r = \tan\theta_r \cdot erfc(\frac{\tau}{2\sigma_0} \cot\theta_r) \quad (83)$$
$$\vec{v} = \hat{k}_r - \hat{k}_i, \quad v_{xy} = \sqrt{v_x^2 + v_y^2} \quad (84)$$
$$\hat{s}_i = \frac{\hat{k}_i \times \hat{n}}{|\hat{k}_i \times \hat{n}|}, \quad \hat{p}_i = \hat{s}_i \times \hat{k}_i \quad (85)$$
$$\hat{s}_r = \frac{\hat{k}_r \times \hat{n}}{|\hat{k}_r \times \hat{n}|}, \quad \hat{p}_r = \hat{s}_r \times \hat{k}_r \quad (86)$$

where $\bar{n}$ is the index of refraction, ρ_s is the specular reflectivity, Δ is a delta function, $|F|^2$ is the Fresnel reflectivity for unpolarized light [14, p.100] evaluated at the bisecting angle given by $\cos^{-1}(|\hat{k}_r - \hat{k}_i|/2)$, G is a geometrical factor, S is the shadowing/masking factor given in equation (23), and D is a distribution function for the directional diffuse reflection term.

Measuring and Modeling Anisotropic Reflection

Gregory J. Ward
Lighting Systems Research Group
Lawrence Berkeley Laboratory

ABSTRACT

A new device for measuring the spatial reflectance distributions of surfaces is introduced, along with a new mathematical model of anisotropic reflectance. The reflectance model presented is both simple and accurate, permitting efficient reflectance data reduction and reproduction. The validity of the model is substantiated with comparisons to complete measurements of surface reflectance functions gathered with the novel reflectometry device. This new device uses imaging technology to capture the entire hemisphere of reflected directions simultaneously, which greatly accelerates the reflectance data gathering process, making it possible to measure dozens of surfaces in the time that it used to take to do one. Example measurements and simulations are shown, and a table of fitted parameters for several surfaces is presented.

General Terms: algorithms, measurement, theory, verification. **CR Categories and Descriptors:** I.3.7 Three-dimensional graphics and realism, I.6.4 Model validation and analysis. **Additional Keywords and Phrases:** reflectance, Monte Carlo, raytracing, shading.

1. Introduction

Numerous empirical and theoretical models for the local reflection of light from surfaces have been introduced over the past 20 years. Empirical and theoretical models have the same goal of reproducing real reflectance functions, but the respective approaches are very different.

An empirical model is simply a formula with adjustable parameters designed to fit a certain class of reflectance functions. Little attention is paid to the physical derivation of the model, or the physical significance of its parameters. A good example of an empirical model is the one developed by Sandford [Sandford85]. This is a four parameter model of isotropic reflection, where the parameters must be fit to a specific set of reflectance measurements. While two of these parameters correspond roughly to measurable quantities such as total reflectance and specularity, the other two parameters have no physical significance and are merely shape variables that make the specular lobe of the model more closely match the data.

In contrast to an empirical model, a theoretical model attempts to get closer to the true distribution by starting from physical theory. A good example of a theoretical model is the one derived recently by He et al [He91]. This is also a four parameter isotropic model, but all four parameters have some physical meaning and can in principle be measured separately from the surface reflectance distribution. In practice, however, it is usually necessary to fit even a theoretical model to measurements of reflectance because the physical parameters involved are difficult to measure. This is the case in the He-Torrance model, since measurements of the requisite surface height variance and auto-

correlation distance variables are impractical for most surfaces. Thus, the physical derivation of such a model serves primarily to inspire greater confidence, and is not necessarily a practical advantage when it comes to fitting measured data. As in all scientific disciplines, if the theory does not fit the data, then the theory must be discarded, not the data.

But where is the data? There is almost no published data on surface reflectance as a function of angle, and what little data is available is in the form of plane measurements of isotropic surfaces with no rotational variance in their reflectance functions. Thus, we have little to compare our reflectance models to, and no real assurance that they are valid. This means that we may once again be falling back on the "if it looks reasonable then it's OK" philosophy that has misdirected computer graphics so often in the past.

Why is the oldest specular model, the one introduced by Phong in 1975 [Phong75], still the most widely used to this day? This model is neither theoretically plausible nor empirically correct. Any renderings that use the straight Phong model are most likely wrong because the model is not physical, and more light may be emitted than is received (for example). The sole virtue of the Phong model is its mathematical simplicity.

Simplicity is indispensable in computer graphics. Simplicity is what permits fast renderings and hardware implementations. Without it, a reflectance model is little more than a novelty. Even a relatively straightforward model such as the one developed by Torrance and Sparrow [Torrance67] and tailored for rendering applications by Blinn [Blinn77] and later Cook [Cook82] has been underutilized in computer graphics due to its moderately complex form. More recent introductions by Poulin and Fournier [Poulin90] as well as He et al [He91] are even more complex. What is really needed for computer graphics is a simple reflectance model that works reasonably well for most materials.

Our goal in this paper is not to present the ultimate mathematical model of reflectance, but to provide a simple formula that is physically valid and fits measured reflectance data. Here we will present both a new method for measuring isotropic and anisotropic reflectance distributions and a mathematical model that fits these data with both accuracy and simplicity.

2. Definition of the BRDF

The interaction of light with a surface can be expressed as a single function, called the *bidirectional reflectance distribution function*, or BRDF for short [Nicodemus77]. This is a function of four angles, two incident and two reflected, as well as the wavelength and polarization of the incident radiation. For the sake of simplicity, we will leave wavelength and polarization out of our equations, but keep in mind that they are contained implicitly in the function ρ_{bd}, which is defined in terms of incident and reflected radiance by the following integral:

Author's address: 1 Cyclotron Rd., 90-3111, Berkeley, CA 94720.
E-mail: GJWard@lbl.gov

$$L_r(\theta_r,\phi_r) = \int_0^{2\pi} \int_0^{\pi/2} L_i(\theta_i,\phi_i)\, \rho_{bd}(\theta_i,\phi_i;\theta_r,\phi_r)\, \cos\theta_i\, \sin\theta_i\, d\theta_i\, d\phi_i \qquad (1)$$

where: ϕ is the azimuthal angle measured about the surface normal

$L_r(\theta_r,\phi_r)$ is the reflected radiance (watts/steradian/meter^2)

$L_i(\theta_i,\phi_i)$ is the incident radiance

$\rho_{bd}(\theta_i,\phi_i;\theta_r,\phi_r)$ is the BRDF (steradian^{-1})

The function ρ_{bd} is *bidirectional* because the incident and reflected directions can be reversed and the function will return the same value. This arises from the fact that the physics of light is the same run backwards as forwards, which is why light-backwards ray tracing works [Whitted80].

3. Measuring the BRDF of a Surface

A device for measuring BRDFs is called a *gonioreflectometer*. The usual design for such a device incorporates a single photometer that is made to move in relation to a surface sample, which itself moves in relation to a light source, all under the control of a computer. Because BRDFs are in general a function of four angles, two incident and two reflected, such a device must have four degrees of mechanical freedom to measure the complete function. This requires substantial complexity in the apparatus and long periods of time to measure a single surface. A typical gonioreflectometer arrangement, designed by Murray-Coleman and Smith [Murray-Coleman90], is shown in Figure 1.
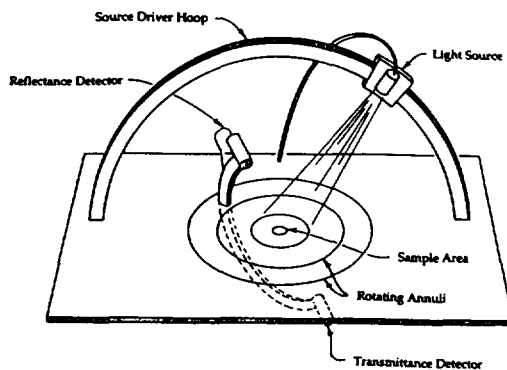
Figure 1. A conventional gonioreflectometer with movable light source and photometer.

As an alternative to building such a gonioreflectometer, there are several labs in North America where one can send a surface sample for BRDF characterization. For a few hundred dollars, one can get a three plane measurement of an isotropic material at four or five angles of incidence. (An isotropic material has a BRDF that is independent of rotation about the the normal. Therefore, only one ϕ_i direction is sampled.) Unfortunately, a comprehensive BRDF measurement of an anisotropic surface typically costs a few *thousand* dollars. (An anisotropic material reflects light differently at different angles of rotation, thus multiple ϕ_i directions must be sampled.) Because of the difficulty and expense of the BRDF measurements themselves, only the very richest research programs can afford their own data. This data is essential, however, for the correct modeling of surface reflectance.

3.1. An Imaging Gonioreflectometer

The Lighting Systems Research Group at Lawrence Berkeley Laboratory has developed a relatively simple device for measuring BRDFs that uses imaging technology to obtain results more quickly and at a lower cost than conventional methods. This *imaging gonioreflectometer* has been developed over the past three years and represents an important advance towards the more practical characterization of BRDFs for lighting simulation and computer graphics. It is our hope that other laboratories and research institutions will construct their own versions of this apparatus and thereby make BRDF measurement a more common and economical practice.

The basic arrangement of the LBL imaging gonioreflectometer is shown in Figure 2†. The key optical elements are a half-silvered hemisphere or hemi-ellipsoid and a charge-coupled device (CCD) camera with a fisheye lens. Combined, these elements take care of the two degrees of freedom handled by a mechanically controlled photometer in a conventional gonioreflectometer. Light reflected off the sample surface in
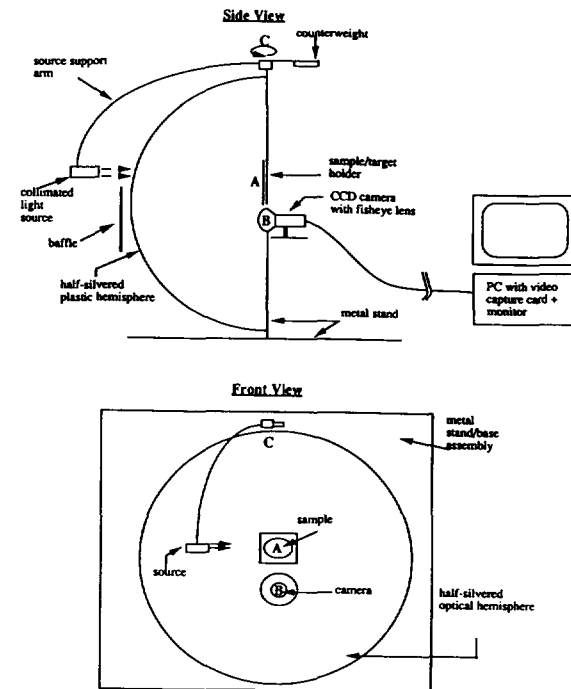
Figure 2. The LBL imaging gonioreflectometer.

holder A is collected by the hemispherical mirror and reflected back into the fisheye lens and onto the CCD array B. By focusing the lens at one half the hemisphere radius, a near perfect imaging of the reflected angles takes place. (See ray diagram in Figure 3.) Because of this highly efficient collector arrangement, the light source does not have to be very bright to obtain a good measurement, and can thus be optimized for collimation to get the best possible angular resolution. In our device, a 3-watt quartz-halogen lamp is used with an optically precise parabolic reflector to produce a well collimated beam. White light is preferable for photopic measurements, although an array of colored filters may be used to measure the spectral dependence of the BRDF. The hemisphere is half-silvered to allow the light beam to illuminate the sample, and an exterior baffle shields the camera from stray radiation. This unique arrangement of light source and optics allows retroreflection (light reflected back towards the light source) and transmission to be measured as well.

The incident θ_i and ϕ_i angles are controlled mechanically by pivoting the light source arm at point C and the sample holder at point A, respectively. In our current prototype, the light source is moved by a computer-controlled motor during data collection, and the sample is rotated manually. Because the hemisphere of reflected directions is captured in a single image, data collection proceeds quite rapidly and a complete BRDF can be recorded in a few minutes, including time for manual rotation of the sample.

3.2. Calibration and Data Reduction

All measurements are made relative to a standard diffuse sample and a background measurement. The background measurement is made with the source on but without any sample in the holder (using the dark room behind to simulate a black body), and is subtracted from the other measurements to reduce the effects of stray and ambient light. The standard sample measurement is used as a basis for obtaining absolute reflectance values using the following simple equation at each image point:

$$\rho_{bd} = \frac{V_{measured} - V_{background}}{V_{standard} - V_{background}} \cdot \frac{\rho_{standard}}{\pi} \qquad (2)$$

where:

$\rho_{standard}$ is the total diffuse reflectance of the standard sample

The ability to measure absolute BRDF values directly is an important feature of the imaging gonioreflectometer. Most other devices rely on auxiliary measurements of directional reflectance (ie. total reflectance for
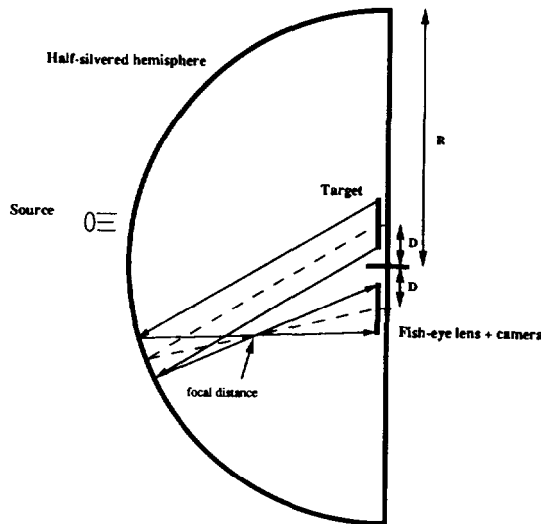
Figure 3. Imaging gonioreflectometer geometry. Light reflected by the sample in a specific direction is focused by the hemisphere or hemi-ellipsoid through a fisheye lens onto a CCD imaging array.

light incident at some (θ_i, ϕ_i)) and numerical integration to arrive at absolute quantities.

Recovering the reflected angles from pixel locations in the captured image is accomplished in two steps. The first step is to determine the mapping from image point locations to the lens incident direction. This is a function of the particular fisheye lens used, the camera, and the video capture board. Since this mapping varies so much from one implementation to the next and is easily measured, we will not discuss it any further here. The second step is to compute the target reflection angles from these camera incident angles. Figure 3 shows the geometry involved, and after a bit of trigonometry one can derive the following approximation:

$$r_c = D\sin\phi_c\sin\theta_c + \sqrt{D^2\sin^2\phi_c\sin^2\theta_c + R^2 - D^2}$$

$$\theta_r = \cos^{-1}\left[\frac{r_c\cos\theta_c}{\sqrt{r_c^2\cos^2\phi_c\sin^2\theta_c + (r_c\sin\phi_c\sin\theta_c - 2D)^2 + r_c^2\cos^2\theta_c}}\right] \quad (3)$$

$$\phi_r = \tan^{-1}\left[r_c\sin\phi_c\sin\theta_c - 2Dr_c\cos\phi_c\sin\theta_c\right]$$

where:

θ_r is polar angle relative to target

ϕ_r is azimuthal angle relative to target, right is $0°$

θ_c is polar angle relative to camera

ϕ_c is azimuthal camera angle, right is $0°$

R is radius of sphere or approximate radius of ellipsoid

D is one half the separation between target and camera centers

r_c is an intermediate result which is the distance from camera to reflector

notes:

The arctangent in the above equation should be computed using the signs of the numerator and denominator to get a range of $360°$. Many math libraries provide a function named atan2 for this purpose.

The above equations are a good approximation both for hemispherical and hemi-ellipsoidal reflectors as long as D is small in relation to R.
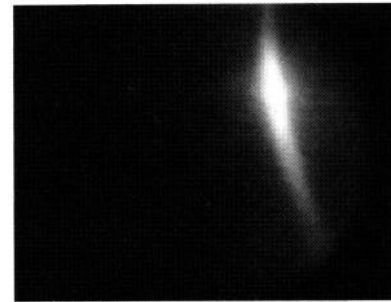
Figure 4. An image captured by the gonioreflectometer from an unfinished aluminum sample.

The image captured by our gonioreflectometer for a piece of unfinished aluminum illuminated at $(\theta_i, \phi_i)=(30°, 0°)$ is shown in Figure 4. Although the image was reduced before data reduction to a resolution of 108 by 80 pixels, there is still much more information than is needed for an accurate lighting simulation. Also, since two or more f-stops may be used to capture the full dynamic range of the BRDF, there is often redundant information where the useful ranges of exposures overlap. We therefore apply a program to eliminate crowding of data points and insure that the peak is recorded at a high enough angular resolution while the rest of the usable distribution is recorded at a uniform density. A data fitting program can then be used to match the reduced data set to a specific reflectance model.

3.3. Measurement Limitations

Our current implementation of the imaging gonioreflectometer has two main limitations in its measurement abilities. First, we are limited in our ability to measure the reflectance function near grazing angles, due to the size and shape of our reflecting hemisphere and the size of our sample. Our present hemisphere is formed from acrylic plastic and its optical properties are less than perfect, especially near the edges. It should be possible to partially overcome this limitation by placing the sample at right angles to its current configuration and illuminating it through the target holder, but this has not yet been tried. The ultimate solution would be to go to a larger, more precise hemisphere and a larger sample target.

The second limitation is our inability to measure more polished surfaces with sharp specular peaks. Again, the optical precision of our hemisphere is a problem, but so is the finite collimation of our light source. A highly uniform, collimated light source is required for the measurement of polished surfaces. That is why many commercial gonioreflectometers employ a laser, despite the laser's inability to yield spectrally balanced measurements. By using an incandescent source with an even smaller filament, it should be possible to measure more polished surfaces without resorting to a laser.

Note that the BRDF of a perfectly smooth surface is not directly measurable by any gonioreflectometer, since it is a Dirac delta function with an infinite value at a single point. Measuring such a BRDF of such a surface is not required however, since the physics of smooth surfaces are well understood and measurements of total reflectance are adequate for their characterization.

4. Modeling Anisotropic Reflectance

Armed with a device that can measure anisotropic reflectance functions economically, we need a mathematical model that can be fit to our newfound data. Using the data directly is impractical because it requires too much memory, and oftentimes the data is noisy and not complete enough to cover the entire domain of the BRDF. We could represent the BRDF as a sum of 100 or so terms in a spherical harmonic series, but this would also be expensive in terms of computation time and of memory [Cabral87][Sillion91]. We would prefer a model that fits the data with as few parameters as possible. Ideally, these parameters would be either physically derived or meaningful so that they could be set manually in the absence of any data at all.

Many models have been suggested for isotropic reflection, but only a few models have been published for the more general anisotropic case. Kajiya published a fairly robust method for deriving BRDFs of metals

from surface microstructure [Kajiya85]. However, his approach is not amenable to fitting measured reflectance data because the parameter space is too large (ie. all possible surface microstructures) and the BRDFs take too long to compute. Poulin and Fournier developed a model based on cylindrical scratches that is better suited [Poulin90]. but their model is restricted to a specific microstructure with cross-sectional uniformity, and its evaluation is still somewhat expensive.

Our goal is to fit our measured reflectance data with the simplest empirical formula that will do the job. If we can develop a model with physically meaningful parameters without adding undue complexity, so much the better.

4.1. The Isotropic Gaussian Model

The Gaussian distribution has shown up repeatedly in theoretical formulations of reflectance [Beckmann63][Torrance67][Cook82], and it arises from certain minimal assumptions about the statistics of a surface height function. It is usually preceded by a Fresnel coefficient and geometrical attenuation factors, and often by an arbitrary constant. Since the geometric attenuation factors are typically difficult to integrate and tend to counteract the Fresnel factor anyway, we have replaced all of these coefficients with a single normalization factor that simply insures the distribution will integrate easily and predictably over the hemisphere.

$$\rho_{bd,iso}(\theta_i,\phi_i;\theta_r,\phi_r) = \frac{\rho_d}{\pi} +$$
$$\rho_s \cdot \frac{1}{\sqrt{\cos\theta_i \cos\theta_r}} \cdot \frac{\exp[-\tan^2\delta/\alpha^2]}{4\pi\alpha^2} \qquad (4)$$

where:

ρ_d is the diffuse reflectance

ρ_s is the specular reflectance

δ is the angle between vectors $\hat{n}$ and $\hat{h}$ shown in Figure 5

α is the standard deviation (RMS) of the surface slope

notes:

The ρ values may have some spectral dependence, and this dependence may vary as a function of angle so long as $\rho_d + \rho_s$ (the total reflectance) is less than 1. Thus, Fresnel effects may be modeled if desired.

The normalization factor, $\frac{1}{4\pi\alpha^2}$, is accurate as long as α is not much greater than 0.2, when the surface becomes mostly diffuse.

The main difference between this isotropic Gaussian reflectance model and that of Phong is its physical validity. For example, most Phong implementations do not have the necessary bidirectional characteristics to constitute a valid BRDF model. It is clear by inspection that the above formula is symmetric with respect to its incident and reflected angles. Without this symmetry, a BRDF model cannot possibly be physical because the simulated surface reflects light differently in one direction than the other, which is forbidden by natural law. Also, without proper normalization, a reflectance model does not yield correct energy balance and thus cannot produce physically meaningful results. Even
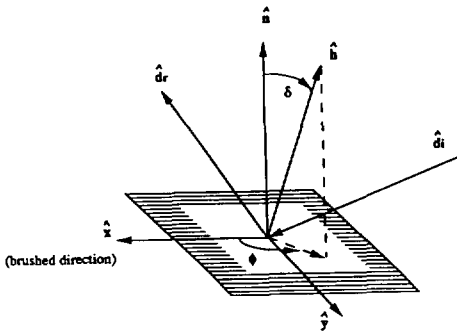
the model introduced recently by He et al [He91] with its rigorous physical derivation does not seem to pay close enough attention to normalization. Specifically, the so-called ambient term in the He-Torrance model is added without regard to the overall reflectance of the material, which by nature of the model is very difficult to compute. Comparisons were not made in He's paper between the reflectance model and absolute BRDF measurements (the data was scaled to match the function), thus normalization was not even demonstrated empirically. The fact that normalization was not adequately treated in He's otherwise impeccable derivation shows just how much normalization is overlooked and under-valued in reflectance modeling. The simplicity of the model presented here is what allows us to incorporate built-in normalization and has other desirable features as well, such as permitting quick evaluation for data reduction and Monte Carlo sampling.

4.2. The Anisotropic (Elliptical) Gaussian Model

It is relatively simple to extend the Gaussian reflectance model to surfaces with two perpendicular (uncorrelated) slope distributions, α_x and α_y. The normalized distribution is as follows:

$$\rho_{bd}(\theta_i,\phi_i;\theta_r,\phi_r) = \frac{\rho_d}{\pi} +$$
$$\rho_s \cdot \frac{1}{\sqrt{\cos\theta_i \cos\theta_r}} \cdot \frac{\exp[-\tan^2\delta(\cos^2\phi/\alpha_x^2 + \sin^2\phi/\alpha_y^2)]}{4\pi\alpha_x\alpha_y} \qquad (5a)$$

where:

ρ_d is the diffuse reflectance

ρ_s is the specular reflectance

α_x is the standard deviation of the surface slope in the $\hat{x}$ direction

α_y is the standard deviation of the surface slope in the $\hat{y}$ direction

δ is the angle between the half vector, $\hat{h}$ and the surface normal, $\hat{n}$.

ϕ is the azimuth angle of the half vector projected into the surface plane.

A computationally convenient approximation for ρ_{bd} is:

$$\rho_{bd}(\theta_i,\phi_i;\theta_r,\phi_r) = \frac{\rho_d}{\pi} +$$
$$\rho_s \cdot \frac{1}{\sqrt{\cos\theta_i \cos\theta_r}} \cdot \frac{1}{4\pi\alpha_x\alpha_y}\exp\left[-2\frac{\left[\frac{\hat{h}\cdot\hat{x}}{\alpha_x}\right]^2 + \left[\frac{\hat{h}\cdot\hat{y}}{\alpha_y}\right]^2}{1 + \hat{h}\cdot\hat{n}}\right] \qquad (5b)$$

where:

$$\hat{h}\cdot\hat{x} = \frac{\sin\theta_r\cos\phi_r + \sin\theta_i\cos\phi_i}{||\vec{h}||}$$

$$\hat{h}\cdot\hat{y} = \frac{\sin\theta_r\sin\phi_r + \sin\theta_i\sin\phi_i}{||\vec{h}||}$$

$$\hat{h}\cdot\hat{n} = \frac{\cos\theta_r + \cos\theta_i}{||\vec{h}||}$$

$$||\vec{h}|| = \left[2 + 2\sin\theta_r\sin\theta_i(\cos\phi_r\cos\phi_i + \sin\phi_r\sin\phi_i) + 2\cos\theta_r\cos\theta_i\right]^{1/2}$$

For vector calculations, the following substitutions are used:

$$\vec{h} = \hat{d}_r + \hat{d}_i$$

$$\hat{h} = \frac{\vec{h}}{||\vec{h}||}$$

$$\cos\theta_r = \hat{d}_r\cdot\hat{n}$$

$$\cos\theta_i = \hat{d}_i\cdot\hat{n}$$

where:

$\hat{d}_r$ is the reflected ray direction (away from surface)

$\hat{d}_i$ is the incident ray direction (away from surface)

$\hat{x}$ is a unit vector in the surface plane

$\hat{y}$ is a unit vector in the surface plane perpendicular to $\hat{x}$

Figure 5. Angles and vectors used in reflection equations. The incident light arrives along vector $\hat{d}_i$ and is measured or simulated in direction $\hat{d}_r$. The polar angle between the half vector $\hat{h}$ and the surface normal $\hat{n}$ is δ. The azimuthal angle of $\hat{h}$ from the direction $\hat{x}$ is ϕ.
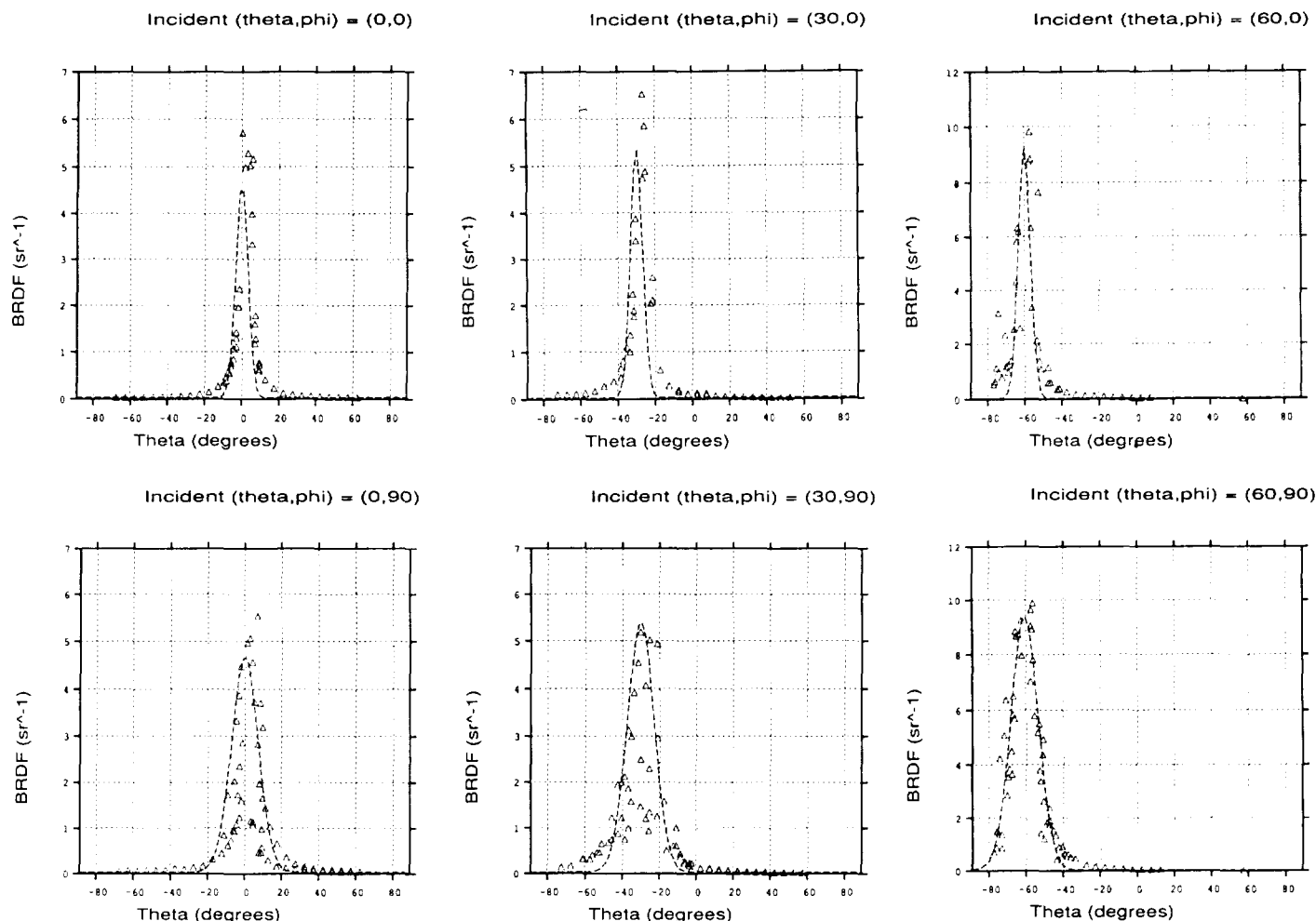
Figure 6. Measured data and elliptical Gaussian fit for unfinished aluminum. Unfinished aluminum exhibits anisotropy from rolling during its manufacture.

As in the isotropic case, the normalization of the above anisotropic model is such that the total surface reflectance will equal the diffuse reflectance coefficient, ρ_d, plus the "rough specular" or "directional-diffuse" coefficient, ρ_s. The two other model parameters, α_x and α_y, represent the standard deviation of the surface slope in each of two perpendicular directions. Thus, all four of the model's parameters have physical meaning and can be set independently of measured data to produce a valid reflectance function. As long as the total reflectance, $\rho_d + \rho_s$, is less than 1 and the two α's are not too large, Equation 5 will yield a physically valid reflectance function.

The elliptical nature of our model arises from the two perpendicular slope distributions, and is apparent in the exponent of Equation 5a. A similar elliptical reflectance model was developed by Ohira and described by Yokoi and Toriwaki [Yokoi88], but this model was derived from that of Phong and likewise lacks physical meaning. By starting with a valid, normalized function, it is much easier to fit the model parameters to physical measurements as well as other specifications such as appearance.

Our simple four parameter model fits well the data we have gathered from anisotropic surfaces such as varnished wood and unfinished (rolled) or brushed metals. Because of its simplicity, it is easy to apply a least squares error minimization method to fit a set of parameters to measured data automatically. Automatic data fitting is essential to the economic modeling of surface reflectance for any significant database of materials. Figure 6 shows an example fit to the BRDF of an unfinished aluminum sample. Although the full hemisphere of reflected data was measured at 21 incident angles, it is difficult to visualize the 21 corresponding 3-dimensional point plots. We therefore present here only a slice of the data in the incident plane at 6 angles. The results section (6) lists the fitted parameters for this material as well as some other example surfaces.

5. Rendering Anisotropic Surfaces

The challenge to applying a new reflectance model to computer graphics is to approximate the luminance equation (1) in a manner that is unbiased and has low variance [Kajiya86]. Unfortunately, unbiased techniques (ie. pure Monte Carlo) tend to have high variance, while low variance approaches (ie. closed-form approximations) tend to be biased. To satisfy these conflicting requirements, we use a hybrid deterministic and stochastic ray tracing technique [Cook84][Cook86]. A strictly deterministic calculation of the highlight contribution of sources, similar to the widely used Whitted approximation [Whitted80], fails to pick up indirect semispecular contributions as demonstrated in Figure 7a. (Note that the crescent shape of the highlight is due to longitudinal anisotropy and not the light source.) Conversely, relying solely on stochastic sampling causes the highlights from sources to show high variance in the form of excessive noise, even with 16 samples per pixel (Figure 7b). By combining the two techniques, using a deterministic solution for source contributions and a stochastic sampling for indirect contributions, we get a clean result without compromising accuracy. Figure 7c was calculated using the hybrid technique and the same number of samples as Figure 7b. Both figures took approximately the same time to compute. (Figure 7a took less time since no sampling was required.)

The hybrid approach described reduces to the following equation:

$$L(\theta_r, \phi_r) = I\frac{\rho_d}{\pi} + L_s \rho_s + \sum_{i=1}^{N} L_i \omega_i \cos\theta_i \rho_{bd}(\theta_i, \phi_i; \theta_r, \phi_r) \qquad (6)$$

where:

I is the indirect irradiance at this point (a constant ambient level or the result of a diffuse interreflection or radiosity calculation)

L_s is the radiance value in the Monte Carlo sample direction given in Equation 7 below

L_i is the radiance of light source i

ω_i is the solid angle (in steradians) of light source i

N is the number of light sources

ρ_{bd} is the elliptical Gaussian function defined in Equation 5

In applying this technique, it is very important not to bias the sample by overcounting the specular component. Bias is easily avoided by associating a flag with the stochastically sampled specular ray. If the ray hits a light source whose contribution is being included in a closed form calculation, then the ray is not counted. Few rays are wasted in this way, since light sources occupy a small amount of the visual space in most scenes.

5.1. Stochastic Sampling of Elliptical Gaussian

Because of its simplicity, the elliptical Gaussian model adapts easily to stochastic sampling techniques. Using standard Monte Carlo integral conversion methods [Rubenstein81], we can write the following formulas for obtaining uniformly weighted sample directions for each L_s ray in Equation 6:

$$\delta = \left[\frac{-\log(u_1)}{\cos^2\phi/\alpha_x^2 + \sin^2\phi/\alpha_y^2}\right]^{\frac{1}{2}} \qquad (7a)$$

$$\phi = \tan^{-1}\left[\frac{\alpha_y}{\alpha_x}\tan(2\pi u_2)\right] \qquad (7b)$$

where:

δ, ϕ are the angles shown in Figure 5

u_1, u_2 are uniform random variables in the range (0,1]

notes:

The tangent and arctangent in the Equation 7a should be computed carefully so as to keep the angle in its starting quadrant.

Uniformly weighted sample rays sent according to the above distribution will correctly reproduce the specified highlight. This is much more efficient than either distributing the samples evenly and then weighting the result, or using other techniques, such as rejection sampling, to arrive at the correct scattering. Readers familiar with Monte Carlo sampling techniques will immediately appreciate the advantage of having a formula for the sample point locations -- something that is impossible with more complicated reflectance models such as He-Torrance.

6. Results

Figure 8a shows a photograph of a child's varnished wood chair with a small desk lamp immediately behind and above it. This arrangement results in a large anisotropic highlight in the seat of the chair. Figure 8b shows the closest simulation possible using a deterministic isotropic reflectance model. Figure 8c shows a hybrid simulation with the elliptical Gaussian model. Notice how the hybrid rendering technique reproduces not only the highlight from the light source, but also the semispecular reflection from the back wall in the seat of the chair.

Figure 9 shows a table with anisotropic reflections in the wood varnish and the two candle holders. The lid of the silver box shown is also anisotropic, and demonstrates the use of local control to affect the reflectance properties of an anisotropic surface. A wave function determines the orientation of the brushed direction in the box lid, producing characteristic highlights. There are four low level light sources in the scene, the two candles on the table, an overhead light source above and to the right, and the moon shining in through a window.
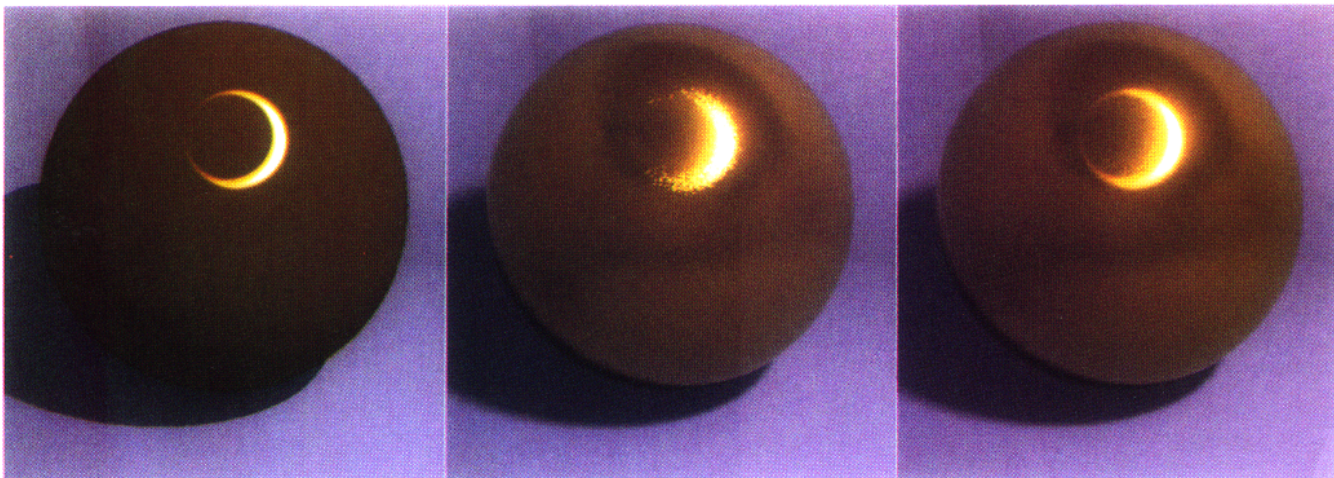
Figure 7a, 7b, 7c. Alternative rendering techniques for anisotropic reflection. 7a on the left shows deterministic technique with no sampling. 7b center shows strict Monte Carlo sampling approach. 7c on the right shows hybrid deterministic and stochastic method.
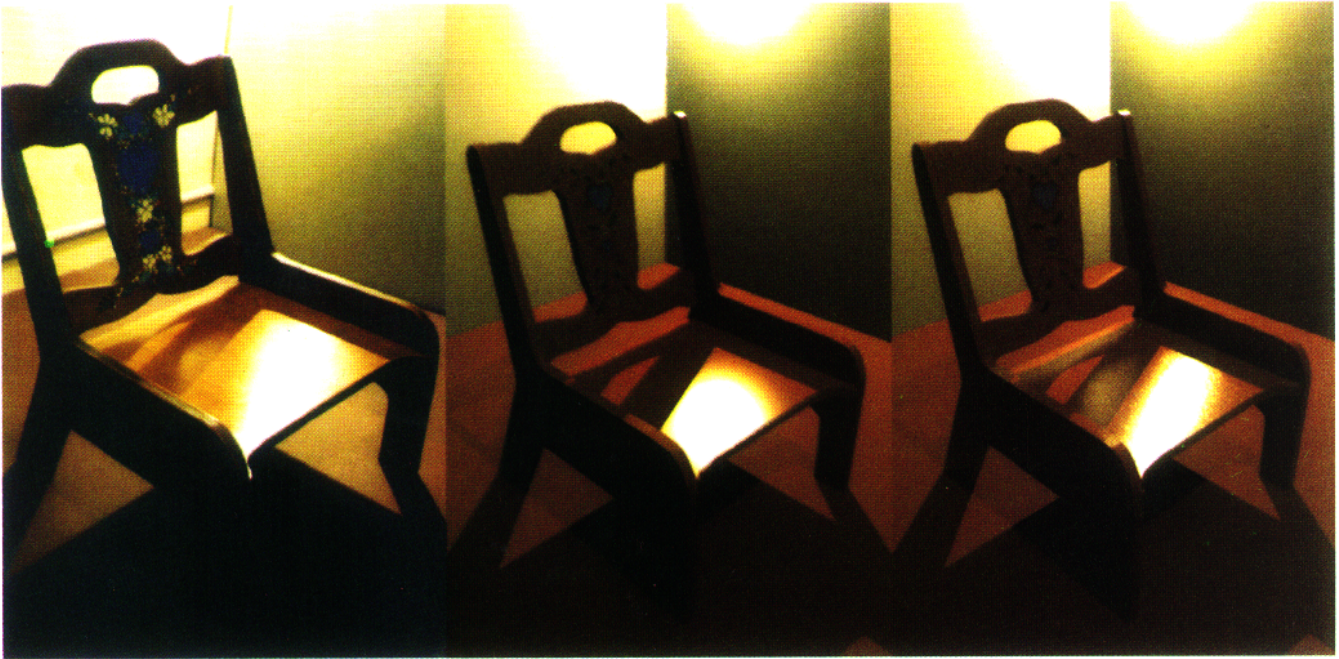
Figure 8a, 8b, 8c. Varnished wood comparison. 8a on the left shows a photograph of a child's chair. 8b center shows a simulation of the chair using the isotropic Gaussian model given in Section 4.1 with a strictly deterministic calculation. (This is similar to the appearance one might obtain using a normalized Phong reflectance model.) 8c on the right shows a hybrid deterministic and stochastic simulation of the chair using the elliptical Gaussian model from Section 4.2.

Figure 9. A table scene with anisotropic reflection in metallic and varnished wood surfaces.

The following table gives a short list of surfaces and their elliptical Gaussian fits. Color was not measured for any of the surfaces. The materials in the second half of the table are isotropic, so the two α values are the same, and Equation (4) can be used.

Material	ρ_d	ρ_s	α_x	α_y
rolled brass	.10	.33	.050	.16
rolled aluminum	.1	.21	.04	.09
lightly brushed aluminum	.15	.19	.088	.13
varnished plywood	.33	.025	.04	.11
enamel finished metal	.25	.047	.080	.096
painted cardboard box	.19	.043	.076	.085
white ceramic tile	.70	.050	.071	.071
glossy grey paper	.29	.083	.082	.082
ivory computer plastic	.45	.043	.13	.13
plastic laminate	.67	.070	.092	.092

We have also measured the reflectance functions of various painted surfaces. We found the "flat" Latex paint we tested to be very nearly diffuse, at least for incident angles up to 60°. Therefore, we present only the results from our measurements of "semi-gloss" and "gloss" Latex. Our ρ_d was around 0.45 for both the semi-gloss and the gloss paints. The value for ρ_s of the semi-gloss Latex was around 0.048 for all surfaces, and the gloss Latex had a slightly higher average of 0.059. Although ρ_d changes dramatically with the color of paint, the value for ρ_s remains fairly constant since it is determined by the index of refraction of the paint base. The values for α_x and α_y are also unaffected by paint color, but since they depend on the exact microstructure of the painted surface, they vary with the application method and the underlying material, as shown in the following two tables.

(α_x, α_y) for Latex Semi-Gloss, $\rho_s \approx 0.048$			
	brushed	rolled	sprayed
metal	(.037, .064)	(.045, .068)	(.041, .055)
sheetrock	(.078, .12)	(.083, .12)	(.096, .11)
wood	(.097, .24)	(.12, .26)	(.099, .26)

(α_x, α_y) for Latex Gloss, $\rho_s \approx 0.059$			
	brushed	rolled	sprayed
metal	(.037, .063)	(.054, .080)	(.038, .054)
sheetrock	(.10, .10)	(.12, .12)	(.10, .10)
wood	(.13, .22)	(.13, .20)	(.12, .17)

7. Conclusion

We have presented an economical new device for measuring BRDFs, and a simple reflectance model that fits a large class of materials. The imaging gonioreflectometer presented here is a working prototype, but improvements are necessary for the measurement of grazing angles and smoother materials. Likewise, the elliptical Gaussian model presented is fast and accurate for many surfaces, but there are still many materials that do not fit our function. In conclusion, although the initial efforts are promising, we hope that this work will stimulate further investigation of empirical shading models. After all, good science requires both theory and data -- one is of little use without the other.

8. Acknowledgements

9. References

[Beckmann63]
Beckmann, Petr, Andre Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces*, Pergamon Press, NY, 1963.

[Blinn77]
Blinn, James F., "Models of Light Reflection for Computer Synthesized Pictures," *Computer Graphics*, Vol. 11, No. 2, July 1977.

[Cabral87]
Cabral, Brian, Nelson Max, Rebecca Springmeyer, "Bidirectional Reflection from Surface Bump Maps," *Computer Graphics*, Vol. 21, No. 4, July 1987.

[Cook82]
Cook, Robert L., Kenneth E. Torrance, "A Reflectance Model for Computer Graphics," *Computer Graphics*, Vol. 15, No. 3, August 1981.

[Cook84]
Cook, Robert L., Thomas Porter, Loren Carpenter, "Distributed Ray Tracing," *Computer Graphics*, Vol. 18, No. 3, July 1984.

[Cook86]
Cook, Robert L., "Stochastic Sampling in Computer Graphics," *ACM Transactions on Graphics*, Vol. 5, No. 1, January 1986.

[He91]
He, X., K.E. Torrance, F.X. Sillion, D.P. Greenberg, "A Comprehensive Physical Model for Light Reflection," *Computer Graphics*, Vol. 25, No. 4, July 1991.

[Kajiya85]
Kajiya, James T., "Anisotropic Reflection Models," *Computer Graphics*, Vol. 19, No. 3, July 1985.

[Kajiya86]
Kajiya, James T., "The Rendering Equation," *Computer Graphics*, Vol. 20, No. 4, August 1986.

[Murray-Coleman90]
Murray-Coleman, J.F., A.M. Smith, "The Automated Measurement of BRDFs and their Application to Luminaire Modeling," *Journal of the Illuminating Engineering Society*, Winter 1990.

[Nicodemus77]
Nicodemus, F.E., J.C. Richmond, J.J. Hsia, *Geometrical Considerations and Nomenclature for Reflectance*, U.S. Department of Commerce, National Bureau of Standards, October 1977.

[Phong75]
Phong, B., "Illumination for Computer Generated Pictures," *Communications of the ACM*, Vol. 18, No. 6, June 1975.

[Poulin90]
Poulin, Pierre, Alain Fournier, "A Model for Anisotropic Reflection," *Computer Graphics*, Vol. 24, No. 4, August 1990.

[Rubenstein81]
Rubenstein, R.Y., *Simulation and the Monte Carlo Method*, J. Wiley, New York, 1981.

[Sandford85]
Sandford, Brian P., David C. Robertson, "Infrared Reflectance Properties of Aircraft Paints," *Proceedings IRIS Targets, Backgrounds, and Discrimination*, 1985.

[Sillion91]
Sillion, Francois, James Arvo, Donald Greenberg, "A Global Illumination Solution for General Reflectance Distributions," *Computer Graphics*, Vol. 25, No. 4, July 1991.

[Torrance67]
Torrance, K.E., E.M. Sparrow, "Theory for Off-Specular Reflection from Roughened Surfaces," *Journal of the Optical Society of America*, Vol. 57, No. 9, September 1967.

[Whitted80]
Whitted, Turner, "An Improved Illumination Model for Shaded Display," *Communications of the ACM*, Vol. 23, No. 6, June 1980, pp. 343-349.

[Yokoi88]
Yokoi, Shigeki, Jun-ichiro Toriwaki, "Realistic Expression of Solids with Feeling of Materials," *JARECT*, Vol. 18, 1988.

Predicting Reflectance Functions from Complex Surfaces

Stephen H. Westin
James R. Arvo
Kenneth E. Torrance

Program of Computer Graphics
Cornell University
Ithaca, New York 14853

Abstract

We describe a physically-based Monte Carlo technique for approximating bidirectional reflectance distribution functions (BRDFs) for a large class of geometries by directly simulating optical scattering. The technique is more general than previous analytical models: it removes most restrictions on surface microgeometry. Three main points are described: a new representation of the BRDF, a Monte Carlo technique to estimate the coefficients of the representation, and the means of creating a milliscale BRDF from microscale scattering events. These allow the prediction of scattering from essentially arbitrary roughness geometries. The BRDF is concisely represented by a matrix of spherical harmonic coefficients; the matrix is directly estimated from a geometric optics simulation, enforcing exact reciprocity. The method applies to roughness scales that are large with respect to the wavelength of light and small with respect to the spatial density at which the BRDF is sampled across the surface; examples include brushed metal and textiles. The method is validated by comparing with an existing scattering model and sample images are generated with a physically-based global illumination algorithm.

CR Categories and Subject Descriptors: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism.
Additional Key Words: spherical harmonics, Monte Carlo, anisotropic reflection, BRDF

1 Introduction

Since the earliest days of computer graphics, experimenters have recognized that the realism of an image is limited by the sophistication of the model of local light scattering [3, 12]. Non-physically-based local lighting models, such as that of Phong [12], although computationally simple, exclude many important physical effects and lack the energy consistency needed for global illumination calculations. Physically-based models [2, 5, 15] reproduce many effects better, but cannot
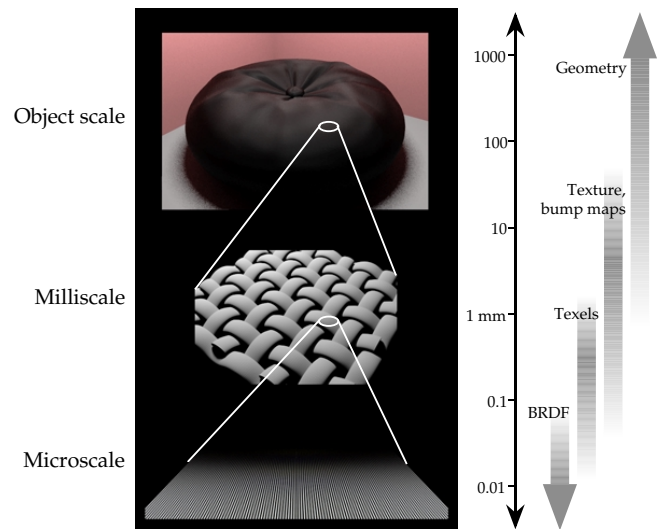
Figure 1: Applicability of Techniques

model many surfaces, such as those with anisotropic roughness. Models that deal with anisotropic surfaces [8, 11] fail to assure physical consistency.

This paper presents a new method of creating local scattering models. The method has three main components: a concise, general representation of the BRDF, a technique to estimate the coefficients of the representation, and a means of using scattering at one scale to create a BRDF for a larger scale. The representation used makes it easy to enforce the basic physical property of scattering reciprocity, and its approximation does not require discretizing scattering directions as in the work of Kajiya [8] and Cabral et al. [1].

The method can predict scattering from any geometry that can be ray-traced: polygons, spheres, parametric patches, and even volume densities. Previous numerical techniques were limited to height fields, and analytical methods have been developed only for specific classes of surface geometry. The new method accurately models both isotropic and anisotropic surfaces such as brushed metals, velvet, and woven textiles.

Figure 1 shows several representations used in realistic rendering, along with approximate scale ranges where each is applicable. At the smallest scale (size $\ll$ 1 mm), which we call *microscale*, the BRDF accurately captures the appearance of a

surface. As individual surface features become larger than one pixel, texture maps, bump maps, and texels can be used to show surface features. At the largest scale, *object scale*, the geometry must be modeled explicitly, for example with polygons or parametric patches.

The applicability of each representation ultimately depends on the context: the upper limit of applicable scale is determined by the frequency of sampling across a surface, and the lower limit is determined by the integration area for each sample; this is often the surface area represented by a pixel. When rendering, say, an interior scene, objects as small as a pencil must be modeled at object scale; when simulating the view from orbit, however, objects as large as trees and buildings can be modeled within the BRDF, so we can think of them as microscale geometry, or *microgeometry*. The advent of global illumination methods (e.g. [6, 18]) has created another concept of scale: these methods generally use a coarser characterization of scattering for indirect illumination, but demand careful attention to energy consistency and physical accuracy.

The method of this paper is applicable wherever the BRDF is an adequate model of surface geometry. It uses an analytical BRDF model for scattering at one scale of roughness, the *microscale*, simulating geometric optical scattering at a larger scale, the *milliscale*. Milliscale scattering embodies large-scale roughness effects (roughness size $\gg$ wavelength of light, λ), and any smooth surface effects (roughness size $\approx \lambda$ or $< \lambda$) are modeled by the microscale BRDF, which can include wave optics effects.

The next three sections present the heart of the technique: the BRDF representation, the Monte Carlo estimator, and the means of estimating a milliscale BRDF from the microscale description of surface roughness.

2 Wheels Within Wheels: Representing the BRDF with Spherical Harmonics

A general scattering function for unpolarized light is a function of four variables, $\rho_{bd}(\theta_i, \phi_i, \theta_r, \phi_r) : S^2 \times S^2 \mapsto \mathcal{R}$, where S^2 is the unit sphere, θ_i, ϕ_i are the elevation and azimuth angles of incidence, and θ_r, ϕ_r are the corresponding angles of reflection (Figure 2). For a BRDF, ρ_{bd} is zero whenever θ_i or $\theta_r > \frac{\pi}{2}$. The BRDF can take on highly arbitrary shapes [5, 16], so a very general method is needed to represent it. Fortunately, a BRDF
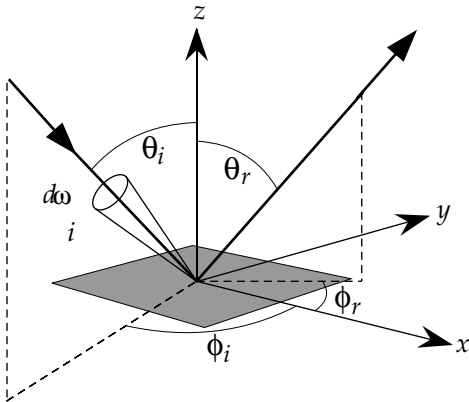
Figure 2: Scattering Angles

is generally smooth, making it a good candidate for representation by smooth orthogonal functions. Previous authors have used spherical harmonics to represent scattering functions [1, 9, 13], since they form a complete basis set of smooth functions over the sphere. Kajiya [9] used spherical harmonics to derive an analytical scattering function; Cabral et al. [1] and Sillion et al. [13] used them as a numerical approximation to the BRDF. The representation used in this paper is an extension of Sillion's technique; it provides an accurate, concise embodiment of the general BRDF.

2.1 Overview of Spherical Harmonics

Any square-integrable function over the sphere can be exactly represented by an infinite sum of spherical harmonic basis functions, $Y_{lm}(\theta, \phi)$, of varying order, *l*, and degree, *m*:

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} C_{lm} Y_{lm}(\theta, \phi). \tag{1}$$

As with a Fourier representation, we can approximate *f* by truncating the series to a finite number of terms. For convenience, we organize this finite collection of basis functions into a vector by the convention of encoding both order and degree with a single subscript. Thus

$$f(\theta, \phi) \approx \sum_{k=0}^{n} C_k Y_k(\theta, \phi) = \mathbf{C} \cdot \mathbf{Y}(\theta, \phi). \tag{2}$$

Each coefficient C_k is defined by the inner product of $f(\theta, \phi)$ with the corresponding spherical harmonic basis function:

$$
\begin{aligned}
C_k &= \int_0^{2\pi} \int_0^{\pi} f(\theta, \phi) Y_k(\theta, \phi) \sin\theta \, d\theta \, d\phi \\
&= \langle Y_k | f \rangle.
\end{aligned}
\tag{3}
$$

This follows directly from the orthogonality of the basis functions [17].

2.2 Representing the BRDF

If we fix the incident direction (θ_i, ϕ_i), the BRDF is a function of two variables, (θ_r, ϕ_r), and the representation in Equation 2 suffices. To account for variation of the BRDF with incident direction, the coefficient vector $\mathbf{C}$ in Equation 2 can be thought of as a function of the incident direction. If a surface has isotropic roughness, as assumed in [1] and [13], the scattering function ρ_{bd} is independent of rotation about the surface normal. In this case,

$$\rho_{bd}(\theta_i, \phi_i, \theta_r, \phi_r) = \rho_{bd}(\theta_i, 0, \theta_r, \phi_r - \phi_i). \tag{4}$$

Each coefficient C_k is a function of θ_i alone, which can be calculated for a number of selected values of θ_i and interpolated for all θ_i [1, 13]. In general, however, a BRDF is a function of ϕ_i as well as of $(\theta_i, \theta_r, \phi_r)$, so a richer representation is needed.

2.3 Extension to Anisotropic Surfaces

For an anisotropic surface both θ_i and ϕ_i must be considered, and none of the previous representations suffices [1, 13]. Each coefficient C_k in Equation 2 is thus a function of θ_i and ϕ_i:

$$\rho_{bd}(\theta_i, \phi_i, \theta_r, \phi_r) \approx \sum_{k=0}^{n} C_k(\theta_i, \phi_i) Y_k(\theta_r, \phi_r). \tag{5}$$

Each coefficient function, $C_k(\theta_i, \phi_i)$, is defined by the inner product of $\rho_{bd}(\theta_i, \phi_i, \cdot, \cdot)$ with the corresponding spherical harmonic basis function:

$$C_k(\theta_i, \phi_i) = \langle \rho_{bd} \mid Y_k \rangle_{refl} \tag{6}$$

where the subscript "*refl*" denotes integration over the reflected hemisphere. Reciprocity makes the dependence of ρ_{bd} on (θ_i, ϕ_i) exactly like its dependence on (θ_r, ϕ_r). Since spherical harmonics concisely represent the latter dependence, we also use them to represent the dependence on (θ_i, ϕ_i), expressing each coefficient function in terms of spherical harmonics. Each element of our vector $\mathbf{C}$ of coefficients is now represented in turn by a vector of coefficients, giving us a matrix $\mathbf{M}$ to represent the BRDF. Each element of the matrix $\mathbf{M}$ is given by

$$m_{jk} = \left\langle Y_j \mid \langle \rho_{bd} \mid Y_k \rangle_{refl} \right\rangle_{in} \tag{7}$$

where the subscripts "*in*" and "*refl*" denote integration over the incident and reflected hemispheres, respectively. Evaluation of the BRDF becomes

$$\begin{aligned} \rho_{bd}(\theta_i, \phi_i, \theta_r, \phi_r) &\approx \sum_{j=0}^{N} \sum_{k=0}^{N} Y_j(\theta_i, \phi_i) m_{jk} Y_k(\theta_r, \phi_r) \\ &= \mathbf{Y}^T(\theta_i, \phi_i) \mathbf{M} \mathbf{Y}(\theta_r, \phi_r), \end{aligned} \tag{8}$$

where $\mathbf{Y}(\theta, \phi)$ is the column vector of basis functions evaluated at (θ, ϕ).

2.4 Reciprocity

An important physical constraint on the BRDF is *reciprocity*, which states that

$$\rho_{bd}(\theta_i, \phi_i, \theta_r, \phi_r) = \rho_{bd}(\theta_r, \phi_r, \theta_i, \phi_i) \tag{9}$$

for all angles of incidence and reflection [14]. If the matrix $\mathbf{M}$ is symmetric, then

$$\mathbf{Y}^T(\theta_i, \phi_i) \mathbf{M} \mathbf{Y}(\theta_r, \phi_r) = \mathbf{Y}^T(\theta_r, \phi_r) \mathbf{M} \mathbf{Y}(\theta_i, \phi_i) \tag{10}$$

and the approximation in Equation 8 satisfies Equation 9. By assuring that we compute a symmetric matrix $\mathbf{M}$, we can enforce exact reciprocity; previous approaches [1, 8, 11, 13] afforded, at best, approximate reciprocity.

2.5 Storage Reduction and Filtering

The matrix $\mathbf{M}$ can be quite large; tens of thousands of elements are typical. Since our BRDF representation, like that of [13], is based on spherical harmonics, we can adapt two techniques from that work to reduce the number of coefficients

(and corresponding basis functions) needed: the first technique causes half the coefficients to vanish, and the second reduces the high-frequency content of the BRDF, reducing the number of coefficients needed to achieve an acceptably accurate approximation. Since we deal only with scattering to one hemisphere, we can complete the other hemisphere with an arbitrary function. We chose a function that reduces the size of the representation: $\rho_{bd}(\theta_i, \phi_i, \pi - \theta_r, \phi_r) = -\rho_{bd}(\theta_i, \phi_i, \theta_r, \phi_r)$; this causes half of the coefficients (those with $l + m$ even in the real form of spherical harmonics) to be zero; they can be omitted from the representation, reducing the matrix size by $\frac{3}{4}$. To economize further, we represent $\rho_{bd} \cos \theta_i \cos \theta_r$ instead of ρ_{bd}; multiplication by $\cos \theta_i$, together with the completion described above, forces C^1 continuity at the equator and drastically reduces ringing. To maintain symmetry of the matrix **M**, we also multiply by $\cos \theta_r$. Representing $\rho_{bd} \cos \theta_i \cos \theta_r$ assures that Equation 9 is still satisfied. We omit this implementation detail from the following discussion.

As with a Fourier representation of a function, simply truncating all coefficients with index $l > l_{max}$ will cause ringing in the approximation, called the Gibbs phenomenon. To reduce this, we attenuate higher frequencies, as did Cabral et al. [1], by progressively reducing the magnitude of coefficients with $l_{filter} < l \le l_{max}$, where l_{filter} is an empirically-determined threshold. The magnitude is reduced according to a half-Gaussian with empirically-determined width.

3 Monte Carlo Estimation of the Coefficient Matrix

If we bombard a specimen with incident rays from an arbitrary direction $U = (\theta_i, \phi_i)$, the BRDF can be expressed as

$$\rho_{bd}(U, V) = \frac{p(U, V) \langle R(U, V) \rangle}{\cos \theta_r} \qquad (11)$$

where a ray from direction U will scatter into $V = (\theta_r, \phi_r)$ with a probability density $p(U, V)$, and $\langle R(U, V) \rangle$ is the mean attenuation of all rays incident from direction U and scattered in direction V.

In order to obtain a spherical harmonics coefficient, we must integrate the product $\rho_{bd} Y_k$ over the hemisphere.

$$\begin{aligned} C_k(U) &= \int_{S^2} \rho_{bd}(U, V) Y_k(V) dV \\ &= \int_{S^2} \langle g_k(U, V) \rangle \, p(U, V) dV \end{aligned} \qquad (12)$$

where

$$g_k(U, V) = \frac{R(U, V)}{\cos \theta_r} Y_k(V). \qquad (13)$$

Unfortunately we have no analytical expression for p or R; we can, however, use a Monte Carlo simulation to estimate the integral in Equation 12. The integral can be interpreted as the expected value of $g_k(U, V)$, where V is a random variable with probability density function $p(U, V)$. If we define

$$G_k(U) = \frac{1}{N} \sum_{n=1}^{N} g_k(U, V_n) \qquad (14)$$

where V_n are random samples distributed according to p, then the expected value of G_k is $C_k(U)$; G_k is said to be an *estimator*

of the integral [10]. The rays departing from the specimen in direction V will have mean attenuation $\langle R(U, V) \rangle$; this attenuation must be multiplied by $Y_k(V)/\cos \theta_r$ to give the expected value g for the estimator.

This leaves another integration, that with respect to U:

$$m_{jk} = \int_{S^2} C_k(U) Y_j(U) dU. \qquad (15)$$

This integration can also be handled via Monte Carlo, this time as *quadrature*, a discrete approximation to an integral. This is handled similarly, with the estimator

$$\widetilde{m}_{jk} = \frac{1}{N} \sum_{n=1}^{N} C_j(U_n) Y_j(U_n) \qquad (16)$$

where the U_n are uniformly distributed over the incident hemisphere. These two sampling processes, each approximating an integral in two dimensions, can be combined into one process to approximate the four-dimensional integral desired.

$$\widetilde{m}_{jk} = \frac{1}{N} \sum_{n=1}^{N} g_k(U_n, V_n) Y_j(U_n) \qquad (17)$$

where the U_n are distributed uniformly and the V_n are distributed according to p. $\widetilde{m}_{jk}$ is an unbiased estimator of m_{jk}.

The simulation yields $\widetilde{M}$, an approximation to the symmetric matrix **M**, and does not guarantee symmetry, so reciprocity of the BRDF is not guaranteed. We average the upper triangle and the lower triangle of $\widetilde{M}$ to obtain a symmetric matrix $\frac{1}{2}(\widetilde{M} + \widetilde{M}^T)$ which is used to compute ρ_{bd}. The two triangles are independent unbiased estimates of the BRDF; by averaging them to obtain a symmetric matrix, we also reduce the variance of our estimate of **M**.

4 From Microscale to Milliscale

The BRDF can be used to model features ranging from microscale to milliscale for visible light, as shown in Figure 1. This section explains how to use microscale scattering events to calculate a milliscale BRDF. The section starts with basic BRDF definitions, describes the individual microscale scattering events, then explains how individual Monte Carlo events are incorporated into the milliscale model to obtain an aggregate BRDF.

At the microscale, arbitrary reflection models may be employed, including ideal specular, ideal diffuse, and directional diffuse models. One illustrative case is where the microgeometry is composed of planar ideal specular surfaces; this is equivalent to geometric optics models based on microfacets, such as the Torrance-Sparrow model [15].

We use ray tracing to model scattering events, as suggested by Cabral et al. [1]. The ray tracer must be carefully designed to assure physically accurate results. Each ray has a certain amount of energy associated with it; microscale reflection will attenuate this energy and perhaps divide it among multiple rays at each bounce. All calculations involve energy flux density until a ray finally exits the model; then the energy is converted to radiance, the proper quantity for the BRDF, by dividing by $\cos \theta_r$. The radiance distribution is averaged over the specimen surface to create a milliscale BRDF.
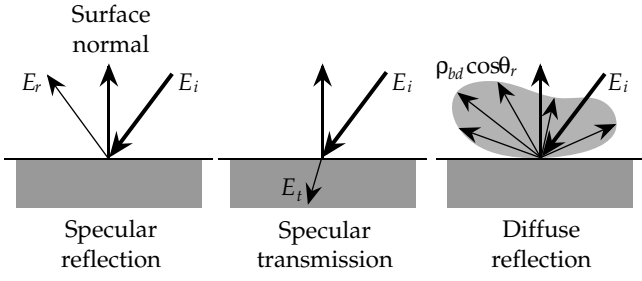
Figure 3: Local Scattering Modes

4.1 Incident Energy and the BRDF

We are estimating the BRDF ρ_{bd}, which is expressed at a given wavelength as

$$\rho_{bd}(\theta_r, \phi_r, \theta_i, \phi_i) = \frac{dI_r(\theta_r, \phi_r)}{dE_i(\theta_i, \phi_i)} \quad (18)$$

where dI_r is the reflected radiance and dE_i is the incident energy flux density, the incident energy per unit time per unit area. This equation holds at both micro- and milliscales. It becomes simpler to evaluate if we hold the denominator (incident energy flux density) constant and vary the incident angles θ_i, ϕ_i. Then

$$\rho_{bd}(\theta_r, \phi_r, \theta_i, \phi_i) = \frac{dI_r(\theta_r, \phi_r)}{dE_i} \quad (19)$$

where dE_i is the (constant) incident energy flux density.

Incident radiance I_i is defined as the incident energy flux density per unit projected area per unit solid angle

$$I_i = \frac{dE_i}{\cos\theta_i d\omega_i}. \quad (20)$$

Thus

$$dE_i(\theta_i, \phi_i) = I_i(\theta_i, \phi_i) \cos\theta_i d\omega_i. \quad (21)$$

The factor $\cos\theta_i$ converts receiving area to projected area, accounting for the dependence of projected surface area on θ_i.

The method allows different local scattering modes, three of which are shown in Figure 3. The next three sections describe how these modes are modeled.

4.2 Specular Reflection

The BRDF at the microscale may contain an ideal specular component $\hat{\rho}_s$. Whenever a ray hits such a microfacet, we model the transfer by spawning a ray in the specular direction as in classical ray-tracing [19]. The energy flux density of this ray is determined by the equation

$$dE_r = \hat{\rho}_s(\theta_i)dE_i \quad (22)$$

where dE_i is the flux density of the incident ray, θ_i is the incident elevation with respect to the local facet, and $\hat{\rho}_s$ is the microscale specular reflection coefficient for the facet.

4.3 Specular Transmission

The method may be used to model microgeometries that include transparent materials. Whenever a ray encounters a smooth interface between media of different refractive indices, we must calculate the energy transfer through the interface. Neither energy flux density nor radiance is preserved at the interface [4], since solid angles are altered, but the distribution of transmitted rays accounts for this. We also must model any attenuation of the ray as it passes through a transparent medium; for a uniform medium, the ray is attenuated by $e^{-\kappa s}$ where s is the path distance and κ is an extinction coefficient determined by the material.

4.4 Directional-Diffuse Reflection

The most complex transfer takes place when a ray strikes a facet that shows directional-diffuse scattering. When a ray hits such a facet, we send out n rays to the hemisphere above the facet and weight them according to $\hat{\rho}_{bd}$; this serves as a discrete approximation of scattering according to the ideal-diffuse and directional-diffuse parts of the BRDF. The total energy transfer is determined by

$$
\begin{aligned}
dE_r(\theta_r, \phi_r) &= dI_r \, d\omega_r \cos\theta_r \\
&= dE_i \, \hat{\rho}_{bd}(\theta_i, \phi_i, \theta_r, \phi_r) \, d\omega_r \cos\theta_r \quad (23)
\end{aligned}
$$

where dE_r is the reflected energy flux density in a particular direction, $\hat{\rho}_{bd}$ is the diffuse part (including directional-diffuse) of the microscale BRDF, and $d\omega_r$ is the solid angle of reflection. The angles (θ_r, ϕ_r) give the reflection direction with respect to the local facet. We multiply by $d\omega_r \cos\theta_r$ to convert the radiance given by $\hat{\rho}_{bd}$ to energy flux density for the next scattering event.

In our implementation reflected rays are cast randomly into the hemisphere above the local (microscale) surface; they are distributed uniformly over this hemisphere, so each ray represents a solid angle of

$$d\omega_r = \frac{2\pi}{n} \quad (24)$$

where 2π is the total solid angle of the hemisphere and n is the number of reflected rays shot.

4.5 Integrating Over Milligeometry

We have described the possible microscale events of a single ray striking a point on the surface, but we must integrate over the specimen to obtain the aggregate BRDF. Just as the Monte Carlo integration was extended in Section 3 to accomodate the two dimensions of the incident hemisphere, it can be extended further to integrate over a two-dimensional specimen surface. We can keep the incident energy flux density constant by keeping both the total incident flux and the receiving surface area constant. We do this by shooting a constant number of rays (energy flux) and by distributing them over a constant surface area. The simplest way to do this is to select a fixed region of the surface, as shown in Figure 4, and to distribute the samples uniformly over this region at each incident angle. The direction of each ray is determined by the incident angles (θ_i, ϕ_i) with respect to the mean surface; its origin will be calculated so that the ray will strike the notional plane of the surface, shown in Figure 4 in red, at the chosen sample point.

Figure 4: Target Area

The surface region chosen should be

- large with respect to the lateral geometric features of the surface, to assure a good statistical average of large-scale scattering;

- large with respect to the *vertical thickness* of the surface geometry; and

- a subset of the total surface geometry, since geometry outside the nominal surface region will be important at high incident angles.

When a ray leaves the specimen area, we update the approximate matrix $\widetilde{\mathbf{M}}$ by adding $\mathbf{Y}(U)\mathbf{Y}^T(V)R/\cos\theta_r$. This matrix represents the BRDF ρ_{bd}. We integrate over the portion of the surface that is visible from the reflection direction (ϕ_r, θ_r), projected onto the mean surface.

4.6 Efficiency Considerations

We can reduce the computation needed to maintain the matrix $\widetilde{\mathbf{M}}$ by holding the incident direction U constant for several reflected directions V, updating the matrix only once for each distinct U. This happens automatically when several randomly distributed rays are spawned at each intersection, as in directional-diffuse scattering. In addition, we choose several target points on the surface for each U, further amortizing the cost of updating the matrix. Updating the matrix then becomes a triple sum

$$\widetilde{\mathbf{M}} = \frac{1}{N_i N_p N_b} \sum_{n=1}^{N_i} \mathbf{Y}(U_n) \left\{ \sum_{m=1}^{N_p} \sum_{l=1}^{N_b} \mathbf{Y}^T(V_{ml}) \frac{R_{nml}}{\cos\theta_r} \right\} \quad (25)$$

where R_{nml} is the attenuation of a ray from incident direction U_n reflected in direction V_l from target point P_m on the surface. N_i is the number of incident directions used, N_p is the number of sample positions across the specimen for each incident direction, and N_b is the number of exit rays resulting from a single incident ray. This approach reduces the number of evaluations of the spherical harmonics basis functions; for $N_i N_p N_b$ samples to update the matrix, $\mathbf{Y}(U)$ is evaluated only N_i times, while $\mathbf{Y}^T(V)$ is evaluated $N_i N_p N_b$ times. The greatest savings, however, comes in matrix adds; we need only perform N_i matrix additions; the other updates simply add vectors and require far less computation.

4.7 Convergence Measure

Since the exact matrix $\mathbf{M}$ is symmetric, we can use the asymmetry of our estimate as a measure of convergence in approximating the true BRDF. We calculate the error Q as

$$Q = \left\| \widetilde{\mathbf{M}} - \widetilde{\mathbf{M}}^T \right\| \quad (26)$$

where

$$\|\mathbf{A}\| = \frac{1}{N^2} \sum_{j=1}^{J} \sum_{k=1}^{J} \left| \mathbf{A}_{ij} \right| \quad (27)$$

where J is the size of the matrix $\mathbf{A}$. This is perhaps not as informative as a direct estimate of the variance of each coefficient, but is much cheaper to compute and tends to decline as $\frac{1}{\sqrt{N}}$, which suggests that it is directly proportional to the variance.

5 Results

We now show several applications of the technique. We obtain BRDF's for surfaces textured at milliscale. At the microscale, the BRDF can be ideal specular, ideal diffuse, or an analytical BRDF that includes wave optics effects. The technique can also be used recursively by using the results of one simulation as the microscale BRDF in another simulation.

All images shown in this section were generated by Monte Carlo ray tracing; the grainy texture of the images is caused by the Monte Carlo integration used to compute global illumination. Other global illumination and rendering techniques might have been used, such as that of Sillion et al. [13].

We first consider a flat Gaussian-rough surface for which, at the microscale, the surface is an ideal specular reflector. We can compare the results of the new method with the results of an existing analytical model for such a surface [5], thus giving some verification of the new technique.

5.1 Initial Verification: An Isotropic Surface

Wave optics effects were not included, except for the Fresnel coefficient for each microfacet. Reflection is governed by geometric optics; shadowing and masking effects of the surface are included because of the occlusion calculations in the ray tracer.

Gaussian height fields were generated by FFT filtering of white noise, and the resulting points were connected by triangles, each of which was modeled as a mirror. To integrate over a specimen large compared to the roughness height, an area of 8×8 millimeters was used. To assure adequate representation of the surface, a total of 524,288 polygons was used. The model was created in four sections of 131,072 polygons, each generated with a different random number seed, to represent a square patch of surface 4mm wide. The roughness length parameters of the surface were $\sigma = 10\mu m$ vertically and $\tau = 65\mu m$ horizontally (Figure 5). The specimen patch actually used was 3.13mm wide in the center of the geometric model; this assured that all incident rays would intersect the "sides" of the patch at least 2σ away from the notional plane.

Incident ray angles were restricted to 88° to keep the effective roughness greater than 460 nm, the shortest wavelength employed. This keeps behavior in the regime where geometric optics is valid; were the wavelength to approach the size
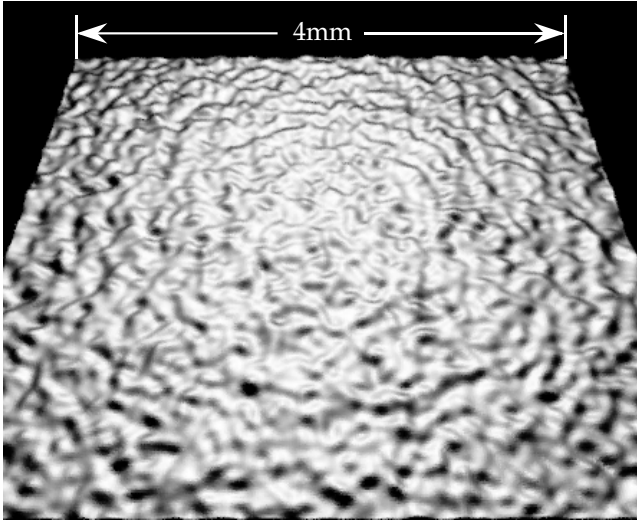
Figure 5: Gaussian Surface

Figure 6: Comparison with Previous Model

of surface features, wave-related effects would begin to affect the scattering. Results are plotted as solid lines in Figure 6 for incident angles $\theta_i = 0°, 30°, 45°, 60°, 75°$. Dashed lines show results from the model of He [5], which assumes a Gaussian rough surface and allows for wave optics effects. The He model is shown in the limit of large surface roughness, $\sigma \gg \lambda$, where wave optics effects should be negligible. The simulation agrees quite well with the analytical model for reflection angles less than about 80°; the divergence at greater angles is disturbing, but not very significant in terms of energy values. Recall that the BRDF ρ_{bd} gives a radiance value dI_r; the energy dE_r scattered in any reflected direction (θ_r, ϕ_r) is proportional to $dI_r \cos\theta_r$, reducing the effect of the error at high angles of reflection. We believe that the error results because we approximate $\rho_{bd} \cos\theta_r$. If we assume that error in approximating this function is roughly constant over the hemisphere, dividing by $\cos\theta_r$ to recover ρ_{bd} will magnify the error near the horizon (i.e. as $\theta_r \rightarrow \frac{\pi}{2}$).

5.2   Simple Anisotropy

We can use the method to create an anisotropic milliscale BRDF by using an isotropic analytical microscale BRDF model; we rely on He's analytical model for microgeometric effects, and use the new technique to model larger-scale anisotropy. Figure 7 shows, at the top, a model of parallel cylinders of slightly rough aluminum. In the left side of the figure, the cylinders are oriented with axes perpendicular to the screen; in the right side the axes are parallel to the screen. The bottom half of the figure shows a similar scene, but with two flat plates replacing the arrays of cylinders. Both plates use a BRDF generated from parallel cylinders like those in the top half of the figure. In the left half of the figure, the axis of anisotropy was oriented perpendicular to the screen; in the right half, it is oriented left-right.

The scattering patterns are similar; when viewed from a distance, the images look the same. The microscale BRDF is important for generating the upper images; the milliscale BRDF is used for the lower figures. Note how the surface orientation affects the appearance, revealing the anisotropic behavior of the reflected light. This is further illustrated in
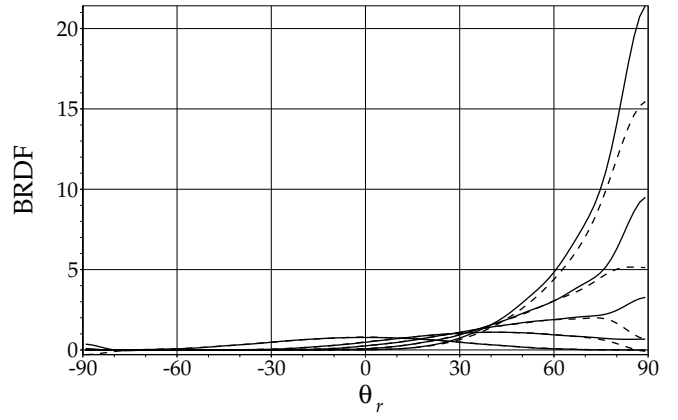
Figure 8, where the same object is rendered with two BRDFs for brushed aluminum, one isotropic and one anisotropic.

Figures 9 and 10 show an aluminum automotive wheel and an aluminum teapot created using this anisotropic BRDF. The polishing scratches were oriented as from rotation, about the vertical axis of the teapot and about the hub of the wheel. The energy-consistency of the BRDF, not guaranteed by previous approaches, allows an accurate global illumination solution.
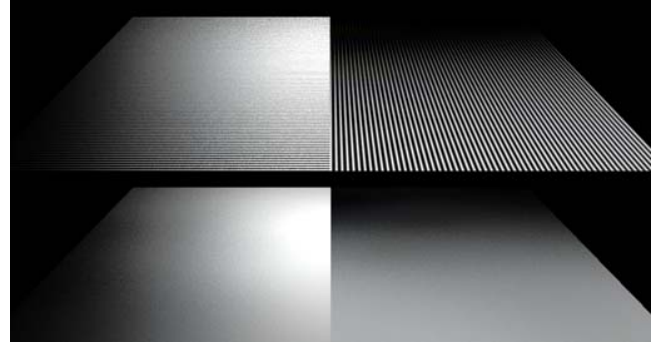
Figure 7: Parallel Cylinder Model of Anisotropic Surface
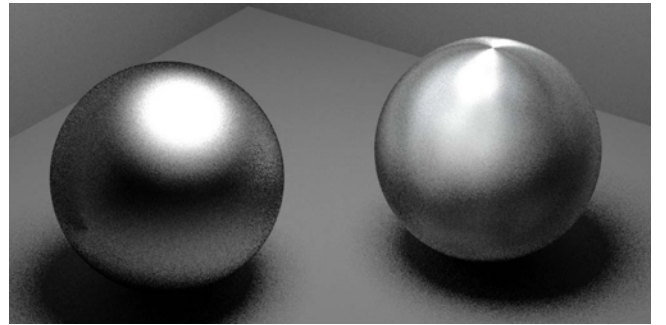
Figure 8: Isotropic and Anisotropic Aluminum

Figure 10: Anisotropic Aluminum Teapot

Figure 12: Velvet Doughnut
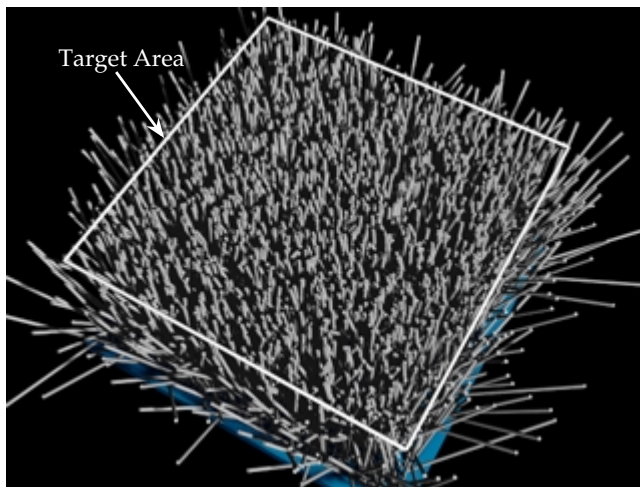
Figure 11: Microscale Geometry for Velvet
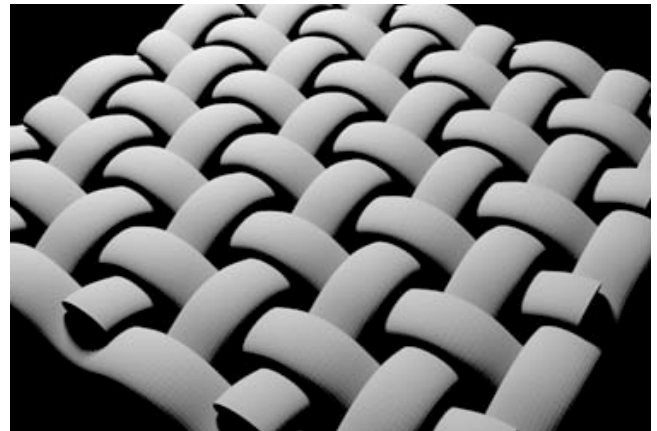
Figure 13: Microscale Structure of Cloth Model

5.3 Velvet

A more complex microgeometry is that of velvet: it consists of many roughly parallel specular fibers extending from a fabric base. This was modeled as a forest of narrow cylinders, with the angle of each cylinder perturbed randomly (Figure 11). The target area for incident rays is shown at the top of the fibers. The fibers are shown as ideal diffuse for clarity; in the BRDF simulation, the fibers were transparent ideal-specular plastic. Whenever a ray intersected a fiber, it was either reflected (with probability equal to the Fresnel reflectivity) or transmitted; when it intersected the base plane, it was absorbed. Figure 12 shows an image made using the resulting BRDF.

5.4 Woven Cloth

The method can also be used recursively to model several scales of roughness; this is demonstrated by modelling woven cloth as shown in Figure 1. At the milliscale, the
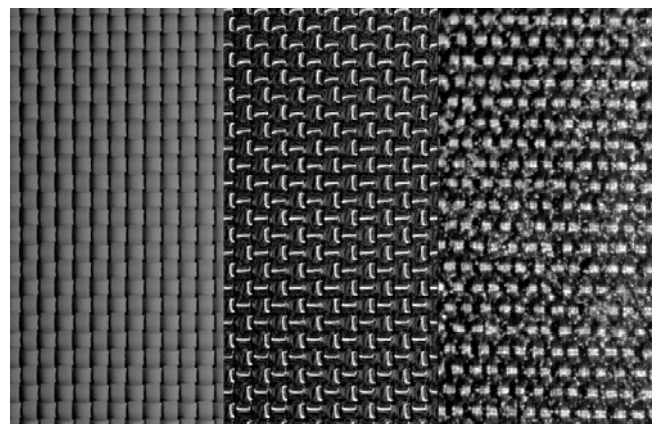
Figure 14: Cloth Microscale Geometry and Real Cloth

Figure 9: Anisotropic Aluminum Wheel
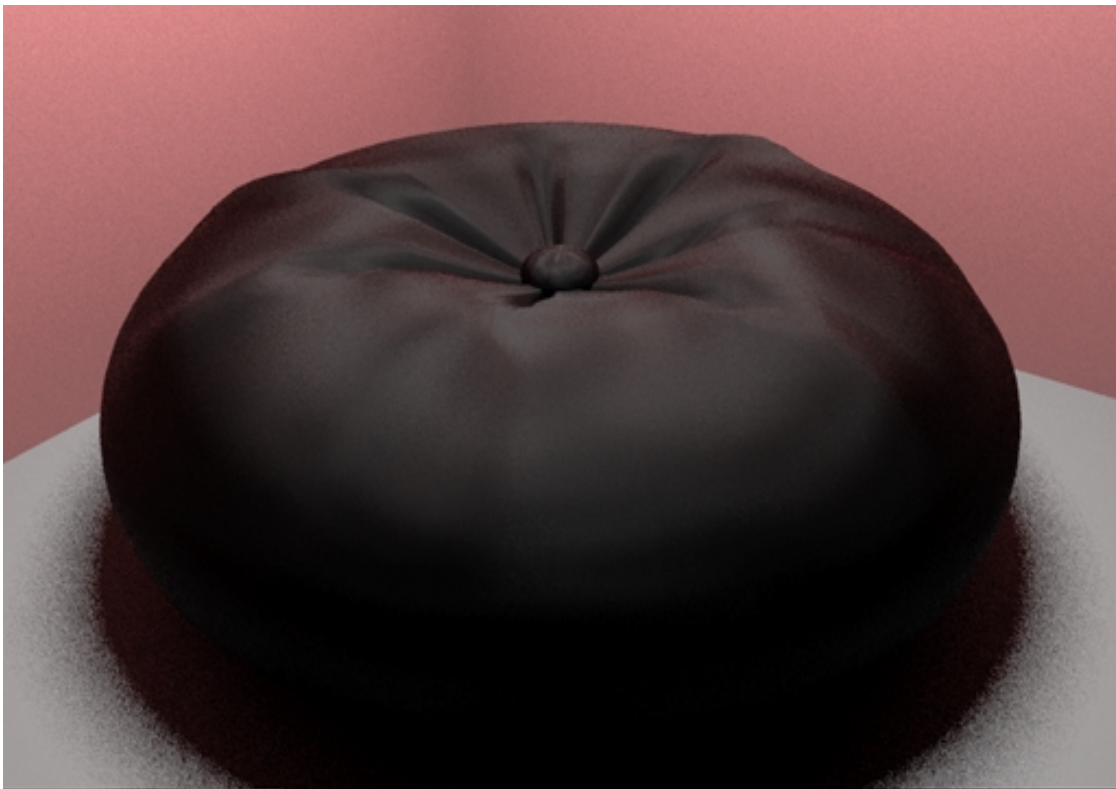
Figure 15: Nylon Cushion

weave pattern of the cloth was modeled as shown in Figure 13, and an anisotropic BRDF was used to model the scattering from individual fibers in the threads. The scattering from the surface of each thread (microgeometry) was modeled by the same geometry used in Section 5.2, but using a Fresnel reflectance function to simulate black synthetic fibers. Figure 14 has three parts: on the left, the cloth microgeometry is shown with an ideal-diffuse BRDF; in the center, it is shown with the thread BRDF, and on the right is a magnified photograph of actual cloth. Figure 15 shows a cushion upholstered in black nylon, rendered using the BRDF obtained from this process.

6   Conclusion

Three main points are described in this paper: a new representation of the BRDF, a Monte Carlo technique to estimate the coefficients of the representation, and the means of creating a milliscale BRDF from microscale scattering events. These allow the prediction of scattering for essentially arbitrary geometries. BRDFs for complex surfaces can be simulated hierarchically by using the result of one simulation in generating the BRDF for the next larger scale.

The new representation is concise and well-suited for use in rendering and global illumination calculations. The technique of [13] can be easily extended to accommodate the new representation. Its ease of evaluation suits it for other global illumination methods such as stochastic ray tracing [7, 18] as well.

The Monte Carlo integration used here enables us to model the scattering of many surfaces which have hitherto been impossible to model in computer graphics, producing accurate models for anisotropic surfaces and surfaces with transparent elements.

7   Acknowledgements

References

[1]   Cabral, B., Max, N., and Springmeyer, R., Bidirectional reflection functions from surface bump maps. In *Proceedings of SIGGRAPH '87*(July 27-31, 1987, Anaheim, California), *Computer Graphics* 21, 4 (July 1990), 273-281.

[2]   Cook, R. L, and Torrance, K. E. A reflectance model for computer graphics, *ACM Transactions on Graphics* 1, 1 (January 1982), 7-24.

[3]   Gouraud, H. Continuous shading for curved surfaces. *IEEE Transactions on Computers* 20, 6 (June 1971), 623-628.

[4]   Hecht, E., and Zajac, A. *Optics*. Addison-Wesley, 1974.

[5]   He, X. D., Torrance, K. E., Sillion, F. X., and Greenberg, D. P. A comprehensive physical model for light reflection. In *Proceedings of SIGGRAPH '91*(July 28-August 2, 1991, Las Vegas, Nevada), *Computer Graphics* 25, 4 (July 1991), 175-186.

[6]   Goral, C. M., Torrance, K. E., Greenberg, D. P., and Battaile, B. Modeling the interaction of light between diffuse surfaces. In *Proceedings of SIGGRAPH '84*(July 23-27, 1984, Minneapolis, Minnesota), *Computer Graphics* 18, 3 (July 1984), 213-222.

[7]   Kajiya, J. The rendering equation. In *Proceedings of SIGGRAPH '86*(August 18-22, 1986, Dallas, Texas), *Computer Graphics* 20, 4 (August 1986), 143-150.

[8]   Kajiya, J. Anisotropic reflectance models. In *Proceedings of SIGGRAPH '85*(July 22-26, 1985, San Francisco, California), *Computer Graphics* 19, 4 (July 1985), 15-21.

[9]   Kajiya, J., and Von Herzen, B. Ray tracing volume densities. In *Proceedings of SIGGRAPH '84*(July 23-27, 1984, Minneapolis, Minnesota), *Computer Graphics* 18, 3 (July 1984), 165-174.

[10]   Kalos, M, H., and Whitlock, P. A. *Monte Carlo Methods*. John Wiley & Sons, 1986.

[11]   Poulin, P., and Fournier, A. A model for anisotropic reflection. In *Proceedings of SIGGRAPH '90* (August 6-10, 1990, Dallas, Texas), *Computer Graphics* 24, 4 (August 1990), 273-282.

[12]   Phong, B-T. Illumination for computer generated pictures., *Communications of the ACM* 18, 6 (June 1975), 311-317.

[13]   Sillion, F. X., Arvo, J., Westin, S. H., and Greenberg, D. P. A global illumination solution for general reflectance distributions. In *Proceedings of SIGGRAPH '91* (July 28-August 2, 1991, Las Vegas, Nevada), *Computer Graphics* 25, 4 (July 1991), 187-196.

[14]   Siegel, R., and Howell, J. R. *Thermal Radiation Heat Transfer*. Hemisphere Publishing, New York, 1981.

[15]   Torrance, K. E. and Sparrow, E. M. Theory for off-specular reflection from roughened surfaces. In *Journal of the Optical Society of America* 57, 9 (September 1967) 1105-1114.

[16]   Torrance, K. E. and Sparrow, E. M. Off-specular peaks in the directional distribution of reflected thermal radiation. In *Journal of Heat Transfer* (May 1966) 223-230.

[17]   Wallace, P. R. *Mathematical Analysis of Physical Problems*. Dover Publications, Mineola, N. Y. 1984.

[18]   Ward, G. J., Rubinstein, F. M. and Clear, R. D. A ray tracing solution for diffuse interreflection. In *Proceedings of SIGGRAPH '88* (August 1-5, 1988, Atlanta, Georgia), *Computer Graphics* 24, 4 (August 1988), 85-92.

[19]   Whitted, T. An improved illumination model for shaded display. In *Communications of the ACM* 23, 6 (June 1980) 343-349.

A Survey of Shading and Reflectance Models

Christophe Schlick

LaBRI [1] *, 351 cours de la libération, 33405 Talence (FRANCE)*
schlick@labri.u-bordeaux.fr

Abstract : *Since the beginning of computer graphics, three decades ago, a great number of models intended to describe the behaviour of light on a given point of a surface have been proposed. Almost every author uses his own terminology and/or notations. To understand clearly the similarities and the differences between existing models, reformulating them with unified notations is essential. Such a work has been done by Hall [12] in 1986. This paper is a new survey of shading and reflectance model, including the most recent models. Moreover, after the lengthy ennumeration, some original models are proposed, which try to embed interesting features of previously disjointed work into new formulations.*
Keywords : Survey, Local Illumination, Reflectance Models, Shading Models, BRDF

1 Introduction

The goal of realistic rendering is to create computer-generated pictures of synthetic scenes almost undistinguishable from photographies of real environments. This goal implies to simulate as close as possible the behaviour of light, starting from luminaires, travelling through the scene, interacting with different objects, and finally reaching the camera.

Such a simulation involves two completely different algorithmic processes. The first one, *global illumination* tries to collect the contributions of all parts of the environment which are illuminating a given point of the scene. The second one, *local illumination*, has to compute the transformation that occurs at this point between incoming and outcoming light.

This survey focuses only on the second aspect. Its goal is first to propose some simple notations to express local illumination models (LIM, for short). Then many of the numerous LIM proposed in the literature are recalled and reformulated using these notations. Finally, several new models are proposed which collect some strong points of previous work. As usual,[12, 35] existing LIM will be divided in two main families.

The first family contains ad hoc empirical models. These models are usually computationally efficient and some of them can provide very realistic-looking pictures, but without any exact value of energy or intensity. Several domains, where such quantitative informations are not required, are perfect application fields for empirical models (special effects in movies, video art, commercials).

In the second family, we find physically-based theoretical models. These models provide quantitative values that have shown to be in good adequation with experimental data. Therefore they are well adapted to applications for which a close simulation of real phenomena is essential (lighting industry, interior decoration, architecture).

Three main proposals about terminology and notation are made in the present paper. First, in order to make a clear distinction between empirical and theoretical models, the first ones will be called *shading models*, and the second ones *reflectance models*. Second, to distinguish easily wavelength dependent and independent variables, every term which is function of the wavelength will be subscripted by λ. Such a term has to be defined and/or computed, theoretically for every wavelength of the visible spectrum, and practically for a given number of samples (three in trichromatic models, more in spectral models[11]). Finally, simple notations for vectors and angles involved in the geometry of local illumination will be used; they are presented in Figure 1 and Table 1.

---

[1] Laboratoire Bordelais de Recherche en Informatique (*Université Bordeaux I* and *Centre*
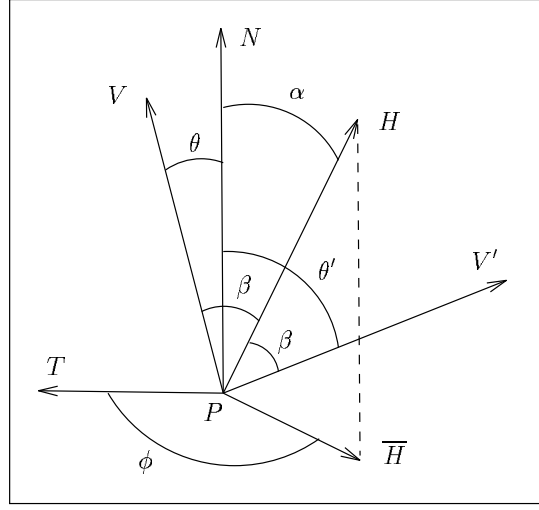
Figure 1: Angles and vectors for local illumination

V	=	Outcoming direction of light
V'	=	Incoming direction of light
N	=	Surface normal vector
T	=	Surface tangent vector*
H	=	Bisector vector of V and V'
$\overline{H}$	=	Projection of $H \perp N$

t	=	$(H \cdot N)$
u	=	$(H \cdot V)$
v	=	$(V \cdot N)$
v'	=	$(V' \cdot N)$
w	=	$(T \cdot \overline{H})$

α	=	$\cos^{-1} t$
β	=	$\cos^{-1} u$
θ	=	$\cos^{-1} v$
θ'	=	$\cos^{-1} v'$
φ	=	$\cos^{-1} w$

Table 1: Angles and vectors for local illumination

* The notion of tangent vector which makes sense only for anisotropic surfaces will be detailled later.

2 Definitions and Properties

The interaction of light with a surface is usually described by relating incoming and outcoming radiances at a given point P on the surface. For theoretical models (*ie* reflectance models with our terminology), this expression involves a function $R_\lambda(P, V, V')$ called *bidirectional reflectance distribution function* (BRDF, for short). For empirical models (*ie* shading models), we propose to use, by analogy, a function $S_\lambda(P, V, V')$ called *bidirectional shading function* (BSF, for short).

Several authors[13, 20] have proposed formulations using only the BRDF for both reflectance and shading models. But, as we will see, some shading models are so far from physics that the use of BRDF is neither natural nor simple and tends to confuse the model.

2.1 BRDF

For reflectance models, relation between outcoming and incoming radiances is given by :

$$L_\lambda(P, V) = \int_{2\pi} R_\lambda(P, V, V') \; L_\lambda(P, \text{-}V') \; (N \cdot V') \; dV' \tag{1}$$

where

- $L_\lambda(P, V)$ is the reflected radiance leaving point P in direction V

- $L_\lambda(P, \text{-}V')$ is the incident radiance reaching point P from direction $\text{-}V'$

- $R_\lambda(P, V, V')$ is the BRDF of the surface at point P between directions V and V'

- dV' is a differential solid angle surrounding direction V'

- V, V' and N are unit vectors so $(N \cdot V')$ is the cosine of the angle between N and V'

The reflected radiance is the integral, for all possible directions (*ie* hemisphere above the surface) of incident radiances, scaled by the BRDF and the projected solid angle. Such a formulation is well adapted to rendering algorithms (radiosity,[7] path tracing,[18] two pass methods[30]) that perform a true integration and for which the solid angle information is available.

The BRDF has got two important properties that result directly from physics of light[2]. First, due to the *Helmholtz Reciprocity Rule*, R_λ is symmetric relative to V and V' :

$$\forall V \quad \forall V' \quad R_\lambda(P, V, V') = R_\lambda(P, V', V) \tag{2}$$

Second, due to the *Energy Conservation Law*, R_λ obeys the following normalization condition :

$$\forall V \quad \int_{2\pi} R_\lambda(P, V, V') \, (N \cdot V') \, dV' \leq 1 \tag{3}$$

2.2 BSF

For shading models, relation between outcoming and incoming radiances is given by :

$$L_\lambda(P, V) = \sum_{i=1}^{l} S_\lambda(P, V, V_i) \, L_\lambda(P, \text{-}V_i) \tag{4}$$

where

- $L_\lambda(P, V)$ is the reflected radiance leaving point P in direction V

- $L_\lambda(P, \text{-}V_i)$ is the incident radiance reaching point P from direction $\text{-}V_i$

- $S_\lambda(P, V, V_i)$ is the BSF of the surface at point P between directions V and V_i

This time, the reflected radiance is simply a discrete sum of l different radiances coming from l different directions weighted by the BSF. Such a formulation works well for rendering algorithms (scan-line projective methods, ray-tracing[37]) in which only a given number of contributions are considered (point light sources, directional light sources, mirror direction).

An important condition that has to be fulfilled by shading models is the normalization property :

$$\forall V \quad \forall V' \quad S_\lambda(P, V, V') \leq 1 \tag{5}$$

which states that the reflected radiance of a given luminaire cannot be larger than its incident radiance. But note that with several light sources, the total reflected radiance may be larger than 1. Therefore, clamping (or global rescaling which avoids color shifts) of the radiance is usually needed before visualization.

3   Shading Models

Every empirical LIM proposed in computer graphics is based on some qualitative knowledge coming from physics of light, which may be summarized as follows :

- Two kinds of surfaces can be distinguished according to the way they reflect light.

- On one hand, there are *diffuse surfaces* for which light is reflected in every direction.

- The limit case of diffuse surfaces, so-called perfectly diffuse or *Lambert surfaces*, is obtained when light is equally reflected in every direction (Lambert provide the quantitative law for such surfaces in 1760).

- On the other hand, there are *specular surfaces* for which light is reflected only in a small area around the mirror direction.

- The limit case of specular surfaces, so-called perfectly specular or *Fresnel surfaces*, is obtained when the light is reflected only in the exact mirror direction (light reflection on such surfaces was comprehensively explained by Fresnel in 1815).

3.1 Diffuse and Specular Reflection

Early work on shading models has been done during the seventies at University of Utah. A simple model which accounts for Lambert surfaces (the BSF depends only on the cosine of the incident angle) was used by several people[5, 10] around 1970 :

$$S_\lambda(v') = C_\lambda \ v' \tag{6}$$

where

- $C_\lambda \in [0, 1]$ represents the ratio of incident radiance at wavelength λ that is reflected by the surface

In 1975, Phong[23] proposed the first model in computer graphics able to deal with non-lambertian surfaces. In this model, the BSF is expressed as a linear combination of a diffuse part and a specular one :

$$S_\lambda(t, v') = d \ C_\lambda \ v' + s \ C'_\lambda \ t^n \quad \text{with} \quad d + s = 1 \tag{7}$$

where

- d (resp. s) $\in [0, 1]$ is the ratio of the surface behaving as a diffuse (resp. specular) reflector

- C_λ (resp. C'_λ) $\in [0, 1]$ is the ratio of incident light at wavelength λ reflected by the diffuse (resp. specular) reflector

- $n \in [1, \infty)$ represents the shininess of the specular reflector (the higher the value of n, the more the surface looks shiny)

3.2 Plastic and Metallic Appearance

Phong has noted that for a lot of materials, the specular highlight is unsatured (*ie* has the same color as the light source). Therefore in his original implementation, he used $C'_\lambda = 1$ for every wavelength λ. But, as remarked by Cook,[8] having $C'_\lambda = 1$ gives a kind of plastic appearance to the surface. In order to get a visual impression of metallic surface, $C'_\lambda = C_\lambda$ is a much better choice.

To enable rendering of surfaces with both plastic and metallic appearance, Hall & Greenberg[11] proposed to add in the Phong model, the wavelength-dependent Fresnel factor $F_\lambda(u)$ introduced by Cook & Torrance[8] (see Section 4.4) :

$$S_\lambda(t, u, v') = d \ C_\lambda \ v' + s \ F_\lambda(u) \ t^n \tag{8}$$

The expression of the Fresnel factor (see Equation 28) is very complex, especially when using it with an empirical model for which speed and hardware implementation are desired. Marsh[21] has presented a simple empirical solution to the plastic vs. metallic appearance problem :

$$S_\lambda(t, v') = d \ C_\lambda \ v' + s \ (rC_\lambda + 1 - r) \ t^n \tag{9}$$

where

- $r \in [0, 1]$ is the ratio of the surface behaving as a metallic reflector (and consequently $1 - r$ is the ratio of the plastic reflector).

Several authors[20, 27] have noticed that when $n = 0$, the specular reflector is very similar to the diffuse one. Therefore, we propose to go a step further with Marsh's expression by unifying the diffuse and the specular reflector :

$$S_\lambda(t) = r \ C_\lambda \ t^n + (1 - r) \ C'_\lambda \ t^{n'} \tag{10}$$

or more generally by defining the surface as a weigthed sum of multiple reflectors :

$$S_\lambda(t) = \sum_{i=1}^{k} r_i \ C_{i_\lambda} \ t^{n_i} \quad \text{with} \quad \sum_{i=1}^{k} r_i = 1 \tag{11}$$

where

- k is the number of reflectors
- $r_i \in [0, 1]$ is the weight of the i$^{\text{th}}$ reflector
- $C_{i_\lambda} \in [0, 1]$ is the ratio of incident light at wavelength λ reflected by the i$^{\text{th}}$ reflector
- $n_i \in [1, \infty)$ defines the shininess of the i$^{\text{th}}$ reflector

Note that if such an approach may appear new for shading models, a similar technique has been used for decades in the fields of astrophysics and participating media with the Henyey-Bernstein[16] phase function.

3.3 Highlight Function

Factor t^n in Equation 7 has been introduced by Phong after an empirical analysis of light reflection on specular surfaces. This factor controls the size of the bright spots that appear on such surfaces, and thus it is sometimes called *highlight function*. The highlight function is by far the most time consuming element in the Phong model. For that reason, many alternatives have been presented to this function since 1975 in order to reduce its computation cost.

Trowbridge & Reitz[34] presented a simple highlight function (later introduced in computer graphics by Blinn[4]) which has been rarely implemented in rendering environments, despite its advantages (simplicity and low cost). This function uses a polar equation of an ellipse to express the behaviour of the specular reflector. Changing the excentricity of the ellipse enables to control precisely the size of specular highlights.

Bishop & Weimer[3] have proposed to tabulate the function t^n (one table for each value of n) with a sufficient set of samples and interpolate the missing ones. When using linear interpolation, the computation cost is very low, but such an interpolation often creates visible Mach bandings which can only be eliminated by taking larger tables or higher order interpolations, leading to a memory/speed tradeoff.

Several solutions have also be proposed which act on angle α and not on its cosine. For instance, Blinn[4] proposed a highlight function that is a gaussian function of α. To speed-up the calculation, this gaussian can eventually be approximated by its Parzen window.[24] Another possibility is to replace the exponentiation $\cos^n \alpha$ by a polynomial $\rho(\alpha)$. This polynomial can be a Chebyshev approximation, as proposed by Poulin & Fournier,[24] or a piece-wise quadratic function, as proposed by Kuijk & Blake.[19] A main limitation for all these methods compared to solution working on the cosine, is that a call to `acos` (almost as expensive as an exponentiation) is needed, whereas t is directly obtained by the dot product $(N \cdot H)$.

As previous papers have observed, there is no need for great accuracy when approximating t^n, since the Phong model is empirical and not intended for physical rendering : its only purpose is to give a visual impression of specularity by adding highlights on objects. Therefore, every function that evokes a similar impression can be used instead of the exponentiation. We have proposed a solution[26, 28] which uses a rational fraction and requires only 1 division, 1 multiplication, 1 subtraction and 1 addition :

$$S_\lambda(t, v') = d \ C_\lambda \ v' + s \ C'_\lambda \ \frac{t}{t - pt + p} \tag{12}$$

where :

- $p \in [1, \infty)$ represents the shininess of the specular reflector. Empirical tests have shown that for a given n in the original Phong model, a quite similar behaviour of the specular reflector in the new model is obtained with $p = n^{3/2}$.

3.4 Anisotropic Reflection

When the radiance $L(P, V)$ does not change while the surface is rotated around its normal vector at point P, the surface is called *isotropic*. Otherwise the surface is called *anisotropic* and, with our notations, it means that the BSF is function of angle φ (or of its cosine w). Such a surface is usually modelized by supposing that there are straight scratches on it. The direction of these scratches are used to define a tangent vector T (and thus a binormal vector $B = T \times N$) on the surface. Arbitrary scratches can be generated by texturing methods, such as frame mapping,[17] which enable to modify the local frame (tangent, normal, binormal) across the surface.

Very few shading models accounting for anisotropy have been proposed. In 1981, Ohira[39] described an extension of the Phong model intended to simulate light reflection on surfaces like brushed metals. The same model has later been used by Yamana and Suenaga[40] to render human hair. The basic idea of the model is to replace the rotational symmetry around the normal vector by an elliptical asymmetry :

$$S_\lambda(t, v', w) = d \; C_\lambda \; v' + \tag{13}$$

$$s \; C'_\lambda \; \left[\frac{a^2 b^2 c^2 t^2}{a^2 b^2 t^2 + (1-t^2)(b^2 c^2 w^2 + c^2 a^2(1-w^2))} \right]^{n/2}$$

where

- a (resp. b and c) $\in [0, 1]$ is the radius of the ellipsoid defining the highlight function, along the tangent (resp. binormal and normal) directions.

Note that the original formulation used by Ohira contains some errors which may generate strange visual effects when implemented (for instance, $a = b = c = 1$ does not give the original Phong model), Equation 13 is the corrected expression. Moreover, having four parameters a, b, c and n to define elliptic anisotropic reflection is a bit redundant. Using an similar idea to the one proposed by Ward[35] (see Section 4.5), there is a more straightforward extension to anisotropy for the Phong model, which requires only two parameters and a much simpler expression :

$$S_\lambda(t, v', w) = d \; C_\lambda \; v' + s \; C'_\lambda \; t^p \tag{14}$$

$$\text{with} \quad p = \frac{mn}{m - mw^2 + nw^2}$$

where

- m (resp. n) $\in [1, \infty]$ represents the shininess of the surface in the tangent (resp. binormal) direction. Thus $m = n$ defines an isotropic reflector and $m > n$ (resp. $m < n$) defines scratches in the tangent (resp. binormal) vector direction. The more m and n are different, the more the anisotropy is visible.

A last empirical model for anisotropic reflection has been proposed by Fournier[9] in which a weighted normalized sum of k repetitive applications of an isotropic model is computed, using each time another normal vector (for instance, two symmetric vectors for V-groove straight scratches) :

$$S_\lambda(t, v') = \sum_{i=1}^{k} r_i \; S_{i_\lambda}(t_i, v'_i) \quad \text{with} \quad \sum_{i=1}^{k} r_i = 1 \tag{15}$$

The advantage of this model is that anisotropy can easily be obtained with an isotropic (eventually hardware) shader. The drawback is that many calculations are wastefully recomputed at each iteration.

3.5 Proposal for a Simple Shading Model

The main benefit of an enumeration using a uniform notation as done above, is that strength and weakness of every model appear clearly. The last step that remains to do is to embed the strong points of several

models into a unique formulation. Therefore, using Equations 11, 12 and 14, we propose a new shading model including multiple reflectors behaviour, anisotropic reflection and computation efficiency :

$$S_\lambda(t, w) = \sum_{i=1}^{k} r_i \; C_{i_\lambda} \; \frac{t}{t - p_i t + p_i} \tag{16}$$

$$\text{with} \quad p_i = \frac{m_i n_i}{m_i - m_i w^2 + n_i w^2} \quad \text{and} \quad \sum_{i=1}^{k} r_i = 1$$

where

- k, $r_i \in [0, 1]$ and $C_{i_\lambda} \in [0, 1]$ have the same meaning as in Equation 11
- $m_i \in [1, \infty)$ defines the shininess of the ith reflector in the tangent direction
- $n_i \in [1, \infty)$ defines the shininess of the ith reflector in the binormal direction

4 Reflectance Models

For centuries, the only knowledge in physics of light has been the *mirror law* that predicts the direction of reflection on a mirror. According to the legend, by gathering this law and his work on conics, Archimedes proposed the use of parabolic mirrors to set fire on Marcellus' fleet during the siege of Syracuse in 212 BC.

The refraction of light (*ie* modification of the light direction at the boundary of two different optical media) was more difficult to understand. It is only at the beginning of the seventhieth century that Snell and Descartes have proposed, independently, an explanation which involves the ratio of light propagation speed in the two media.

The first law in physics of light able to provide quantitative values (*ie* energy distributions) was introduced in 1760 by Lambert who discovered the famous *cosine law* that expresses light reflection on perfectly diffuse surfaces (now called lambertian surfaces, for that reason). On the other side, reflection of light on perfectly specular surfaces has been comprehensively described, in a quantitative way, by Fresnel in 1815 (using the transversal wave model for light proposed by Huyghens), and generalized to any kind of electromagnetic wave by Maxwell in 1866.

One has to wait until the second half of the twentieth century (consequently to the invention of the radar) to see the publication of some theories[2, 33, 32, 1] able to explain the behaviour of electromagnetic waves on rough surfaces. These theories are more or less complex, according to the number of physical phenomena they account for (self-shadowing, diffraction, interference, polarization, finite electric conductivity and magnetic permeability). Moreover, all these theories are limited to reflection on rough surfaces; several phenomena that occur during transmission of light through a rough surface are still unexplained.[13]

4.1 Reflectance Models for Ray Optics

In physics, there are usually two different approaches to explain phenomena involved in light transport. In the first theory, called *wave optics* or physical optics, light is considered as an electromagnetic wave for which the wavelength is in the visible spectrum, approximatively [380nm,780nm]. In the second theory, called *ray optics* or geometrical optics, light is supposed to be composed of non-interacting straight rays, each of them carrying a certain amount of energy.[13] Using ray optics simplifies noticeably the expression of light transport but, as a counter part, specific phenomena due to the wave aspect of light (interference, diffraction, polarization) are totally neglected.

Another classical approximation in physics (proposed originally by Bouguer around 1750) is to consider a rough surface as being composed of so-called *microfacets* which are small planar surfaces. Only microfacets for which the normal vector is in direction H (see Figure 1) contribute to the reflection between V and V'. When the size of the microfacets is large compared to the wavelength, the ray optics assumption is generally considered as valid.

The first theoretical LIM in computer graphics has been introduced by Cook & Torrance[8] and is based on the ray optics simplification. This model was developed upon work previously done in physics by Beckmann & Spizzichino[2] and Torrance & Sparrow.[33] As the BSF in the Phong model, the BRDF in the Cook-Torrance model is expressed as a linear combination of a diffuse part and a specular one, but it depends here on five different angles :

$$R_\lambda(\alpha, \beta, \theta, \theta', \varphi) = \frac{d}{\pi} C_\lambda + \frac{s}{4\pi v v'} D(\alpha, \varphi) G(\theta, \theta') F_\lambda(\beta) \tag{17}$$

where

- $d \in [0, 1]$, $s \in [0, 1]$ and $C_\lambda \in [0, 1]$ have the same meaning as in Equation 7 (in particular, it means that $d + s = 1$)

- $D(\alpha, \varphi) \in [0, \infty)$ is the microfacets slope distribution function which defines the fraction of the facets that are oriented in direction H

- $F_\lambda(\beta) \in [0, 1]$ is the Fresnel factor which describes how light is reflected by each smooth microfacet

- $G(\theta, \theta') \in [0, 1]$ is the geometric attenuation coefficient which expresses the ratio of light that is not self-obstructed by the surface

In fact, due to the presence of the Fresnel factor, the Cook-Torrance model is more exactly a kind of hybred between ray optics and wave optics — a pure ray optics model might have used C'_λ instead of $F_\lambda(\beta)$.

One important point to notice is that Equation 17 is only valid when it fulfills the normalization condition (Equation 3). Therefore, it implies a condition on the distribution function[2] :

$$\int_0^{\pi/2} \int_0^{2\pi} D(\alpha, \varphi) \cos \alpha \sin \alpha \, d\alpha \, d\varphi = \pi \tag{18}$$

If we assume an isotropic behaviour of the surface (*ie* the BRDF is invariant by rotation around the normal vector) then Equation 18 becomes :

$$\int_0^{\pi/2} D(\alpha) \, 2 \cos \alpha \sin \alpha \, d\alpha = 1 \tag{19}$$

Moreover, if we express the angular dependence of the three factors (D, F and G) in terms of their cosine, Equation 19 may be rewritten as :

$$\int_0^1 2t \, D(t) \, dt = 1 \tag{20}$$

and Equation 17 becomes :

$$R_\lambda(t, u, v, v') = \frac{d}{\pi} C_\lambda + \frac{s}{4\pi v v'} D(t) \, G(v, v') \, F_\lambda(u) \tag{21}$$

In the next paragraphs, different formulations that have been proposed, both in physics and computer graphics, for the slope distribution function, the geometric attenuation coefficient and the Fresnel factor, will be detailled.

4.2 Slope Distribution Function

In the original work of Torrance & Sparrow,[33] the distribution function $D(t)$ was a gaussian function of angle α that did not obey Equation 20. When Cook & Torrance introduced the model in computer graphics, they proposed the formulation derived by Beckmann & Spizzichino[2] in the case where the ray optics assumption is valid :

$$D(t) = \frac{1}{m^2 t^4} \, e^{\frac{t^2-1}{m^2 t^2}} \tag{22}$$

where

8

- m is the root mean square slope of the microfacets (theoretically $m \in [0, \infty)$ but, in fact, it almost never exceeds 0.5 for real surfaces)

When the surface is very rough (large values for m), orientations of microfacets are very dispersed. When the surface is smoother (small values for m) microfacet normals H come closer to the average normal N. For perfectly smooth surfaces ($m = 0$), the distribution $D(t)$ becomes a Dirac function.

Equation 22 has at least two main advantages. First, as noticed by Cook, it does not involve any arbitrary constant but only a physically meaningful value m that can be experimentally measured. Second, it obeys Equation 20, whatever the value of m, which means that the model is energy conservative.

Equation 22 results from some simple assumptions about the microfacets distribution. More complete expressions for $D(t)$ exist (see Section 4.7) which provide much better adequation to experimental data. Therefore, because Equation 22 is more or less empirical, why not try to find a simpler and less expensive expression, as far as it obeys Equation 20 ? A possible solution[26, 29] that fulfills, as Equation 22, $D(0) = 0$ and $D(1) = 1/m^2$ is :

$$D(t) = \frac{t^2}{(mt^4 - rt^4 + r)^2} \quad \text{with} \quad r = \frac{1}{2m} \tag{23}$$

which needs only 1 division, 4 multiplications and 1 addition using an optimized implementation.

4.3 Geometric Attenuation Coefficient

Several expressions for the geometric attenuation (sometimes called self-shadowing) coefficient $G(v, v')$ have also be proposed in the literature.[33, 25, 31] In their paper, Cook & Torrance proposed to use the formulation derived originally by Torrance & Sparrow[33] :

$$G(t, u, v, v') = \min \left[1, 2\frac{tv}{u}, 2\frac{tv'}{u} \right] \tag{24}$$

This expression results from several coarse approximations about the surface geometry and therefore does not meet experimental results : its first derivative is not continuous, it is not invariant by rotation around the normal vector and it is totally independent of the surface roughness.

The formulation proposed by Smith[31] (introduced in the computer graphics field by He & *al.*[15]) is not subjected to these restrictions and has been experimentally validated. Moreover, it depends only on v and v', and is separable in v and v' :

$$G(v, v') = G(v) \; G(v') \tag{25}$$

After several equivalences, the original expression of $G(v)$ can be written more compactly :

$$G(v) = \frac{g}{g + 1} \tag{26}$$

$$\text{with} \quad g = \sqrt{h\pi} \left(2 - \text{erfc} \sqrt{h} \right) \quad \text{and} \quad h = \frac{v^2}{2m^2 (1 - v^2)}$$

Despite its complicated formulation, the shape of the function is quite simple. We have found[26, 29] that it can be approximated by a very simple expression :

$$G(v) = \frac{v}{v - kv + k} \quad \text{with} \quad k = \sqrt{\frac{2m^2}{\pi}} \tag{27}$$

Several tests have shown[26] that the use of Equation 27 instead of Equation 26 for the expression of $G(v, v')$ in the Cook-Torrance model, provides pictures which are totally undistinguishable from the originals (less than 1% error).

4.4 Fresnel Factor

The Fresnel factor $F_\lambda(u)$ expresses the reflection of light on the well-oriented microfacets (*ie* the facets whose normal vector is in direction H). For a non-polarized electromagnetic wave, its formulation is[38] :

$$F_\lambda(u) = \frac{1}{2} \frac{(a-u)^2 + b^2}{(a+u)^2 + b^2} \left[\frac{(a+u-1/u)^2 + b^2}{(a-u+1/u)^2 + b^2} + 1 \right] \tag{28}$$

$$a^2 = \frac{1}{2} \left(\sqrt{(n_\lambda^2 - k_\lambda^2 + u^2 - 1)^2 + 4n_\lambda^2 k_\lambda^2} + n_\lambda^2 - k_\lambda^2 + u^2 - 1 \right)$$

$$b^2 = \frac{1}{2} \left(\sqrt{(n_\lambda^2 - k_\lambda^2 + u^2 - 1)^2 + 4n_\lambda^2 k_\lambda^2} - n_\lambda^2 + k_\lambda^2 - u^2 + 1 \right)$$

where

- n_λ is the ratio of the refraction indices above and below the surface

- k_λ is the extinction coefficient of the surface.

An interesting characteristic of the Fresnel factor is that $F_\lambda = 1$ at a grazing incidence ($\beta = \pi/2$ *ie* $u = 0$) whatever the wavelength λ. Another characteristic is that for dielectric materials (for which $k_\lambda = 0$), the expression becomes much simpler :

$$a = \sqrt{n_\lambda^2 + u^2 - 1} \quad \text{and} \quad b = 0 \tag{29}$$

One difficulty that has prevented a large use of F_λ in rendering environments comes from the fact that n_λ and k_λ are seldom known. Some experimental data exist[22] but usually one can only find a unique value $\overline{n}$ and $\overline{k}$ for a wavelength in the middle of the visible spectrum ($\lambda = 589$ nm, a sodium lamp). On the other hand, a data which has been measured for thousands of materials is the spectral distribution f_λ of the Fresnel factor at normal incidence[38] ($\beta = 0$ *ie* $u = 1$). When $\overline{n}$, $\overline{k}$ and f_λ are the only known data, Cook & Torrance have proposed the following approximation[8] :

- Compute $\overline{F}(u)$ and $\overline{f} = \overline{F}(1)$ with $\overline{n}$ and $\overline{k}$, using Equation 28

- For each wavelength λ, compute $F_\lambda(u)$, using Equation 30

$$F_\lambda(u) = f_\lambda + (1 - f_\lambda) \frac{\overline{F}(u) - \overline{f}}{1 - \overline{f}} \tag{30}$$

This approximation not only solves the lack of experimental data, but also speeds-up the calculation because the complete expression of $F_\lambda(u)$ is evaluated only once, for an average n_λ and k_λ. But even so, the computation of the Fresnel factor remains expensive and further optimization should be possible.

By plotting the Fresnel factor for different materials,[13] one can see that the shape of the curves does not vary very much according to the kind of material. The main difference is the value f_λ where the curve arrives at $u = 1$. Therefore, a step further in the approximation process could be to make $F_\lambda(u)$ only dependent on f_λ. We have found[26, 29] that a simple polynomial interpolation provides an excellent approximation (less than 1% error) :

$$F_\lambda(u) = f_\lambda + (1 - f_\lambda)(1 - u)^5 \tag{31}$$

4.5 Anisotropic reflection

As for shading models, very few reflectance models dealing with anisotropic reflection have been proposed (except several non analytical models that will be discussed in Section 4.8). Poulin & Fournier[24] have proposed a model in which anisotropic orientations of the microfacets are simulated by adding or substracting groups of microscopic cylinders on the surface. When changing size and orientation of the cylinders, the model enables to create various structures for surface geometry, but as admitted by one of

the authors,[9] it does not provide a greater variety of visual effects than the very empirical model he also proposes (see Equation 15). This fact tends to prove that, after all, the elliptical model used in shading models could also be interesting for reflectance models.

Following that idea, Ward[35] has proposed a simple anisotropic reflectance model based on two assumptions : a gaussian model is used for the specular part of the BRDF and an elliptical model is used for the anisotropic part :

$$R_\lambda(t, v, v', w) = \frac{d}{\pi} C_\lambda + \frac{s}{4\pi\sqrt{vv'}} C'_\lambda \ D(t, w) \tag{32}$$

$$\text{with} \quad D(t, w) = \frac{1}{mn} \ e^{\frac{t^2-1}{t^2}(\frac{w^2}{m^2} + \frac{1-w^2}{n^2})}$$

where

- $d \in [0, 1]$, $s \in [0, 1]$, $C_\lambda \in [0, 1]$ and $C'_\lambda \in [0, 1]$ have the same meaning as in Equation 7

- $D(t, w) \in [0, \infty)$ is the anisotropic slope distribution function

- $m \in [0, 0.5]$ (resp. $n \in [0, 0.5]$) is the RMS slope of the surface in the tangent vector (resp. binormal vector) direction

Compared to the Cook-Torance, the most noticeable difference is the suppression of the geometric attenuation coefficient and the Fresnel factor. In fact, with his model, Ward does search for experimental justification (the model was presented with a device that enables measurements of BRDFs at low cost) and not for complete theoretical correctness. Therefore, the four parameters (d, s, m and n) in Equation 32 are usually not defined by hand, but by a least squares error minimization technique in order to fit experimental results as close as possible. In order to get a fast convergence for the least squares algorithm, Ward has to work with a simple model; thus the suppression of $G(v, v')$ and $F_\lambda(u)$.

Note also the inclusion of the square root on the denominator which is used to avoid infinite values of the BRDF at grazing reflection angles ($v \to 0$). This unbounded growing is normally prevented by the geometrical attenuation coefficient in the Cook-Torance model. As stated by Ward, this ad hoc solution for normalization provides fairly good energy conservation when $m^2 \ll 1$ and $n^2 \ll 1$.

Using Ward's idea, a more complete extension to anisotropic reflection of the Cook-Torrance model may be imagined :

$$R_\lambda(t, v, v', w) = \frac{d}{\pi} C_\lambda v' + \frac{s}{4\pi vv'} D(t, w) \ G(v, v') \ F_\lambda(u) \tag{33}$$

$$\text{with} \quad D(t, w) = \frac{1}{mnt^4} \ e^{\frac{t^2-1}{t^2}(\frac{w^2}{m^2} + \frac{1-w^2}{n^2})}$$

Note the addition of a t^4 factor on the denominator (compared to Equation 32) which enables $D(t, w)$ to fulfill Equation 18, for each value of the RMS slopes, and thus insures rigourous energy conservation.

4.6 Proposal for a Simple Reflectance Model

As we have done for shading model, interesting features of different reflectance models for ray optics should be embeded in a unique formulation. Following the same process as in Section 3, a first step is to unify the diffuse and the specular behaviours, in order to enable multiple reflectors as in Equation 11. This implies to find a slope distribution function $D(t, w)$ which varies continuously from a constant distribution (perfectly diffuse) to a Dirac distribution (perfectly specular) according to some roughness parameter. The following function includes this feature (constant when $m = n = 1$, Dirac when $m = n = 0$) and obeys Equation 18 whatever the values of m and n :

$$D(t, w) = \frac{1}{mn(p - pt^2 + t^2)^2} \quad \text{with} \quad p = \frac{w^2}{m^2} + \frac{1-w^2}{n^2} \tag{34}$$

where

- m (resp. n) $\in [0, 1]$ is the roughness in the tangent (resp. binormal) direction

Using that distribution function, we propose a reflectance model, including energy conservation, reciprocity rule, various material appearances, self-shadowing, Fresnel effects, anisotropic reflection, and computational efficiency :

$$R_\lambda(t, u, v, v', w) = \sum_{i=1}^{k} \frac{r_i}{4\pi v v'} D_i(t, w) G_i(v, v') F_{i_\lambda}(u) \tag{35}$$

where

- k and r_i have the same meaning as in Equation 11

- $D_i(t, w) \in [0, \infty)$ is given by Equation 34

- $G_i(v, v') \in [0, 1]$ is given by Equation 27

- $F_{i_\lambda}(u) \in [0, 1]$ is given by Equation 31

4.7 Reflectance Models for Wave Optics

The first model based on wave optics for light reflection on a rough surface has been proposed in 1963 by Beckmann & Spizzichino[2] and accounts comprehensively for diffraction and interference phenomena. In that model, the reflected electromagnetic scalar field is obtained by performing an analytical integration of the Kirchoff diffraction equation, using a statistical average over microfacet orientations. In fact, three different expressions are derived according to the relative size of the wavelength and the microfacets. An extension has later been presented by Stogryn[32] which computes the reflected electromagnetic vector field (instead of the scalar field), and therefore enables to account for light polarization and Fresnel theory. Bahar & Chakrabarti[1] have proposed quite a similar technique, including some approximation functions to get better adequacy to computer graphics.

The most complete reflectance model using wave optics has been proposed by He *et al.*[15] in the computer graphics field. This model includes almost every physical phenomena involved in light reflection on rough surface (polarization, diffraction, interference, smaller roughness for grazing rays) and can be applied to many different materials. Detailling this model and its numerous equations is out of scope here. We will limit us to point out that for non-polarized light, the overall form of its expression is very close to Equation 17. The main differences are the addition of a coherent reflection term (which explains the mirror-like reflection that may appear even on rough surfaces) and a more complete — and much more complex — slope distribution function that accounts for many characteristics of the material (*ie* not only the RMS slope).

It should be noted that if wave optics can be included quite easily in local illumination models, accounting efficiently and comprehensively for wave effects in global illumination is still an open problem.

4.8 Non Analytical Reflectance Models

Several reflectance models have been proposed which are based on a completely different approach. Instead of providing an analytical formulation, these models use an algorithmic process to compute the values of the BRDF for a couple of incoming/outcoming directions. This algorithmic evaluation can either be computed and stored (using sampled hemispheres[17] or spherical harmonics[6, 36]) as a preprocessing, or calculated directly during the rendering process.[14]

The first model using this approach has been presented in 1985 by Kajiya[17], in which a numerical integration of the Kirchoff scalar equation is done for each couple of directions, providing an anisotropic extension of the Beckmann model.

Extending an idea initially proposed by Cabral *et al.*[6], Westin *et al.*[36] have presented a very general technique using a Monte-Carlo ray tracing technique to simulate local behaviour of light on surfaces with almost any kind of microscopic geometrical structure (stone, brushed metal, velvet, weaving). While Westin *et al.* have worked mainly on multiple surface scattering created by complex microfacets configurations, Hanrahan & Krueger[14] have proposed a similar approach (using also a Monte-Carlo technique) which simulates subsurface scattering on layered surfaces (plastic, skin, pigmented or varnished materials).

It should be noted that multiple surface and subsurface scattering are usually considered as being the origin of diffuse reflection. Therefore, Westin's and Hanrahan's simulation techniques appear as somewhat complementary methods, which should be combined and may, ultimately, provide some alternatives to the Lambert model (that is rarely validated by experiments) for diffuse reflection. This could be an interesting direction for future research.

5   Conclusion

Local illumination models, which are intended to describe the behaviour of light on a surface, can be divided in two main families : empirical *shading* models and theoretical *reflectance* models. The first ones are adapted to rendering algorithms that consider only a given number of light contributions for each point (scan-line projective methods, ray-tracing). The second ones need to compute a solid angle for each contribution and therefore can only be used by algorithms that perform a true integration of the rendering equation (radiosity, path-tracing, two-pass methods).

In this paper, we have first reformulated local illumination models with uniform notations in order to observe their similarities and differences. Then, some interesting features of existing models have been identified, and finally, several new models, which collect these features in a unique formulation, have been proposed. Note that this paper has focused on the survey aspect, therefore the new propositions have been described shortly; more detailled explanations as well as a more complete model can be found elsewhere.[29]

Aknowledgments

References

[1] E.Bahar, S.Chakrabarti, *Full Wave Theory applied to Computer Aided Graphics*, Computer Graphics & Applications, v7, n7, p46-60, 1987

[2] P.Beckmann, A.Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces*, Pergamon Press, 1963

[3] G.Bishop, D.Weimer, *Fast Phong Shading*, SIGGRAPH 86, Computer Graphics, v20, n4, p103-106, 1986

[4] J.Blinn, *Models of Light Reflection for Computer Synthesized Pictures*, SIGGRAPH 77, Computer Graphics, v11, n4, p192-198, 1977

[5] W.Bouknight, *A Procedure for Generation of Three Dimensional Half-Toned Computer Graphics Presentations*, Com. of the ACM, v13, n9, p527-536, 1970

[6] B.Cabral, N.Max, R.Springmeyer, *Bidirectional Reflection Functions from Surface Bump Maps*, SIGGRAPH 87, Computer Graphics, v21, n4, p273-281, 1987

[7] M.Cohen, D.Greenberg, *The Hemi-Cube : a Radiosity Solution for Complex Environments*, SIGGRAPH 85, Computer Graphics, v19, n4, p31-40

[8] R.Cook, K.Torrance, *A Reflectance Model for Computer Graphics*, SIGGRAPH 81, Computer Graphics, v15, n4, p307-316, 1981

[9] A.Fournier, *Private Communication*

[10] H.Gouraud, *Computer Display of Curved Surfaces*, Trans. on Computers, v20, p623-629, 1971

[11] R.Hall, D.Greenberg, *A Testbed for Realistic Image Synthesis*, Computer Graphics & Applications, v3, n8, p10-20, 1983

[12] R.Hall, *Characterization of Illumination and Shading Techniques*, Visual Computer, v2, p268-277, 1986

[13] R.Hall, *Illumination and Color in Computer Generated Imagery*, Springer-Verlag, 1989

[14] P.Hanrahan, W.Krueger, *Reflection from Layered Surfaces due to Subsurface Scattering*, SIGGRAPH 93, Computer Graphics, v27, n3, p159-169, 1993

[15] X.He, K.Torrance, F.Sillion, D.Greenberg, *A Comprehensive Physical Model for Light Reflection*, SIGGRAPH 91, Computer Graphics, v25, n4, p187-196, 1991

[16] L.Henyey, J.Greenstein, *Diffuse Reflection in the Galaxy*, Journal of Astrophysics, v93, p70-77, 1941

[17] J.T.Kajiya, *Anisotropic Reflection Models*, SIGGRAPH 85, Computer Graphics, v19, n4, p15-22, 1985

[18] J.T.Kajiya, *The Rendering Equation*, SIGGRAPH 86, Computer Graphics, v20, n4, p143-150, 1986

[19] A.Kuijk, E.Blake, *Faster Phong Shading via Angular Interpolation*, Computer Graphics Forum, v8, n4, p315-324, 1989

[20] R.Lewis, *Making Shaders more Physically Plausible*, Fourth Eurographics Workshop on Rendering (Paris, France), Eurographics Association, p47-62, 1993

[21] D.Marsh, *An Efficient Ray-Tracing Renderer*, Tech. Report 87/360, University California San Diego, 1987

[22] E.D.Palik, *Handbook of Optical Constants of Solids*, Academic Press, 1985

[23] B.T.Phong, *Illumination for Computer Generated Pictures*, Com. of the ACM, v18, n6, p449-455, 1975

[24] P.Poulin, A.Fournier, *A Model for Anisotropic Reflection*, SIGGRAPH 90, Computer Graphics, v24, n4, p273-282, 1990

[25] M.Sancer, *Shadow-Corrected Electromagnetic Scattering from a Randomly Rough Surface*, Trans. on Antennas and Propagation, v17, n5, p577-585, 1969

[26] C.Schlick, *Divers Eléments pour une Synthèse d'Images Réalistes*, PhD Thesis (in french), Université Bordeaux I, 1992

[27] C.Schlick, *A Customizable Reflectance Model for Everyday Rendering*, Fourth Eurographics Workshop on Rendering (Paris, France), Eurographics Association, p73-83, 1993

[28] C.Schlick, *A Fast Alternative to Phong's Specular Model*, to appear in Graphics Gems IV (P.Heckbert Editor), Academic Press, 1994

[29] C.Schlick, *An Inexpensive BDRF Model for Physically Based Rendering*, Tech. Report 94/795, LaBRI, 1994 (submitted for publication to EUROGRAPHICS'94)

[30] F.Sillion, C.Puech, *A General Two-Pass Method integrating Specular and Diffuse Reflection*, SIGGRAPH 89, Computer Graphics, v23, n4, p335-344, 1989

[31] B.G.Smith, *Geometrical Shadowing of a Random Rough Surface*, Trans. on Antennas and Propagation, v15, n5, p668-671, 1967

[32] A.Stogryn, *Electromagnetic Scattering from Rough Finitely Conducting Surfaces*, Radio Science, v2, n4, p415-428, 1967

[33] K.E.Torrance, E.M.Sparrow, *Theory for off-specular Reflection from Roughened Surfaces*, Journal of the Optical Society of America, v57, n9, p1105-1114, 1967

[34] T.Trowbridge, K.Reitz, *Average Irregularity Repesentation of Roughened Surfaces*, Journal of the Optical Society of America, v65, n5, p531-536, 1975

[35] G.J.Ward, *Measuring and Modeling Anisotropic Reflection*, SIGGRAPH 92, Computer Graphics, v26, n2, p265-272, 1992

[36] S.Westin, J.Arvo, K.Torrance, *Predicting Light Reflection from Complex Surfaces*, SIGGRAPH 92, Computer Graphics, v26, n2, p255-264, 1992

[37] T.Whitted, *An Illumination Model for Shaded Display*, Com. of the ACM, v23, n6, p343-349, 1980

[38] G.Wyszecky, W.Stiles, *Color Science*, Willey, 1967

[39] S.Yokoi, J.Toriwaki, *Realistic Expression of Solids with Feeling of Materials*, JARECT, v.18, 1988

[40] T.Yamana, Y.Suenaga, *Hair Image Generation using Anisotropic Reflection*, IMAGE'COM 90 (Bordeaux, France), p189-193, 1990

An Inexpensive BRDF Model for Physically-based Rendering

Christophe Schlick

LaBRI [1]

351 cours de la Libération, 33405 Talence, FRANCE

`schlick@labri.u-bordeaux.fr`

Abstract : *A new BRDF model is presented which can be viewed as an kind of intermediary model between empirism and theory. Main results of physics are observed (energy conservation, reciprocity rule, microfacet theory) and numerous phenomena involved in light reflection are accounted for, in a physically plausible way (incoherent and coherent reflection, spectrum modifications, anisotropy, self-shadowing, multiple surface and subsurface reflection, differences between homogeneous and heterogeneous materials). The model has been especially intended for computer graphics applications and therefore includes two main features :* simplicity *(a small number of intuitively understandable parameters controls the model) and* efficiency *(the formulation provides adequation to Monte-Carlo rendering techniques and/or hardware implementations).*

Keywords : Physically-Based Rendering, Bidirectional Reflectance Distribution Function, Optimization

1 Introduction

Computation of a reflectance model is the heart of every rendering method because it provides the illumination of objects in the scene, and therefore the color of pixels in the image. Reflectance models currently in use can be divided in two main families : empirical models and theoretical ones. Empirical models [12, 3, 6] are computationally efficient but are lacking of physical validity (energy conservation law, for instance) and thus do not provide plausible values of energy or intensity. In fact, they are generally only used to create bright spots on surfaces in order to add some tridimensional information which helps to understand the image. Therefore they are limited to applications where good-looking pictures are sufficient (computer generated imagery for movies or commercials). On the other side, theoretical models [5, 9, 20] involve higher computational costs but provide quantitative values that have shown to be in good adequation with experimental data. Therefore they are well adapted to applications for which physically-based rendering is essential (simulation for lighting industry or architecture).

This paper proposes a kind of intermediary model between empirical and theoretical models. In Section 2, simple notations are presented and used to reformulate several existing reflectance models. Section 3 focuses on some unsatisfactory points that can be found in these models. In Section 4, a general purpose optimization technique is detailled and several low-cost alternatives to expensive terms involved in existing models are proposed. Finally, Section 5 presents the new reflectance model, which uses that optimization technique to combine several interesting features of previously disjointed work into an inexpensive formulation well-suited to computer graphics.

2 Background

The interaction of light with a surface is described by relating incoming and outcoming radiances at a given point P on the surface. This expression usually involves a function $R_\lambda(P, V, V')$ called *bidirectional reflectance distribution function* (BRDF, for short) :

$$L_\lambda(P, V) = \int_{V' \in \mathcal{V}} R_\lambda(P, V, V') \ L_\lambda(P, -V') \ (N \cdot V') \ dV' \tag{1}$$

- $L_\lambda(P, V)$ is the reflected radiance leaving point P in direction V

- $L_\lambda(P, -V')$ is the incident radiance reaching point P from direction $-V'$

- $R_\lambda(P, V, V')$ is the BRDF of the surface at point P between directions V and V'

- $\mathcal{V}$ is the set of possible directions for incident light (*ie* hemisphere above the surface)

- dV' is a differential solid angle surrounding direction V'

- V, V' and N are unit vectors so $(N \cdot V')$ is the cosine of the angle between N and V'

The reflected radiance is the integral, for all possible directions, of incident radiances scaled by the BRDF and the projected solid angle. It should be noted that such a formulation is well adapted to rendering algorithms (radiosity, path tracing, two pass methods) that perform effectively a true integration and for which the solid angle information is available. Other rendering algorithms (direct illumination methods, ray tracing) consider only a discrete sum of light contributions, and therefore are unable to provide close simulation of real phenomena as required by physically-based rendering.

Equation 1 is a monochromatic equation expressed for a given wavelength λ. In the present paper, we use the following notation convention : every term that is function of the wavelength will be subscripted by λ. Such a term has to be defined and/or computed, theoretically for every wavelength of the visible spectrum, and practically for a given number of samples (three in trichromatic models, up to twenty in spectral models [6]).

The BRDF has got two important properties that result directly from physics of light [2]. First, due to the *Helmholtz Reciprocity Rule*, R_λ is symmetric relative to V and V' :

$$\forall\ V \in \mathcal{V} \quad \forall\ V' \in \mathcal{V} \quad R_\lambda(P, V, V') = R_\lambda(P, V', V) \tag{2}$$

Second, due to the *Energy Conservation Law*, R_λ has to fulfill the normalization condition :

$$\forall\ V \in \mathcal{V} \quad \int_{V' \in \mathcal{V}} R_\lambda(P, V, V')\ (N \cdot V')\ dV' \le 1 \tag{3}$$

V	$=$	Outcoming direction of light
V'	$=$	Incoming direction of light
N	$=$	Surface normal vector
T	$=$	Surface tangent vector
H	$=$	Bisector vector of V and V'
$\overline{H}$	$=$	Projection of $H \perp N$

t	$=$	$(H \cdot N)$	α	$=$	$\cos^{-1} t$
u	$=$	$(H \cdot V)$	β	$=$	$\cos^{-1} u$
v	$=$	$(V \cdot N)$	θ	$=$	$\cos^{-1} v$
v'	$=$	$(V' \cdot N)$	θ'	$=$	$\cos^{-1} v'$
w	$=$	$(T \cdot \overline{H})$	φ	$=$	$\cos^{-1} w$
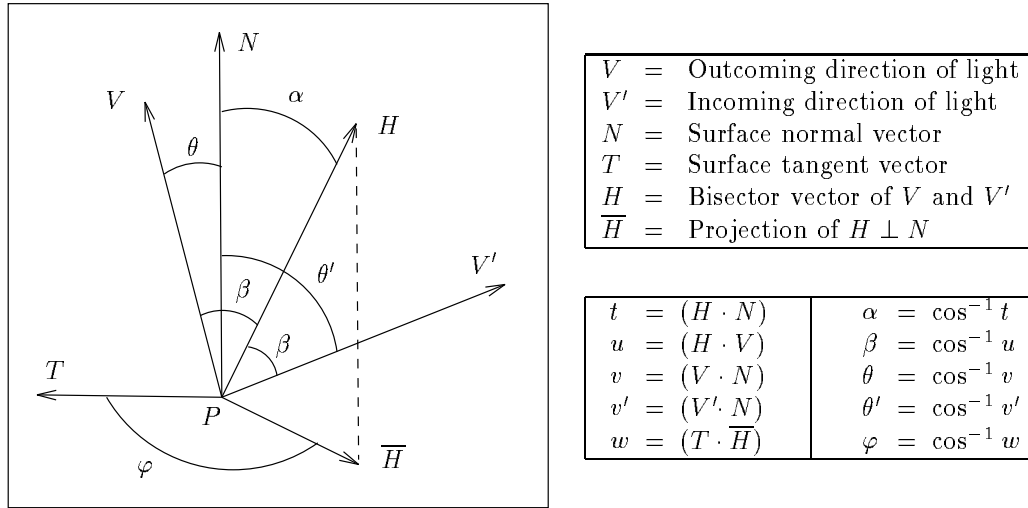
Figure 1 : Angles and vectors for definition of BRDF

According to the shape of the BRDF, two kinds of surfaces are traditionally distinguished :

Diffuse surfaces : The light is reflected in every direction. The limit case — *perfectly diffuse surfaces* or *Lambert surfaces* — is obtained when the BRDF becomes a constant function (*ie* the light is equally reflected in every direction).

Specular surfaces : The light is reflected only in a small area around the mirror direction. The limit case — *perfectly specular surfaces* or *Fresnel surfaces* — is obtained when the BRDF becomes a Dirac function (*ie* the light is reflected in one single direction).

Figure 1 presents the notations that will be used to formulate BRDF models throughout the paper. A complete review of all the models that have been proposed in the litterature is out of scope here (the interested reader may refer to [17] where such a survey has been done), we will focus especially on the three theoretical models that have been used to define our new model.

2.1 The Cook-Torrance Model

The first theoretical reflectance model has been introduced in the computer graphics field by Cook & Torrance [5] using work previously done in physics by Beckmann & Spizzichino [2] and Torrance & Sparrow [19] about the reflection of electromagnetic waves on rough surfaces. In that model, a surface is supposed to be composed of so-called *microfacets* which are small smooth planar elements. Only microfacets for which the normal vector is in direction H (see Figure 1) contribute to the reflection between V and V'. The BRDF depends on five different angles and is expressed as a linear combination of a diffuse reflector and a specular one :

$$R_\lambda(\alpha, \beta, \theta, \theta', \varphi) = \frac{d}{\pi} C_\lambda + \frac{s}{4\pi v v'} F_\lambda(\beta) \ G(\theta, \theta') \ D(\alpha, \varphi) \quad \text{with} \quad d + s = 1 \tag{4}$$

- d (resp. s) $\in [0, 1]$ is the ratio of the surface behaving as a diffuse (resp. specular) reflector

- $C_\lambda \in [0, 1]$ is the ratio of light at wavelength λ, reflected by the diffuse reflector (a complete set of C_λ, one for each wavelength sample, defines the color of the diffuse reflector)

- $D(\alpha, \varphi) \in [0, \infty)$ is the microfacets slope distribution function which defines the fraction of the facets that are oriented in direction H

- $F_\lambda(\beta) \in [0, 1]$ is the Fresnel factor which defines the ratio of light at wavelength λ, reflected by each microfacet (a complete set of F_λ defines the color of the specular reflector)

- $G(\theta, \theta') \in [0, 1]$ is the geometrical attenuation coefficient which expresses the ratio of light that is not self-obstructed by the surface

One important point to notice is that Equation 4 is only valid when it fulfills the normalization condition (Equation 3). Therefore, it implies a condition on the slope distribution function [2] :

$$\int_0^{\pi/2} \int_0^{2\pi} D(\alpha, \varphi) \ \cos \alpha \sin \alpha \ d\alpha \ d\varphi = \pi \tag{5}$$

If we assume an isotropic behaviour of the surface (*ie* the BRDF is invariant by rotation around the normal vector) then Equation 5 becomes :

$$\int_0^{\pi/2} D(\alpha) \ 2 \cos \alpha \sin \alpha \ d\alpha = 1 \tag{6}$$

Moreover, if we express the angular dependence of the three factors (D, F and G) in terms of their cosines, Equation 6 may be rewritten as :

$$\int_0^1 2t \ D(t) \ dt = 1 \tag{7}$$

and finally Equation 5 becomes :

$$R_\lambda(t, u, v, v') = \frac{d}{\pi} D_\lambda + \frac{s}{4\pi v v'} F_\lambda(u) \ G(v, v') \ D(t) \tag{8}$$

2.2 The He-Torrance-Sillion-Greenberg Model

A more complete model for BRDF has been proposed by He *et al.* [9] which accounts comprehensively for every physical phenomena (polarization, diffraction, interference, conductivity, smaller roughness for grazing rays) involved in light reflection on rough surfaces. When restricted to unpolarized light (the usual case in computer graphics), it can be rewritten in a form that mimics Equation 8 :

$$R_\lambda(t, u, v, v') = \frac{d}{\pi} C_\lambda + \frac{s}{4\pi v v'} F_\lambda(u) \ G(v, v') \ D(t) + \frac{s}{v' \ dV'} F_\lambda(v) \ G(v, v') \ A(v, v') \ \Delta \qquad (9)$$

Such a formulation hides the complexity of the model (for instance, the expression of the slope distribution function $D(t)$ has discouraged many potential users to implement it) but lets clearly appear one of the main difference with the Cook-Torrance model : there is an additional term in the linear combination which corresponds to coherent reflection on the mean plane of the surface (*ie* not the microfacets). Coherent reflection has been used for years in computer graphics (it is the fundamental principle of recursive ray tracing) but only for perfectly specular surfaces; He *et al.* have shown that this term exists also for non-smooth surfaces, though it decreases rapidly when the roughness increases, due to the presence of the roughness attenuation coefficient $A(v, v')$.

2.3 The Ward Model

Relatively few reflectance models have been proposed for anisotropic reflection (*ie* the BRDF is function of angle φ and thus of its cosine w) and usually they involve high computational costs [10, 4, 13, 21]. Ward has presented a simple model [20] in which the rotational symmetry of isotropic BRDF is replaced by an elliptical asymmetry of varying excentricity :

$$R_\lambda(t, v, v', w) = \frac{d}{\pi} C_\lambda + \frac{s}{4\pi\sqrt{vv'}} C'_\lambda \ D(t, w) \quad \text{with} \quad D(t, w) = \frac{1}{mn} \ e^{\frac{t^2 - 1}{t^2}(\frac{w^2}{m^2} + \frac{1 - w^2}{n^2})} \qquad (10)$$

Intuitively, the model considers scratches on the surface (oriented along the tangent vector of the local frame) leading to different roughnesses (defined by m and n) when considering directions parallel or perpendicular to the scratches : the more m and n are different, the more anisotropy is created.

3 Unsatisfactory Points

By examining existing reflectance models, one can find several points that appear somewhat unsatisfactory. For instance, the BRDF is formulated as a linear combination with constant weights between a diffuse part and a specular one. The justification usually given by the authors is that, for a large class of materials, diffuse and specular components come from different physical phenomena, and thus they may have different colors. One classical example is a plastic surface on which light can be reflected either by the uncolored substrat in a coherent way (*ie* surface reflection is specular) or by the colored pigments beneath the surface in an incoherent way (*ie* subsurface reflection is diffuse) [5].

But, as noticed by Shirley [15], such a linear combination with constant weights is incorrect because proportions of diffuse and specular components are usually not constant but function of the incident angle. Taking the example of a varnished wood floor (see Figure 2), one can see that, according to the Fresnel law, for large incident angles most light is reflected specularly by the varnish, whereas for small incident angles, most light penetrates the varnish before beeing reflected diffusely by the wood.
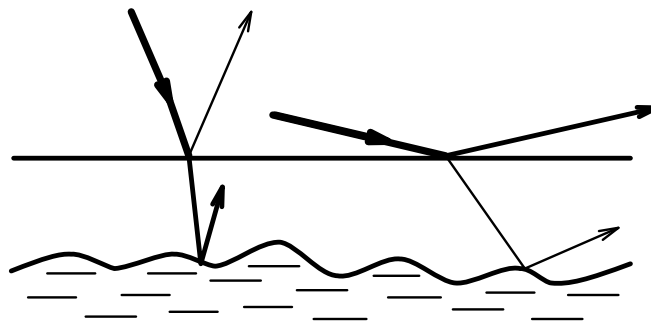
Figure 2 : Influence of the incident angle on surface and subsurface reflection

Beside these *heterogeneous* materials, there are many *homogeneous* ones, for which the diffuse-specular distinction is unnecessary. For such materials (metals, for instance) there is rather a kind of continuum between perfect diffuse and perfect specular behaviours (see Figure 3) according to the roughness of the surface. Therefore a linear combination with constant weights is inadequate again.

Figure 3 : Continuum between diffuse and specular for surface reflection

Another weak point in existing models appears when light reaches or leaves a rough surface where self-obstruction (from one microfacet to another) occurs. Usually, a geometrical attenuation coefficient (G in Equation 8) is used as a multiplicative factor to express the ratio of light which is not subject to that obstruction. But in real life, the remainder of the light (*ie* $1 - G$) is reflected in other directions and not simply blocked. Currently, none of the existing reflectance models does correctly account for that reemission of self-obstructed light.

The last — and perhaps the most — unsatisfactory point is about the accuracy/cost ratio. As said, empirical reflectance models are inexpensive but their lack of physical validity prevents their use in any physically-based rendering system. On the other hand, theoretical models are physically accurate but involve complex mathematical expressions which are computationally expensive and preclude hardware implementations. Moreover, when including such a reflectance model in an image synthesis software, the error generated by other stages of the rendering pipeline (tessellation for geometrical modeling, spectral sampling for optical modeling, directional sampling for global illumination, interpolation at almost every steps) does usually totally cancel the benefit of greater accuracy. In other words, there is no need to compute BRDF at a precision of 0.1%, if directional sampling is only done at 5% and spectral sampling at 15%. A possible solution could be to replace expensive formulas in theoretical reflectance models by some well-chosen low-cost alternative functions. In Section 4, we present a new technique that enables to find such approximations.

4   Optimization by Rational Fraction Approximation

One classical optimization technique (which has been applied several times in computer graphics) to speed up an algorithm that involves the computation of a complex function is to store many sample values of the function in a table, and compute missing values by interpolation (usually linear or cubic). Implementing a whole theoretical reflectance model which such a technique would require numerous tables (in order to account for various surface properties and illuminating conditions) which means high memory costs and difficulties to switch to hardware implementations [10, 4, 21].

A well-known technique in mathematics is to replace such a function by its Taylor expansion, giving a polynomial that can be computed with a handful of multiplications and additions by Horner's rule. Unfortunately a Taylor expansion is only valid nearby the origin ; therefore the approximation is usually only accurate in such a neighbourhood. To overcome this limitation, a possible solution is to use piece-wise Taylor approximants. But creating large ranges of values where the approximation is accurate implies to use many pieces, for which continuity in their derivatives cannot always be insured.

Another classical technique (which exists since the beginning of the century and has been applied to numerous scientific fields) is to use Padé approximants [1] in which a rational fraction is generated according to the Taylor expansion of the function. Compared to pure polynomial approximations, Padé approximants have usually a much better accuracy when leaving the neighbourhood of the origin. Piece-wise Padé approximants have also been proposed but rarely used in practice, because insuring continuity of the derivatives becomes almost impossible.

The previous approximation methods, which deal all with Taylor expansions, have got two strong limitations. First, the Taylor expansion of the function has to be known, this is not always possible even with numerical techniques. Second, specific properties of the function are generally not preserved. For instance, if we want to approximate a statistical distribution function f on a range $[a, b]$ (which by definition has to fulfill $\int_a^b f(t)dt = 1$) by one of the previous approximation techniques, the approximant has virtually no chance to fulfill that condition too, giving something that is mathematically incorrect.

4.1 Principle

We propose here another method that we simply call *rational fraction approximation*. This method differs from the Padé approximation technique by the fact that we do not use Taylor expansions to find the coefficients of the numerator and denominator polynomials. The idea is to study the function that we want to approximate, in order to find what we call *kernel conditions*.

A kernel condition can be any intrinsic characteristic of the function : value at a given point either of the function or of one of its derivative, integral or differential equation it obeys to... The detection and the choice of kernel properties can be done in several ways, by using its mathematical definition, by picking some of its remarkable values or even by plotting the function and examinating the graph. In fact, the only generic method for finding kernel conditions is to answer the question : "For me what are the characteristics of the function that every approximation should fulfill ?"

Once the kernel conditions have been found, coefficients of the rational fraction are simply obtained by identifying the function and its approximation for each kernel condition. This gives a system of n equations and n unknowns where n is the number of conditions.

For instance, let's suppose that we want to approximate the function $f(x) = \sin x$ on the range $[0, \pi/2]$. By plotting the corresponding curve or by specific knowledge on the function, we can find at least four characteristics that appear to be essential for every approximation function :

$$f(0) = 0 \quad f'(0) = 1 \quad f(\pi/2) = 1 \quad f'(\pi/2) = 0$$

Because there are four kernel conditions, we search for a rational fraction containing four independent parameters (a, b, c and d). For instance :

$$\forall x \in [0, \pi/2] \quad \overline{f}(x) = \frac{x^2 + ax + b}{cx + d} \tag{11}$$

When we express the kernel conditions with that function[2] , it leads to a system of four equations :

$$\begin{cases} b = 0 & \text{(condition 1)} \\ a = d & \text{(condition 2)} \\ a + \pi/2 = c + 2a/\pi & \text{(condition 3)} \\ c\pi^2 + 4\pi a + 4a^2 = 0 & \text{(condition 4)} \end{cases}$$

giving finally the following approximation of $f(x) = \sin x$:

$$\forall x \in [0, \pi/2] \quad \overline{f}(x) = x\frac{u + x}{u + vx} \quad \text{with} \ \ u = -\pi^2/4 \ \ \text{and} \ \ v = \pi + u \tag{12}$$

Plotting the original function and the approximation (see Figure 4) enables to control visually the similarities between the curves. In order to test our approximations in a quantitative way, we have also developed a statistical test (evaluation of the function and its approximation for one million random values) which provides two measures : ε = relative error and γ = speed-up factor (in fact, several other testing processes have been developped, see [16] for more complete results). For instance, with our sine example, we have obtained $\varepsilon = 1.4\%$ and $\gamma = 230\%$

---

[2]Notice that the fraction $(ax + b)/(cx + d)$ does not have four independent parameters, because one of them can be eliminated by simultaneous division of the numerator and the denominator.
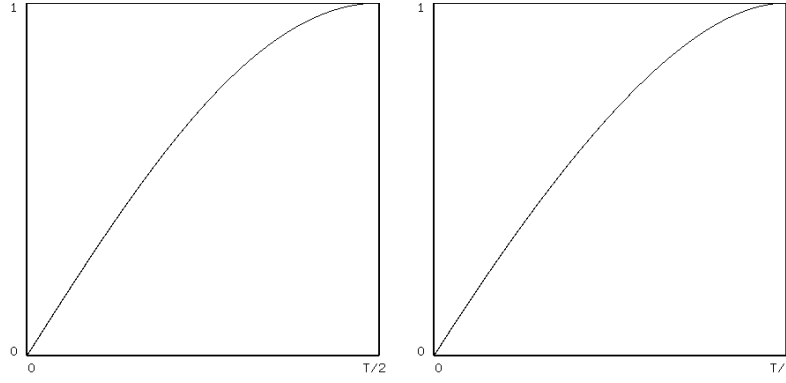
Figure 4 : $f(x) = \sin x$ where $x \in [0, \pi/2]$
LEFT : True function   RIGHT : Rational approximation

4.2 Approximation of the Fresnel Factor

The Fresnel factor $F_\lambda(u)$ expresses the reflection of light on the well-oriented microfacets (*ie* the facets whose normal vector is H). For a non-polarized electromagnetic wave, its formulation is [15] :

$$F_\lambda(u) = \frac{1}{2} \frac{(a-u)^2 + b^2}{(a+u)^2 + b^2} \left[\frac{(a+u-1/u)^2 + b^2}{(a-u+1/u)^2 + b^2} + 1 \right] \tag{13}$$

$$a^2 = \frac{1}{2} \left(\sqrt{(n_\lambda^2 - k_\lambda^2 + u^2 - 1)^2 + 4n_\lambda^2 k_\lambda^2} + n_\lambda^2 - k_\lambda^2 + u^2 - 1 \right)$$

$$b^2 = \frac{1}{2} \left(\sqrt{(n_\lambda^2 - k_\lambda^2 + u^2 - 1)^2 + 4n_\lambda^2 k_\lambda^2} - n_\lambda^2 + k_\lambda^2 - u^2 + 1 \right)$$

where n_λ is the ratio of the refraction indices above and below the surface and k_λ is the extinction coefficient of the surface. An interesting characteristic of the Fresnel factor is that $F_\lambda = 1$ at a grazing incidence ($\beta = \pi/2$ so $u = 0$) whatever the wavelength λ.

One difficulty that precludes a general use of F_λ in every rendering environment comes from the fact that n_λ and k_λ are seldom known. Some experimental values exist [11] but usually one can only find a single value $\overline{n}$ and $\overline{k}$ for a wavelength in the middle of the visible spectrum. On the other hand, a data which has been measured for thousands of materials is the spectral distribution f_λ of the Fresnel factor at normal incidence ($\beta = 0$ so $u = 1$). When $\overline{n}$, $\overline{k}$ and f_λ are the only known data, Cook & Torrance have proposed the following approximation [5] :

- Compute $\overline{F}(u)$ and $\overline{f} = \overline{F}(1)$ with $\overline{n}$ and $\overline{k}$, using Equation 13

- For each wavelength λ, compute $F_\lambda(u)$, using Equation 14

$$F_\lambda(u) = f_\lambda + (1 - f_\lambda) \frac{\overline{F}(u) - \overline{f}}{1 - \overline{f}} \tag{14}$$

This approximation not only solves the lack of experimental data, but also speeds-up the calculation because the complete expression of $F_\lambda(u)$ is evaluated only once, for an average n_λ and k_λ. But even so, the computation of the Fresnel factor remains expensive and further optimization should be possible.

By examining Figure 5, one can see that the shape of the curves does not vary very much according to the kind of material. The main difference is the value f_λ where the curve arrives at $u = 1$. Therefore, a step further could be to make $F_\lambda(u)$ only dependent on f_λ. By choosing the kernel conditions :

$$F_\lambda(0) = 1 \qquad F_\lambda(1) = f_\lambda \qquad F'_\lambda(1) = 0 \qquad F''_\lambda(1) = 0$$

we have found

$$F_\lambda(u) = f_\lambda + (1 - f_\lambda)(1 - u)^5 \tag{15}$$

which costs only 4 multiplications and 2 additions in an optimized implementation. Our statistical testing process shows that the approximation can be computed almost 32 times faster with less than 1% error :
$\varepsilon = 0.6\%$ and $\gamma = 3180\%$
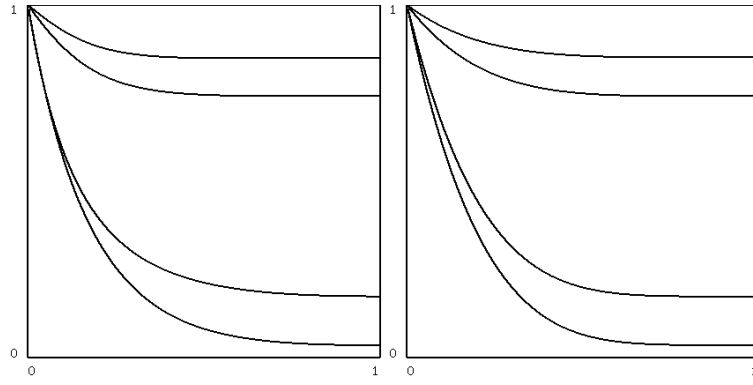
Figure 5 : $F_\lambda(s)$ by Fresnel for diamant, glass, copper and gold
LEFT : True function   RIGHT : Rational approximation

4.3 Approximation of Geometrical Attenuation Coefficient

Several expressions for the geometrical attenuation coefficient $G(v, v')$ have been proposed in physics [19, 18, 14]. In their original paper, Cook & Torrance used the formulation derived in [19] :

$$G(t, u, v, v') = \min\left[1, 2\frac{tv}{u}, 2\frac{tv'}{u}\right] \tag{16}$$

This expression results from coarse approximations about the surface geometry and therefore does not meet experimental results : its first derivative is not continuous, it is not invariant by rotation around the normal vector and it is independent of the surface roughness. The formulation proposed by Smith [18] (introduced in the computer graphics field by [9]) is not subjected to these restrictions and has been experimentally validated. Moreover, it depends only on v and v', and is separable in v and v' :

$$G(v, v') = G(v) \ G(v') \tag{17}$$

After several equivalences, the original expression of $G(v)$ can be written more compactly :

$$G(v) = \frac{g}{g + 1} \quad \text{with} \quad g = \sqrt{h\pi} \ (2 - \mathrm{erfc}\sqrt{h}) \quad \text{and} \quad h = \frac{v^2}{2m^2(1 - v^2)} \tag{18}$$

where m is the root mean square (RMS) slope of the microfacets (theoretically $m \in [0, \infty)$ but in fact, it almost never exceeds 0.5 for real surfaces). Despite its complicated form, the shape of the function (see Figure 6) is quite simple. To characterize it, we choose the following kernel conditions :

$$G(0) = 0 \qquad G(1) = 1 \qquad G'(0) = \sqrt{\frac{\pi}{2m^2}}$$

that leads to a very simple expression :

$$G(v) = \frac{v}{v - kv + k} \quad \text{with} \quad k = \sqrt{\frac{2m^2}{\pi}} \tag{19}$$

By precomputing k and $1 - k$, $G(v)$ needs only 1 division, 1 multiplication and 1 addition. Except for a small neighbourhood of $v = 1$, the two curves on Figure 6 are very similar. That visual feeling is confirmed by the testing process : $\varepsilon = 1.8\%$ and $\gamma = 2870\%$
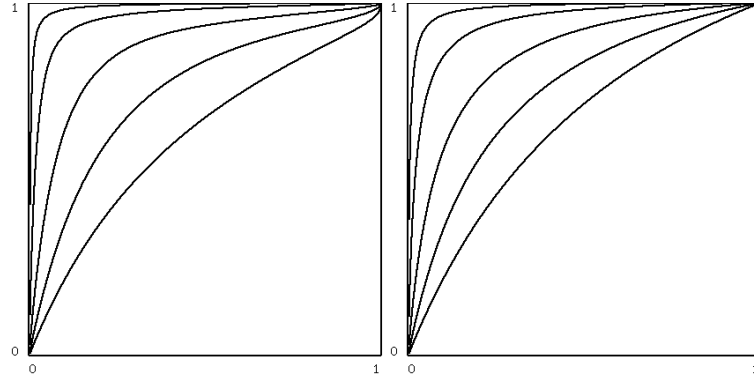
Figure 6 : $G(v)$ by Smith for $m = 0.01, 0.05, 0.1, 0.25, 0.5$
LEFT : True function   RIGHT : Rational approximation

4.4   Approximation of Slope Distribution Function

Among the different formulations of the slope distribution function that exist in the literature [2, 19, 3], only the one proposed by Beckmann fulfills Equation 7. Moreover, compared to others, this formulation depends only on the RMS slope m of the microfacets and does not introduce any arbitrary constant :

$$D(t) = \frac{1}{m^2 t^4} \; e^{\frac{t^2-1}{m^2 t^2}} \tag{20}$$

When the surface is rough (large values for m), orientations of microfacets are very dispersed. When the surface is smoother (small values for m) microfacet normals H come closer to the average normal N. And for perfectly smooth surfaces (m is null) $D(t)$ becomes a Dirac function (see Figure 7).

The normalization condition (Equation 7) is an obvious kernel condition for $D(t)$. Another important characteristic (already noticed by Beckmann) is that $D(t)$ is almost null for $t < 1-m$. And finally, $D(1)$ gives a last kernel condition :

$$\forall t \in [0, 1-m] \;\; D(t) = 0 \qquad D(1) = \frac{1}{m^2} \qquad \int_0^1 D(t)\, 2t\, dt = 1$$

Having an integral equation as a kernel condition complicates somewhat the process of finding a good approximation. A solution that works well for many cases is to search a rational fraction having a u'/u^2 form, for which an analytic integration can be done. Therefore, we propose :

$$\forall t \in [1-m, 1] \quad D(t) = \frac{m^3 x}{t\,(mx^2 - x^2 + m^2)^2} \quad \text{with} \quad x = t + m - 1 \tag{21}$$

An optimized implementation of $D(t)$ costs only 1 division, 4 multiplications and 2 additions. The curves drawn in polar coordinates are shown on Figure 7. The quantitative results are somewhat less good than for the two other approximations but still interesting : $\varepsilon = 2.7\%$ and $\gamma = 1650\%$
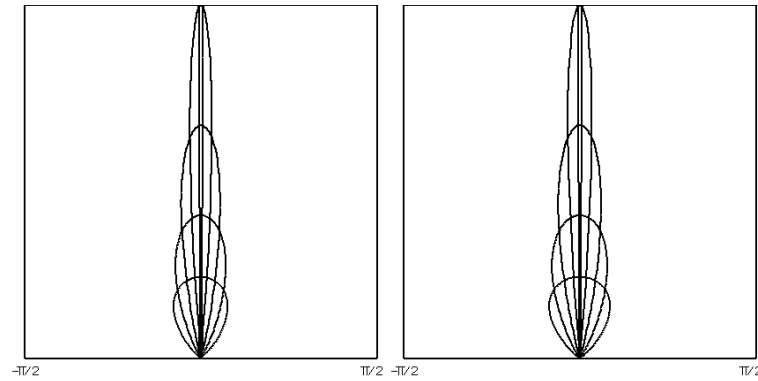
Figure 7 : $D(t)$ by Beckmann for $m = 0.01, 0.05, 0.1, 0.25, 0.5$
LEFT : True function   RIGHT : Rational approximation

When collecting the results of Equations 15, 19 and 21, one gets a kind of approximated/optimized Cook-Torrance model (more than 20 times faster with less than 3% error[3] compared to an implementation using the true formulas). Because it uses only basic arithmetic operations $(+ - */)$ the approximated model is well-suited to hardware implementation and represents a possible answer to the accuracy/cost ratio inadequacy of current BRDF models (see Section 3).

5   A New BRDF Model

The rational fraction approximation scheme enables to speed-up the computation of reflectance models but does not provide a solution for the other unsatisfactory points discussed in Section 3. With regards to that discussion, an appealing BRDF model should include the following features :

- Main results of physics *(Energy Conservation Law, Helmholtz Reciprocity Rule, Microfacet Theory)* should be fulfilled to enable physically-based rendering

- A continuum between Lambert and Fresnel surfaces should be provided

- A distinction between homogeneous and heterogeneous materials should be made

- Both isotropic and anisotropic behaviours should be accounted for

- Only a small number of simple and meaningful parameters should control the model

- Only expressions with low computational costs should be used

A new model which includes all these features is presented here and can be viewed as an intermediary model between empirism and theory. Before giving the formulation of the BRDF, let's examine how we modelize optical properties of surfaces. In fact, two different kinds of materials SINGLE/DOUBLE are distinguished in opposition to the classical diffuse/specular separation[4] :

- SINGLE : Materials having homogeneous optical properties (metal, glass, paper, cotton)

- DOUBLE : Materials having heterogeneous optical properties (plastic, skin, stratified or varnished or painted surfaces) usually composed of a more or less translucide layer over an opaque one, each of them being SINGLE materials.

We propose to characterize a SINGLE material by a set of parameters (C_λ, r, p) and a DOUBLE material by two sets (C_λ, r, p) and (C'_λ, r', p'), one for each layer :

- $C_\lambda \in [0, 1]$: Reflection factor at wavelength λ

- $r \in [0, 1]$: Roughness factor ($r = 0$: perfect specular, $r = 1$: perfect diffuse)

- $p \in [0, 1]$: Isotropy factor ($p = 0$: perfect anisotropic, $p = 1$: perfect isotropic)

The choice of these parameters was motivated mainly by two of their characteristics. First, the role of every parameter can be understood intuitively and therefore easily defined by a non-specialist user. Second, the parameters can also be assigned according to experimental data [11]. Indeed, C_λ can be viewed as the reflectivity at normal incidence f_λ (see Equation 14), r can be related to the RMS slope m of the surface (see Equation 20), and p is in fact the ratio of the RMS slopes m/n between the scratch ($\varphi = 0$) and the ortho-scratch ($\varphi = \pi/2$) directions for an anisotropic surface (see Equation 10).

---

[3] In fact, it is much more precise than the original Cook-Torrance implementation in which Equation 16 is used for G, giving an error of $\varepsilon = 53\%$ with our statistical test.

[4] This idea of a layered surface model has appeared several times in physics (but has never been completely investigated) as well as in a very recent paper by Hanrahan & Krueger in computer graphics [8].

5.1 Formulation

When only geometrical optics is involved (an hypothesis made by almost every rendering technique), the spectral and the directional behaviours of the BRDF can be separated (*ie* rays are reflected in the same direction, whatever their wavelength) in two multiplicative factors S_λ and D. According to the kind of material, we propose the following expression :

$$\begin{cases} \text{SINGLE} & : R_\lambda(t, u, v, v', w) = S_\lambda(u) \ D(t, v, v', w) \\[2mm] \text{DOUBLE} & : R_\lambda(t, u, v, v', w) = S_\lambda(u) \ D(t, v, v', w) + [1 - S_\lambda(u)] \ S'_\lambda(u) \ D'(t, v, v', w) \end{cases} \tag{22}$$

Spectral factor

The simplest expression for the spectral factor is to consider it as a constant function :

$$S_\lambda(u) = C_\lambda \tag{23}$$

But in fact, $S_\lambda(u)$ depends on the incident angle and should obey to the (approximated) Fresnel law :

$$S_\lambda(u) = C_\lambda + (1 - C_\lambda) \ (1 - u)^5 \tag{24}$$

Directional factor

Using the microfacet model and following a similar derivation as [2], one can imagine a straigthforward formulation for the directional factor (extended to anisotropy), in which the dependence on the zenith angle α and on the azimuth angle φ can be separated and expressed by two factors $Z(t)$ and $A(w)$:

$$D(t, v, v', w) = \frac{1}{4\pi v v'} \ Z(t) \ A(w) \tag{25}$$

$D(t, v, v', w)$ in only valid when the product $Z(t)A(w)$ obeys to Equation 5. If we suppose an identical anisotropic behaviour in all four quadrants around the normal vector, Equation 5 can be rewritten as :

$$\int_0^{\pi/2} Z(\alpha) \ \cos \alpha \ \sin \alpha \ d\alpha = \frac{1}{2} \qquad \text{and} \qquad \int_0^{2\pi} A(\varphi) \ d\varphi = 4 \int_0^{\pi/2} A(\varphi) \ d\varphi = 2\pi \tag{26}$$

and thus

$$\int_0^1 2t \ Z(t) \ dt = 1 \qquad \text{and} \qquad \int_0^1 \frac{1}{\sqrt{1 - w^2}} \ A(w) \ dw = \frac{\pi}{2} \tag{27}$$

Using Equation 27 as kernel conditions, as well as other characteristics, we have found simple expressions for Z and A, which are in fact polar equations of ellipses (the former with the pole on the focus, the latter with the pole in the middle) :

$$Z(t) = \frac{r}{(1 + rt^2 - t^2)^2} \qquad \text{and} \qquad A(w) = \sqrt{\frac{p}{p^2 - p^2 w^2 + w^2}} \tag{28}$$

When looking at the resulting curves (see Figure 8), one can notice that when $r = 1$, $Z(t)$ is a constant function (perfect diffuse) and when $r = 0$, $Z(t)$ becomes a Dirac function (perfect specular). The same remark can be made for $A(w)$ which varies continuously between a constant function when $p = 1$ (perfect isotropy) and a Dirac function when $p = 0$ (perfect anisotropy).

Adequation to Monte-Carlo Techniques

One main feature of the expressions for $Z(t)$ and $A(w)$ is that they are well-suited to Monte-Carlo rendering methods [15, 20]. Indeed, a usual technique for a Monte-Carlo process is to generate a stochastic *importance sampling* to improve the convergence [7]. Such an importance sampling for α and φ (limited to the first azimuthal quadrant but easily generalized by duplication) can be simply obtained from two uniform random variables $(a, b) \in [0, 1]^2$ with :

$$\alpha = \text{acos} \ \sqrt{\frac{a}{r - ar + a}} \quad \text{and} \quad \varphi = \frac{\pi}{2} \ \sqrt{\frac{p^2 b^2}{1 - b^2 + b^2 p^2}} \tag{29}$$

11
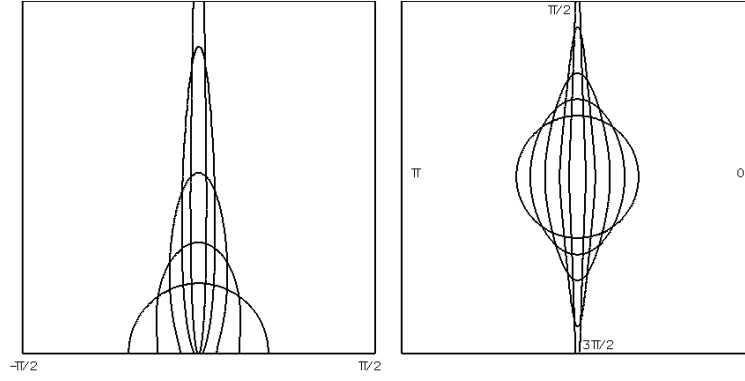
Figure 8 : Directional factor in logarithmic polar coordinates
LEFT : Zenith angle dependence $Z(t)$ for $r = 0.01, 0.05, 0.2, 0.5, 1.0$
RIGHT : Azimuth angle dependence $A(w)$ for $p = 0.01, 0.05, 0.2, 0.5, 1.0$

As in [18], self-obstruction can be included by a geometrical attenuation factor $G(v)G(v')$ where $G(v)$ (resp. $G(v')$) expresses the ratio of reflected (resp. incident) non-obstructed light. But, as discussed in Section 3, we want to account for reemission of self-obstructed light (*ie* $1-G$). Because of the stochastic orientation of microfacets, the direction of light after several reflections is essentially random, therefore it seems logical to reemit the light with $Z(t) = A(w) = 1$:

$$D(t, v, v', w) = \frac{G(v)G(v')}{4\pi vv'} \; Z(t) \; A(w) \quad + \quad \frac{1 - G(v)G(v')}{4\pi vv'} \tag{30}$$

As with the Fresnel factor, the approximated formulation of the Smith factor is used :

$$G(v) = \frac{v}{r - rv + v} \qquad \text{and} \qquad G(v') = \frac{v'}{r - rv' + v'} \tag{31}$$

Diffuse-Specular Continuum

Due to the presence of v and v' on the denominator, Equation 25 does not provide a complete transition from perfect diffuse to perfect specular. A classical result in physics [2, 8] says that there exists no microfacets configuration (even with a constant slope distribution function) that provides a Lambert reflector. Similarly, because coherent reflection is not considered, Equation 25 cannot provide a Fresnel reflector (even with a Dirac slope distribution function). Again, we propose a kind of empirical/theoretical solution, inspired by Equation 9, to create a continuum between these limit situations. The directional factor is defined as a sum of three different reflectors (Lambert Model, Anisotropic Microfacets Model, Fresnel Model), each of them having a specific weight (a, b, c) :

$$D(t, v, v', w) = \frac{a}{\pi} \; + \; \frac{b}{4\pi vv'} \; B(t, v, v', w) \; + \; \frac{c}{v' \; dV'} \; \Delta \quad \text{with} \quad a + b + c = 1 \tag{32}$$

where $B(t, v, v', w)$ is the directional factor of the anisotropic microfacets model given either by Equation 25 or 30 and Δ is a Dirac function (equal to 1 in dV' and 0 otherwise). The weights (a, b, c) could be specified directly by the user, but it would represent three additional parameters per material. We propose rather a (physically plausible) automatic scheme which provides a quadratic interpolation between the three fundamental behaviours of a surface, according to the roughness factor.

$$\text{if} \; (r < 0.5) \; \text{then} \; \{b = 4r(1-r) \; ; \; a = 0 \; ; \; c = 1-b\} \; \text{else} \; \{b = 4r(1-r) \; ; \; c = 0 \; ; \; a = 1-b\} \tag{33}$$

To conclude, notice that combining together Equations 23,24,25,30 and 32 provides in fact six different BRDF models which can be tailored when specific accuracy and/or cost are desired.

5.2 Pictures

In order to show various illumination effects (incidence angles ranging from grazing to normal and varying either fast or slow) a simple test scene similar to [9] has been chosen. Every cylinder on the pictures

12

has been rendered individually at a 256x512 resolution using Monte-Carlo ray-tracing. Top row (resp. bottom row) of Picture 1 illustrates the continuum that is provided between diffuse (resp. isotropic) and specular (resp. anisotropic) reflection by taking four different values for r (resp. p) from 1 to 0.05. To achieve a better understanding of the behaviour of the new model, only direct illumination from a single light source put at the view point is shown. In order to exhibit anisotropy, the cylinder is made of brushed metal, having concentric circular scratches on its top and parallel horizontal scratches on its body. Top row of Picture 2 is similar to the bottom row of Picture 1, but indirect specular illumination from the surrounding environment is shown this time.

Finally, bottom row of Picture 2 illustrates the benefit of accounting for the Fresnel law in the spectral factor. The cylinders are made here of a DOUBLE material, having a transparent specular layer over an opaque (black for the two left cylinders, white for the two right ones) diffuse layer. When using Equation 23 for S_λ (first and third cylinder), the whole cylinder exhibits the same amount of diffuse and specular behaviours (linear combination with constant weights) which is both visually unnatural and physically incorrect. When using Equation 24 (second and fourth cylinders), different reflection behaviours are provided according to the incident angle as predicted by theory and confirmed experimentally [9].

6 Conclusion

A general purpose tool for approximating complicated mathematical expressions by simple rational fractions with low degree polynomials has been proposed. This tool has been used to propose a BRDF model for computer graphics including the following features :

- A distinction is made between materials with homogeneous properties and materials with heterogeneous properties (which are supposed to be composed of two homogeneous layers).

- A two dimensional continuum is insured both between perfect diffuse and perfect specular, and between perfect isotropy and perfect anisotropy.

- The Fresnel factor is introduced in a new way, expressing ratio of surface vs subsurface reflection.

- The obstruction factor is also introduced in a new way, expressing ratio of single vs multiple surface reflection.

- Very few parameters are used to define a material, each of them may be defined intuitively or related to experimental data usually available.

- A formulation of varying complexity is provided, allowing the user to tailor physical accuracy vs computational cost.

Acknowledgements

References

[1] G.A.Baker, P.G.Morris, *Padé approximants*, Encyclopedia of Mathematics, v13-14, Addison Wesley, 1981

[2] P.Beckmann, A.Spizzichino, *Scattering of Electromagnetic from Rough Surfaces*, Pergamon Press, 1963

[3] J.Blinn, *Models of Light Reflection for Computer Synthesized Pictures*, SIGGRAPH 77, Computer Graphics, v11, n4, p192-198, 1977

[4] B.Cabral, N.Max, R.Springmeyer, *Bidirectional Reflection Functions from Surface Bump Maps*, Proc. SIGGRAPH 87, Computer Graphics, v21, n4, p273-281, 1987

[5] R.Cook, K.Torrance, *A Reflectance Model for Computer Graphics*, Proc. SIGGRAPH 81, Computer Graphics, v15, n4, p307-316, 1981

[6] R.Hall, *Illumination and Color in Computer Generated Imagery*, Springer-Verlag, 1989

[7] J.M.Hammerley, D.C.Handscomb, *Monte-Carlo methods*, Wiley Editors, 1964

[8] P.Hanrahan, W.Krueger, *Reflection from Layered Surfaces due to Subsurface Scattering*, SIGGRAPH 93, Computer Graphics, v27, n3, p159-169, 1993

[9] X.He, K.Torrance, F.Sillion, D.Greenberg, *A Comprehensive Physical Model for Light Reflection*, SIGGRAPH 91, Computer Graphics, v25, n4, p187-196, 1991

[10] J.T.Kajiya, *Anisotropic Reflection Models*, SIGGRAPH 85, Computer Graphics, v19, n4, p15-22, 1985

[11] E.D.Palik, *Handbook of Optical Constants of Solids*, Academic Press, 1985

[12] B.T.Phong, *Illumination for Computer Generated Pictures*, Comm. of the ACM, v18, n6, p449-455, 1975

[13] P.Poulin, A.Fournier, *A Model for Anisotropic Reflection*, SIGGRAPH 90, Computer Graphics, v24, n4, p273-282, 1990

[14] M.Sancer, *Shadow-Corrected Electromagnetic Scattering from a Randomly Rough Surface*, IEEE Transactions on Antennas and Propagation, v17, n5, p577-585, 1969

[15] P.Shirley, *Physically Based Lighting Calculations for Computer Graphics*, PhD thesis, University of Illinois at Urbana-Champaign, November 1990

[16] C.Schlick, *Divers Eléments pour une Synthèse d'Images Réalistes*, PhD Thesis, Univ. Bordeaux I, 1992

[17] C.Schlick, *A Survey of Shading and Reflectance Models for Computer Graphics*, to appear in Computer Graphics Forum, 1994

[18] B.G.Smith, *Geometrical Shadowing of a Random Rough Surface*, IEEE Transactions on Antennas and Propagation, v15, n5, p668-671, 1967

[19] K.E.Torrance, E.M.Sparrow, *Theory for Off-Specular Reflection from Roughened Surfaces*, Journal of the Optical Society of America, v57, n9, p1105-1114, 1967

[20] G.J.Ward, *Measuring and Modeling Anisotropic Reflection*, SIGGRAPH 92, Computer Graphics, v26, n2, p265-272, 1992

[21] S.West, J.Arvo, K.Torrance, *Predicting Light Reflection from Complex Surfaces*, SIGGRAPH 92, Computer Graphics, v26, n2, p255-264, 1992
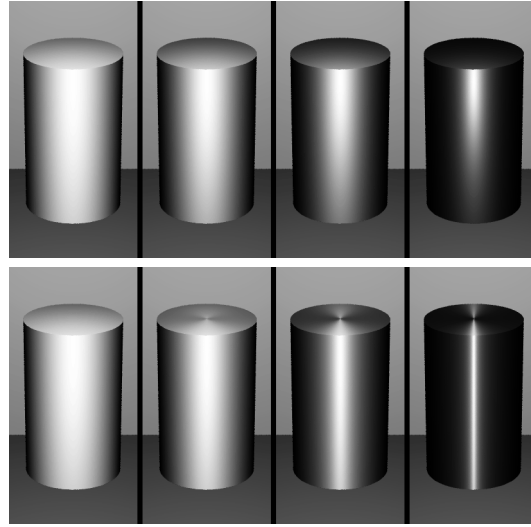
Picture 1: *Direct illumination from light source*
Top row : Continuum between diffuse and specular reflection
Bottom row : Continuum between isotropic and anisotropic reflection
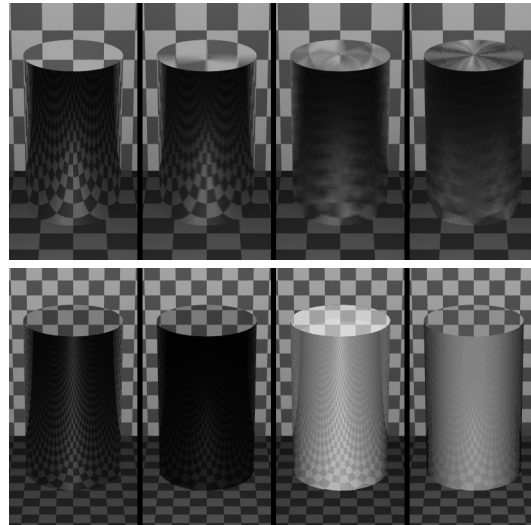
Picture 2: *Direct illumination from light source*
Top row : Continuum between diffuse and specular reflection
Bottom row : Continuum between isotropic and anisotropic reflection

Non-Linear Approximation of Reflectance Functions

Eric P. F. Lafortune   Sing-Choong Foo *   Kenneth E. Torrance   Donald P. Greenberg

Program of Computer Graphics
Cornell University †

Abstract

We introduce a new class of primitive functions with non-linear parameters for representing light reflectance functions. The functions are reciprocal, energy-conserving and expressive. They can capture important phenomena such as off-specular reflection, increasing reflectance and retro-reflection. We demonstrate this by fitting sums of primitive functions to a physically-based model and to actual measurements. The resulting representation is simple, compact and uniform. It can be applied efficiently in analytical and Monte Carlo computations.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism; I.3.3 [Computer Graphics]: Picture/Image Generation

Keywords: Reflectance function, BRDF representation

1 INTRODUCTION

The bidirectional reflectance distribution function (BRDF) of a material describes how light is scattered at its surface. It determines the appearance of objects in a scene, through direct illumination and global interreflection effects. Local reflectance *models* therefore play an essential role in local and global illumination simulations.

The diagram of Figure 1 illustrates the importance of a proper representation of reflectance data. The data originate from physical measurements, from scattering simulations on surfaces, from physically-based reflectance models, or from a set of empirical parameters input by the user. The representation should capture the necessary information in a way that allows it to be used in global illumination algorithms. Several factors contribute to the quality and usefulness of a representation: *accuracy*, *physical correctness* and *computational efficiency*.

First of all, the original data should be represented accurately enough to obtain physically faithful results. However, in practice, precise measurements are often not available. As a very precise representation cannot improve imprecise data, a simpler model that naturally interpolates the data may be preferable. It can also be

*Currently at Blue Sky Studios, Harrison, NY.

†580 Rhodes Hall, Ithaca, NY 14853, USA.
WWW: http://www.graphics.cornell.edu/
E-mail: eric@graphics.cornell.edu

useful to have a model with a limited set of parameters that are intuitive to use. Such parameters provide an easy way to control or to monitor the behavior of the model.

Secondly, the representation should be physically plausible. Reflectance functions are positive, reciprocal and energy-conserving [12]. Preferably, their representations should satisfy these constraints as well, because global illumination algorithms may rely on it.

Thirdly, for actual application in global illumination computations, the ideal model should be computationally efficient. It is usually an element in the larger context of an illumination simulation algorithm. One thus looks for a proper balance between accuracy, memory use and computation times of the various components. In the context of physically-based rendering, it makes little sense to use an overly precise and computationally expensive or memory-hungry model, when small subtleties are overwhelmed by global illumination effects, or when the simulation is relatively inaccurate.

At present, many reflectance models are not physically plausible. More precise physical models are often computationally expensive and geared toward specific types of surfaces. The most expressive models, such as spherical harmonics or wavelet representations, may require significant memory to obtain acceptable representations of even the simplest BRDFs. Yet we want to efficiently represent the relatively complex reflectance of common surfaces such as the wooden table shown in Figure 2. The pictures illustrate the varying specular and diffuse reflectance for different viewing angles.

In this paper we introduce a representation based on a new class of functions with non-linear parameters. While the representation does not offer the arbitrary accuracy that linear basis functions can achieve, it is expressive enough to fit complex reflectance behavior. Importantly, a single function can capture a complete BRDF over its entire domain of incident and exitant directions. It is therefore uniform and compact, as well as computationally efficient.

The next section gives a brief overview of previous work. Section 3 discusses the concept of non-linear approximation. We then present our specific primitive functions for modeling reflectance in Section 4. The qualitative properties of functions are discussed in Section 5, while quantitative fits to complex reflectance functions are presented in Section 6. Section 7 shows more results.

2 PREVIOUS WORK

Previous research focuses on various aspects of reflectance functions: their derivation, their measurement, and their representation. Torrance and Sparrow [22], and Cook and Torrance [3, 4] derived physical models based on geometrical optics, assuming specular V-grooves, and incorporating masking and self-shadowing effects. Their models correctly predict the off-specular reflection that they had previously measured [21]. Extending this work, He *et al.* [9] derived a model based on physical optics. The final representation of the model consists of an ideal diffuse component, a directional-diffuse component and a specular mirror component, which are all expressed by a set of analytic expressions. These can be evaluated numerically, albeit at a fair computational expense. Poulin
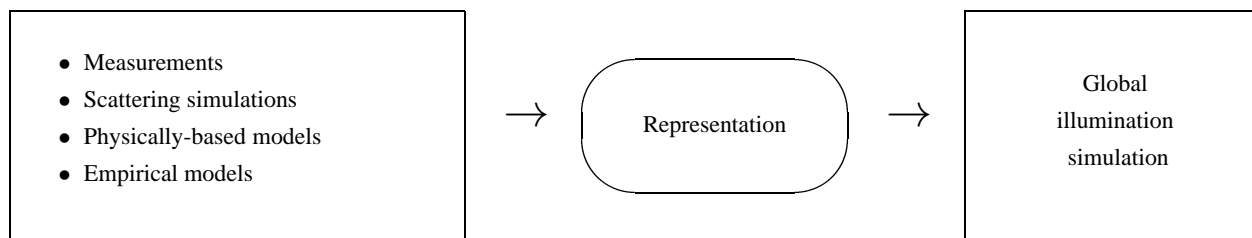
Figure 1: The representation of reflectance data constitutes the essential link between the origin of the raw data and their application in global illumination algorithms.
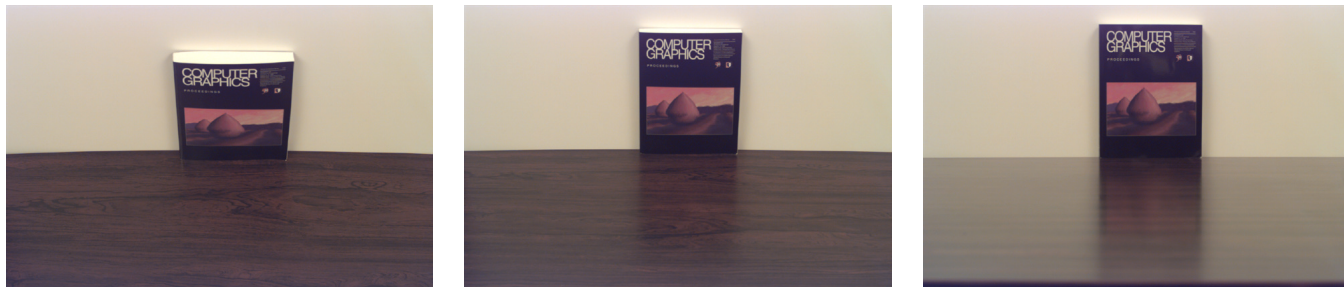

Figure 2: These pictures show a table exhibiting typical increasing specular reflection for increasingly grazing angles. At the same time the diffuse component, which results from subsurface scattering, fades out; the wood-grain texture and color disappear.

and Fournier [16] constructed a model assuming a surface consisting of microscopic cylinders. Oren and Nayar [14] derived a non-Lambertian diffuse model on the basis of diffuse micro-facets.

An alternative approach for deriving theoretical models is to perform a deterministic or Monte Carlo simulation on a surface model at a micro-scale. Kajiya [10] computed anisotropic reflectance functions based on the Kirchoff laws. He proposed storing the results in a table from which the values are linearly interpolated. Cabral [2] also stored reflectance simulation results in a table, but then represented them using spherical harmonics for a rendering step. Westin *et al.* [24] directly estimated the coefficients of the spherical harmonics. Hanrahan and Krueger [8] simulated subsurface scattering and stored the results in a uniform subdivision of the hemisphere. Gondek *et al.* [7] stored results in an adaptive subdivision of the geodesic sphere.

Empirical models, on the other hand, are not constructed from physical first principles. Instead, they capture reflectance effects using basis functions or other generic functions. The functions usually do not have any inherent physical meaning. Their physical validity stems from the theoretical or measured data to which they are fitted. For this purpose the functions should be expressive, while still being compact and efficient to use. Lambert's approximation, which assumes that the reflectance function of a diffuse surface is simply a constant, is widespread and sufficiently accurate for many applications. Phong [15] introduced one of the first more general shading models into computer graphics. Although it was not presented in the context of physically-based rendering, Lewis [12] showed how a physically plausible reflectance function can be derived from it. Ward [23] presented a model based on a Gaussian lobe, stressing its physical plausibility and ease of use. He successfully fitted the model to measurements of various surfaces and presented an equation to sample directions for it, which is important for Monte Carlo applications such as stochastic ray tracing. Schlick [17, 18] presented a model in which the important factors of previous physically-based models are approximated numerically, making it more convenient for use in Monte Carlo algorithms. Fournier [6] experimented with sums of separable functions for representing reflectance models, for application in radios-

ity algorithms. Schröder and Sweldens [19] represented reflectance functions using spherical wavelets. Koenderink *et al.* [11] recently introduced a compact representation based on Zernike polynomials.

Our work falls within the latter category of representations. We take a novel approach, using non-linear approximation with a sum of one or more appropriate functions. In the next section, we explain the general principle of non-linear approximation.

3 NON-LINEAR APPROXIMATION

Approximating functions with linear basis functions is well studied. Some common basis functions are Fourier bases, Chebychev polynomials and piece-wise linear functions. When approximating a function, the coefficients of the basis functions are determined by a set of linear equations. Non-linear approximation, for instance with rational functions or with Gaussians, is somewhat less known. In this approach, the parameters of the approximating functions are not necessarily linear with respect to the original function. They therefore generally have to be determined using non-linear optimization. Figure 3 shows an example of a peaked one-dimensional function that is approximated using the first four terms of a Fourier series and using two Gaussian functions. The Fourier terms vary in amplitude and in phase. Due to the relatively sharp peaks in the original function, their sum is only a rough approximation, which becomes negative at some point. The Gaussians are parametrized by a position, a standard deviation and a size. Their sum approximates the original function much better and remains positive over the interval. Obviously, this is not true in general, for all possible functions. However, the non-linear functions can be chosen such that they span a region of the function space that suits a specific application. Functions can then be approximated using a more compact representation. Furthermore, the parameters can be more intuitive when interpreting or controlling the model.

In the context of modeling BRDFs, more general representations are usually linear, e.g. spherical harmonics [2, 24], sums of separable bicubic polynomials [6] or wavelets [19]. Especially the former representations may require many coefficients, for instance
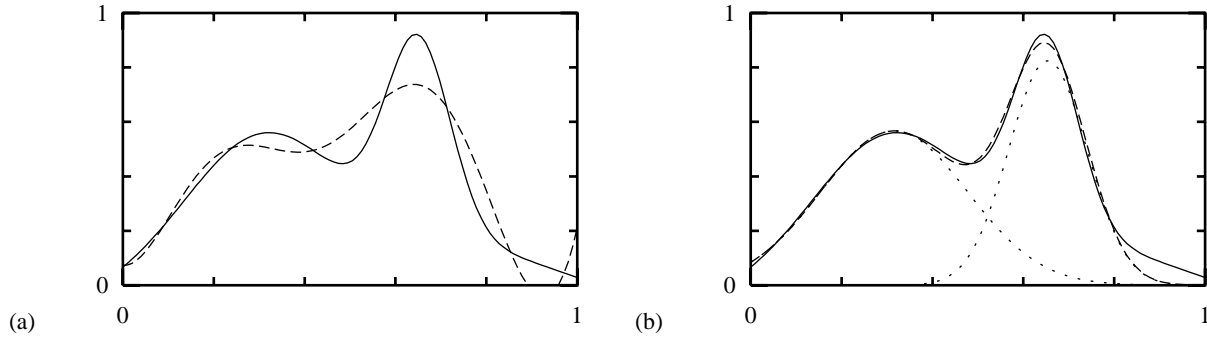
Figure 3: (a) A one-dimensional function (solid line) and its approximation by the first four terms of the Fourier series (dashed line). (b) The same function (solid line) and its approximation by the sum of two unconstrained Gaussians (dashed line). The Gaussians (dotted lines) correspond directly to the main features of the function.

for specular surfaces, which have reflectance functions with high frequencies. On the other hand, many popular models are simple non-linear approximations. The cosine lobe model [12] and the Gaussian model by Ward [23] are probably the most widely used examples, being simple and efficient. Instead of fitting a function in one dimension as in Figure 3, these approximations are defined in the four directional dimensions of the reflectance function.

In this work we take the idea of non-linear approximation a step further, paying close attention to physical plausibility and ensuring computational efficiency.

4 THE GENERALIZED COSINE MODEL

Our representation is a generalization of the cosine lobe model that is based on the Phong shading model. As such, it is intended to approximate the directional-diffuse component and possibly a non-Lambertian diffuse component of a reflectance function. We first discuss the cosine lobe model and then our generalization.

4.1 The Classical Cosine Lobe Model

The original cosine lobe model is attractively simple, but it has a few major shortcomings for representing directional-diffuse reflection. Figure 5 shows the appearance of the model for different viewing angles. The behavior contrasts sharply with the reflectance behavior of most real surfaces, which appear more specular at grazing angles, because the apparent roughness decreases (Figure 2). So why do the reflections in the images of Figure 5 disappear? There are two related reasons. Figure 4a shows how the shape and size of the reflectance lobe remain the same for all incident directions. For grazing angles, up to half the lobe disappears under the surface. Furthermore, the remaining part has to be multiplied by the cosine of the angle with the normal when computing the reflected power. As illustrated in Figure 4b, this results in the albedo (the directional-hemispherical reflectance) decreasing rapidly towards grazing angles. Visually, this means that the directional-diffuse reflection will disappear rather than increase.

In spite of these flaws, the original cosine lobe model is still widely used for illumination simulations. The model is physically plausible: it is reciprocal and conservation of energy can be ensured easily. It is simple and computationally inexpensive to evaluate. It is attractive for Monte Carlo algorithms as one can easily sample directions according to the function. In the context of deterministic algorithms, Arvo [1] showed how irradiance tensors can be applied to analytically compute cosine lobe reflections on surfaces illuminated by diffuse luminaires.

We briefly recall that the original cosine lobe model for a given position and wavelength can be written formally as follows:

$$f_r(\mathbf{u}, \mathbf{v}) \quad = \quad \rho_s C_s \cos^n \alpha, \tag{1}$$

where α is the angle between the exitant direction $\mathbf{v}$ and the mirror direction of the incident direction $\mathbf{u}$, which we will denote by $\mathbf{u}_m$. In order not to burden our notation we will define the power of negative values as 0; the lobe is clamped to 0 for negative cosine values. If we choose C_s to be the normalization factor $(n+2)/(2\pi)$, then ρ_s is a value between 0 and 1, expressing the maximum albedo of the lobe. This maximum is reached for perpendicularly incoming light. The maximum albedo ρ_s and the specular exponent n are the parameters that determine the size and shape of the reflectance function. The cosine can be written as a dot product, and as mentioned in [1], the mirroring around the normal $\mathbf{n}$ can be written using a Householder matrix:

$$\begin{aligned} f_r(\mathbf{u}, \mathbf{v}) \quad &= \quad \rho_s C_s \left[\mathbf{u}_m \cdot \mathbf{v} \right]^n \\ &= \quad \rho_s C_s \left[\mathbf{u}^T (2\mathbf{n}\mathbf{n}^T - \mathbf{I})\mathbf{v} \right]^n. \end{aligned} \tag{2}$$

4.2 The Generalized Cosine Lobe Model

Our model can be regarded as a generalization of the original cosine lobe model. Most known generalizations simply scale the reflectance lobes in some way, violating reciprocity in the process. Changing the model while still satisfying the reciprocity constraint is hard. Physical plausibility, and reciprocity in particular, are therefore important merits of the generalization presented. Yet the representation is conceptually simple and it retains the original advantages for Monte Carlo sampling and analytical evaluation. As a result, it can easily be integrated into existing code.

The essential observation is that Equation 2 can be generalized by replacing the Householder transform together with the normalization factor by a general 3×3 matrix $\mathbf{M}$:

$$f_r(\mathbf{u}, \mathbf{v}) \quad = \quad \rho_s \left[\mathbf{u}^T \mathbf{M} \mathbf{v} \right]^n, \tag{3}$$

where we assume that the direction vectors are defined with respect to a fixed local coordinate system at the surface. This representation provides us with 9 coefficients and an exponent to shape the reflectance function. Of course, certain physical restrictions apply to these parameters. In order for this reflectance function to be reciprocal, the matrix has to be symmetrical: $\mathbf{M} = \mathbf{M}^T$.

We can now apply a singular value decomposition of $\mathbf{M}$ into $\mathbf{Q}^T \mathbf{D} \mathbf{Q}$. This yields the transformation $\mathbf{Q}$ for going to a new local coordinate system, in which the matrix simplifies to the diagonal matrix $\mathbf{D}$. Except for unusual types of anisotropy, the axes
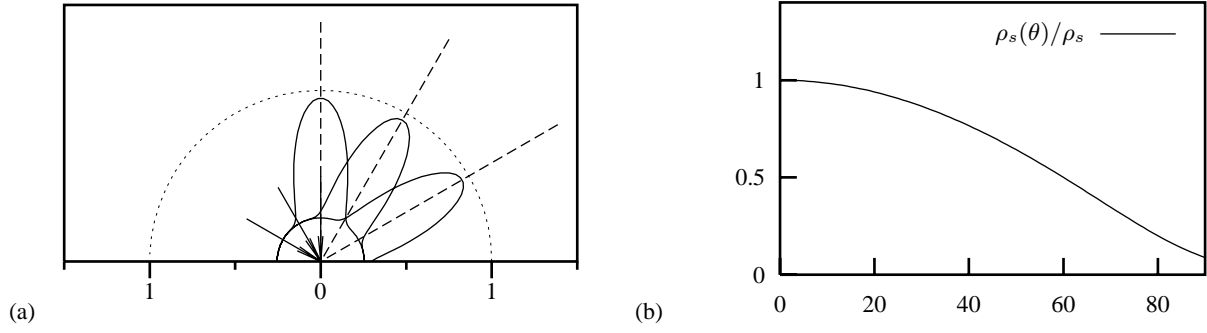
Figure 4: (a) Polar plots of the classical cosine lobe reflectance model ($\rho_s = 0.2$, $n = 20$) with a Lambertian term ($\rho_d = 0.8$) in the incidence plane, for incidence angles $0°$, $30°$ and $60°$. (b) The relative decrease of the albedo of the directional-diffuse term as a function of incidence angle.
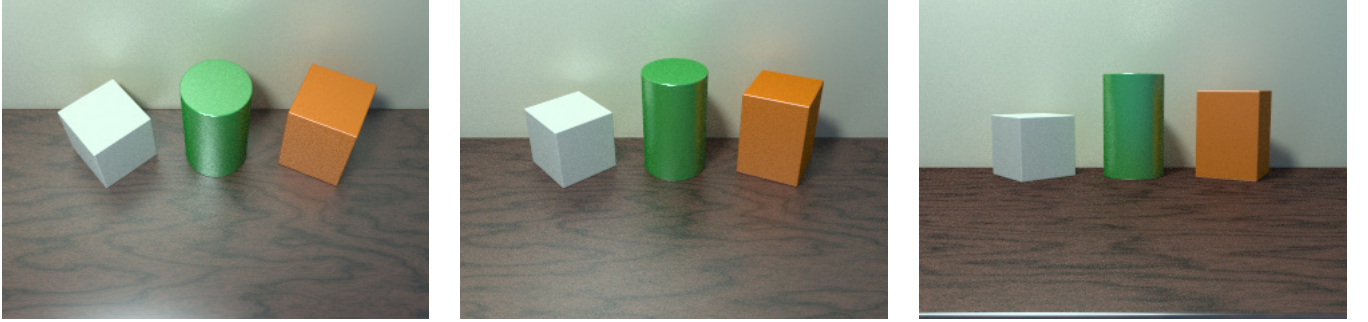
Figure 5: Rendered pictures of a scene with the classical cosine lobe model, for various viewing angles. The glossy reflection on the table disappears at grazing angles, which is exactly the opposite of real surface behavior.
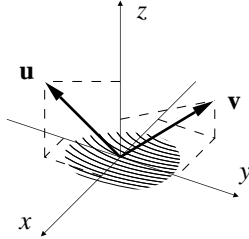
Figure 6: The incident direction **u** and exitant direction **v** are defined in a local coordinate system at the surface. The coordinate system is aligned to the normal and to the principal directions of anisotropy, if any.

are now aligned to the normal and to the principal directions of anisotropy, as illustrated in Figure 6. The diagonal matrix can be seen as weighting the terms of the dot product $\mathbf{u} \cdot \mathbf{v}$:

$$f_r(\mathbf{u}, \mathbf{v}) = \rho_s \left[C_x u_x v_x + C_y u_y v_y + C_z u_z v_z \right]^n. \quad (4)$$

This formulation of the model is the most convenient to use. In the case of isotropic reflection, $C_x = C_y$. The original cosine lobe model is obtained by choosing $-C_x = -C_y = C_z = \sqrt[n]{C_s}$. However, much more expressive functions than the cosine lobe model can be obtained by varying the different parameters, as we will show in more detail in Section 5. Note that the function is defined for all incident and exitant directions. It is thus fully four-dimensional and we apply and fit it as such.

4.3   The Generalized Function as a Cosine Lobe

The generalized function has an elegant and very practical property: for each given incident direction **u** the function can be rewritten as a scaled version of an ordinary cosine lobe. Simply rewriting Equation 3:

$$
\begin{aligned}
f_r(\mathbf{u}, \mathbf{v}) &= \rho_s \|\mathbf{u}^T \mathbf{M}\|^n \left[\frac{\mathbf{u}^T \mathbf{M}}{\|\mathbf{u}^T \mathbf{M}\|} \mathbf{v} \right]^n \\
&= \rho_s C_s(\mathbf{u}) \left[\mathbf{u}' \cdot \mathbf{v} \right]^n \\
&= \rho_s C_s(\mathbf{u}) \cos^n \alpha'. \quad (5)
\end{aligned}
$$

The direction $\mathbf{u}' = (\mathbf{u}^T \mathbf{M}/\|\mathbf{u}^T \mathbf{M}\|)^T$ is a transformed and normalized version of the incident direction **u**, and the angle α' is its angle with **v**. The scaling factor $C_s(\mathbf{u}) = \|\mathbf{u}^T \mathbf{M}\|^n$ is a power of the normalization factor and therefore varies with the incident direction. For the specific case of Equation 4, the direction $\mathbf{u}' = (C_x u_x, C_y u_y, C_z u_z)^T / \sqrt{C_x^2 u_x^2 + C_y^2 u_y^2 + C_z^2 u_z^2}$ and the scaling factor $C_s(\mathbf{u}) = \sqrt{C_x^2 u_x^2 + C_y^2 u_y^2 + C_z^2 u_z^2}^n$. This observation shows how the original cosine lobe function is now generalized in its orientation and its scaling. The changes in orientation and scale are specific results of Equation 3 – if they were just arbitrary, reciprocity would generally not be preserved.

Practically, the equation makes it straightforward to continue using the same Monte Carlo sampling strategies and deterministic evaluation techniques as for the original cosine lobe model. One only needs to substitute the mirror direction $\mathbf{u}_m$ by $\mathbf{u}'$ (or the angle α by α') and scale the results as required. For instance, the albedo $\rho_s(\mathbf{u})$ for each incident direction **u** can be computed analytically, using the procedures presented by Arvo [1]. This is specifically useful to ensure energy conservation.

(a)   (b)
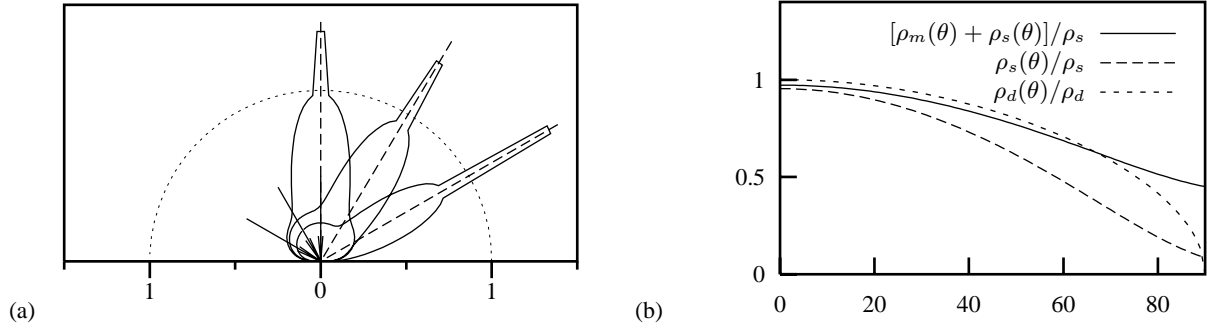
Figure 7: (a) Polar plots of the classical cosine lobe model ($\rho_s = 0.2$, $n = 20$) with a generalized diffuse term ($\rho_d = 0.8$, $n = 0.5$) and an additional mirror term ($R_m = 0.4$). (b) The albedos of the diffuse and directional-diffuse terms, $\rho_d(\theta)$ and $\rho_s(\theta)$ respectively, decrease towards grazing angles; the mirror term $\rho_m(\theta)$ gradually takes over.
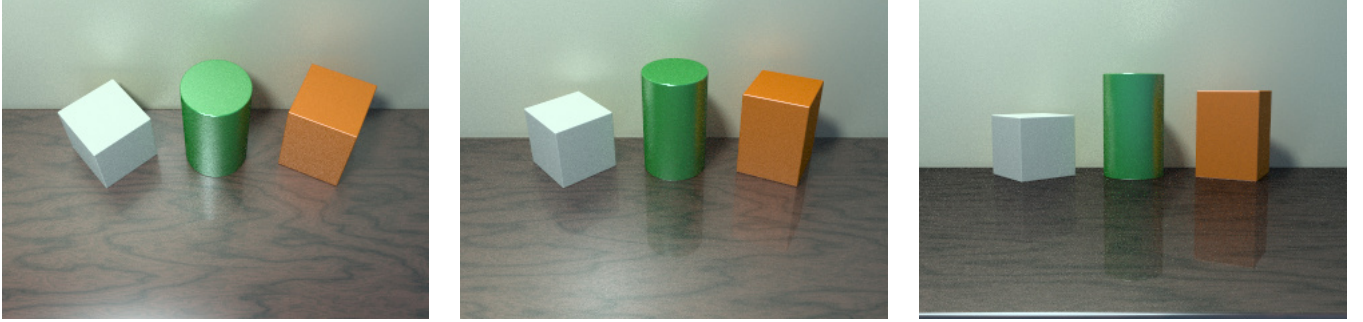
Figure 8: Rendered pictures of a scene with the classical cosine lobe model, now including the mirror term and a generalized diffuse term. The mirror term gradually takes over from the directional-diffuse term, and the diffuse term fades out. Even with these minor changes the table surface already shows a more realistic reflective behavior.

5   QUALITATIVE PROPERTIES

In this section, we illustrate the qualitative properties of our generalized model. We construct a few simple reflectance functions with diffuse, directional-diffuse and specular components, to demonstrate how the model can simulate important aspects of real-life reflectance behavior. Section 6 will then demonstrate the quantitative properties of the model, by fitting sums of primitive functions to a complex physically-based model and to actual measurements.

5.1   Non-Lambertian Diffuse Reflection

An effect apparent in the pictures of Figure 2 is the fading out of the diffuse component for grazing angles. As more light is reflected off the coating of the surface, the subsurface scattering responsible for the diffuse reflection diminishes. The surface looks less saturated and the wood texture disappears. While our generalized cosine lobe model encompasses the Lambertian model (by setting $n = 0$), a more general *rotationally symmetric* diffuse component can be derived from Equation 4, by setting $C_x = C_y = 0$:

$$f_r(\mathbf{u}, \mathbf{v}) \quad = \quad \rho_d C_d \left[u_z v_z \right]^n, \qquad (6)$$

where the normalization factor $C_d = (n + 2)/(2\pi)$, and ρ_d is the parameter between 0 and 1 specifying the maximum albedo. For grazing incident or exitant directions the reflectance decreases proportionally to a power of the cosine of the angle with the normal. This instance actually corresponds to the model presented by Minnaert [13], in the context of modeling the reflectance of the lunar surface. The non-Lambertian diffuse component is plotted in Figure 7a (appearing as the small circular component near the origin), along with directional-diffuse and mirror components that will be

discussed in the next section. Figure 7b shows the behavior of the albedo $\rho_d(\mathbf{u})$ as a function of incidence angle θ, normalized by the parameter ρ_d. Figure 8 illustrates the effect visually: the diffuse component of the table surface fades out for grazing angles.

5.2   Specularity at Grazing Angles

The other important visual effect shown in the pictures of Figure 2 is the increasing specularity of the polished table surface at grazing angles. This behavior can be accounted for by extending the model of a diffuse lobe and a directional-diffuse lobe with a specular mirror term. The directional-diffuse lobe can in the simplest case be an ordinary cosine lobe. The mirror term can be made to reflect a fraction of the power that is not reflected by the directional-diffuse lobe. A simple instance of these two components thus becomes:

$$\begin{aligned} f_r(\mathbf{u}, \mathbf{v}) \quad = \quad & \rho_s C_s \left[\mathbf{u}_m \cdot \mathbf{v} \right]^n \qquad (7) \\ & + (\rho_s - \rho_s(\mathbf{u})) R_m \, \delta(\mathbf{u}_m - \mathbf{v}), \end{aligned}$$

where $\delta(\mathbf{u}_m - \mathbf{v})$ is the Dirac delta function with respect to the canonical measure on the sphere. In this case it is convenient to choose $C_s = (n + 1)/(2\pi)$. The factor $\rho_s - \rho_s(\mathbf{u})$ is the difference between the directional-diffuse scaling factor and the actual albedo for direction $\mathbf{u}$. The parameter R_m expresses the fraction of the power lost in the directional-diffuse lobe that is reflected in the mirror term. In Monte Carlo simulations this can be taken quite literally. One can sample a direction according to the cosine lobe. Any sample is then tested against the cosine of the angle with the normal, with rejection sampling. The fraction R_m of rejected samples is sent into the mirror direction. In analytical computations each of the terms, including the mirror term, can be computed.
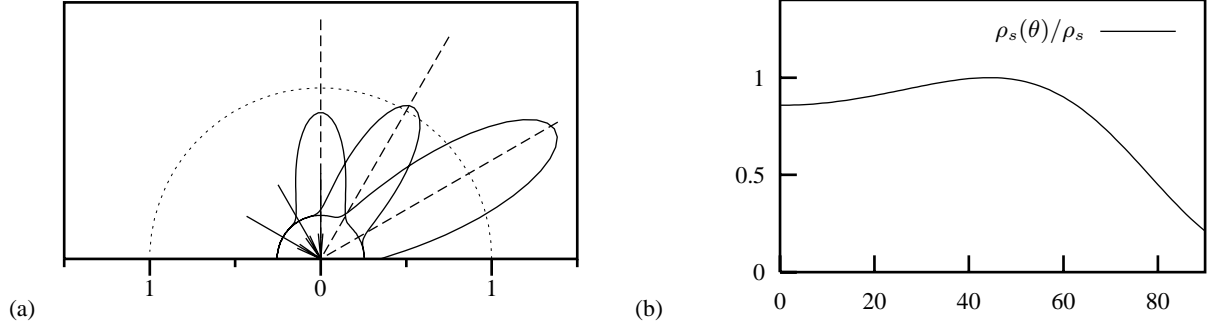
Figure 9: (a) Polar plots of the generalized cosine lobe model ($\rho_s = 0.2$, $n = 20$, $C_z/C_x = 0.95$) with a Lambertian term ($\rho_d = 0.8$). The lobes are slightly off-specular and increase in size towards grazing angles. (b) The albedo of the directional-diffuse term only decreases for larger incidence angles as a result.
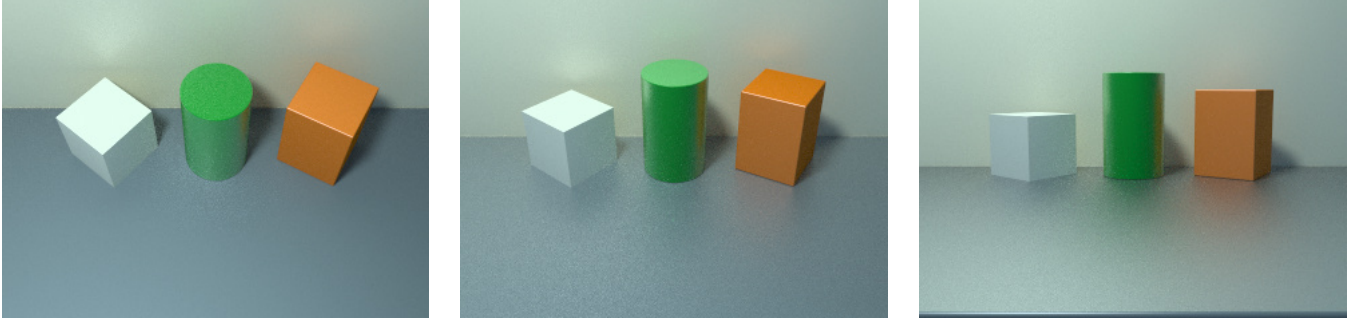
Figure 10: Rendered pictures of a scene with the generalized cosine lobe model. The off-specular directional-diffuse reflectance of the table surface gradually increases for grazing angles.

Figure 7 presents an example function, including the non-Lambertian diffuse reflection that was discussed in the previous section. Note that the mirror term is actually a Dirac delta function; it is broadened here to visualize its behavior. Figure 7b displays the albedos $\rho_s(\theta)$ and $\rho_m(\theta)$ for the directional diffuse and the mirror terms, respectively. Figure 8 then shows the example scene rendered with the extended model.

The results look reasonably realistic because the mirror term is a rough approximation of an actual Fresnel term multiplied by masking-shadowing and roughness factors (e.g. [9]). If it is known, a more accurate approximation can be used by attenuating the mirror term, so that R_m becomes a function of incidence angle.

5.3 Off-Specular Reflection

Application of the model becomes more interesting by varying the individual parameters of Equation 4. Torrance and Sparrow [21] already observed that the directional-diffuse lobe for a given incident direction generally does not reach its maximum for the mirror direction, but rather for a more grazing direction. At the same time the size of the reflectance lobe increases. The original cosine lobe model obviously does not account for these effects. This shortcoming is sometimes overcome by dividing by the cosine of the exitance angle, which breaks reciprocity. In the generalized model, parameters C_z that are smaller than $-C_x = -C_y$ yield a range of off-specular reflection effects, without compromising the physical plausibility. Figure 9 gives an example with moderately increasing reflectance, and Figure 10 shows a set of rendered images. The table surface exhibits off-specular reflection. It looks mostly diffuse from above, while the directional-diffuse component increases for grazing angles.

5.4 Retro-Reflection

Many surfaces not only scatter light in the forward direction, but also backwards, in the direction of the illuminant. This phenomenon is called retro-reflection. The moon surface is an extreme example, where a large fraction of light from the sun is reflected in the incident direction. In the generalized model, a retro-reflective lobe can be represented in the same uniform framework by using a set of parameters C_x, C_y and C_z that are all positive. The reflectance measurements of paint in section 6.2 will illustrate this effect.

5.5 Anisotropy

Anisotropic reflection can be modeled with a single primitive function, by assigning different values to the parameters C_x and C_y. As with the parameter C_z that controls the off-specular reflection, this will pull the reflectance lobes for all incident directions in a preferential direction and scale them. More general anisotropy, e.g. with a splitting lobe, can be obtained by constructing a matrix $\mathbf{M}$ for Equation 3 that is not necessarily symmetrical. Adding a reflectance term with its transpose $\mathbf{M}^T$ then yields a new reciprocal model.

6 QUANTITATIVE PROPERTIES

In this section, we show how the model is also suitable for representing complex real-life reflectance functions. The representation is a sum of several primitive functions of the form of Equation 4. Absorbing the albedo ρ_s in the other parameters, each primitive function i is defined by the parameters $C_{x,i}(= C_{y,i}), C_{z,i}$ and n_i.
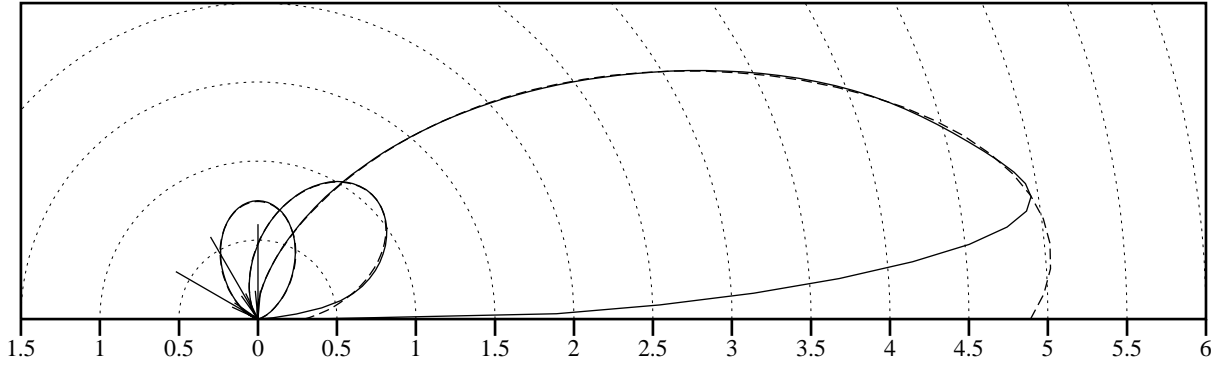
Figure 11: Polar plots of the fitted reflectance model (dashed lines) against the original physically-based model of a roughened aluminum surface (solid lines) in the plane of incidence, for $\theta = 0°, 30°, 60°$, at $500nm$. The reflectance function becomes more off-specular and strongly increases in size towards grazing angles. The sum of generalized cosine functions captures these effects.

Figure 12: Plots of the original physically-based model of roughened aluminum (top row, a) and of the fitted reflectance model (bottom row, b), now multiplied by the cosines of the incidence and exitance angles with the normals, fitted and shown over the entire hemisphere, for various incidence angles.

The model can thus be written as:

$$f_r(\mathbf{u}, \mathbf{v}) = \sum_i [C_{x,i} u_x v_x + C_{y,i} u_y v_y + C_{z,i} u_z v_z]^{n_i}. \quad (8)$$

The model is fitted to the BRDF of aluminum, based on the physically-based reflectance model of He *et al.*, and to the measured BRDF of blue paint. We minimize the mean-square error of the reflectance functions multiplied by the cosines of the incidence and exitance angles with the normal. As the primitive functions are non-linear, a non-linear optimization technique is required to determine the parameters. The Levenberg-Marquardt optimization algorithm has proven to be efficient for this application; computing each approximation requires only a few minutes in a standard numerical package. This is not a serious penalty, as it only has to be done once for each measured material.

In both case studies, we first look at the BRDFs in the incidence plane, and then in the entire function space. In the incidence plane the function space is two-dimensional, depending on the incident polar angle and the exitant polar angle. The entire function space of isotropic BRDFs is three-dimensional, additionally depending on the exitant azimuthal angle.

6.1 Fit to a Physically-Based Model

The reflectance model derived by He *et al.* [9] is generally acknowledged as the most sophisticated model in use in computer graphics. It consists of a Lambertian term, a directional-diffuse term and a mirror term. Here we concentrate on approximating the directional-diffuse term. In our example, the Lambertian term and the mirror term are mostly negligible, but in any case representing and using these terms is straightforward. We present the results for roughened aluminum, as in their original paper for wavelength $\lambda = 500nm$, roughness $\sigma_0 = 0.28\mu m$ and autocorrelation length $\tau = 1.77\mu m$.

Figure 11 shows the results of a fit in the incidence plane, using the sum of three primitive functions. It is important to note that the function has not been fitted for each of the individual lobes, which would be a lot easier, but to the reflectance function as a whole. The fit is visually perfect, except for more grazing angles. In this regime of angles, most of the difference is due to the masking term, which is not present in the representation. These values are less important, however, as they are multiplied in illumination computations by the cosine of the angle between the direction and the surface normal. Additionally, the mirror reflection becomes more important than the directional-diffuse reflection for grazing angles.

Figure 12 shows the results of fitting the approximation to the reflectance function in the entire three-dimensional space of directions. The functions are plotted for three different incidence angles, in a uniform parametrization of the hemisphere [20]. The creases along the diagonals of the square are a result of the parametrization and are not related to the functions. The functions are multiplied by the cosine of the exitance angle with the normal, so that the volumes below the surfaces are proportional to the albedos. Both the shapes of the functions and the albedos match very well.
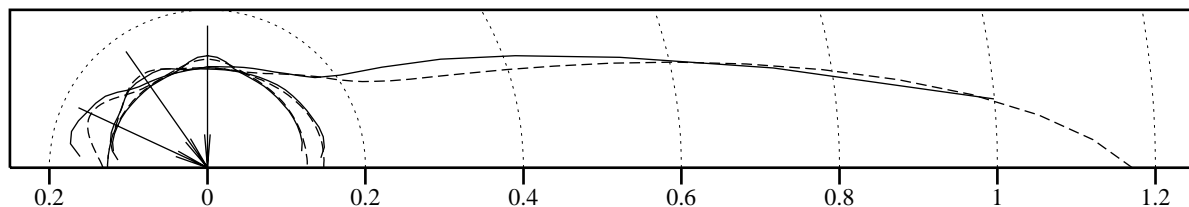
Figure 13: Polar plots of the fitted reflectance model (dashed lines) against the original measured BRDF data of blue paint (solid lines) in the plane of incidence, for $\theta = 0°, 35°, 65°$, at $550nm$. The model successfully reproduces both the increasing retro-reflection and off-specular reflection.
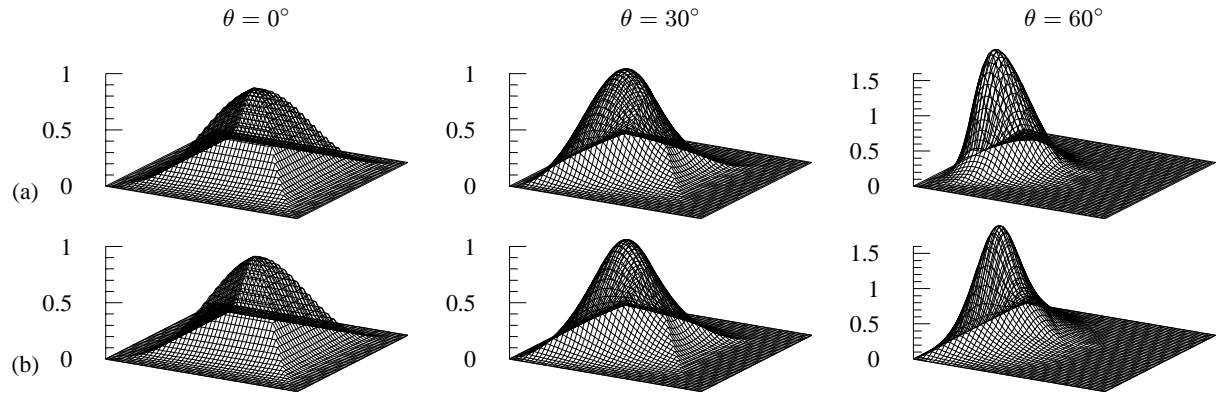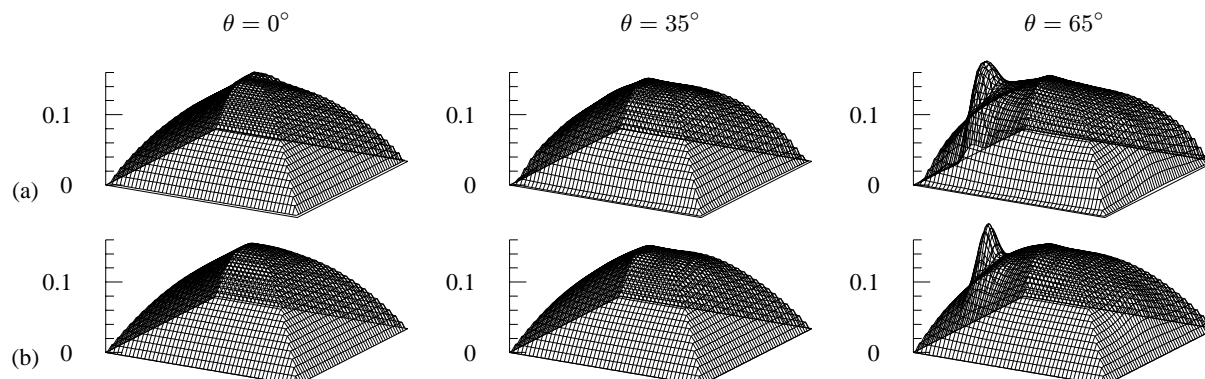

Figure 14: Plots of the original measured model of blue paint (top row, a) and of the fitted reflectance model (bottom row, b), now fitted and shown over the entire hemisphere, for various incidence angles.

6.2 Fit to Reflectance Measurements

The second comparison is with the measured reflectance data of a blue paint sample (spray-painted latex blue paint, Pratt & Lambert, Vapex Interior Wall Base 1, Color #1243, Cal. III) [5]. Figure 13 shows the data and the approximation in the incidence plane at $550nm$, for three incidence angles.

Compared to the strong forward-scattering behavior of the roughened aluminum, the paint is largely diffuse. Due to measurement noise, the data are more irregular. Still, there are important other phenomena. The forward scattering lobe increases rapidly for grazing angles and is very off-specular. The measurements did not include highly grazing angles, for which theory predicts a drop-off. The measurements did show increasing retro-reflection. The approximation, which uses a sum of three directional-diffuse functions and a Lambertian term, captures this effect.

Figure 14 shows the data and the approximation fitted over the three-dimensional space of incident and exitant directions. Table 1 lists the coefficients for this approximation, illustrating how simple and compact the model is. The positive value of C_x for lobe I indicates that it is a retro-reflective lobe, while lobes II and III account for the forward scattering. The ratios of the parameters C_x and C_z give an idea of how off-specular the lobes are and how fast they increase in size for grazing angles. Note that the exponents are not necessarily integers. For Monte Carlo simulations using the model, this is generally not a problem. For analytical computations the exponents would have to be constrained to integer values.

Lobe	$C_x = C_y$	C_z	n
I	0.86	0.77	18.6
II	-0.41	0.018	2.58
III	-1.03	0.70	63.8
Diffuse	0.13		

Table 1: The coefficients of the representation for the three-dimensional fit of Figure 14.

7 RESULTS

We have approximated the measured reflectance data of the blue paint presented in Section 6.2 and of a standardized steel sample (Matte finished steel, Q-Panel Laboratory Products, Q-panel R-46) at 6 discrete wavelengths. The resulting models were then used for global illumination rendering, using a Monte Carlo path tracing program. The implementation required only a few additional lines of code. The reflectance functions are evaluated using Equation 8. For sampling an exitant direction for a given incident direction we construct a probability density function that is a linear combination of the primitive cosine reflectance lobes.

Figure 15 shows a rendering of a simple scene with two spheres, a Q-panel, and two colored light sources, positioned symmetrically with respect to the viewer. A larger white light source above the viewer illuminates the whole scene. The sphere on the left is rendered with a Lambertian diffuse approximation of the measured blue paint, while the sphere on the right is rendered with the generalized reflectance model. The latter sphere has both red and green highlights due to strong forward scattering. These are lacking on the Lambertian sphere. With a light source near the viewer, the right sphere has a slightly flatter appearance due to retro-reflection. The Q-panel has a completely different appearance, displaying a blurry metallic reflection of the colored lights and of the objects. The representation successfully captures these very different reflectance characteristics.

8 CONCLUSIONS

We have introduced an efficient representation for a wide range of bidirectional reflectance distribution functions. It is an interesting alternative for previous models of directional-diffuse reflectance, which required either simplified single-term representations, complex analytical expressions for specific classes of functions, or general but large representations with linear basis functions.
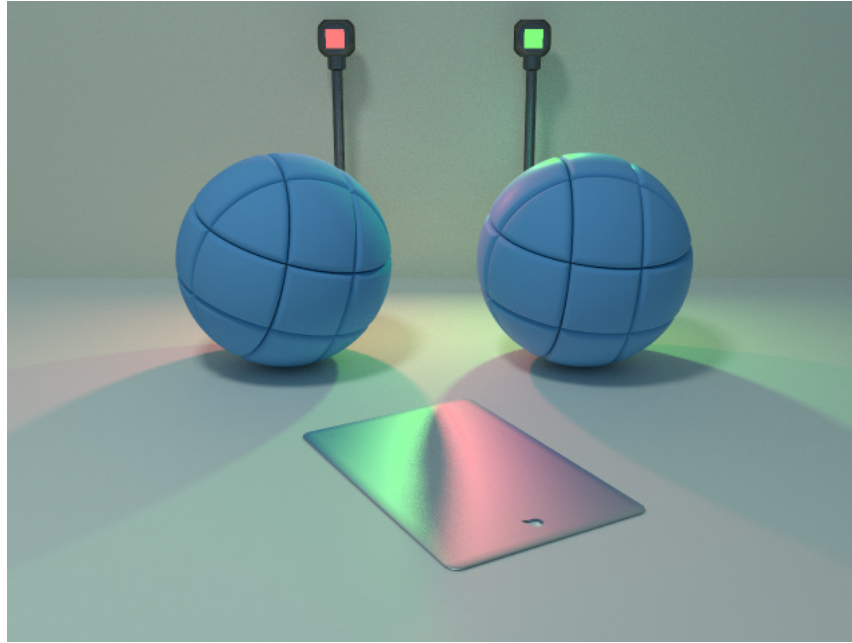
Figure 15: Rendered picture of a scene with two spheres and a Q-panel, illuminated by two colored light sources and one larger white light source. The sphere on the left has a Lambertian approximation of the measured paint reflectance; the sphere on the right is rendered with the non-linear approximation. The Q-panel has the non-linear approximation of the measured steel reflectance.

- The representation is compact. Each primitive function is determined by two or three coefficients and an exponent. Because the representation is memory-efficient, any complex wavelength dependency can be modeled by constructing independent approximations at discrete wavelengths.

- The functions are expressive. They can represent complex reflectance behavior, such as off-specular reflection, increasing directional-diffuse reflectance for grazing angles, retro-reflection and non-Lambertian diffuse reflection in a uniform way.

- The functions handle noise in the raw reflectance data gracefully. They can capture sharp reflectance lobes without suffering from small spurious errors in the data. If the data are sparse, the model interpolates them naturally.

- The functions themselves are physically plausible, irrespective of how they were constructed. They are inherently reciprocal. Energy-conservation can be verified analytically for each incident direction.

- On the algorithmic side, the representation is efficient and easy to use in both local and global illumination algorithms. Its simplicity and uniformity make it practical for implementation in hardware. In Monte Carlo algorithms, reflection directions for a given incident direction can be sampled according to the transformed cosine lobe. In deterministic algorithms, illumination from diffuse emitters can be computed analytically, using a straightforward extension of the calculations for ordinary cosine lobes.

- While the representation cannot approximate all possible reflectance functions to any desired accuracy, it adequately represents a range of measured BRDF data, which usually only have a very limited accuracy. In our tests, we have obtained satisfactory results with as few as three primitive functions to represent directional-diffuse reflections from roughened metals and paints. Broad, glossy reflectance lobes are relatively easy to approximate. Sharp directional-diffuse peaks, such as for smooth metal surfaces, may be harder to represent, due to a strong dependency on the Fresnel factor, which is not explicitly included in the representation.

As future work, we will look into the details of representing anisotropic reflectance measurements with one or more terms of the current model, e.g. to model the effect of splitting reflectance lobes at anisotropic surfaces.

Acknowledgements

References

[1] J. Arvo. Applications of irradiance tensors to the simulation of non-Lambertian phenomena. In *SIGGRAPH 95 Conference Proceedings*, pages 335–342, Los Angeles, California, August 1995.

[2] B. Cabral, N. Max, and R. Springmeyer. Bidirectional reflection functions from surface bump maps. *Computer Graphics*, 21(4):273–281, July 1987.

[3] R.L. Cook and K.E. Torrance. A reflectance model for computer graphics. *Computer Graphics*, 15(4):187–196, July 1981.

[4] R.L. Cook and K.E. Torrance. A reflectance model for computer graphics. *ACM Transactions on Graphics*, 1(1):7–24, January 1982.

[5] S.C. Foo. A gonioreflectometer for measuring the bidirectional reflectance of materials for use in illumination computations. Master's thesis, Cornell University, Ithaca, New York, July 1997.

[6] A. Fournier. Separating reflection functions for linear radiosity. In *Proceedings of the Sixth Eurographics Workshop on Rendering*, pages 383–392, Dublin, Ireland, June 1995.

[7] J.S. Gondek, G.W. Meyer, and J.G. Newman. Wavelength dependent reflectance functions. In *SIGGRAPH 94 Conference Proceedings*, pages 213–220, Orlando, Florida, July 1994.

[8] P. Hanrahan and W. Krueger. Reflection from layered surfaces due to subsurface scattering. In *SIGGRAPH 93 Conference Proceedings*, pages 165–174, Anaheim, California, August 1993.

[9] X.D. He, K.E. Torrance, F.X. Sillion, and D.P. Greenberg. A comprehensive physical model for light reflection. *Computer Graphics*, 25(4):175–186, July 1991.

[10] J. Kajiya. Anisotropic reflectance models. *Computer Graphics*, 19(4):15–21, July 1985.

[11] J.J. Koenderink, A.J. van Doorn, and M. Stavridi. Bidirectional reflection distribution function expressed in terms of surface scattering modes. In *European Conference on Computer Vision*, pages 28–39, 1996.

[12] R.R. Lewis. Making shaders more physically plausible. In *Proceedings of the Fourth Eurographics Workshop on Rendering*, pages 47–62, Paris, France, June 1993.

[13] M. Minnaert. The reciprocity principle in lunar photometry. *Astrophysical Journal*, 93:403–410, 1941.

[14] M. Oren and S.K. Nayar. Generalization of Lambert's reflectance model. In *SIGGRAPH 94 Conference Proceedings*, pages 239–246, Orlando, Florida, July 1994.

[15] B.T. Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975.

[16] P. Poulin and A. Fournier. A model for anisotropic reflection. *Computer Graphics*, 24(4):273–282, August 1990.

[17] Ch. Schlick. A customizable reflectance model for everyday rendering. In *Proceedings of the Fourth Eurographics Workshop on Rendering*, pages 73–83, Paris, France, June 1993.

[18] Ch. Schlick. A survey of shading and reflectance models. *Computer Graphics Forum*, 13(2):121–131, June 1994.

[19] P. Schröder and W. Sweldens. Spherical wavelets: Efficiently representing functions on the sphere. In *SIGGRAPH 95 Conference Proceedings*, pages 161–172, Los Angeles, California, August 1995.

[20] P. Shirley and K. Chiu. Notes on adaptive quadrature on the hemisphere. Technical Report 411, Department of Computer Science, Indiana University, Bloomington, Indiana, 1994.

[21] K.E. Torrance and E.M. Sparrow. Off-specular peaks in the directional distribution of reflected thermal radiation. In *Transactions of the ASME*, pages 1–8, Chicago, Ill., November 1965.

[22] K.E. Torrance and E.M. Sparrow. Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America*, 57(9):1105–1114, September 1967.

[23] G.J. Ward. Measuring and modeling anisotropic reflection. *Computer Graphics*, 26(2):265–272, July 1992.

[24] S.H. Westin, J.R. Arvo, and K.E. Torrance. Predicting reflectance functions from complex surfaces. *Computer Graphics*, 26(2):255–264, July 1992.

Image-Based BRDF Measurement
Including Human Skin

Stephen R. Marschner*   Stephen H. Westin   Eric P. F. Lafortune
Kenneth E. Torrance   Donald P. Greenberg
Program of Computer Graphics
Cornell University

Abstract: We present a new image-based process for measuring the bidirectional reflectance of homogeneous surfaces rapidly, completely, and accurately. For simple sample shapes (spheres and cylinders) the method requires only a digital camera and a stable light source. Adding a 3D scanner allows a wide class of curved near-convex objects to be measured. With measurements for a variety of materials from paints to human skin, we demonstrate the new method's ability to achieve high resolution and accuracy over a large domain of illumination and reflection directions. We verify our measurements by tests of internal consistency and by comparison against measurements made using a gonioreflectometer.

1   Introduction

To render accurate images reliably and easily, the reflectance of surfaces must be simulated accurately. The most direct way to ensure correct simulation is to use physical reflectance measurements. Such measurements can guide the choice of parameters for existing reflectance models, and if they are sufficiently complete they can be used as input for renderers or provide the basis for entirely new models. To completely capture the reflectance of an opaque surface, one must measure the *bidirectional reflectance distribution function* (BRDF). BRDF measurements have traditionally been made with purpose-built devices known as *gonioreflectometers*, which are rare and expensive.

This paper presents a system that measures reflectance quickly and completely without special equipment. The method works by taking a series of photographs of a curved object; each image captures light reflected from many differently oriented parts of the surface. By using a curved test sample and an imaging detector, and by using automated photogrammetry to measure the camera position, we eliminate the precise mechanisms needed to position the source and detector in a conventional gonioreflectometer. By knowing the sample shape and the light source position, we can analyze the photographs to determine the sample's BRDF. With only a light source and a digital camera, objects of known, regular shape can be measured; adding a 3D geometry scanner extends the technique to cover a whole class of surfaces, including human skin, that are impractical to measure by other methods. Although the apparatus is simple and the measurement rapid, the resulting data are accurate and can be very complete, covering the full hemisphere almost to grazing angles.

2   BRDF Background

The BRDF, f_r, completely describes the reflectance of an opaque surface at a single point. Its value measures the ratio of the radiance L exiting the surface in a given

---

direction to the incident irradiance I at a particular wavelength λ from an incident solid angle $d\omega_i$ about a given illumination direction. Representing the incident and exitant directions in spherical coordinates according to Figure 1,

$$f_r(\theta_i, \phi_i, \theta_e, \phi_e, \lambda) = \frac{dL(\theta_e, \phi_e)}{dI(\theta_i, \phi_i)}. \tag{1}$$

The BRDF is thus a function of five variables, but its domain is reduced somewhat by a symmetry called *reciprocity*, which states that reversing the light's path does not change the reflectance:

$$f_r(\theta_1, \phi_1, \theta_2, \phi_2, \lambda) = f_r(\theta_2, \phi_2, \theta_1, \phi_1, \lambda).$$

In this paper we will concentrate on the important class of *isotropic* materials, for which the reflectance is independent of rotating the incident and exitant directions about the surface normal. For these surfaces, the BRDF depends only on $\Delta\phi = \phi_e - \phi_i$, rather than on ϕ_i and ϕ_e separately, which reduces the domain from five to four variables:

$$f_r(\theta_i, \phi_i, \theta_e, \phi_e, \lambda) = f_r(\theta_i, \theta_e, \Delta\phi, \lambda). \tag{2}$$

In computer graphics, the wavelength dependence of BRDF is of interest only for the purposes of determining colors seen by human observers, so the continuous wavelength dimension

Fig. 1. Geometry of surface reflection.

can often be replaced with an appropriate discrete set of three measurements (R, G, B), further reducing the isotropic BRDF to a vector-valued function of three variables.

3   Overview of Method

A straightforward device for measuring isotropic BRDFs is shown in Figure 2a, illustrating the three mechanical degrees of freedom required. A flat sample is illuminated by a light source, and a detector measures the complete distribution of reflected light by moving around the entire hemisphere. To measure a full BRDF, this process must be repeated many times, moving the light source each time to measure a different incidence angle.

(a)   (b)   (c)
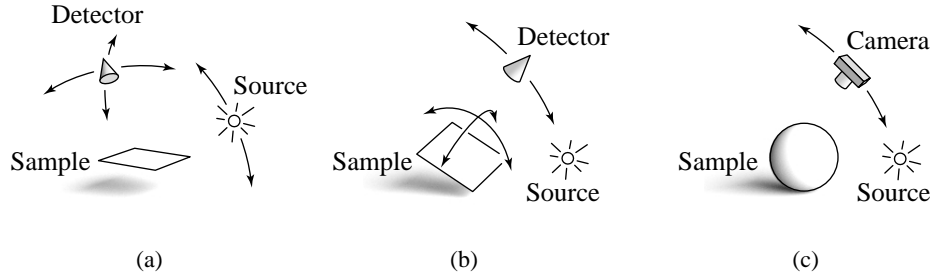
Fig. 2. Three BRDF measurement devices, including our image-based approach (c).

2

Because only the relative positions of the sample, source, and detector are relevant, all the same measurements could be made using the device of Figure 2b, in which the sample rotates with two degrees of freedom but the detector has only one and the source is fixed. The number of degrees of freedom remains the same, and all the same configurations of source, sample, and detector can be achieved.

If we replace the flat sample with a curved one, we can acquire data from many sample orientations simultaneously. Since every part of the sample's surface has a different orientation, we can use a camera to measure different parts of the surface instead of rotating the sample, as shown in Figure 2c. In this device, the two dimensions of the image sensor substitute for the two degrees of freedom of sample rotation. If there is sufficient curvature, we can make all the same measurements as the other devices, and by measuring two degrees of freedom in parallel we can greatly reduce the measurement time while increasing the sampling density.

This is the essence of image-based BRDF measurement: in an image of a curved object taken using a small light source, every pixel is in effect a BRDF measurement. Given a 3D model of the sample, camera, and light source, we can determine the incident and exitant directions for each pixel relative to the surface normal, as well as the irradiance due to the light source. Together with the radiance measured by the camera, these are all the data required to compute the BRDF.

Because a single image will only cover a two-dimensional subset of the possible BRDF configurations, many images are required to measure the whole domain. In the case of an isotropic BRDF, we are filling up the three-dimensional domain of the BRDF by measuring two-dimensional sheets, so we will need a one-dimensional sequence of images, with the camera or light source positioned differently in each.

4   Related Work

Traditionally [19, 21], the three or four angular dimensions of the BRDF are handled by specialized mechanisms that position a light source and a detector at various directions from a flat sample of the material to be measured. Because three or four dimensions must be sampled sequentially, reflectance measurements are time-consuming, even with modern computer controls. Even a sparse sampling of the incident and exitant hemispheres can take several hours.

More recently, image-based methods have been used to speed measurements by gathering many angular samples at once. These methods, including the method presented in this paper, use a two-dimensional detector—the image sensor of a digital camera—to measure a two-dimensional range of angles simultaneously, leaving one or two dimensions of angle to be sampled by sequential measurements.

These can be categorized in two groups: those that attempt to measure the BRDF over its entire bihemispherical domain and those that measure some useful subset. The BRDF over an appropriate subset of the domain can be used to deduce characteristics of the surface microgeometry or to find parameters for a low-dimensional BRDF model.

One example of measuring a subset of the domain is the work of Karner et al. [15], who use images of a planar sample to measure the BRDF over a limited range of interesting angles. They use these data to fit coefficients for a simple reflectance model.

Ikeuchi and Sato [14] estimate reflectance model parameters using a surface model from a range scanner and a single image from a video camera. They use a curved sample to capture a larger range of incidence and exitance angles, but their data are still constrained to the angles provided by the illumination and view directions of a single image. Sato et al. [20] extend this method to deal with spatial variations in BRDF by

acquiring a sequence of images while the sample rotates. The image sequence provides samples along a one-dimensional path for each surface point; a simple reflectance model is fit to these data.

The surface optics literature also includes a number of approaches to measure a subdomain of the BRDF rapidly; these are generally used to deduce physical parameters of the surface itself, such as feature size on integrated circuits [12] or surface roughness [3], and often measure only at a single wavelength.

Ward describes a device [23] that is able to measure the complete BRDF of anisotropic materials. His camera captures the entire exitant hemisphere at once with a hemispherical mirror and a fisheye lens. The source and sample are moved mechanically to cover all incident angles.

More recently, Lu, Koenderink, and Kappers [17] use a cylindrical sample to give broad angular coverage in the incident plane, using multiple images with different source positions to cover all angles.

Like these other image-based systems, the system presented in this paper uses a camera to sample a two-dimensional set of angles in a single measurement, so it shares their advantages in speed and sampling density over traditional approaches. It can be thought of as a combination and extension of the techniques of Ward and Lu et al. By adopting a curved sample, it avoids the fisheye lens and hemispherical mirror of Ward's method and permits measurements much closer to grazing.[1] By using samples with compound curvature, we extend coverage from the incidence plane to the entire BRDF domain. We go beyond both of these techniques in allowing hand-held positioning, which obviates any precision source positioning mechanism, and in extending the technique to arbitrary convex objects. The method of this paper was derived from that in the dissertation of the first author [18], but it works with more general shapes, requires less equipment, and is simpler to use.

The following sections describe the specifics of our system, give the results of measuring several materials, and demonstrate the accuracy of those results by comparing them to measurements from a gonioreflectometer of verified accuracy.

5   Method

Our image-based technique can measure the BRDF of two different classes of objects: simple geometric shapes, for which the 3D shape can be defined analytically, and irregular shapes, for which the 3D model is provided by a range scanner.

Geometric shapes, such as spheres and cylinders, can be modeled and aligned precisely, giving measurements with low error. This approach also requires less equipment, since a range scanner is not required. However, only certain materials can be measured using these shapes—typically only paints or other man-made finishes that can be applied to such an object.

Fig. 3. Schematic of measurement setup.

If a 3D description of the sample shape is available, we can measure any convex object that has a uniform BRDF. Since we no longer have to control the geometry, it becomes possible to measure many more interesting materials. This generality has a cost, however: the limitations of the scanner introduce geometric errors that lead to noise in our results.

---

[1]Ward's device covers angles of up to 45° to 75°, depending on azimuth angle [10].

We use a hand-held camera to photograph the sample from a sequence of positions, with a single stationary light source providing the only illumination. The camera moves from a position next to the light source, which allows measurement of near-retroreflection, to opposite the light source, where we measure grazing-angle reflection (Figure 3). A few additional photographs, described below, are also taken to measure the location and intensity of the light source. In all, a typical measurement session, including the range scan and all the photographs, takes about half an hour.

The equipment we use to make our measurements includes:

- A digital still camera using a 1.5 megapixel CCD sensor with an RGB color filter array (Kodak DCS 420).
- A simple industrial electronic flash, rated at 400 W-sec output (Photogenic Machine Co. EP377).
- A structured-light range scanner, for measurements of irregularly-shaped samples (Cyberware 3030/PS).

From each pixel in each measurement image we derive one sample somewhere in the domain of the BRDF; the locations of the samples are determined by the geometry of the sample's surface and the arrangement of camera, source, and sample. As explained in Section 3, each image measures a two-dimensional set of BRDF configurations, but we take multiple images (typically about 30) from different positions to cover the full three-dimensional BRDF.

5.1 Calibration

Turning the camera images into accurate BRDF measurements requires both geometric and radiometric calibration. Geometric calibration establishes the relative positions of the light source, sample, and camera for each measurement image, and radiometric calibration determines the irradiance due to the source and the relationship between pixel values and radiance reflected from the sample.

Geometric calibration. Geometric calibration is done with photogrammetric techniques, using machine-readable targets that are placed on a structure positioned near the sample [18 (Appendix C)]. These targets are located and identified automatically in each image using ID codes embedded in the targets. The information that must be derived from the target locations includes:

- The position of the light source.
- The camera pose for each measurement image.
- The pose of the sample.

The poses of the camera are found from the image-plane target locations using bundle adjustment[2] [4, 7, 18 (Appendix B)]. Since our targets are recognized automatically and coded with unique ID numbers, no manual intervention is needed to establish correspondence between points in the various images.

There are three sets of targets: the *sample targets*, fixed with respect to the sample, the *source targets*, fixed with respect to the source, and the *stationary targets*, fixed in the room. The positions of the camera in the room are obtained using the stationary

---

[2]Bundle adjustment takes the image-plane projections of m points in n images and computes the m 3D locations of the points and the n camera poses by solving a nonlinear system of equations.

targets. Three extra images that include both the stationary targets and the source targets allow us to extract the position of the source in the room. The sample targets are used to determine the position of the sample relative to the camera positions. With the knowledge of these three relationships, the incident and exitant directions relative to the surface normal can be computed for any point on the sample. When measuring the skin of a human subject, which may change position from one image to the next (see Section 6.4), the sample position is determined separately for each frame, but when measuring inanimate samples the stationary targets are redundant, and are used only to improve position estimates. Gortler et al. [9] also used encoded targets to determine camera pose, but we have extended the technique to find sample and source positions; we also use more targets to cover a wide angular range robustly.

Radiometric calibration. In order to make BRDF measurements for each pixel, we must know the radiance reflected to the camera and the irradiance due to the source. To use a digital camera to measure radiance we must characterize both the optoelectronic conversion function (OECF), which relates the digital count reported for a pixel with the image-plane exposure, and the flat-field response, which relates the image-plane exposure to radiance in the scene. We used a calibrated reference source (Labsphere CSTM-USS-1200) to measure each of these camera characteristics.

To measure the OECF, we removed the camera lens to expose the CCD sensor directly to the source. We used a variable iris aperture and individual control of the four lamps in the source to vary irradiance through a range of more than 1600:1. A previously calibrated digital camera was used as a reference.

To measure the flat-field response, we remounted the lens (which is the principle source of flat-field variation) and took a series of exposures with the source appearing at various positions on the image plane. By fitting a biquadratic function to these images, we approximated the spatial variation across the image plane and were able to compensate for it. This procedure differs from that used previously [18 (Appendix A)] in order to reduce flare associated with the lens used here.

To determine the irradiance at each location on the surface, we approximated the source as a single point.[3] In order for this model to be valid, the source must be small compared to the distance to the sample, and its angular intensity distribution must be uniform. We measured the angular distribution of the source by capturing calibrated images of a flat, uniform surface illuminated by the flash and verified that, with an additional diffuser, it is sufficiently uniform over the range of angles we use. To get the absolute magnitude of the BRDF correct, we measured the intensity of the light source relative to the camera's three color sensitivities by photographing a diffuse white reference sample (a calibrated Spectralon target from Labsphere, Inc.) in a known position.

5.2 Data processing

Processing the measurement images to extract BRDF samples involves two steps. First, the photogrammetric targets are used to determine the geometric arrangement of the sample, camera, source, and reference white target. Second, all this information is given to a *derenderer*, which computes the BRDF values.

---

[3]While the real source only approximates a point, compensating for its solid angle requires a deconvolution process that is not trivial. We follow accepted practice of reporting our raw measurements and the solid angle of the source, which is a circle subtending 1.3×10^{-3} steradians. The solid angle of the camera's aperture, $\approx 6 \times 10^{-5}$ steradians, is negligible by comparison.

We begin by extracting the target positions in each image. This gives us the 2D image positions of the targets visible in each image and their correspondence in different images. This information is used to solve a bundle adjustment system, which computes the poses of all the cameras and the 3D locations of all the targets. It then remains to locate the model of the sample in the same coordinate system. For a cylindrical sample, a cylinder is automatically fit to the 3D locations of the sample targets, which are attached to the sample's surface. If the sample is a sphere, the user manually specifies points on the boundary of the sample in 3 or 4 images, and a tangent sphere is fit to the corresponding rays to define the sample model. For a sample of arbitrary shape, we scan the sample and the sample targets together. The targets can then be automatically recognized in the luminance image produced from the scan and transformed to their 3D positions within the scan. A rigid-body transformation aligns these scanned 3D positions with the 3D positions of the corresponding targets in the bundle adjustment results, putting the scanned 3D geometry in the same coordinate system as the camera and source positions.

The derenderer is derived from a ray-tracing renderer, and its input is a scene description including the cameras, the light source, and a model of the sample. It uses standard rendering techniques [8] to find the intersection point of each pixel's viewing ray with the sample surface and to compute the irradiance due to the source. Rather than using a BRDF value to compute the radiance reflected to the pixel, as a renderer would, the derenderer instead divides the pixel's measured radiance by the irradiance to obtain the BRDF value. The derenderer's output is a list of BRDF samples, each including the incident direction, the exitant direction, and the BRDF value for that configuration. Separate sample sets are generated from the camera's red, green, and blue pixels.

If a range scan is providing the model of the sample, the points from the scanner are tessellated to define the surface for ray intersection. To reduce the effects of scanner noise, we derive a normal to compute the BRDF at each point by fitting a plane to a weighted set of nearby points.

6 Results

We have used our image-based system to measure the BRDFs of several materials. Here, we present three materials: matte gray paint, a squash, and human skin. The matte gray paint, applied to a cylinder, allows us to verify that our BRDF measurements are accurate by comparing them with gonioreflectometer measurements. The squash and human skin demonstrate the measurement of two surfaces impractical to measure in a traditional gonioreflectometer. We have measured other materials ranging from paints to felt, a few of which are shown in rendered images.

For each of the samples, we show measurements in the incidence plane for several values of θ. Plotted with the measurements is a slice of a smooth BRDF reconstructed using local quadratic regression in the BRDF's 3D domain [5]. This technique defines a smooth, continuous function over the entire BRDF domain that follows the samples and interpolates across unsampled areas. Each curve is a slice of a 3D function fit to all the data points, not just a fit to the points visible with it. Since the curve accounts for more points than are shown, it may sometimes diverge slightly from the points. In these plots, backward scattering is on the left and forward on the right; the specular direction is marked by a vertical gray line.

6.1 Gray cylinder

To verify the correctness of our measurements, we painted a section of aluminum tubing (outside diameter 6 inches) with a sprayed gray primer. The resulting sample has a very uniform surface and is well modeled by an ideal cylinder. We measured its geometry and position using a strip of photogrammetric targets along each edge; a typical measurement image is shown in Figure 4. Because a cylinder curves only along one direction, the resulting data lie very near a two-dimensional slice of the three-dimensional (isotropic) BRDF domain; this allows us to concentrate our measurement points on the incidence plane.

Figure 6 summarizes the results of the gray cylinder measurement. Note the low noise and broad coverage—the results seem reliable out to at least $80°$. The raw points shown include measurements both in the forward direction (all θ_e for the θ_i indicated on each plot) and in the reciprocal direction (all θ_i for the indicated θ_e): the scatter shown includes any deviations from reciprocity. The low scatter serves as a first validation of our measurements, since the reciprocal measurements are independent.

We measured a matching flat sample using a gonioreflectometer [6] designed according to ASTM recommendations [1] and verified to an accuracy of 5%. The gonioreflectometer results are plotted with a dashed line. Note the good correspondence to our image-based measurements; this independent measurement further validates the correctness of our method.

6.2 Squash

Having verified the accuracy of our technique, we applied it to more interesting objects. One of these, a squash, illustrates some of the strengths of our method. There is no practical way to obtain a flat sample of this surface to use in a traditional gonioreflectometer.

A typical measurement image is shown in Figure 5. Below the squash, one can see the support structure containing the sample targets (see Section 5.1); the stationary targets can be seen above it. Figure 5 also indicates the approximate subset of the geometry used in the derendering process. The top was truncated by the limits of our 3D scanner; we deleted the lower part to reduce computation time and storage demands. Even this small subset of the available data results in over 300,000 BRDF samples per channel.

The first column of Figure 7 shows the coverage obtained with this sample. The dots are plotted in a polar coordinate system, with radius indicating θ_e or θ_i and angle indicating $\Delta\phi$ ($\Delta\phi = 0$ is at the bottom; the incidence plane, where $\Delta\phi = 0$ or $180°$ is marked by the vertical line). We include both the forward measurements (fixed θ_i) and reciprocal measurements (fixed θ_e); reciprocity allows us to use these points to help fill the hemisphere. The points in each plot fall on rings, each consisting of samples from one measurement image. The rings appear because the angle between the illumination and viewing direction is nearly constant within each image. There are never any data within a small circle around the incident direction because we cannot physically place the camera at the source position to measure exact retroreflection.

Because of the squash's compound curvature, much of the BRDF domain is sampled. There are some gaps in the coverage; had we scanned the entire rounded end of the squash, we would have covered the entire hemisphere well. The reconstructed curves in the incidence-plane plots (Figure 7, second column) show that the dataset as a whole defines a smooth function that describes an interesting and plausible BRDF, with an off-specular forward scattering lobe but also a non-Lambertian base color. The data

8
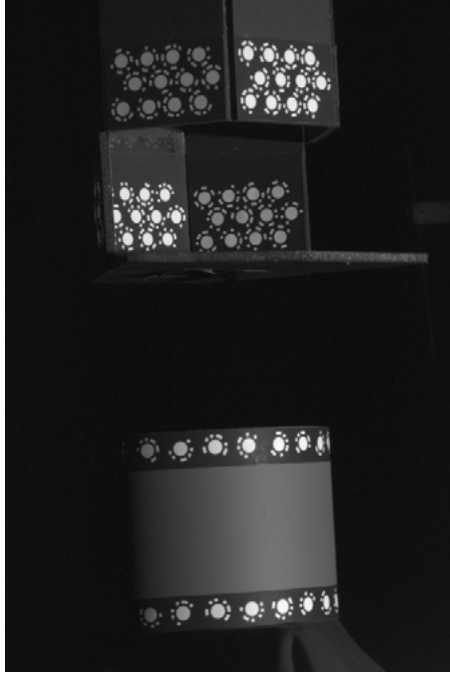
Fig. 4. A typical measurement image from the gray cylinder dataset.
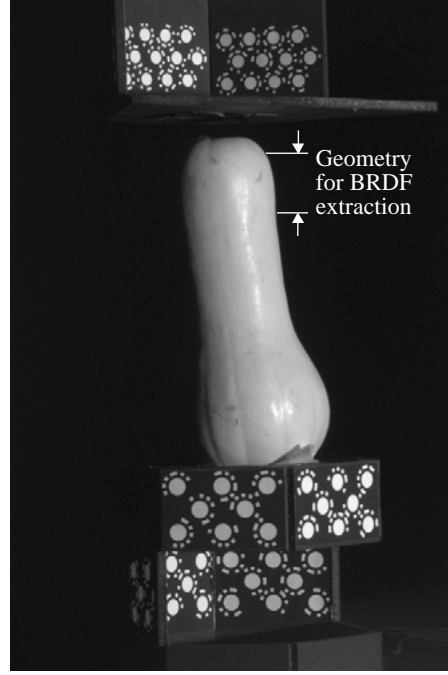
Fig. 5. A typical measurement image from the squash dataset.
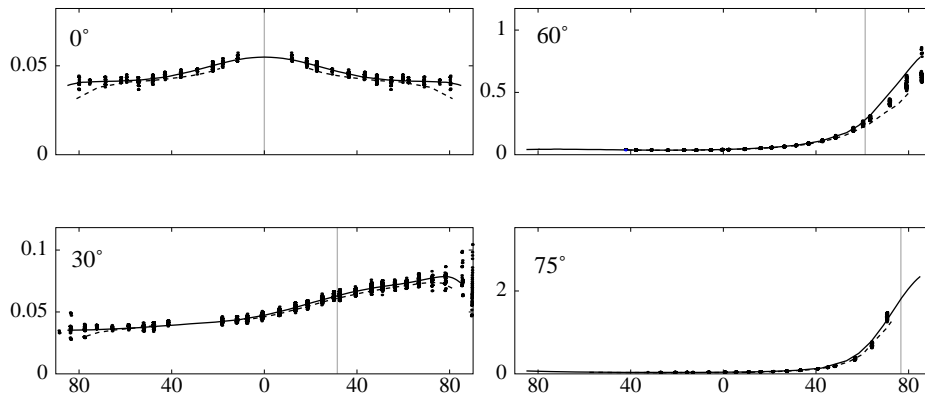
Fig. 6. Summary of results from 29 images of the gray cylinder. Points: raw measurements including reciprocal data. Solid line: local polynomial fit. Dashed line: gonioreflectometer results.

9
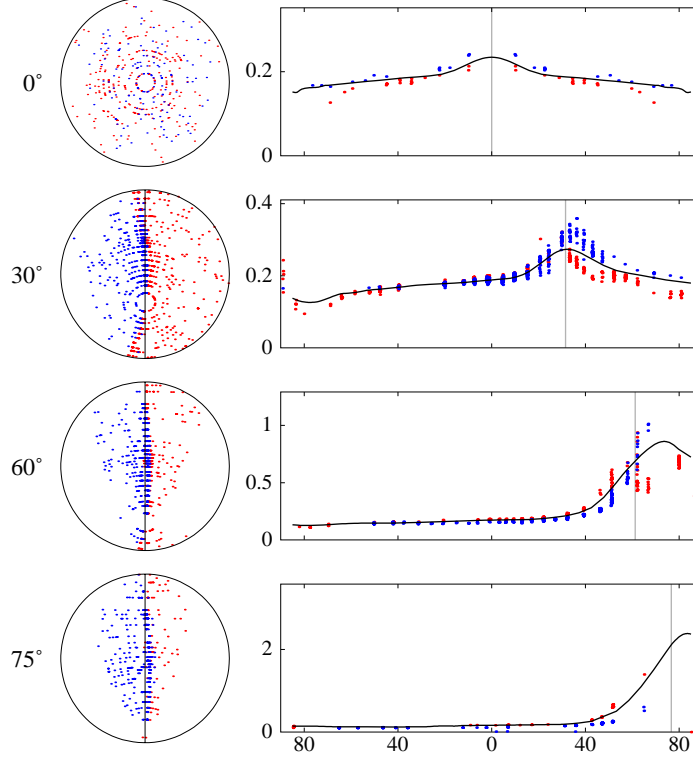
Fig. 7. Summary of results for the squash dataset. Left column: sample coverage; right column: raw data and local polynomial fit.

contain considerably more noise than do the gray cylinder data, as might be expected given the irregular nature of the surface, the noise in the 3D scanner data, and the finite precision of aligning the scanner data with the images. The slight surface blemishes visible in Figure 5 will affect the scatter plots, but have much less influence on the smoothed BRDF, as they cover only a small fraction of the surface.

6.3 Renderings

Plate 1 (see Appendix) shows some visual results of our reflectance measurements. To condense the data for tractable rendering times, the measurements were approximated with the representation presented by Lafortune et al. [16], using three cosine lobes (besides the diffuse term) for each BRDF.[4] Of course, the same data could be used in more sophisticated representations or for studies of surface optics and development of new parametric models.

The scene is rendered with Monte Carlo path tracing. It is illuminated by one overhead light source and two smaller light sources in the background, one on each side of the scene. All object surfaces show reflectances measured by our method: gray primer

---

[4]Because local polynomial reconstruction is slow and difficult to use for stochastic sampling, we did not use it for the renderings.
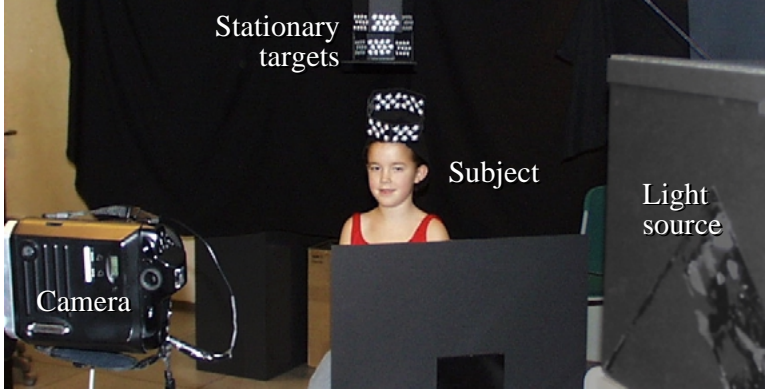
Fig. 8. Measurement setup.

on the floor, an unglazed ceramic on the flowerpot, blue and metallic red paints on the puzzle, and black felt on the hat. Even those surfaces that seem Lambertian in this image display distinctive directional behavior; the floor, for example, shows no visible shadows from the back lights in a Lambertian approximation.

6.4   Human skin

We adapted our method to measure the skin of human subjects. To our knowledge, our measurements are only the second angle-resolved reflectance measurements of living human skin; Cader and Jankowski have used a gonioreflectometer-like device to measure UV reflectance [2]. Our method, however, obtains many more BRDF samples in a short time (typically 20 minutes).

To accommodate a human subject, we attached our sample targets to a baseball cap worn backward by the subject. This fixes a field of targets to the subject's head; the geometry of the targets and the head together is obtained, as before, with the 3D range scanner. We selected a section of the forehead for derendering because it presents a relatively smooth, convex, uniform area of skin that is unlikely to deform during the measurement session. The hat positions the targets so as to make it easy to capture the forehead and all targets in each image.

Since the sample targets are no longer stationary, the stationary targets shown in Figure 8 provide a frame of reference for the positions of the camera, the subject, and the light source. Transforming everything into this frame for derendering allows us to accommodate minor movement of the subject's head without loss of measurement accuracy.

We measured skin BRDFs from several different subjects. Figure 9 shows coverage and incidence-plane slices of one of our data sets. Scatter is remarkably low, given the difficulties of precise geometric alignment and extracting reliable normals from noisy geometric data.

The BRDF itself is quite unusual; at small incidence angles it is almost Lambertian, but at higher angles strong forward scattering emerges. Note that the scale changes by a factor of 25 from the top row to the bottom. This scattering does not seem to correlate with the specular direction, so it cannot be simulated with a Phong function, nor would it be predicted by traditional rough-surface models such as those of Torrance and Sparrow [22] or He [13]. The only predictive model that might match these data is

11
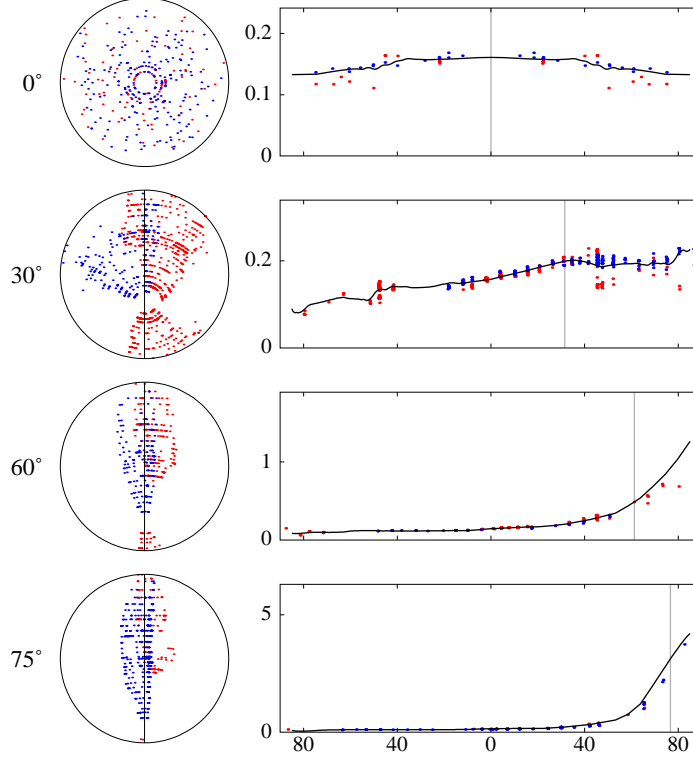
Fig. 9. BRDF of typical skin, showing coverage and scatter in raw data

the Monte Carlo simulation of Hanrahan and Krueger [11]; our data could be used to confirm or refine that method.

The renderings of Plate 2 (see Appendix) show the two extremes of our measurements to date: the BRDFs of a 43-year-old Caucasian male and a 23-year-old male from India, who exhibits not only a different skin color but also noticeably glossier skin.

7 Conclusion

This paper has described a simple technique that can measure the BRDFs of many materials using only a digital camera and a light source. We achieve accuracy rivaling that of a specialized gonioreflectometer but with much greater speed and resolution, and with one twentieth the equipment cost. In addition, the technique is versatile enough to measure living human skin. The technique is rapid because the two dimensions of a camera image sample two angular degrees of freedom instantaneously, leaving only one to be handled by sequential measurement. In a measurement session lasting under half an hour, our system can acquire hundreds of thousands of samples scattered over the full domain of an isotropic BRDF. The resulting data are internally consistent and agree closely with independent measurements.

Our technique demands samples with homogeneous BRDF, as do most traditional gonioreflectometers and almost all image-based techniques. We also require convex

curved samples; this complements the capabilities of more conventional methods, which only work with flat samples. Just as some materials are most readily available as flat samples (e.g. various building materials), others, including most organic objects, are only available in curved samples.

Acknowledgments

References

1. Standard practice for angle resolved optical scatter measurements on specular or diffuse surfaces. ASTM Standard E 1392-96.

2. A. Cader and J. Jankowski. Reflection of ultraviolet radiation from different skin types. *Health Physics*, 74(2):169–172, February 1998.

3. Raymond J. Castonguay. New generation high-speed high-resolution hemispherical scatterometer. In John C. Stover, editor, *SPIE Proceedings*, volume 1995, pages 152–165, July 1993.

4. J. H. Chandler and C. J. Padfield. Automated digital photogrammetry on a shoestring. *Photogrammetric Record*, 15(88):545–559, 1996.

5. J. Fan and I. Gijbels. *Local Polynomial Modeling and Its Applications*. Chapman and Hall, London, 1996.

6. Sing-Choong Foo. A gonioreflectometer for measuring the bidirectional reflectance of material for use in illumination computation. Master's thesis, Cornell University, 1997.

7. C. S. Fraser, M. R. Shortis, and G. Ganci. Multi-sensor system self-calibration. In *Videometrics IV*, pages 2–18. SPIE, October 1995.

8. A. Glassner, editor. *An Introduction to Ray Tracing*. Academic Press, London, 1989.

9. Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. The lumigraph. In *Computer Graphics (SIGGRAPH '96 Proceedings)*, pages 43–54, August 1996.

10. Anat Grynberg and Greg Ward. A new tool for reflectometry. Monograph 161, Lawrence Berkeley Laboratory, July 1990.

11. Pat Hanrahan and Wolfgang Krueger. Reflection from layered surfaces due to subsurface scattering. In *Computer Graphics (SIGGRAPH '93 Proceedings)*, pages 165–174, August 1993.

12. Ziad R. Hatab, John R. McNeil, and S. Sohail H. Naqvi. Sixteen-megabit dynamic random access memory trench depth characterization using two-dimensional diffraction analysis. *Journal of Vacuum Science and Technology B*, 13(2):174–181, March/April 1995.

13. Xiao D. He, Kenneth E. Torrance, Francois X. Sillion, and Donald P. Greenberg. A comprehensive physical model for light reflection. *Computer Graphics (SIGGRAPH '91 Proceedings)*, 25(4):175–186, July 1991.

14. Katsushi Ikeuchi and Kosuke Sato. Determining reflectance properties of an object using range and brightness image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(11):1139–1153, 1991.

15. Konrad F. Karner, Heinz Mayer, and Michael Gervautz. An image based measurement system for anisotropic reflection. *Computer Graphics Forum (Eurographics '96 Proceedings)*, 15(3):119–128, August 1996.

16. Eric P. F. Lafortune, Sing-Choong Foo, Kenneth E. Torrance, and Donald P. Greenberg. Non-linear approximation of reflectance functions. In *Computer Graphics (SIGGRAPH '97 Proceedings)*, pages 117–126, August 1997.

17. Rong Lu, Jan J. Koenderink, and Astrid M. L. Kappers. Optical properties (bidirectional reflectance distribution functions) of velvet. *Applied Optics*, 37(25):5974–5984, September 1998.

18. Stephen R. Marschner. *Inverse Rendering for Computer Graphics*. PhD thesis, Cornell University, 1998.

19. F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis. Geometric considerations and nomenclature for reflectance. Monograph 160, National Bureau of Standards (US), October 1977.

20. Yoichi Sato, Mark D. Wheeler, and Katsushi Ikeuchi. Object shape and reflectance modeling from observation. In *Computer Graphics (SIGGRAPH '97 Proceedings)*, pages 379–387, August 1997.

21. K. E. Torrance and E. M. Sparrow. Off-specular peaks in the directional distribution of reflected thermal radiation. *Transactions of the ASME*, 88:223–230, May 1966.

22. K. E. Torrance and E. M. Sparrow. Theory for off-specular reflection from roughened surfaces. *Journal of Optical Society of America*, 57(9):1105–1114, 1967.

23. Gregory J. Ward. Measuring and modeling anisotropic reflection. *Computer Graphics (SIGGRAPH '92 Proceedings)*, 26(2):265–272, July 1992.
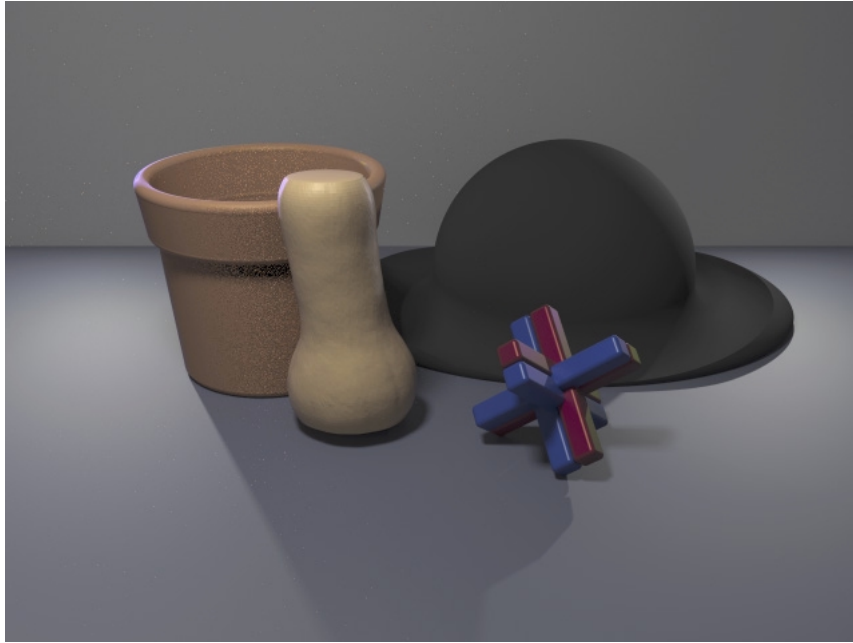
Plate 1. A rendered image showing a scene containing objects made of the measured materials.
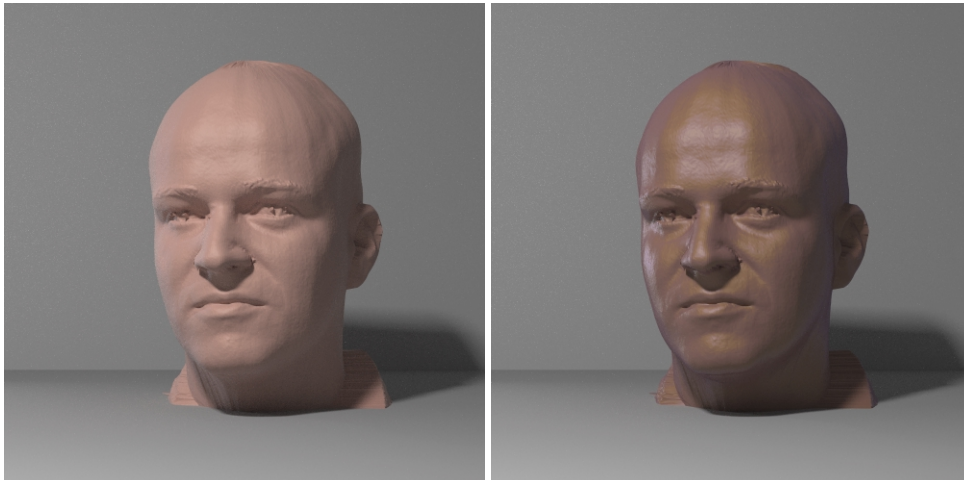

Plate 2. Rendered images showing BRDFs measured from two different subjects.

Diffraction Shaders

Jos Stam[*]

Alias | wavefront

Abstract

The reflection of light from surfaces is a fundamental problem in
computer graphics. Although many reflection models have been
proposed, few take into account the wave nature of light. In this
paper, we derive a new class of reflection models for metallic sur-
faces that handle the effects of diffraction. Diffraction is a purely
wave-like phenomenon and cannot be properly modeled using the
ray theory of light alone. A common example of a surface which ex-
hibits diffraction is the compact disk. A characteristic of such sur-
faces is that they reflect light in a very colorful manner. Our model
is also a generalization of most reflection models encountered in
computer graphics. In particular, we extend the He-Torrance model
to handle anisotropic reflections. This is achieved by rederiving,
in a more general setting, results from surface wave physics which
were taken for granted by other researchers. Specifically, our use
of Fourier analysis has enabled us to tackle the difficult task of an-
alytically computing the Kirchhoff integral of surface scattering.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional
Graphics and Realism—Color, shading, shadowing, and texture J.2
[Physical Sciences and Engineering]: Physics

Keywords: shading models, diffraction, Fourier transform, Kirch-
hoff theory, rough surface scattering, random processes

1 Introduction

The modeling of the interaction of light with surfaces is one of the
main goals of computer graphics. Over the last thirty years many
reflection models have been proposed that have considerably im-
proved the quality of computer graphics imagery. Almost all of
these reflection models are either empirical or based on the ray the-
ory of light. Surprisingly little attention has been devoted to the
purely wave-like character of light. It is well known from physical
optics that ray theory is only an approximation of the more funda-
mental wave theory. Why then has wave theory been so neglected ?
The main reason is that the ray theory is sufficient to visually cap-
ture the reflected field from many commonly occurring surfaces.
This observation is usually true when the surface detail is much
larger than the wavelength of visible light (roughly 0.5 microns
(10^{-6} meters)). Another reason for this neglect is the common
belief that models based on wave theory are computationally too

[*]Alias | wavefront, 1218 Third Ave, 8th Floor, Seattle, WA 98101, U.S.A.
jstam@aw.sgi.com

expensive to be of any use in computer graphics. In this paper we
challenge this point of view by introducing a new class of analytical
reflection models which simulate the effects of *diffraction*. Diffrac-
tion is a purely wave-like phenomenon which cannot be modeled
using the standard ray theory of light. Diffraction occurs when the
surface detail is comparable to the wavelength of light. A common
example of a surface that produces visible diffraction patterns is the
compact disk (CD). By rotating a CD under a steady light source,
one can fully appreciate the visual complexity of diffraction. To
capture these subtle changes in color and intensity requires a wave-
like description of light. In this paper we derive analytical reflection
models based on the wave theory that capture the effects of diffrac-
tion. In addition, our model is both easy to implement as a standard
"shader" and computationally efficient. The derivation which leads
to our new model, however, is not simple. This is because the wave
theory is mathematically much more complex than the ray theory
of light.

Scanning through the computer graphics literature, we found
only a few references which explicitly use the wave description
of light. In 1981 Moravec proposed solving the global illumina-
tion problem using the wave theory of light [11]. For his method
to give acceptable results, both a very fine resolution (on the order
of the wavelength of light) and a large ensemble of simulations (to
model incoherent natural light sources) are required. This makes his
approach unsuitable for practical computer graphics applications.
Later in 1985, Kajiya proposed to numerically solve the Kirchhoff
integral[1] to simulate the light reflected from anisotropic surfaces
[9]. His approach, although less ambitious than Moravec's, suffers
from the same limitations. In this context it would appear to be
more promising to solve directly for the coherence functions asso-
ciated with the waves, which are second order statistical averages
of the wave fields. Some work in this area has been pursued by Tan-
nenbaum et al. [20]. The coherence functions can also be employed
to define generalized radiances [23].

A more practical use of the wave theory in computer graphics is
to employ it to derive analytical reflection models. This approach,
which has a long history in the applied optics literature, e.g., [2],
was first seriously introduced to computer graphics by Bahar and
Chakrabarti [1]. Using Bahar's full wave theory, they were able to
fit analytical distributions to their computations for surfaces having
a large isotropic surface roughness. The full wave theory has the
advantage over the Kirchhoff theory in that it takes into account the
global shape of the object. However, in practice analytical expres-
sions are known only for simple objects such as spheres. Also the
global shapes of surfaces in computer graphical models are usu-
ally much larger than the wavelength of light. Later in 1991, He
and collaborators derived a general reflection model based on the
electro-magnetic wave theory to predict the reflection of light from
isotropic surfaces of any surface roughness [8]. At about the same
time, a very similar model was proposed in the computer vision
literature by Nayar [13]. As in Kajiya's work, these two models
are essentially based on the Kirchhoff approximation of surface re-
flection [2]. We also note that Blinn already used some asymptotic
results from Beckmann's monograph [3]. However, Blinn's model
does not account for wave-like effects.

---

[1]This integral will be defined more precisely below.
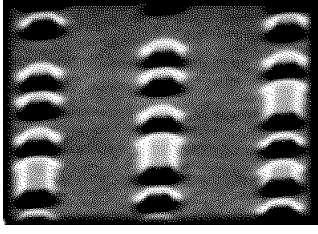
Figure 1: Close-up "view" of the micro-geometry of the surface of a compact disk.

Figure 2: Basic geometry of the surface wave reflection problem.

Although the analytical models just discussed are based on wave theory, none of them is able to capture the visual complexity of the light reflected off of a compact disk, for example. The main reason is that these models assume the surface detail to be isotropic, i.e., the surface "looks the same" in every direction. Interesting diffraction phenomena, however, occur mostly when the surface detail is highly *anisotropic*, viz. non-isotropic. Fig. 1 shows that this is certainly the case for the CD. Other examples include brushed metals and colorful diffraction gratings. In computer graphics, both empirical and ray optics models have been proposed to model the reflection from anisotropic surfaces [15, 17, 22]. However, since these models are not based on wave theory, they failed to capture the effects of diffraction. To the best of our knowledge, reflection models that handle colorful diffraction effects have not appeared in the computer graphics literature or in any commercially available graphics software before. The phenomenon of diffraction was used, however, by Nakamae et al. to model the fringes caused when viewing bright light sources through the pupil and eyelashes [12].

In this paper, we derive various analytical anisotropic reflection models using the scalar Kirchhoff wave theory and the theory of random processes. In particular, we show that the reflected intensity is equal to the spectral density of a simple function $p = e^{i\alpha h}$ of the (random) surface height h. We show that the spectral density can be computed for a large class of surfaces not considered in previous models. We believe that our approach is novel, since the "classic" monographs on scattering from statistical surfaces do not mention such an approach [2, 14].

Diffraction should not be confused with the related phenomenon of *interference*. Interference produces colorful effects due to the phase differences caused by a wave traversing thin media of different indices of refraction, e.g., a soap bubble. Interference effects, unlike diffraction, can be modeled using the ray theory of light alone [7].

To fully understand the derivations in this paper the reader should have a background in Fourier analysis, distribution theory and random processes. Due to a lack of space we refer the reader not versed in these areas to the relevant literature, e.g., [16, 24]. The reader might also want to consult the longer version of this paper available on the CDROM proceedings which contains appendices summarizing the main results from these disciplines. The remainder of this paper is organized as follows. A reader who is interested solely in implementing our new shaders can go directly to Section 6 where the model is stated "as is". Section 2 summarizes the main results from wave theory which are required in this paper. Section 3 presents our derivation. Subsequently, Sections 4 and 5 present several applications of our new reflection model. Section 6 addresses implementation issues and can be read without any advanced mathematical knowledge. Section 7 discusses several results created using our new shaders. Finally, Section 8 concludes, outlining possible directions for future research.
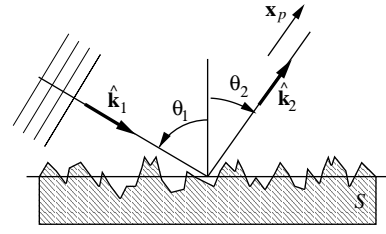
2 Wave Theory and Computer Graphics

In this section we briefly outline some results and concepts from the wave theory necessary to understanding the derivation of our reflection model. We employ the so-called "scalar wave theory of diffraction" [4]. In this approximation the light wave is assumed to be a complex valued scalar disturbance ψ. This theory completely ignores the polarization of light, so its results are therefore restricted to unpolarized light. Fortunately, most common light sources such as the sun and light bulbs are totally unpolarized. The waves generated by these sources also have the property that they fluctuate very rapidly over time. Typical frequencies for such waves are on the order of $10^{14} \ s^{-1}$. In practice this means that we cannot take accurate "snapshots" of a wave. Light waves are thus essentially random and only statistical averages of the wave function have any physical significance. The averaging, denoted by $\langle . \rangle$, can be interpreted either as an average over a long time period or equivalently (via ergodicity) as an ensemble average. An example of a statistical quantity associated with waves is the *irradiance*, $I = \langle |\psi|^2 \rangle$.

We also assume that the waves emanating from the source are stationary. This means that the wave is a superposition of independent monochromatic waves. Consequently, we can restrict our analysis to a wave having a definite wavelength λ associated with it. For visible light, the wavelengths range from the ultraviolet (0.3 microns) to the infrared (0.8 microns) region. Each of these waves satisfies a Helmholtz's wave equation:

$$\nabla^2 \psi + k^2 \psi = 0,$$

where k is the *wavenumber* equal to the reciprocal of the wavelength, $k = 2\pi/\lambda$.

The main task in the theory of diffraction is to solve this wave equation for different geometries. In our case we are interested in computing the reflected waves from various types of surfaces. More precisely, we want to compute the wave ψ_2 equal to the reflection of an incoming planar monochromatic wave $\psi_1 = e^{ik\hat{\mathbf{k}}_1 \cdot \mathbf{x}}$ traveling in the direction $\hat{\mathbf{k}}_1$ from a surface S. Fig. 2 illustrates this situation. The equation relating the reflected field to the incoming field is known as the *Kirchhoff integral*. This equation is a formalization of Huygen's well-known principle that states that if one knows the wavefront at a given moment, the wave at a later time can be deduced by considering each point on the first wave as the source of a new disturbance. This principle implies that once the field on the surface is known, the field everywhere else away from the surface can be computed. The field on the surface is usually related to the incoming field ψ_1 using the *tangent plane* approximation. For a planar surface, the wave theory predicts that a fraction F of the incoming light is specularly reflected. The fraction F is equal to the Fresnel factor for unpolarized light (see p. 48 of [4]). The tangent approximation states that the wave field on the surface is equal to the incoming field plus the field reflected off of the tangent plane at the surface point. Using this relation and the assumption that the "observation point" is sufficiently far removed from the surface, the

Kirchhoff integral is ([2], p. 22):

$$\psi_2 = \frac{ike^{ikR}}{4\pi R}\left(F\mathbf{v} - \mathbf{p}\right) \cdot \int_S \hat{\mathbf{n}}\, e^{ik\mathbf{v}\cdot\mathbf{s}}\, d\mathbf{s}, \tag{1}$$

where R is the distance from the center of the patch to the receiving point $\mathbf{x}_p$, $\hat{\mathbf{n}}$ is the normal of the surface at $\mathbf{s}$ and the vectors

$$\mathbf{v} = \hat{\mathbf{k}}_1 - \hat{\mathbf{k}}_2 \ \text{ and } \ \mathbf{p} = \hat{\mathbf{k}}_1 + \hat{\mathbf{k}}_2.$$

The vector $\hat{\mathbf{k}}_2$ is equal to the unit vector pointing from the origin of the surface towards the point $\mathbf{x}_p$. To obtain this result it is also assumed that the Fresnel coefficient F is replaced by its average value over the normal distribution of the surface and can thus be taken out of the integral. Eq. 1 is the starting point for our derivation. We will show below that it can be evaluated analytically for a large class of interesting surface profiles. Before we do so, we will also outline how the reflected wave is related to the usual reflection nomenclature used in computer graphics.

In computer graphics, the reflected properties are often modeled using the bidirectional reflection distribution function (BRDF) which is defined as the ratio of the reflected radiance to the incoming irradiance. In this paper we will provide in every case the BRDF corresponding to our reflection model. The relationship between the BRDF and the waves can be shown to be [21]:

$$\text{BRDF} = \lim_{R\to\infty} \frac{R^2}{A\cos\theta_1}\frac{\langle|\psi_2|^2\rangle}{\langle|\psi_1|^2\rangle\cos\theta_2}, \tag{2}$$

where A is the area of the surface and θ_1 and θ_2 are the angles that the vectors $\hat{\mathbf{k}}_1$ and $\hat{\mathbf{k}}_2$ make with the vertical direction (see Fig. 2).

3 Derivation

In this section we demonstrate that the Kirchhoff integral of Eq. 1 can be computed analytically. In this paper, as in related work, we restrict ourselves to the reflection of waves from height fields. We assume that the surface is defined as an elevation over the (x, y) plane. Each surface point is then parameterized by the equation

$$\mathbf{s} \to \mathbf{s}(x, y) = (x, y, h(x, y)), \tag{3}$$

where $h(x, y)$ is a (random) function. The normal to the surface at each point then admits an analytical expression in terms of the partial derivatives h_x and h_y of the height function:

$$\hat{\mathbf{n}}\, d\mathbf{s} \to \hat{\mathbf{n}}(x, y)\, d\mathbf{s} = (-h_x(x, y), -h_y(x, y), 1)\, dxdy.$$

Introducing the notation $\mathbf{v} = (u, v, w)$, it then follows directly that the integral in Eq. 1 acquires the following form:

$$\mathbf{I}(ku, kv) = \int\int (-h_x, -h_y, 1)\, e^{ikwh} e^{ik(ux+vy)}\, dxdy. \tag{4}$$

The integrand can be further simplified by noting that:

$$(-h_x, -h_y, 1)e^{ikwh} = \frac{1}{ikw}(-p_x, -p_y, ikwp),$$

where

$$p(x, y) = e^{ikwh(x,y)}. \tag{5}$$

We now use the common assumption (e.g., [2, 8]) that the integration can be extended over the entire plane. This assumption is usually justified on the grounds that the surface detail is much smaller than the distances over which the surface is viewed. In

doing so we observe that the integral of Eq. 4 is now a two-dimensional Fourier transform:

$$\mathbf{I}(ku, kv) = \int\int \frac{1}{ikw}(-p_x, -p_y, ikwp)e^{ik(ux+vy)}dxdy.$$

This important observation can be implemented. Let $P(ku, kv)$ be the Fourier transform of the function p. We observe that differentiation with respect to x (resp. y) in the Fourier domain is equivalent to a multiplication of the Fourier transform by $-iku$ (resp. $-ikv$). This leads to the simple relationship

$$\mathbf{I}(ku, kv) = \frac{1}{w}\, P(ku, kv)\, \mathbf{v}.$$

We have thus related the integral of Eq. 1 directly to the Fourier transform of the function p. Now, since

$$\left(F\mathbf{v} - \mathbf{p}\right) \cdot \mathbf{v} = 2F\left(1 - \hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_2\right),$$

the scattered wave of Eq. 1 is equal to

$$\psi_2 = \frac{ike^{ikR}}{2\pi R}\frac{F(1 - \hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_2)}{w}P(ku, kv). \tag{6}$$

This result shows that the scattered wave field is proportional to the Fourier transform of a simple function of the surface height. Consequently, from Eq. 2, it follows that the BRDF is

$$\text{BRDF} = \frac{k^2 F^2 G}{4\pi^2 A w^2}\,\langle|P(ku, kv)|^2\rangle, \tag{7}$$

where

$$G = \frac{(1 - \hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_2)^2}{\cos\theta_1 \cos\theta_2}. \tag{8}$$

This result and the derivation that leads to it are remarkably simple when compared to derivations that do not employ the Fourier transform, e.g., [2]. More importantly, this treatment is more general, since we have not made any assumptions regarding the function P yet.

We now specialize our results for a homogeneous random function [16]. Homogeneity is a natural assumption since we are interested in the bulk reflection from a large portion of the surface having a certain profile. For example, the portion of the CD depicted in Fig. 1 could have been taken from any part of the CD. However, and this is important, we do not assume that the surface is isotropic. This is mainly where we depart from previous wave physics models in computer graphics. Referring again to Fig. 1 we observe that the CD is clearly not isotropic.

From the definition of the function p (Eq. 5) it follows immediately that this function is also homogeneous. In particular, its correlation function depends only on the separation between two locations:

$$C_p(x', y') = \langle p^*(x, y)p(x + x', y + y')\rangle - |\langle p\rangle|^2,$$

independently of the location (x, y). The Fourier transform of the correlation function is known as the *spectral density* ([16], p. 338):

$$S_p(u, v) = \int\int C_p(x', y')e^{i(ux'+vy')}dx'dy'.$$

The spectral density is a non-negative function which gives the relative contribution of each wavenumber (u, v) to the entire energy. We now show that the average in Eq. 7 is directly related to the spectral density. Indeed, let $\xi = (x, y)$, $\xi' = (x', y')$ and $\zeta = (ku, kv)$, then

$$\langle|P(\zeta)|^2\rangle = \langle P^*(\zeta)P(\zeta)\rangle = \int\int\langle p^*(\xi)p(\eta)\rangle e^{-i\zeta\cdot\xi}e^{i\zeta\cdot\eta}d\xi d\eta.$$
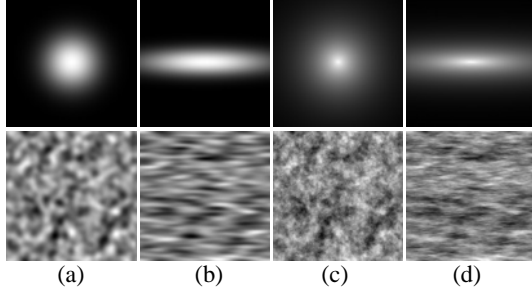
Figure 3: Effect of the correlation function on the appearance of a random surface. The pictures at the top show plots of different correlation functions with a realization of the corresponding random surface below. The surface types are: (a) isotropic Gaussian, (b) anisotropic Gaussian, (c) isotropic fractal and (d) anisotropic fractal.

With the change of variable $\eta = \xi + \xi'$, this integral becomes

$$
\iint \langle p^*(\xi) p(\xi+\xi') \rangle e^{i\zeta \cdot \xi'} d\xi d\xi' =
$$

$$
\int d\xi \int (C_p(\xi') + |\langle p \rangle|^2) e^{i\zeta \cdot \xi'} d\xi' = A \left(S_p(\eta) + 4\pi^2 \delta(\zeta) \right),
$$

where δ is the two-dimensional Dirac delta function. Consequently, the average in Eq. 7 is a function of the spectral density of the function p:

$$
\frac{1}{A} \langle |P(ku, kv)|^2 \rangle = S_p(ku, kv) + 4\pi^2 |\langle p \rangle|^2 \delta(ku, kv).
$$

Substituting this result back into Eq. 7 we get:

$$
\text{BRDF} = \frac{F^2 G}{w^2} \left(\frac{k^2}{4\pi^2} S_p(ku, kv) + |\langle p \rangle|^2 \delta(u, v) \right), \quad (9)
$$

where we have used the fact that $\delta(ku, kv) = \delta(u, v)/k^2$ [24]. Eq. 9 is the main theoretical result of this paper. It shows that the reflection from a random surface is proportional to the spectral density of the random function e^{ikwh}. In the next two sections we apply this result to the derivation of reflection models for various types of surfaces.

4 Anisotropic Rough Surfaces

4.1 General Case

Every surface depicted in Fig. 3 is a realization of a *Gaussian random process*. These processes are entirely defined by their corresponding correlation function depicted in the upper part of Fig. 3. From the figure it is clear that the correlation function determines the general appearance of the random surface. Radially symmetrical correlation functions correspond to isotropic surfaces, c.f., surfaces (a) and (c), while the derivative of the correlation function at the origin also determines smoothness of the surfaces. Consequently, surfaces (a) and (b) are smooth, while surfaces (c) and (d) have a fractal appearance. In this section we further clarify the fact that the reflection from these surfaces is intimately related to the correlation function. Gaussian random processes have the nice property that their characteristic functions admit analytical expressions [2]. These functions are exactly what we require in order to compute the spectral density S_p and the variance $|\langle p \rangle|^2$ appearing in Eq. 9. Indeed, for Gaussian random processes these quantities

are related to their surface height counterparts as follows. Firstly, we have the following identities ([16], p. 255):

$$
\langle p \rangle = \langle e^{ikwh} \rangle = e^{-g/2} \quad \text{and} \quad (10)
$$

$$
C_p(x, y) = e^{-g} \left(e^{g C_h(x,y)} - 1 \right), \quad (11)
$$

where $g = (kw\sigma_h)^2$, and σ_h is the standard deviation of the height fluctuations. Secondly, the spectral density S_p is the Fourier transform of the correlation function C_p ([16], p. 338). To compute this Fourier transform analytically we can use the expansion of the exponential function into an infinite series [2]:

$$
e^{g C_h(x,y)} = \sum_{m=0}^{\infty} \frac{g^m}{m!} C_h(x, y)^m.
$$

Then using the linearity of the Fourier transform, we can compute the spectral density as

$$
S_p = \mathcal{F}\{C_p\} = e^{-g} \sum_{m=1}^{\infty} \frac{g^m}{m!} \mathcal{F}\{(C_h)^m\}. \quad (12)
$$

This requires the computation of the Fourier transform of the surface correlation to a power m. We now give analytical results for the two correlation functions corresponding to the surfaces depicted in Fig. 3. These surfaces are defined by the following two correlation functions:

$$
C_1(x, y) = e^{-\frac{x^2}{T_x^2} - \frac{y^2}{T_y^2}} \quad \text{and} \quad C_2(x, y) = e^{-\sqrt{\frac{x^2}{T_x^2} + \frac{y^2}{T_y^2}}}.
$$

In all cases, the *correlation lengths* T_x and T_y control the anisotropy of the surface. Fig. 3.(a) and (b) both correspond to the correlation function C_1. This function is infinitely smooth at the origin, which accounts for the smoothness of the corresponding surfaces. In Fig. 3.(a) $T_x = T_y$ and the surfaces are isotropic. Most previous wave-based models considered only the isotropic case. Fig. 3.(c) and (d) correspond to the correlation function C_2. The corresponding surfaces have a fractal appearance. They are thus good models for very rough materials. In the results section we will see that these surfaces give rise to reflection patterns which are visually different from the smooth case.

For each correlation function, we can compute its Fourier transforms to a power m analytically. They are equal to

$$
D_1^m = \frac{\pi T_x T_y}{m} e^{-\frac{U^2 + V^2}{4m}} \quad \text{and} \quad D_2^m = \frac{2\pi T_x T_y m}{(m^2 + U^2 + V^2)^{3/2}}, \quad (13)
$$

respectively, where $U = kuT_x$ and $V = kvT_y$. By substituting these expressions back into the infinite sum of Eq. 12, we get an analytical expression for the BRDF:

$$
\text{BRDF} = \frac{F^2 G}{w^2} e^{-g} \left(\frac{k^2}{4\pi^2} \sum_{m=1}^{\infty} \frac{g^m}{m!} D^m + \delta(u, v) \right), \quad (14)
$$

where D^m is any one of the functions of Eq. 13.

4.2 Discussion

In this section we demonstrate that most previous models in computer graphics are special cases of our new shading model.
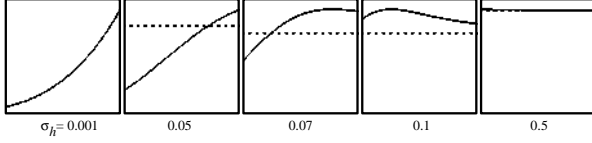
Figure 4: Plots of the BRDF for k ranging from the infrared $(8.06\mu^{-1})$ to the ultraviolet region $(16.53\mu^{-1})$. The reflection is in the specular direction: $\theta_1 = \theta_2 = 45^o$. The plots show the effect of the standard deviation σ_h on the color of the reflection. For low deviations the reflection is bluish, while for higher roughness it tends to flatten out. The dashed line is the geometrical optics approximation.

Born Approximation

When $g << 1$, the infinite sum appearing in Eq. 12 can be truncated to its first term. This is equivalent to the approximation $e^{ikwh} \approx 1 + ikwh$ often taken in physical theories. This approximation should be valid whenever the scales of the surfaces are much smaller than the wavelength of light.

$$\mathrm{BRDF}_{Born} = F^2 G \ e^{-g} \left(\frac{\sigma_h^2 k^4}{4\pi^2} S_h(ku, kv) + \delta(u, v) \right).$$

This result is described in the *Handbook of Optics* [5]. Notice that the BRDF is dependent on the fourth power of the inverse of the wavelength. This means that generally "bluish" light is more strongly scattered than "reddish" light. These surfaces should therefore have a bluish appearance. An interesting feature of this approximation is that one can actually "see" the spectral density of the random surface in its highlight, i.e., any of the plots in Fig. 3 (top).

Geometrical Optics

In the opposite limit when $g >> 1$, an approximate expression for the sum of Eq. 12 can also be derived. This case corresponds to a situation usually encountered in computer graphics when the surface detail is much larger than the wavelength of light. For large g, the Fourier integral only depends on the behavior of the function e^{gC_h} near the origin (see [2, 1] for details):

$$e^{gC_h(x,y)} \approx e^g e^{-g(x^2/T_x^2 + y^2/T_y^2)}.$$

The Fourier transform of this function can be computed analytically and is equal to:

$$S_p(ku, kv) = \frac{\pi T_x T_y}{g} \ e^{-\frac{U^2}{4g}} \ e^{-\frac{V^2}{4g}}.$$

The BRDF in this case is equal to ($e^{-g} \approx 0$):

$$\mathrm{BRDF}_{geom} = \frac{F^2 G}{4\pi w^4 r_x r_y} \ e^{-\frac{u^2}{4w^2 r_x^2}} \ e^{-\frac{v^2}{4w^2 r_y^2}}, \qquad (15)$$

where $r_x = \sigma_h/T_x$ and $r_y = \sigma_h/T_y$. This distribution is a generalization of the isotropic distributions found in the Blinn and Cook-Torrance models where there is only one roughness parameter "m". In fact, our model closely resembles Ward's anisotropic reflection model [22]. As in the Cook-Torrance model, BRDF_{geom} is only dependent on the wavelength of light through the Fresnel factor F, as there is no other explicit dependence on wavelength: k does not explicitly appear in the distribution.

rect(x) rect(y)   g(x)   g(x) rect(x) rect(y)
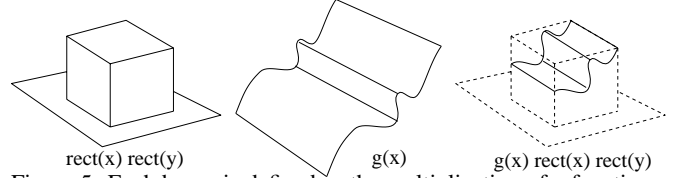
Figure 5: Each bump is defined as the multiplication of a function $g(x, y)$ with the product of box-like functions.
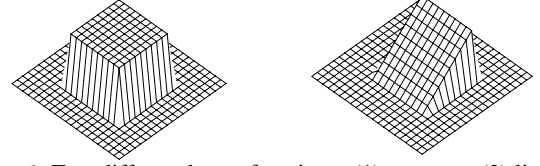
Figure 6: Two different bump functions: (1) constant, (2) linear in one coordinate.

Isotropic Distributions

The He-Torrance [8] and the Nayar [13] reflection models are obtained when our model is restricted to the class of isotropic surfaces corresponding to Fig. 3.(a). Using our result for the correlation function C_1 with $T_x = T_y$, we essentially recover both of these models. It is worth noting that one of the versions of the He-Torrance model handles polarization effects while our model doesn't. This is because they used the vector valued version of the Kirchhoff integral. However, in practice it seems He-Torrance have only used their unpolarized version to create the pictures accompanying their paper. The dependence on wavelength (as in our model) is a function of the Fresnel factor F and the function $k^2 S_p(ku, kv)$. In Fig. 4 we illustrate the dependence of this function on wavenumber k for different surface deviations σ_h. The reflection goes from a k^2 dependence to a flat spectrum. Notice that in the midrange we actually get a small yellowish hue. The figure also demonstrates that for $\sigma_h > 0.5$ the geometrical optics model, shown as a dashed line, is a very good approximation. In practice we have found that whenever $g > 10$ the pictures generated with the geometrical optics approximation are visually indistinguishable from pictures generated using the exact model.

5   Diffraction from Periodic-like Surfaces

We now turn to an application that most clearly demonstrates the power of our new reflection model.

Many surfaces have a micro-structure that is made out of similar "bumps". A good example is a compact disk which has small bumps that encode the information distributed over each "track". Fig. 1 is a magnified view of the actual surface of a compact disk. Notice in particular that the distribution of bumps is random along each track but that the tracks are evenly spaced. In this section we derive general formulae for certain shapes of bumps, and then specialize the results for a CD-shader.

symbol	description	size
h_0	height of a bump	$0.15 \ \mu m$
a	width of a bump	$0.5 \ \mu m$
b	length of a bump	$1 \ \mu m$
Δx	separation between the tracks	$2.5 \ \mu m$
ν_y	density of bumps on each track	$0.5 \ (\mu m)^{-1}$

Table 1: Typical dimensions of a compact disk.

We assume that the surface is given by a superposition of bumps:

$$h(x, y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} b(x - x_n, y - y_m), \qquad (16)$$

where the locations (x_n, y_m) are assumed to be either regularly spaced or randomly (Poisson) distributed. To handle the two cases simultaneously, we assume that x_n is evenly spaced and that y_n is Poisson distributed. Extensions to the case where both locations are evenly spaced or where both are Poisson distributed should be obvious from our results. Let Δx be the constant spacing between the x-locations: $x_n = n\Delta x$. The random Poisson distribution of the locations y_m is entirely specified by a density ν_y of bumps per unit length. The function $b(x, y)$ appearing in Eq. 16 is a "bump function": a function with (small) finite support. We will assume that the bump function has the following simple form:

$$b(x, y) = h_0 \, g(x/a)\mathrm{rect}(x/a)\mathrm{rect}(y/b), \qquad (17)$$

where a, b and h_0 define the width, length and height of each bump respectively ($a \leq \Delta x$). Typical values of these parameters for a CD are provided in Table 1. The function rect equals one on the interval $[-1/2, 1/2]$ and zero elsewhere. Fig. 5 illustrates our definition of a bump. Our derivation is valid for arbitrary g, however, we provide an analytical expression only for the following two functions:

$$g^0(x) = 1 \quad \text{and} \quad g^1(x) = 1/2 + x. \qquad (18)$$

The bumps corresponding to these functions are depicted in Fig. 6. The function g^0 is a good approximation of the bumps found on a CD and the function g^1 can be used to model diffraction gratings.

The function $p(x, y)$ defined by Eq. 5 in our case is equal to:

$$p(x, y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \phi((x - x_n)/a, (y - y_m)/b), \qquad (19)$$

where $\phi(x, y) = e^{i\alpha g(x)}\mathrm{rect}(x)\mathrm{rect}(y)$ and $\alpha = kwh_0$. We dropped a constant term "1" that accounts for the space between the bumps and only adds a delta spike in the specular direction.

A simple computation shows that the Fourier transform of the function $p(x, y)$ is equal to

$$P(u, v) = \sigma_x(u)\sigma_y(v) \, ab \, \Phi(au, bv),$$

where $\Phi(u, v)$ is the Fourier transform of $\phi(x, y)$ and

$$\sigma_x(u) = \sum_{n=-\infty}^{\infty} e^{iun\Delta x} \quad \text{and} \quad \sigma_y(v) = \sum_{m=-\infty}^{\infty} e^{ivy_m}. \qquad (20)$$

To compute the spectral density of Eq. 9 we note that:

$$S_p(u, v) = (ab)^2 |\Phi(au, bv)| \, |\sigma_x(u)|^2 \, S_{\sigma_y}(v).$$

The spectral density and the average of the sum of random Poisson distributed locations are both equal to the density ν_y (see [16] p. 561):

$$S_{\sigma_y}(v) = \nu_y \quad \text{and} \quad \langle \sigma_y \rangle = \nu_y.$$

The sum of evenly spaced location x_n is a bit harder to deal with. First we need the following two results from the theory of distributions (see pp. 54-55 of reference [24]):

$$\sum_{n=-\infty}^{\infty} e^{iun} = 2\pi \sum_{n=-\infty}^{\infty} \delta(u - 2\pi n) \quad \text{and} \quad \delta(sz + t) = \frac{1}{s}\delta(z + t/s),$$

where $s > 0$ and t are real numbers. The first of these two equalities is known as "Poisson's summation formula". Using these results we can express the square of the sum σ_x in terms of delta distributions only:

$$|\sigma_x(u)|^2 = \frac{(2\pi)^2}{\Delta x^2} \sum_{n=-\infty}^{\infty} \delta(u - 2\pi n/\Delta x).$$

We can now compute the spectral density S_p by putting all these computations together:

$$S_p(ku, kv) = b^2\nu_y\lambda \sum_{n=-\infty}^{\infty} |\Phi_n(kv)|^2\delta(u - n\lambda/\Delta x), \qquad (21)$$

where

$$|\Phi_n(kv)|^2 = \frac{a^2}{\Delta x^2}|\Phi(2\pi na/\Delta x, kv)|^2. \qquad (22)$$

The function $|\Phi|^2$ can be computed analytically for the two bumps depicted in Fig. 6:

$$|\Phi^0(au, bv)|^2 = 2(1 - \cos(\alpha))\mathrm{sinc}^2(au/2)\mathrm{sinc}^2(bv/2), \quad (23)$$

$$|\Phi^1(au, bv)|^2 = \big(\mathrm{sinc}^2(\alpha_0/2) - 2\mathrm{sinc}(\alpha_0/2)\times$$
$$\mathrm{sinc}(au/2)\cos(\alpha/2) + \mathrm{sinc}^2(au/2)\big)\mathrm{sinc}^2(bv/2), \quad (24)$$

where $\alpha_0 = \alpha + au$. Putting all these pieces together we get the following expression for the BRDF:

$$\mathrm{BRDF} = \frac{F^2 G}{w^2}b^2\nu_y \sum_{n=-\infty}^{\infty} |\Phi_n(kv)|^2\delta(u - n\lambda/\Delta x)(k + \nu_y\delta(u)).$$

6 Implementation

We have implemented our reflection models as various shaders in our MAYA animation system. Any model created in that package can be rendered using our new shaders. The fact that our shaders have been included in a commercial product should be a sufficient proof of their practicality.

As in [9], we model the anisotropy of the surface by assigning an orthonormal frame at each point of the surface. In the case of a parametric surface, the most natural choice for this frame is to take the normal and the two vectors tangent to the iso-parameter lines. We have also added an additional rotation angle to the frame around the normal. When this angle is texture mapped, it allows us to create effects such as brushed metal (Fig. 8.(a)).

The general form of our shader is

$$\mathrm{BRDF} = |F(\theta_1')|^2 \, G(\mathbf{k}_1, \mathbf{k}_2) \, S(\mathbf{k}_1, \mathbf{k}_2) \, (D(\mathbf{v}, \lambda) + r\mathrm{Env}),$$

where F is the Fresnel factor [6], S is a shadowing function [8], G is a geometrical factor defined by Eq. 8 in Section 3 and D is a distribution function that is related to the micro-geometry of the surface. The function "Env" returns the color in the mirror direction of $\mathbf{k}_2$ from an environment map and the factor r accounts for how much the surface reflects direct illumination. The vector $\mathbf{v} = (u, v, w)$ is the angle midway between $-\mathbf{k}_1$ and $\mathbf{k}_2$. The Fresnel factor is evaluated at the angle θ_1' that the direction $\mathbf{k}_1$ makes with the vector $\mathbf{v}$. The Fresnel factor varies with the index of refraction of the metallic surface and is wavelength dependent [6]. We do not use the He-Torrance shadowing function since it is restricted to isotropic surfaces. Instead, we employ a model introduced by Sancer [18]. For convenience, we have included this model in Appendix A. The distribution D is the most important component of our model and is now described in more detail.

In the previous sections we have derived distribution functions for both the random surfaces depicted in Fig. 3 and for periodic-like profiles such as the one shown in Fig. 1. When the surface is random, the distribution is defined by the three parameters σ_h, T_x and T_y. The variance σ_h^2 models the average height fluctuations of the surface and the parameters T_x and T_y model the amount of correlation of the micro-surface in the directions of the local frame. See Section 3 for further details on these quantities. When $T_x = T_y$, the surface is isotropic. In the most general case, the distribution D is computed by the infinite sum appearing in Eq. 14. In Appendix B, we provide a stable implementation of this sum. As pointed out in Section 4.2, the sum is very well approximated by the geometrical optics approximation of Eq. 15, when $g = (kw\sigma_h)^2$ is large (see also Fig. 4). The factor "r" is equal to $\exp(-g)$. The smoother the surface, the more indirect illumination is directly reflected off of it.

The implementation of periodic-like profiles giving rise to colorful diffraction patterns is different. When evaluating the distribution D, the values u and v (and w) are determined by the incoming and outgoing angles. The incoming light is usually assumed to be an incoherent sum of many monochromatic waves whose number is proportional to the distribution $L(\lambda)$ of the light source. To determine the intensity and the color of the light reflected in the outgoing direction, we first compute the wavelengths λ_n for which $L(\lambda)$ is non zero and for which the delta spikes in Eq. 21 are non-zero. This only occurs when $\lambda_n = \Delta x u / n$ and $n \neq 0$. When $n = 0$, all wavelengths contribute intensities in the specular direction $u = 0$. In general, visible light is comprised only of waves with wavelengths between $\lambda_{min} = 0.4\mu m$ and $\lambda_{max} = 0.7\mu m$. This means that the indices n are constrained to lie in the range $\Delta x u [1/\lambda_{max}, 1/\lambda_{min}]$ if $u > 0$ and in the range $\Delta x u [1/\lambda_{min}, 1/\lambda_{max}]$ when $u < 0$. Once these wavelengths are determined, the red, green and blue components of the distribution D are computed as follows

$$D_{rgb} = b^2 \nu_y \sum_{n=N_{min}}^{N_{max}} \frac{1}{\lambda_n} \text{Spec}_{rgb}(\lambda_n) L(\lambda_n) \left| \Phi_n \left(\frac{2\pi v}{\lambda_n} \right) \right|^2 ,$$

where Spec_{rgb} is a function that for each wavelength returns the corresponding color. This function can either be constructed from psychophysical experiments or simply set by an animator as a "ramp". In our implementation we constructed a ramp function from standard RGB response curves. See Eq. 22 for a definition of the function Φ_n.

7 Results

Once the shaders were implemented in MAYA, it was an easy task to generate results demonstrating the power of our new shading model. In Fig. 7 we show the effect of some of the parameters of our model on the appearance of the surfaces. In each rendering we chose to have a spectrally flat Fresnel factor to demonstrate the dependence of the distribution on wavelength. For the Gaussian correlations the reflection is more bluish for small roughness and becomes whiter for larger roughness, in accordance with the analysis of Section 4.2. The reflection from fractal surfaces is quite interesting: bluish for small roughness, then yellowish for intermediate roughness and finally white for large roughness. The third row of spheres exhibits the effect of the separation and twist angle parameters of our diffraction shader. We used a different texture map for the twist angle of each one of the three "diffraction cones" at the bottom of Fig. 7.

Fig. 8 shows several renderings created in this manner. In each case we have texture mapped the directions of anisotropy to add more interesting visual detail. Fig. 8.(a) demonstrates that this can be employed to create a "brushed metal" look. In Fig. 8.(b) we

textured both the roughness and the degree of anisotropy of the surface. Fig. 8.(c) is a picture of a CD illuminated by a directional light source. Notice that all the highlights appear automatically in the correct places when the data from Table 1 is used. Fig. 8.(d) is an example of the use of our diffraction grating model. Notice all the subtle coloring effects that result (especially when viewing the corresponding animation). These colorful effects would be hard to model by trial and error without properly modeling the wave properties of light.

The effects of the anisotropy and of diffraction are most pronounced in an animation when moving either the object or the light sources. For this reason we have included some animations on the CDROM proceedings.

8 Conclusions

In this paper we have proposed a new class of reflection models that take into account the wave-like properties of light. For the first time in computer graphics, we have derived reflection models that properly simulate the effects of diffraction. We have shown that our models can be easily implemented as standard shaders in our MAYA animation software. Our derivations, while mathematically involved, are simpler and more general than previously published results in this area. In particular, our use of the Fourier transform has proven to be a very powerful tool in deriving new reflection models.

In future work, we hope to extend our model to an even wider class of surfaces by relaxing some of the assumptions in our model. Presently, our model only accounts for the reflection from metallic surfaces and ignores multiple-scattering. It would be interesting to derive more general models that take into account subsurface scattering by waves. It seems unlikely that the effects of multiple scattering might be captured by an analytical model. An alternative would be to fit analytical models to either the results from a Monte-Carlo wave simulation or to experimentally measured data.

As well, we wish to extend our work to the computation of the fluctuations of the intensity field [10]. In this manner we can compute exact texture maps for given surface profiles. We could achieve this by deriving analytical expressions for the higher order statistics of the reflected intensity field. More specifically, we hope to extend our previous work on stochastic rendering of density fields to surfaces [19].

Acknowledgments

A A Shadowing Function

The shadowing function used in He's model applies only to isotropic surfaces. For this reason we have used a different model derived by Sancer [18]. The shadowing function is valid for a Gaussian random surface having a correlation function C_h and standard deviation σ_h:

$$S = \begin{cases} (C_1 + 1)^{-1} & \text{if } u = v = 0 \text{ and } \theta_1 \leq \theta_2 \\ (C_2 + 1)^{-1} & \text{if } u = v = 0 \text{ and } \theta_2 \leq \theta_1 \\ (C_1 + C_2 + 1)^{-1} & \text{else} \end{cases} ,$$

where

$$C_i = \sqrt{\frac{2|\beta_i|}{\pi}} \tan \theta_i \exp \left(-\frac{\cot^2 \theta_i}{2|\beta_i|} \right) - \mathrm{erfc} \left(\frac{\cot \theta_i}{\sqrt{2|\beta_i|}} \right)$$

$$\beta_i = \sigma_h^2 \left(C_{h,xx} \cos^2 \phi_i + C_{h,xy} \sin 2\phi_i + C_{h,yy} \sin^2 \phi_i \right),$$

where $i = 1, 2$ and $C_{h,xx}$ is the second derivative with respect to x of the correlation function at the origin. Since the derivatives of the correlation function depend on the correlation lengths T_x and T_y, this clearly shows that this shadowing function takes into account the anisotropy of the surface.

B   Computing Infinite Sums

The following piece of code will compute the distribution of reflected light from the surface:

```
compute_D (lambda,u,v,w,s_h,Tx,Ty)
  k = 2*PI/lambda;
  g = k*s_h*w; g *= g;
  if ( g>10 )
    return D_geom(u,v,w,s_h/Tx,s_h/Ty);
  tmp=1; sum=log_g=0;
  for(m=1;abs(tmp)>EPS||m<3*g;m++) {
    log_g += log(g/m); tmp = exp(log_g-g);
    sum += tmp*D(m,k*u,k*v,Tx,Ty);
  }
  return lambda*lambda*sum
```

The function `D()` is any one of the functions of Equation 13. This routine is a stable implementation of the infinite sum appearing in Equation 14. A naive implementation of the sum results in numerical overflows. The condition "`m<3*g`" is there to make sure that we do not exit the loop too early. This is an heuristic which has worked well in practice.

References

[1] E. Bahar and S. Chakrabarti. Full-Wave Theory Applied to Computer-Aided Graphics for 3D Objects. *IEEE Computer Graphics and Applications*, 7(7):46–60, July 1987.

[2] P. Beckmann and A. Spizzichino. *The Scattering of Electromagnetic Waves from Rough Surfaces*. Pergamon, New York, 1963.

[3] J. F. Blinn. Models of Light Reflection for Computer Synthesized Pictures. *ACM Computer Graphics (SIGGRAPH '77)*, 11(3):192–198, August 1977.

[4] M. Born and E. Wolf. *Principles of Optics. Sixth (corrected) Edition*. Cambridge University Press, Cambridge, U.K., 1997.

[5] E. L. Church and P. Z. Takacs. *Chapter 7. Surface Scattering*. In *Handbook of Optics (Second Edition). Volume I: Fundamentals, Techniques and Design*. McGraw Hill, New York, 1995.

[6] R. L. Cook and K. E. Torrance. A Reflectance Model for Computer Graphics. *ACM Computer Graphics (SIGGRAPH '81)*, 15(3):307–316, August 1981.

[7] J. S. Gondek, G. W. Meyer, and J. G. Newman. Wavelength dependent reflectance functions. In *Computer Graphics Proceedings, Annual Conference Series, 1993*, pages 213–220, 1994.

[8] X. D. He, K. E. Torrance, F. X. Sillion, and D. P. Greenberg. A Comprehensive Physical Model for Light Reflection. *ACM Computer Graphics (SIGGRAPH '91)*, 25(4):175–186, July 1991.

[9] J. T. Kajiya. Anisotropic Reflection Models. *ACM Computer Graphics (SIGGRAPH '85)*, 19(3):15–21, July 1985.

[10] W. Krueger. Intensity Fluctuations and Natural Texturing. *ACM Computer Graphics (SIGGRAPH '88)*, 22(4):213–220, August 1988.

[11] H. P. Moravec. 3-D Graphics and the Wave Theory. *ACM Computer Graphics (SIGGRAPH '81)*, 15(3):289–296, August 1981.

[12] E. Nakamae, K. Kaneda, and T. Nishita. A Lighting Model Aiming at Drive Simulators. *ACM Computer Graphics (SIGGRAPH '90)*, 24(4):395–404, August 1990.

[13] S. K. Nayar, K. Ikeuchi, and T. Kanade. Surface Reflection: Physical and Geometrical Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):611–634, July 1991.

[14] J. A. Ogilvy. *Theory of Scattering from Random Rough Surfaces*. Adam Hilger, Bristol, U.K., 1991.

[15] Tomohiro Ohira. A Shading Model for Anisotropic Reflection. *Technical Report of The Institute of Electronic and Communication Engineers of Japan*, 82(235):47–54, 1983.

[16] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Systems Science Series, New York, 1965.

[17] P. Poulin and A. Fournier. A Model for Anisotropic Reflection. *ACM Computer Graphics (SIGGRAPH '90)*, 24(4):273–282, August 1990.

[18] M. I. Sancer. Shadow Corrected Electromagnetic Scattering from Randomly Rough Surfaces. *IEEE Transactions on Antennas and Propagation*, AP-17(5):577–585, September 1969.

[19] J. Stam. Stochastic Rendering of Density Fields. In *Proceedings of Graphics Interface '94*, pages 51–58, Banff, Alberta, May 1994.

[20] D. C. Tannenbaum, P. Tannenbaum, and M. J. Wozny. Polarization and Birefringency Considerations in Rendering. In *Computer Graphics Proceedings, Annual Conference Series, 1994*, pages 221–222, July 1994.

[21] K. Tomiyasu. Relationship Between and Measurement of Differential Scattering Coefficient (σ^0) and Bidirectional Reflectance Distribution Function (BRDF). *SPIE Proceedings. Wave Propagation and Scattering in Varied Media*, 927:43–46, 1988.

[22] G. J. Ward. Measuring and Modelling Anisotropic Reflection. *ACM Computer Graphics (SIGGRAPH'92)*, 26(2):265–272, July 1992.

[23] E. Wolf. Coherence and Radiometry. *Journal of the Optical Society of America*, 68(1), January 1978.

[24] A. H. Zemanian. *Distribution Theory and Transform Analysis: An Introduction to Generalized Functions, with Applications*. Dover, New York, 1987.
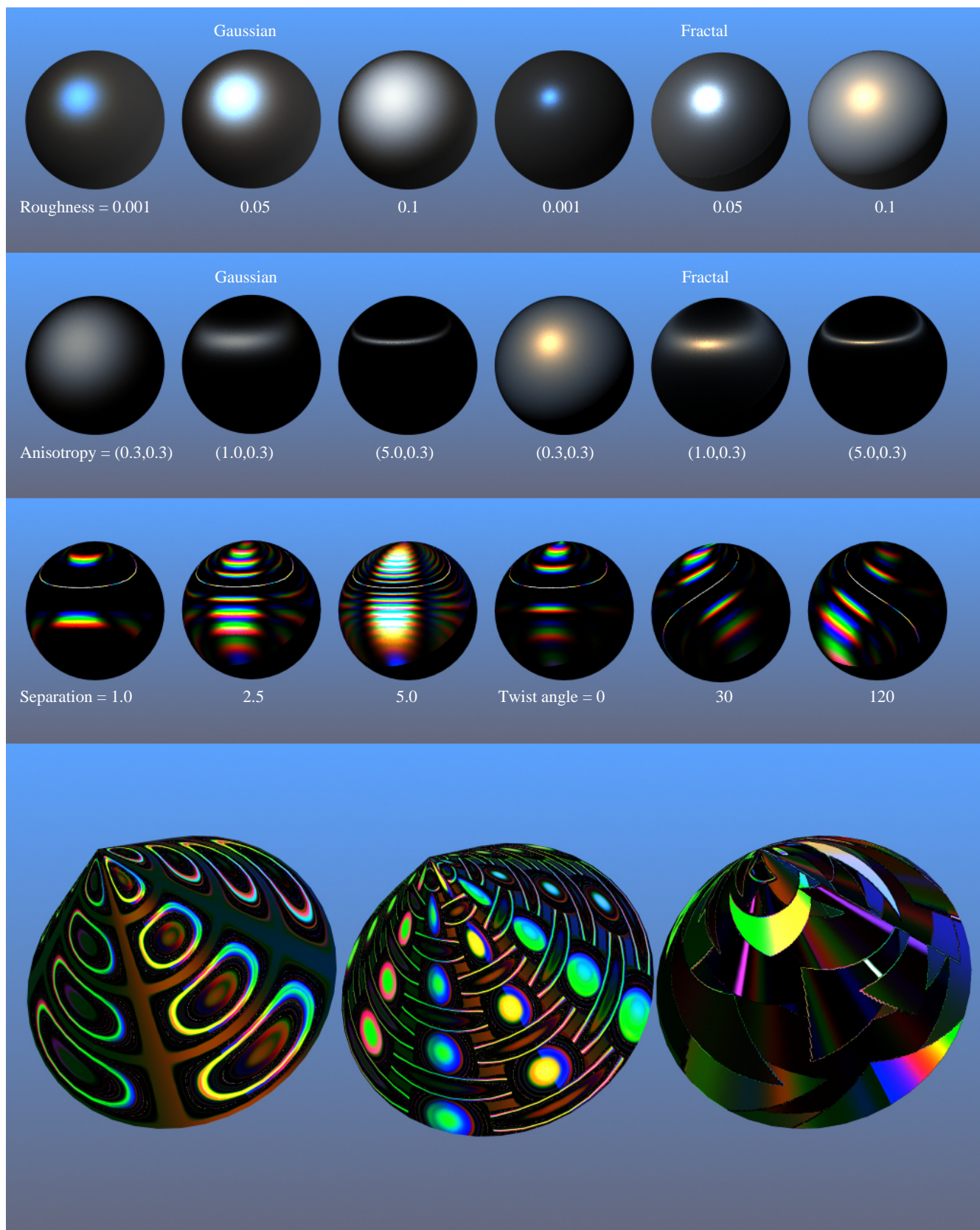
Figure 7: Effect of some of the parameters.
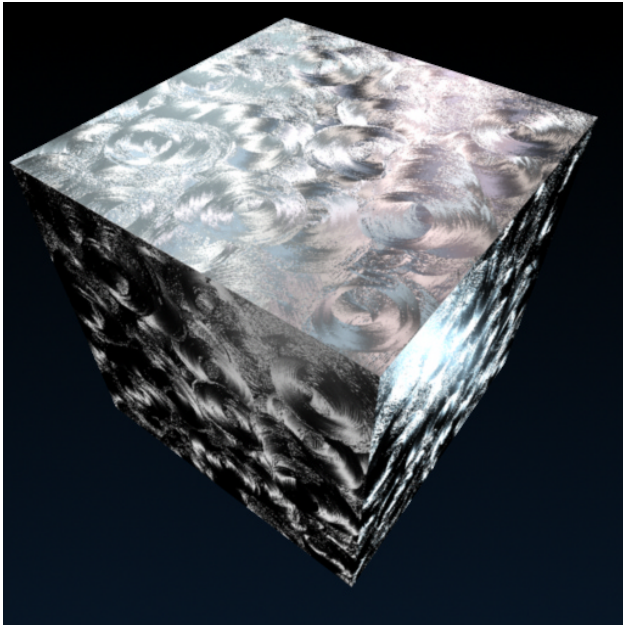
(a)
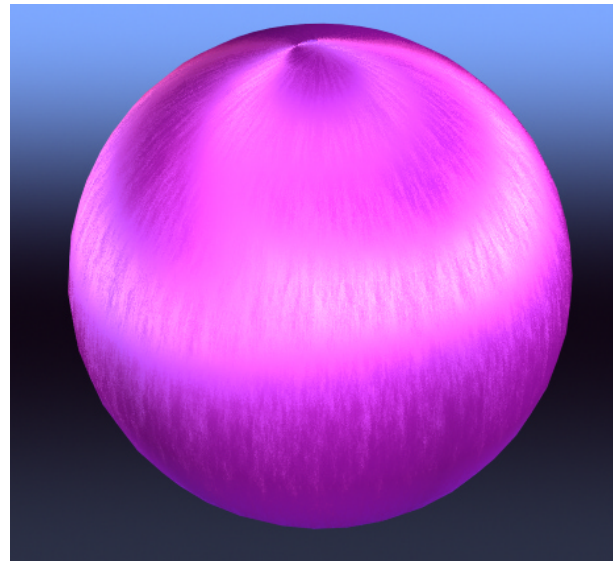
(b)

(c)
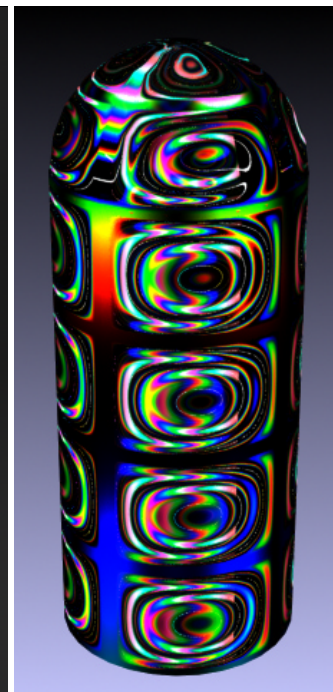
(d)

Figure 8: More pictures.

Reflectance Models with Fast Importance Sampling

László Neumann, Attila Neumann, László Szirmay-Kalos

Department of Control Engineering and Information Technology,
Technical University of Budapest,
Budapest, Műegyetem rkp. 11, H-1111, HUNGARY
Email: szirmay@iit.bme.hu, neumann@mail.datanet.hu

Abstract

We introduce a physically plausible mathematical model for a large class of BRDFs. The new model is as simple as the well-known Phong model, but eliminates its disadvantages. It gives a good visual approximation for many practical materials: coated metals, plastics, ceramics, retro-reflective paints, anisotropic and retro-reflective materials, etc. Because of its illustrative properties it can be used easily in most commercial software and because of its low computational cost it is practical for virtual reality. The model is based on a special basic BRDF definition, which meets the requirements of reciprocity and of energy conservation. Then a class of BRDFs is constructed from this basic BRDF with different weight functions. The definition of such weight functions requires the user to specify the profile of the highlights, from which the weight function is obtained by derivation. It is also demonstrated how importance sampling can be used with the new BRDFs.

Keywords: *Rendering, reflection models, BRDF, albedo, anisotropy, retro-reflective models, coated metals, energy mirror, Monte-Carlo quadrature, importance sampling.*

1. Introduction

A central problem in photorealistic rendering is the study of the optical material models characterised by *Bi-directional Reflectance Distribution Functions*, or BRDFs.

BRDF models can be classified according to how they correspond to the real physical phenomenon of light-surface interaction.

Global illumination algorithms require that the BRDFs do not violate physics. Such shading models must satisfy both reciprocity and energy balance, and are called *physically plausible*[1]. Physically plausible models can be constructed empirically using simple functions and enforcing properties of the reciprocity and energy balance, or they can be derived from more detailed physical models describing the microstructure of surfaces or the scattering of electromagnetic waves[2].

The classical Lambert model of diffuse reflection and the Torrance-Sparrow model[3] of specular reflection are physically founded. Other physically based models used so far are the Cook-Torrance[4] and He-Torrance[5] models, which are good approximations for metals and plastics but are very expensive computationally. The Cook-Torrance model has been generalised for anisotropic surfaces by Kajiya[6]. The Cook-Torrance model has also been simplified and made appropriate for practical implementations by Ward[7] and Schlick[20].

The original Phong[8], Phong-Blinn[9] and Whitted[10] models, on the other hand, are empirical models. These models are not physically plausible since they do not fulfil the law of reciprocity and the principle of the conservation of energy. The physically plausible Oren-Nayar model[11] characterises different materials with an appropriate retro-reflective component, including ceramics, walls, foams, etc. Painted coatings are often fitted to the Beard-Maxwell empirically based model[12]. The NIST database also applies this model to represent measurement data. Recently, Lafortune et.al.[13] proposed the combination of general cosine models to obtain a versatile family of BRDFs.

Based on the preliminary version of this paper[14], Tobler et.al.[15] proposed a new BRDF that is particularly good for producing artistic contour effects.

During rendering, the BRDF models are usually used for two different purposes. On the one hand, in local illumination formulae the output radiance in a direction is computed from the incident radiance and from the surface orientation, which requires the evaluation of the BRDF for a single view vector and light vector pair. On the other hand, to improve Monte-Carlo random walk methods by applying importance sampling, directions should be generated that follow a probability density of the BRDF multiplied by a cosine function.

A practically useful BRDF must support both tasks. Unfortunately, physically based BRDFs contain many complex formulae, thus not only applying accurate importance sampling is hopeless, but even their simple evaluation is rather time consuming.

This is one reason why the Phong model is still in common use despite of its deficiencies. In global illumination algorithms, it is usually modified to satisfy reciprocity[16, 1] and to allow simple importance sampling[17]. The main disadvantage of the physically plausible Phong model is evident for "grazing angles" where it becomes dark. If the exponent parameter is high, this darkening is true from two different aspects. At grazing angles both the radiance and the total reflected power sharply decrease. The missing power becomes noticeable in images generated by global illumination algorithms.

This paper intends to develop a class of BRDF models. The proposed models are physically plausible, but not physically based. It can be viewed as a mathematical construction that is simple to compute and is particularly efficient for importance sampling.

Specific members of the new class can be defined in a new way. The computational cost of the new BRDFs is comparable to that of the Phong model. The model is able to approximate the visual properties of metals, plastics, reflective, retro-reflective and anisotropic materials.

2. Basic BRDF

The complex models to be proposed by this paper are based on a simple, elementary BRDF model. This *basic BRDF* is just a mathematical tool, we do not intend to use it directly for rendering.

In the formulae the classical notations are used: $\mathbf{L}$ and $\mathbf{V}$ are the unit vectors to the light source and to the viewpoint, $\mathbf{N}$ is the unit normal vector of the surface element. The $\mathbf{L}'$ and $\mathbf{V}'$ vectors are the results of mirroring the $\mathbf{L}$ and $\mathbf{V}$ vectors across the normal vector $\mathbf{N}$. The subscript P indicates the orthogonal projection of the given unit vectors - or their endpoints - onto the base plane as shown in figure 1.
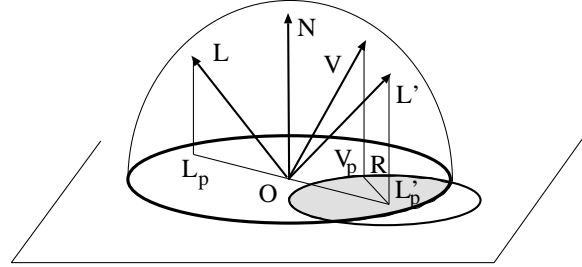
Figure 1: *Geometry of the BRDF definition*

The basic BRDF is based on simple geometric constructs, namely on circles on the base plane. Before giving the definition of the proposed basic BRDF, two notations are introduced.

Firstly, let $C(\mathbf{L}'_P, R)$ denote the common part of two circles, namely the unit circle with centre in $\mathbf{O}$ called the base circle and the circle of radius R and centre $\mathbf{L}'_P$.

Secondly, let the $m(\mathbf{L}, \mathbf{V})$ metric be the distance of the $\mathbf{L}'_P$ and $\mathbf{V}_P$ points. It can be shown that

$$m(\mathbf{L}, \mathbf{V}) = |\mathbf{V}_P - \mathbf{L}'_P| = |\mathbf{h} - (\mathbf{N} \cdot \mathbf{h})\mathbf{N}| \qquad (1)$$

where $\mathbf{h} = \mathbf{L} + \mathbf{V}$. It must be emphasised that $\mathbf{h}$ is not the same as the half-way unit vector denoted usually by $\mathbf{H}$, since $\mathbf{h}$ is *not normalised*.

Note that function m is symmetric, that is

$$m(\mathbf{L}, \mathbf{V}) = m(\mathbf{V}, \mathbf{L})$$

since the distance of $\mathbf{L}'_P$ and $\mathbf{V}_P$ is the same as the distance between $\mathbf{V}'_P$ and $\mathbf{L}_P$.

Using these notations, the basic BRDF corresponding to parameter R ($0 \le R \le 2$) is defined by:

$$f^*_r(R, \mathbf{L}, \mathbf{V}) = \begin{cases} c(R) \text{ if } m(\mathbf{L}, \mathbf{V}) \le R, \\ \\ 0 \text{ otherwise.} \end{cases} \qquad (2)$$

where $c(R)$ is a constant depending on R.

Note that $m(\mathbf{L}, \mathbf{V}) \le R$ means that the projection of $\mathbf{V}$ is in the circle of radius R, which is centred by the mirror of projection $\mathbf{L}$, i.e.

$$\mathbf{V}_P \in C(\mathbf{L}'_P, R).$$

Thus the BRDF also depends on R, which is shown by including it as the first parameter. The star character is to emphasise the fact that this is a basic BRDF from which other BRDFs will be constructed.
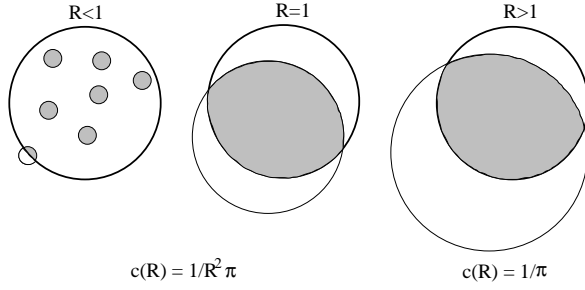
Figure 2: *Possible relations to the base circle*

The basic BRDF is a constant $c(R)$ if the projection of the mirroring of **L** onto the centre of the directional sphere and the projection of **V** are not further than R.
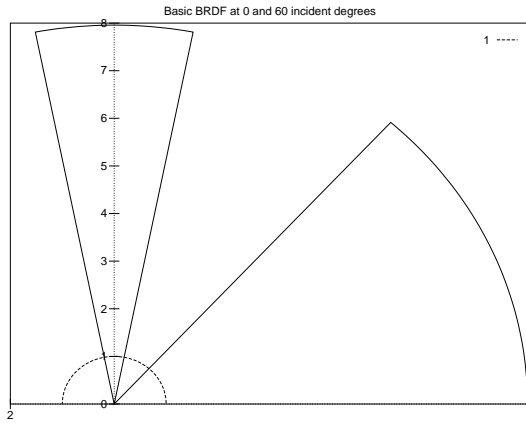
Figure 3: *Basic BRDF for $R = 0.2$*

The computation of the basic BRDF function is quite simple:

```
BasicBRDF(R, V, N, L)
    h = L + V
    m² = |h − (N · h)N|²
    if (m² ≤ R²) then return c(R)
    else                 return 0
end
```

2.1. Reciprocity

The basic BRDF is reciprocal, which is represented by identity

$$f_r^*(R, \mathbf{L}, \mathbf{V}) = f_r^*(R, \mathbf{V}, \mathbf{L})$$

since according to definition (2) it is equivalent to

$$m(\mathbf{L}, \mathbf{V}) = m(\mathbf{V}, \mathbf{L}).$$

2.2. Energy conservation

2.2.1. Albedo function

The power reflected into the half-space from the point light source of unit irradiance situated in direction **L** is called the direction-dependent *albedo*, or the *directional hemispherical reflectance*:

$$a(\mathbf{L}) = \int_\Omega f_r(\mathbf{L}, \mathbf{V}) \, (\mathbf{N} \cdot \mathbf{V}) \, d\omega_\mathbf{V}. \qquad (3)$$

To satisfy energy conservation, the reflected power cannot exceed the incoming power, that is

$$a(\mathbf{L}) \leq 1. \qquad (4)$$

The albedo of the basic BRDF can be easily determined. In formula (3) the multiplication with $(\mathbf{N} \cdot \mathbf{V}) = \cos \omega_\mathbf{V}$ corresponds to the *projection onto the base plane*. Thus, according to definitions (2) and (3) the albedo is proportional to the area of $C(\mathbf{L}_P', R)$, that is

$$a_R(\mathbf{L}) = c(R) \cdot \text{Area}(C(\mathbf{L}_P', R)). \qquad (5)$$

The area of $C(\mathbf{L}_P', R)$ is maximum when $\mathbf{L} = \mathbf{N}$, i.e. $\mathbf{L}_P' = \mathbf{O}$. If $R < 1$, this means that the circle is not truncated. If $R \geq 1$, then the common part coincides with the base circle. Since

$$\text{Area}(C(\mathbf{L}_P', R)) \leq \text{Area}(C(\mathbf{O}, R)), \qquad (6)$$

if in equation (2) we choose

$$c(R) = \frac{1}{\text{Area}(C(\mathbf{O}, R))} = \frac{1}{(\pi \cdot \min(1, R^2))}, \qquad (7)$$

then the albedo

$$a_R(\mathbf{L}) = \frac{\text{Area}(C(\mathbf{L}_P', R))}{\text{Area}(C(\mathbf{O}, R))} \qquad (8)$$

will not exceed 1, that is, the basic BRDF will conserve energy.
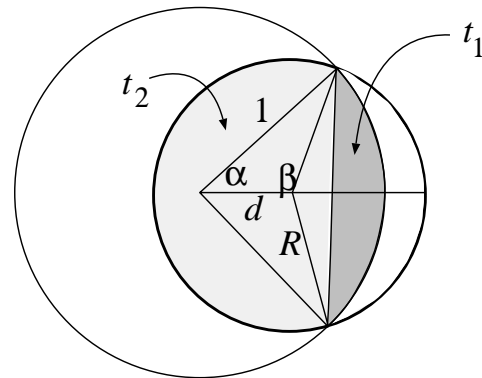
Figure 4: *Computation of the intersection of two circles*

To compute $\text{Area}(C(\mathbf{L}'_P, R))$, the intersection of two circles must be determined. Let the distance of the two circles be d. If $d < 1 - R$ or $d > 1 + R$, then the solution is trivial. For the intersecting case, we can use the following formulae that uses the notation of figure 4:

$$\text{Area}(C(\mathbf{L}'_P, R)) = t_1 + t_2 =$$

$$1 \cdot (\alpha - \cos\alpha \cdot \sin\alpha) + R^2 \cdot (\beta - \cos\beta \cdot \sin\beta). \quad (9)$$

The angle α and β can, in turn, be determined using the theorem of cosine angles:

$$d^2 + 1^2 - 2d \cdot \cos\alpha = R^2$$
$$d^2 + R^2 - 2d \cdot R \cdot \cos\beta = 1. \quad (10)$$

The direction-dependent albedo of the basic BRDF will reach the imposed maximum of 1 for vertical illumination, i.e. for $\mathbf{L} = \mathbf{N}$.
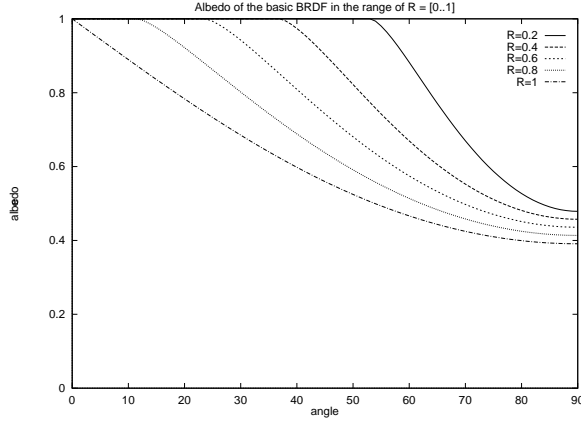
Figure 5: *Albedo of the basic BRDF for small R values ($R \leq 1$)*

When $R < 1$, then the albedo equals to the ratio of the area of the possibly truncated circle and the area of the complete circle of radius R. For $R \geq 1$ it is the ratio of the area of the possibly truncated base circle and the area of the whole base circle (π). Thus, we always have $a_R(\mathbf{N}) = 1$, furthermore it is easy to show that

$$0.391 < \frac{2}{3} - \frac{\sqrt{3}}{2\pi} \leq a_R(\mathbf{L}) \leq 1$$

holds. The minimum point is at horizontal illumination ($(\mathbf{N} \cdot \mathbf{L}) = 0$) and for $R = 1$.

For small values of R, the albedo is generally 1, and even at the margins of the base circle it hardly decreases under half. This is due to the fact that in

this case the circle is generally not truncated, only for illumination at grazing angles.

The $R = 0$ situation is the marginal case corresponding to a special mirror. In this case $C(R)$ becomes a Dirac-delta function.

The $R = 2$ case corresponds to the ideal diffuse model, where also $a(\mathbf{L}) \equiv 1$.
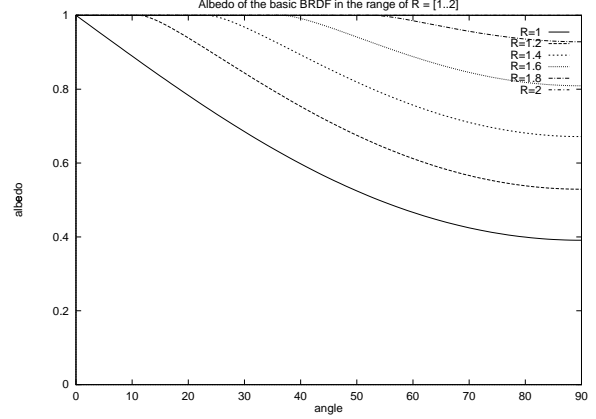
Figure 6: *Albedo of the basic BRDF for large R values ($1 \leq R \leq 2$)*

2.2.2. Mean albedo

In addition to the direction-dependent albedo defined by equation (3), we can also introduce the notion of the *mean albedo*. This direction-independent and dimensionless value is equal to the "reflectivity" for diffuse materials.

For non-diffuse materials, it can be defined as the weighted mean value of direction-dependent albedos:

$$a_{\text{mean}} = \frac{1}{\pi} \cdot \int_{\Omega} a(\mathbf{L}) \, (\mathbf{N} \cdot \mathbf{L}) \, d\omega_{\mathbf{L}}, \quad (11)$$

where $\mathbf{L}$ is the running vector of the directional hemisphere and the weighting factor is $(\mathbf{N} \cdot \mathbf{L}) = \cos\omega_{\mathbf{L}}$.

In the case of uniform illumination on the half-space, the value of the mean albedo is equal to the ratio of the reflected and the incident power. Figure 7 illustrates the values of a_{mean} for the basic BRDFs as a function of R. The mean albedo is minimal at $R = 1$.

The BRDF and the albedo characterise a material from two different aspects. The BRDF is used to describe the "reflected radiance" and it is a scalar which depends on both $\mathbf{L}$ and $\mathbf{V}$. The albedo depending only on direction $\mathbf{L}$ serves for the description of the "reflected power". The mean albedo (which does not depend on $\mathbf{L}$ and $\mathbf{V}$) characterises the average reflectivity of the BRDF.
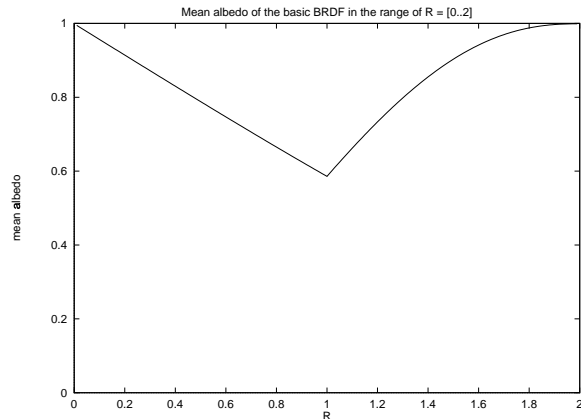
Figure 7: *Mean albedo of the basic BRDF*

Examining these characteristics, we can see the essential difference between the behaviour of the ideal mirror obtained by the new BRDF and that obtained as a limit case from the $(\mathbf{N} \cdot \mathbf{H})^n$ Phong-Blinn model without diffuse component[18].

For the ideal metal mirror, both the mean and the direction dependent albedos are 1, that is $a_{\mathrm{mean}} = 1$ and $a(\mathbf{L}) \equiv 1$.

However, for the "Phong-mirror", which is the result of letting n go to infinity, we have $1 \geq a(\mathbf{L}) \geq 0$ and $a_{\mathrm{mean}} = 0.5$. In the case of a Phong mirror, the reflected energy tends to zero for grazing angles.

The mirror case of the new model is not equivalent to a metal mirror since at grazing angles the metallic mirrors reflect the total energy by increasing the BRDF by a factor of $1/(\mathbf{N} \cdot \mathbf{L})$, while in the new model the BRDF is constant and the solid angle of the reflectance is increased by $1/(\mathbf{N} \cdot \mathbf{L})$. Thus the new model is a ideal "energy mirror" but not ideal "radiance mirror" for higher incident angles.

2.3. Importance sampling

Suppose that we want to calculate the radiance of a point in direction $\mathbf{V}$ due to the illumination of the scene. The input and output *radiances* $[\mathrm{Wm}^{-2}\mathrm{sr}^{-1}]$ are denoted by L^{in} and L^{out}, respectively. The radiance in the viewing direction $\mathbf{V}$ is:

$$L^{\mathrm{out}}(\mathbf{V}) = \int_{\Omega} L^{\mathrm{in}}(\mathbf{L}) \, f_r(\mathbf{L}, \mathbf{V}) \, (\mathbf{N} \cdot \mathbf{L}) \, d\omega_{\mathbf{L}}. \quad (12)$$

This integral can be approximated by the Monte-Carlo method. Using importance sampling to reduce the variance, the incoming $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_M$ directions are not uniformly distributed, but their probability

density distribution function is at least approximately proportional to the integrand. Since $L^{\mathrm{in}}(\mathbf{L})$ is usually not known a-priori, we can sample $\mathbf{L}$ with probability proportional to the

$$R(\mathbf{L}, \mathbf{V}) = f_r(\mathbf{L}, \mathbf{V}) \, (\mathbf{N} \cdot \mathbf{L})$$

weighted BRDF function. This function has an interesting illustrative content, according to the "inverse situation". If there were a light source in the viewing direction, then the distribution of the energy on the half-space after the reflection would follow the function R.

Function $R \cdot d\omega_{\mathbf{L}}$ also defines the conditional probability that a photon is reflected into the solid angle $d\omega_{\mathbf{L}}$ given that it comes from $\mathbf{V}$ [19].

Due to the fact that f_r is reciprocal, we have:

$$a(\mathbf{V}) = \int_{\Omega} f_r(\mathbf{L}, \mathbf{V}) \, (\mathbf{N} \cdot \mathbf{L}) \, d\omega_{\mathbf{L}}, \quad (13)$$

which is exactly the value of the albedo function in viewing direction $\mathbf{V}$. For sampling of $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_M$ directions, we have to use the

$$d_{\mathbf{V}}(\mathbf{L}) = \frac{f_r(\mathbf{L}, \mathbf{V}) \cdot (\mathbf{N} \cdot \mathbf{L})}{a(\mathbf{V})} \, d\omega_{\mathbf{L}} \quad (14)$$

probability density function. The approximation of the integral in equation (12) is:

$$L^{\mathrm{out}}(\mathbf{V}) = a(\mathbf{V}) \cdot E\left[L^{\mathrm{in}}(\mathbf{L})\right] \approx \frac{a(\mathbf{V})}{M} \cdot \sum_{k=1}^{M} L^{\mathrm{in}}(\mathbf{L}_k). \quad (15)$$

In the case of isotropic BRDFs the albedo functions can be tabulated in a one-dimensional table. When calculating multiple interreflections, the main computational problem is the generation of the directions from equation (14) following distribution $d_{\mathbf{V}}$.

2.4. Importance sampling for the basic BRDF

2.4.1. Sampling in special cases

For the introduced basic BRDFs, we can easily generate directions with a distribution of equation (14).

Let us consider first a circle of radius $R < 1$ and assume that this circle does not intersect the base circle. In this case $C(\mathbf{V}'_P, R)$ is identical with the circle with centre in $\mathbf{V}'_P$ and of radius R. The value of the albedo is 1 because the circle is not truncated.

Let us generate M number of 2D points inside this circle with a uniform distribution. These will be the $(\mathbf{L}_k)_P$ $(k = 1, \ldots, M)$ projection points in the plane of projection. Since the projection automatically includes the "cosine" factor in $d(\mathbf{L})$, re-projecting these

points onto the unit hemisphere's surface, we obtain the $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_M$ directions of the desired distribution.
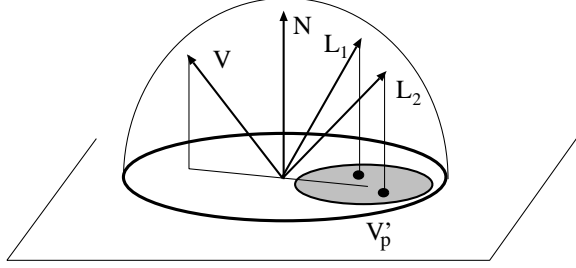
Figure 8: *Sampling when $R < 1$ and the circles do not intersect*

For a given unit viewing vector $\mathbf{V}$, the projection on the base plane and the mirroring are realised in the following way:

$$\mathbf{V}'_P = -\mathbf{V}_P = -(\mathbf{V} - (\mathbf{N} \cdot \mathbf{V})\mathbf{N}). \tag{16}$$

The formula for re-projecting the $\mathbf{L}_P$ sample point which is inside the circle of centre $\mathbf{V}'_P$ and of radius R onto the hemisphere is:

$$\mathbf{L} = \mathbf{L}_P + \sqrt{1 - (\mathbf{L}_P \cdot \mathbf{L}_P)} \cdot \mathbf{N}. \tag{17}$$
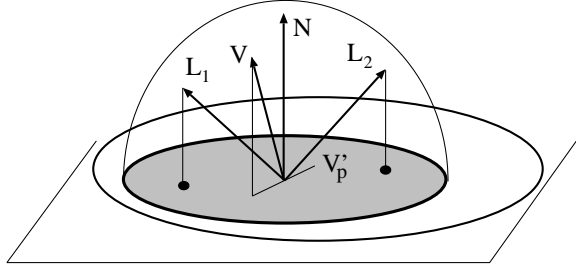
Figure 9: *Sampling when $R \geq 1$ and the circles do not intersect*

The situation is similar if $R > 1$ and the circle contains the complete unit base circle. Now the uniformly distributed points can be generated on the whole base circle. This case corresponds to the generation of directions following a "cosine distribution" for diffuse materials.

2.4.2. Generating uniformly distributed points in a circle

To generate uniformly distributed points inside a circle, we can either use an area preserving mapping of the square onto the circle, or apply rejection sampling. In rejection sampling the points are generated in a square that encloses the circle and those points that are outside the circle are ignored.

2.4.3. Sampling in the general case

In the cases discussed so far, the domain where the sample points are generated was a complete circle and the value of the albedo was 1. In the general case the $C(\mathbf{V}'_P, R)$ region is the common part of two circles. To realise importance sampling, we have to generate uniformly distributed points inside this region.
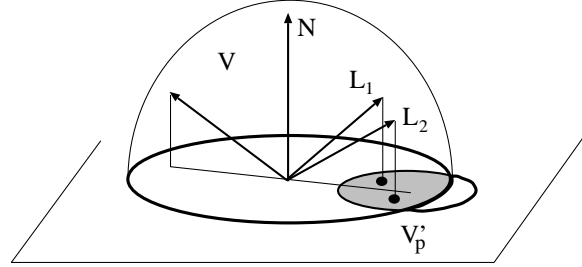
Figure 10: *Sampling when the circles intersect*

To solve this problem, *rejection sampling* can be applied. It means that uniform samples are generated in one of the circles and then checked whether or not the other circle contains this sample. If it does not, the sample is rejected and a new sample is generated.

To reduce the number of rejected samples, the smaller circle is worth selecting as a domain for sample generation and the bigger as the region of rejection sampling.

The following subroutine finds an $\mathbf{L}$ sample from direction $\mathbf{V}$ using a basic BRDF of parameter R, and also provides the albedo to weight the radiance obtained using direction $\mathbf{L}$:

```
DOUBLE GenerateDirectionWithBasicBRDF(R, V, N, L)
    V'_P = -(V - (N · V)N)
    found = TRUE
    repeat
      if R < 1 then
        Generate point L_P in C(V'_P, R)
        if L_P is outside C(O, 1) then found = FALSE
      else
        Generate point L_P in C(O, 1)
        if L_P is outside C(V'_P, R) then found = FALSE
      endif
    until found
    L = L_P + √(1 - (L_P · L_P)) · N
    a = C(V'_P, R)/(π · min(1, R²))
    return a
end
```

2.4.4. Sampling with Russian roulette

In the case of highly reflective materials R is small and thus $C(\mathbf{V}'_P, R)$ is typically a complete circle. The com-

putation of the common part of two circles is required only at the margins.

For other materials, truncation of the circles must be considered several times. However, if we use the Russian roulette, we have to check only whether or not the points are inside the circle, the albedo (the area of the common part) should not be computed explicitly. Let

$$C_{\min}(\mathbf{V}'_P, R) = \begin{cases} C(\mathbf{V}'_P, R), & \text{if } R < 1, \\ \\ C(\mathbf{O}, 1), & \text{otherwise.} \end{cases} \qquad (18)$$

Let us generate in the $C_{\min}(\mathbf{V}'_P, R)$ region exactly M uniformly distributed 2D points. Now, the weighting factor of $(\mathbf{L}_k)_P$ $(k = 1, \ldots, M)$ will be 0 or 1 (i.e. accepted or rejected) according to whether or not the point is inside the "other circle" too. The approximation of the half-space integral is:

$$L^{\text{out}}(\mathbf{V}) = E\left[L^{\text{in}}(\mathbf{L})\right] \approx \frac{1}{M} \cdot \sum_{k=1}^{M} L^{\text{in}}(\mathbf{L}_k), \qquad (19)$$

where only the accepted points are considered in the sum, that is, which have a weighting factor of 1. This means that the number of terms can be at most M, but is often less. This method is somewhat simpler, but its variance is not the possible minimum because it does not carry out the exact importance sampling. The weight factors of the elementary rays are 0 or 1, only their expected value is equal to the $a_R(\mathbf{V})$ albedo.

The algorithm that generates either direction $\mathbf{L}$ from direction $\mathbf{V}$, or reports that the sample should be rejected by returning a zero value is the following:

```
BOOL GenerateDirectionWithBasicBRDF(R, V, N, L)
    V'_P = -(V - (N · V)N)
    if R < 1 then
        Generate point L_P in C(V'_P, R)
        if L_P is outside C(O, 1) then return 0
    else
        Generate point L_P in C(O, 1)
        if L'_P is outside C(V'_P, R) then return 0
    endif
    L = L_P + √(1 - (L_P · L_P)) · N
    return 1
end
```

3. Mixture of basic BRDFs

3.1. General description

The basic BRDF is not a realistic material model by itself, but only a mathematical construction since it is discontinuous (figure 3). Therefore we intend to create
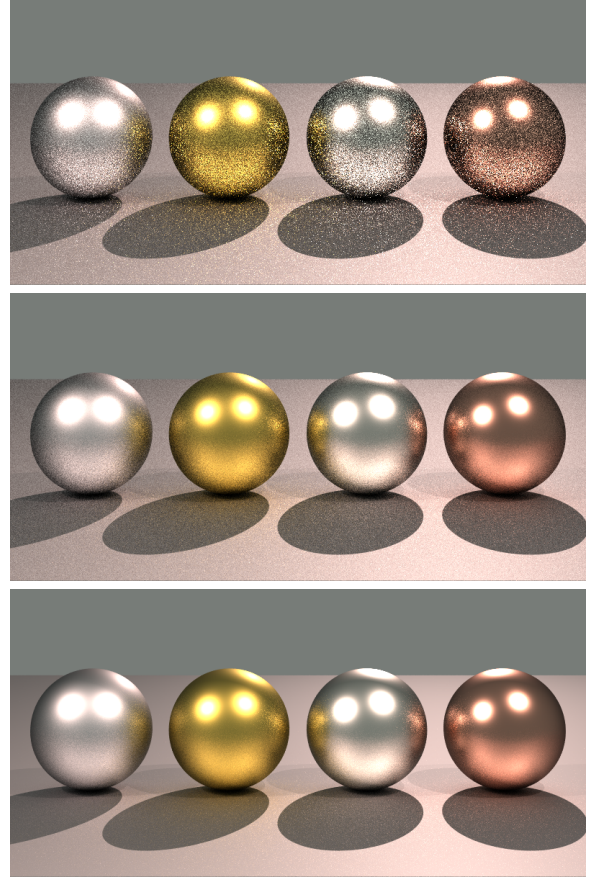
Figure 11: *Metallic spheres generated by Monte-Carlo ray tracing with 50 samples per pixel (top: importance sampling according to the cosine angle only (51 min); middle: Russian roulette based importance sampling of section 2.4.4 (46 min); bottom: normal importance sampling of section 2.4.3 (54 min)*

real BRDFs using a mixture of the basic ones. Let the weighting function be $p(R)$ that is expected to be "normalised" ($\int p(R)dR = 1$).

The *mixture BRDF* is then:

$$f_r(\mathbf{L}, \mathbf{V}) = \int_0^2 p(R) \cdot f_r^*(R, \mathbf{L}, \mathbf{V}) \, dR. \qquad (20)$$

Weighting function $p(R)$ can be considered as a probability density function of one variable in the $0 \le R \le 2$ interval.

The mixture BRDFs evidently meet the requirements of reciprocity and energy conservation. So the BRDFs are physically plausible models if their value is

non-negative. A sufficient but not necessary condition is that the weighting function is non-negative.

The singularities of Dirac-delta type cases must be examined separately, namely when the integral of $p(R)$ in the respective point is a finite, non-zero value, and $p(R)$ is "infinite".

Two types of singularities are important in BRDF modelling. The $R = 0$ case corresponds to a special ideal mirror, and $R = 2$ corresponds to the ideal diffuse model. Separating these, we have:

$$\int_0^2 p(R)dR = p_{\text{Mirror}} + p_{\text{Diffuse}} + \int_{+0}^{2-0} p(R)\, dR = 1 \quad (21)$$

3.2. Albedo for a given $p(R)$ function

Since mixing is a linear operation, the albedo a^p and the mean albedo a^p_{mean} of the mixture BRDF can be computed as follows:

$$a^p(\mathbf{L}) = \int_0^2 p(R) \cdot a_R(\mathbf{L})\, dR, \quad (22)$$

$$a^p_{\text{mean}} = \int_0^2 p(R) \cdot (a_R)_{\text{mean}} dR. \quad (23)$$

According to equation (21) for the albedos of the mirror and of the Lambertian model, we have:

$$a_{\text{Mirror}} = p_{\text{Mirror}}, \qquad a_{\text{Diffuse}} = p_{\text{Diffuse}}. \quad (24)$$

4. Modified importance sampling for a given $p(R)$

4.1. Reducing the general case to basic BRDFs

Using equations (12) and (20), we transform the half-space integral into a double integral and then reverse the order of integration:

$$L^{\text{out}}(\mathbf{V}) = \int_\Omega L^{\text{in}}(\mathbf{L})\, f_r(\mathbf{L}, \mathbf{V})\, (\mathbf{N} \cdot \mathbf{L})\, d\omega_{\mathbf{L}} =$$

$$\int_\Omega L^{\text{in}}(\mathbf{L}) \cdot \left[\int_0^2 p(R)\, f_r^*(R, \mathbf{L}, \mathbf{V})\, dR \right] (\mathbf{N} \cdot \mathbf{L})\, d\omega_{\mathbf{L}} =$$

$$\int_0^2 p(R) \cdot \left[\int_\Omega L^{\text{in}}(\mathbf{L})\, f_r^*(R, \mathbf{L}, \mathbf{V})\, (\mathbf{N} \cdot \mathbf{L})\, d\omega_{\mathbf{L}} \right] dR =$$

$$\int_0^2 p(R) \cdot L_R^{\text{out}}(\mathbf{V})\, dR. \quad (25)$$

This integral is approximated by a sum using Monte-Carlo quadrature. For the calculation of the required R_k values, we apply importance sampling according to the $p(R)$ function. Using the approximate value of the integral (12), we have:

$$L^{\text{out}}(\mathbf{V}) = \int_0^2 p(R) \cdot L_R^{\text{out}}(\mathbf{V})\, dR \approx \frac{1}{M} \cdot \sum_{k=1}^M L_{R_k}^{\text{in}}(\mathbf{V}),$$
$$(26)$$

similarly to equation (25), with the exception that here each incoming direction is represented by a different radius R_k.

Let the cumulative probability distribution function of p be the following function:

$$F(R) = \int_0^R p(t)\, dt, \quad (0 \le R \le 2, \quad 0 \le F(R) \le 1).$$
$$(27)$$

In practice, we may calculate F in a closed form, or using numerical methods. Importance sampling requires the inverse of F, which is usually computed numerically and the resulting values are stored in a sufficiently large one-dimensional array.

The algorithm of generating a direction $\mathbf{L}$ to direction $\mathbf{V}$ is as follows:

```
BOOL GenerateDirectionWithMixtureBRDF(V, N, L)
    Generate a random value r in [0, 1]
    R = F⁻¹(r)
    return GenerateDirectionWithBasicBRDF(R, V, N, L)
end
```

5. Generalisation of the basic BRDF

5.1. Generalisation of the $\mathbf{L} \to \mathbf{L}'$ mirroring transformation

The $\mathbf{L}'$ vector in definition (2) was obtained from the original $\mathbf{L}$ vector by mirroring. Instead of mirroring, basic BRDFs can be defined using other $\mathbf{L} \to T(\mathbf{L})$ transformations too. To fulfil the requirement of reciprocity, the $T(\mathbf{L})$ transformation has to satisfy the following identity:

$$T(T(\mathbf{L})) = \mathbf{L}. \quad (28)$$

This requirement is satisfied by mirroring since mirroring two times gives back the original point.

Another trivial transformation is the identity, that is $T(\mathbf{L}) = \mathbf{L}$.

5.2. The retro-reflective model

For the $T(\mathbf{L}) = \mathbf{L}$ transformation, the $m(\mathbf{L}, \mathbf{V})$ metric from definition (2) becomes:

$$m(\mathbf{L}, \mathbf{V}) = |\mathbf{L}_P - \mathbf{V}_P| = |\mathbf{g} - (\mathbf{N} \cdot \mathbf{g})\mathbf{N}|, \qquad (29)$$

where $\mathbf{g} = \mathbf{L} - \mathbf{V}$. For such models, the radiance of the reflected light is maximum in the incidence direction. This model is called the *retro-reflective model*. In practice, materials exhibiting such properties are the beaded screens or the signal paints used on the highways. The marginal cases of this retro-reflective model are those mirror-constructions which are used in laser-based telemeters that behave practically as ideal mirrors, but reflect exactly in the incident direction.

5.3. Generalisation of the metric

In definition (2) the $m(\mathbf{L}, \mathbf{V})$ metric was an Euclidean distance in the projection plane. The metric can be generalised as follows:

$$m(\mathbf{L}, \mathbf{V}) = \|\mathbf{V}, T(\mathbf{L})\|. \qquad (30)$$

In equation (30) $\|\cdot, \cdot\|$ is an arbitrary metric for two unit vectors on the hemisphere. For the Euclidean metric on the base plane and for basic BRDFs defined in equation (2), we have:

$$\|\mathbf{V}, T(\mathbf{L})\| = |\mathbf{V}_P - T(\mathbf{L})_P|. \qquad (31)$$

5.4. Anisotropic models

The use of anisotropic models can often substantially improve realism in rendering, especially in the case of metals[6, 20]. Formula (31) provides useful generalisation possibilities for practice. Using an "elliptic norm" on the base plane, anisotropic BRDFs are obtained. Let us define an orthonormal co-ordinate system on the surface element representing the given BRDF, by considering the $(\mathbf{u}, \mathbf{v})$ vectors which are perpendicular to each other and to the $\mathbf{N}$ unit normal vector too. The metric m can be defined as follows:

$$m(\mathbf{L}, \mathbf{V}) = \sqrt{\left(\frac{B}{A} \cdot (\mathbf{z} \cdot \mathbf{u})\right)^2 + (\mathbf{z} \cdot \mathbf{v})^2}, \qquad (32)$$

where $\mathbf{z} = \mathbf{V}_P - T(\mathbf{L})_P$ and $B/A \leq 1$ is the ratio of the axes of the ellipse, and the major and the minor axis of the ellipse are taken along vectors $\mathbf{u}$ and $\mathbf{v}$, respectively. This metric m contracts the distance component along vector $\mathbf{u}$. When $B/A = 1$, we have the same case as in equation (1).

Let $E(\mathbf{Q}, \mathbf{v}, b)$ denote the common part of the ellipse with centre in $\mathbf{Q}$, with a minor axis of length b and of the unit base circle. In this case:

$$\mathbf{V}_P \in E(\mathbf{L}'_P, \mathbf{v}, b) \Longleftrightarrow \mathbf{L}'_P \in E(\mathbf{V}'_P, \mathbf{v}, b), \qquad (33)$$

which means that the model meets the requirement of reciprocity.

The maximum value of b is determined for that case when the $E(\mathbf{Q}, \mathbf{v}, b)$ ellipse contains the complete unit base circle, for every $\mathbf{Q}$ inside the base circle. The maximal $b = 2$ case happens for illumination which is along the minor axis and is horizontal. The value of b can be in $[0, 2]$, the length of the major axis in $[0, 2A/B]$.

Parameter b corresponds to the earlier R parameter. The elementary BRDF depends now only on b:

$$f_r^*(b, \mathbf{L}, \mathbf{V}) = \begin{cases} c(b), & \text{if } \mathbf{V}_P \in E(\mathbf{L}'_P, \mathbf{v}, b), \\ \\ 0 \text{ otherwise.} \end{cases} \qquad (34)$$

The albedo of the BRDF in formula (34) is:

$$a_b(\mathbf{L}) = \frac{\text{Area}(E(\mathbf{L}'_P, \mathbf{v}, b))}{\text{Area}(E(\mathbf{O}, \mathbf{v}, b))}. \qquad (35)$$

The albedo is maximal for $\mathbf{L} = \mathbf{N}$. The constant in equation (34) for every value of b is taken so that for a vertical illumination the albedo function is 1.

The computation of this constant can be separated into two cases. When $b \leq B/A$, i.e. when the major axis is less than 1, the base circle contains the whole ellipse with centre $\mathbf{Q} = \mathbf{O}$. On the other hand, when the major axis is greater than 1, the intersections should be computed.

$$c(b) = \begin{cases} 1/(ab\pi), & \text{if } b \leq B/A, \\ \\ 1/\text{Area}(E(\mathbf{O}, \mathbf{v}, b)), & \text{if } B/A < b \leq 2. \end{cases} \qquad (36)$$

In importance sampling, the anisotropic model can be used similarly to the isotropic model. Points are generated inside a circle that has radius b, then a scaling transformation is applied that transforms the circle to the ellipse.

The anisotropic model can also be used to define mixture BRDFs:

$$f_r(\mathbf{L}, \mathbf{V}) = \int\limits_0^2 p(b) \cdot f_r^*(b, \mathbf{L}, \mathbf{V}) \, db = \int\limits_{m(\mathbf{L}, \mathbf{V})}^2 p(b) \cdot c(b) \, db \qquad (37)$$

Anisotropic models can be defined with other norms
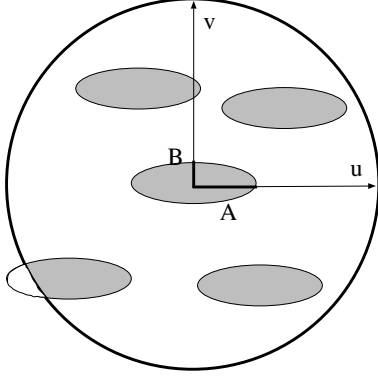
Figure 12: *Elliptical BRDFs*
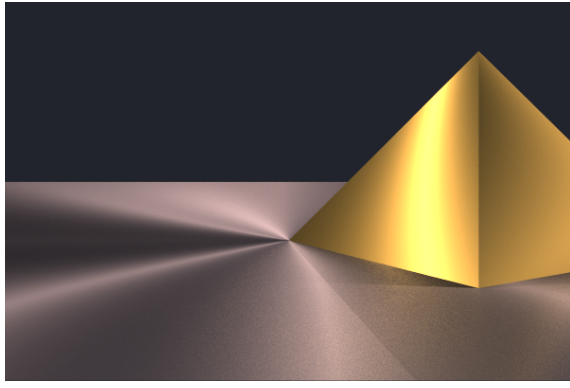
Figure 13: *A scene containing anisotropic metals*

too, for example with the p-th power norm:

$$m(\mathbf{L}, \mathbf{V}) = \left[\left(\frac{B}{A} (\mathbf{z} \cdot \mathbf{u}) \right)^p + (\mathbf{z} \cdot \mathbf{v}))^p \right]^{1/p} \quad (38)$$

6. BRDF definition with a given scalar function

A BRDF maps two vector variables to a scalar value, thus always involves dimension reduction. The exact phase where the dimension reduction happens can vary in different BRDFs. The dimension reduction also means that each BRDF can be associated with equivalence classes of vector pairs, in a way that two vector pairs provide the same BRDF value if they are in the same class. In Phong-Blinn BRDFs, for example, all $\mathbf{L}, \mathbf{V}$ pairs are equivalent if they have the same $\mathbf{N} \cdot \mathbf{H}$ scalar product. In the new model, the dimensional reduction takes place when computing the $m(\mathbf{L}, \mathbf{V}) = |\mathbf{h} - (\mathbf{N} \cdot \mathbf{h})\mathbf{N}|$ metric.

In this section we shall examine how a BRDF can be defined with an arbitrary function of one variable

and how its corresponding $p(R)$ function can be determined. Let us consider an arbitrary, non-negative $\xi(m)$ scalar function defined in the $[0, 2]$ interval, which will define the shape of the reflection lobes or the characteristics of the BRDF. Using this, we define a $f_r^\xi(\mathbf{L}, \mathbf{V})$ function:

$$f_r^\xi(\mathbf{L}, \mathbf{V}) = \xi(m(\mathbf{L}, \mathbf{V})) = \xi(m), \quad m \in [0, 2] \quad (39)$$

Function $\xi(m)$ is multiplied with an appropriate $C_{\max}$ constant in order to fulfil energy conservation requirement, thus the

$$f_r(\mathbf{L}, \mathbf{V}) = f_r^{C_{\max}\xi}(\mathbf{L}, \mathbf{V}) = C_{\max} \cdot f_r^\xi(\mathbf{L}, \mathbf{V})$$

function will be a plausible BRDF.

Let us recall definition (20) of the mixture BRDFs. The lower limit of the integral can be modified, since according to definition (2), if $R < m(\mathbf{L}, \mathbf{V})$, the value of $f_r(R, \mathbf{L}, \mathbf{V})$ is zero. When $R \geq m$, its value is the $c(R)$ defined by equation (7). Thus, the following formula is equivalent to formula (20) for the definition of the mixture BRDF:

$$f_r(\mathbf{L}, \mathbf{V}) = \int\limits_{m(\mathbf{L},\mathbf{V})}^{2} p(R) \cdot f_r^*(R, \mathbf{L}, \mathbf{V}) \, dR =$$

$$\int\limits_{m(\mathbf{L},\mathbf{V})}^{2} p(R) \cdot c(R) \, dR = \int\limits_{m(\mathbf{L},\mathbf{V})}^{2} p(R) \cdot \pi \cdot \min(1, R^2) \, dR.$$

$$(40)$$

Computing the derivative according to variable m, we obtain:

$$p(R) = -C_{\max} \cdot \frac{d\xi(m)}{dm}(R) \cdot \pi \cdot \min(1, R^2). \quad (41)$$

The derivative of function ξ must be taken with a minus sign, since m is the lower limit of the integration.

The $C_{\max}$ factor should be set to make $p(R)$ a probability density and consequently the BRDF energy conserving, thus

$$\frac{1}{C_{\max}} = -\int\limits_{0}^{2} \frac{d\xi(m)}{dm}(R) \cdot \pi \cdot \min(1, R^2) \, dR \quad (42)$$

for decreasing ξ functions.

The maximum of the albedo of this BRDF is 1. To obtain darker materials, the BRDF can be multiplied with a factor that is less than 1.

In a software implementing this approach ξ can be provided interactively by the user as an arbitrary curve using a graphics interface (the curve defines

a one variable function). Only the ratio of the ideal mirror component (polishing) must be specified separately (see equation (21)).

Note that ξ is not required to be monotone decreasing, but the condition of non-negativity must be imposed in any case. For ξ functions having monotone increasing parts, the computation of the maximal albedo and of the normalising factor is more difficult than equation (42) and can be accomplished by numerical methods using formula (22).

6.1. Analytically integrable cases

In this section we will present some monotone decreasing ξ functions which are suitable for applications and for which the $C_{\max}$ multiplier from equation (42) can be derived analytically.

Let G be the primitive function of ξ, and H be the primitive of G (with arbitrary constants!) i.e. $G = \int \xi$ and $H = \int G$.

Then, it can be easily deduced that:

$$C_{\max} = \frac{1}{2\pi(G(1) - H(1) + H(0))}. \qquad (43)$$

This formula is true even if $\xi(2) > 0$.

6.1.1. Exponential function

This section uses the exponential function as a scalar function to define the highlight profile:

$$\xi(m) = e^{-s \cdot m}. \qquad (44)$$

The normalisation factor is

$$C_{\max} = \frac{s^2}{2\pi(1 - (1 + s)e^{-s})}, \qquad (45)$$

and the BRDF is

$$f_r^{(s)}(\mathbf{L}, \mathbf{V}) = \frac{s^2}{2\pi(1 - (1 + s)e^{-s})} \cdot e^{-s \cdot m}. \qquad (46)$$

Using formula (41), we can obtain the following expression for the required weighting function p:

$$p(R) = -C_{\max} \cdot \frac{d\xi(m)}{dm}(R) \cdot \pi \cdot \min(1, R^2) =$$

$$\frac{s^3 \cdot e^{-s \cdot R} \cdot \min(1, R^2)}{2(1 - (1 + s)e^{-s})}. \qquad (47)$$

The probability densities, the inverse of the probability distribution functions, the albedos, and the shapes of the BRDF lobes are shown in figure 14, figure 15, figure 16 and figure 17, respectively, for different s parameters.

For $s > 10$, the model is already highly specular. A

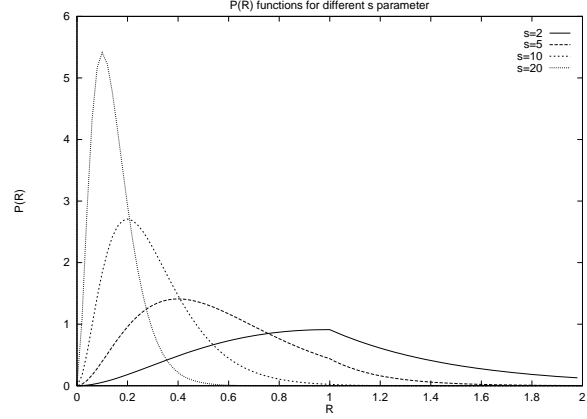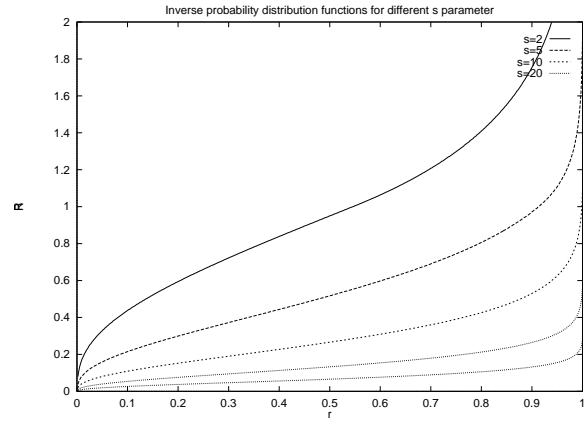**Figure 14:** *Probability density function $p(R)$ for different s values*

Figure 15: *Inverse of the probability distribution functions $F^{-1}(r)$ for different s values*
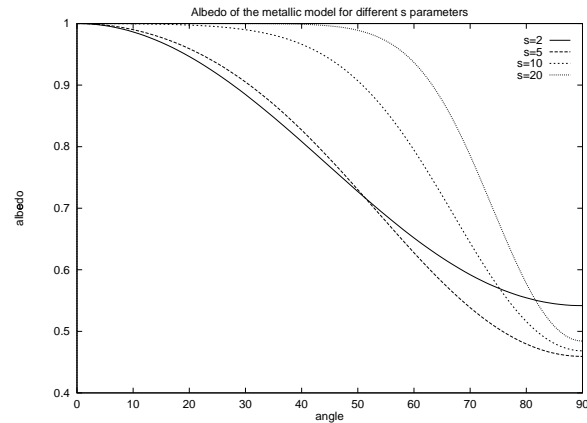

Figure 16: *Albedos of the mixed BRDFs for different s values*

very good approximation of the above formula for this parameter domain is:

$$f_r^{(s)}(\mathbf{L}, \mathbf{V}) = \frac{s^2}{2\pi} \cdot e^{-s \cdot m(\mathbf{L}, \mathbf{V})}. \qquad (48)$$

when the corresponding p weighting function is:

$$p(R) = \frac{s^2}{2} \cdot R \cdot e^{-s \cdot R} \cdot \min(1, R^2). \qquad (49)$$

Let us consider formula (40) and an $m < 1$ value:

$$f_r(\mathbf{L}, \mathbf{V}) = \int\limits_0^2 p(R) \cdot c(R) \, dR = I_{m,1} + I_{1,2} =$$

$$\pi \cdot \int\limits_m^1 p(R) \cdot R^2 \, dR + \pi \cdot \int\limits_1^2 p(R) \, dR. \qquad (50)$$

In the case of a vertical illumination we have $m \leq 1$ independently of $\mathbf{V}$, thus even for the horizontal viewing direction there will be a non-zero "diffuse part". This is component $I_{1,2}$ in equation (50) which is the integral over the $[1, 2]$ interval. Thus, for the truly specular materials not only $\xi(2) = 0$ but $\xi(m) = 0$ for $m \geq 1$ must be imposed.

To illustrate this, the table 1 contains the corresponding values of the $C_{\max}$ and s parameters of the $f_r(\mathbf{L}, \mathbf{V}) = C_{\max} \cdot e^{-s \cdot m}$ function, $I_{1,2}$ and $a_{\text{Diffuse}} = p_{\text{Diffuse}} = C_{\max} \cdot \pi \cdot e^{-2s}$.

7. Real materials

7.1. Metals and the new model

The BRDFs of metals increase towards infinity for grazing angles. This is due to the fact that for smooth metals the product of the BRDF and the cosine term is approximately constant, thus the incident and the outgoing radiance is approximately equal. The proposed model does not follow this behaviour since the maximum of the BRDF values is constant for any incident angle.

To further explain this, let us compare the behaviour at the limiting mirror case of the Phong, of the new model and of the real metals. Assume that the Fresnel function is approximately 1 (this is a good approximation for silver mirror, for instance), and the shininess — which is represented by s in the new model and by the exponent in the Phong model — goes to infinity. If the observer looks at the mirror perpendicularly and the illumination is also perpendicular, then both the new model and the Phong model provide the same radiance and reflected power as a

real metal. However at grazing angles, while a real mirror would reflect the incoming radiance and the output power would be equal to the input power, for the Phong model both the radiance and the output power tend to zero. Interestingly, the output radiance of the new model also goes to zero at grazing angles, but the output power equals to the input power at arbitrarily close to the horizontal illumination. Thus, the new model is an energy mirror in this limiting case, but is not an ideal radiance mirror. This is "better" than the Phong model which is neither radiance nor energy mirror. The radiance mirror property means that the illumination of point light sources are reflected like real metals even at grazing angles. The energy mirror property, on the other hand, means that the distributed (e.g. skylight) illumination is reflected in a way as real metals since for uniformly distributed illumination, the perceived radiance is proportional to the albedo (equation (13).

	Phong $(n \to \infty)$	new BRDF $(s \to \infty)$	real metal mirrors
a_{mean}	1/2	1	1
$a(\theta \to 90°)$	0	1	1
$L^{\text{out}}(\theta \to 90°)$	0	0	L^{in}
$f_r(\theta \to 90°)$	constant	constant	∞

Table 2: *Comparison the limiting case of the Phong model, of the new model and of real metal mirrors at grazing angles*

7.2. Coated metals, metallic paints

We concluded that the new model is not a perfect metallic model because the BRDF of metals goes to infinity at grazing angles while the BRDF of the new model is bounded but its reflection lobe gets wider in order not to absorb all energy. Considering real materials, metals coated by a lacquer layer[21] behave like this. Since in lacquer the attenuation of the power is an exponential function of the length travelled, for higher incident angles, both the reflected radiance and power will tend to zero at grazing angles. If the lacquer also contains the particles of the basic metal, then the behaviour of the material will be in between the metal and lacquered metals. Absorption, scattering and reflection can occur simultaneously in the lacquer layer. Thus for higher incident angles the metal particles, especially those that are close to the surface, will be responsible for reflecting the incoming photons. The object acts as energy mirror even if the reflected radiance decreases. Consequently the new model seems

s	$C_{\max}$	$I_{1,2}$	a_{Diffuse}	
0	$1/\pi$	0	1	LAMBERTIAN
0.25	0.3754	0.20316	0.71527	
0.5	0.4411	0.33071	0.50979	
1.	0.6023	0.44002	0.25608	
2	1.0718	0.39401	0.06167	
4	2.8032	0.15834	0.00295	
8	10.216	0.01076	0.00000	
16	40.743	0.00001	0.00000	HIGHLY SPECULAR
32	162.97	0.00000	0.00000	
64	651.89	0.00000	0.00000	
128	2607.5	0.00000	0.00000	
256	10430.	0.00000	0.00000	
512	41721.	0.00000	0.00000	ALMOST MIRROR

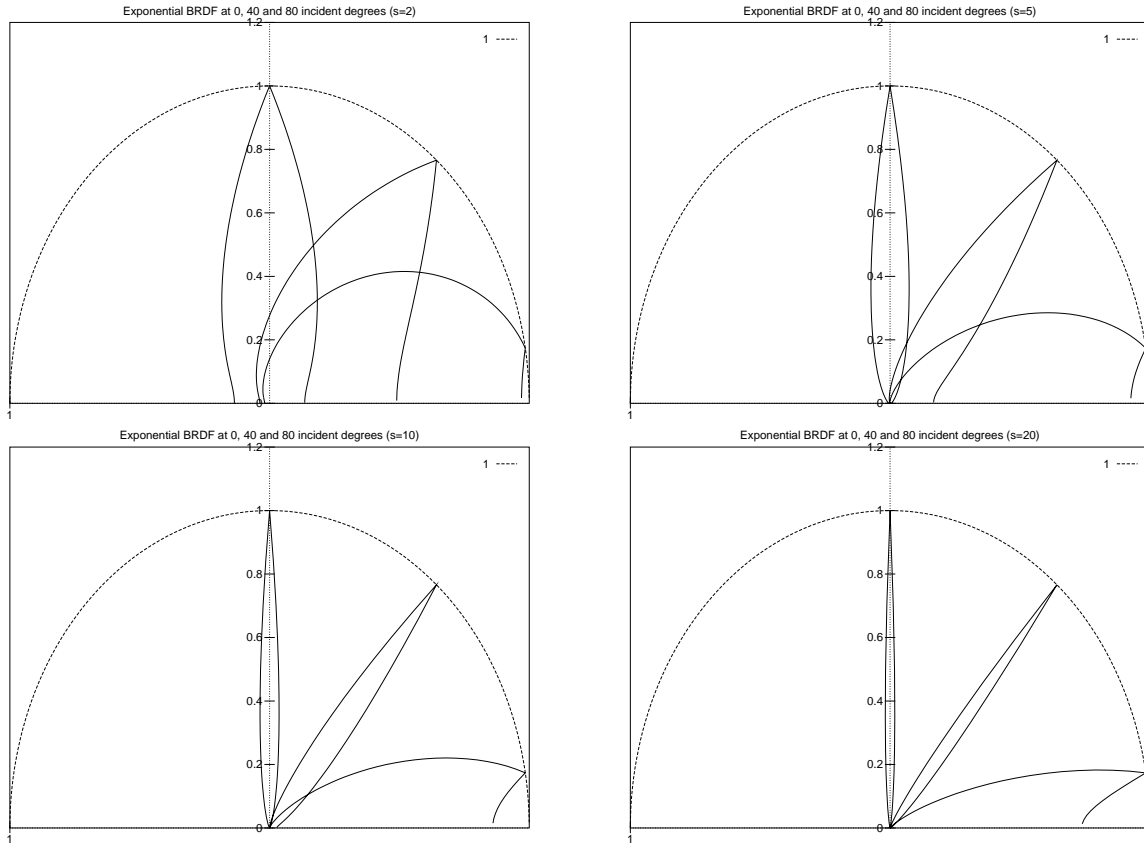**Table 1:** *Parameters of the $f_r = C_{\max} \cdot e^{-s \cdot m}$ BRDF*

Figure 17: *BRDF using exponential weight function*

to be appropriate for representing coated metals and metallic paints.

Since the diffuse component of the metals is practically negligible, the metallic paints are approximated as follows

$$C_{\max} \cdot \xi(m) \cdot \text{Fresnel}(\theta'), \tag{51}$$

where $C_{\max} \cdot \xi(m)$ is a BRDF. For metals, the $\xi(m)$ function has a quite small diffuse component. We can use, for example, the $\xi(m) = (s^2/2\pi) \cdot e^{-s \cdot m}$ function, assuming that $s > 10$.

The wavelength dependent Fresnel function describes the reflection of the optically smooth material as a function of the incident angle.

In the Cook-Torrance model, as well as in other mirror-particle models, for a given $(\mathbf{L}, \mathbf{V})$ pair, the expectation of the normal vectors of the reflecting facets is the $\mathbf{H}$ vector, thus the angle of incidence is $\theta_{\mathbf{H}} = \arccos(\mathbf{N} \cdot \mathbf{H})$.

This approach is slightly modified in the following way. The distance of the $\mathbf{H}_P = (\mathbf{V}_P + \mathbf{L}_P)/2$ vector from the $\mathbf{O}$ origin is used to evaluate the Fresnel function.

$$|\mathbf{H}_P - \mathbf{O}| = \frac{|\mathbf{L}_P - \mathbf{V}_P|}{2} = \frac{|\mathbf{g} - (\mathbf{N} \cdot \mathbf{g})\mathbf{N}|}{2} = \frac{m(\mathbf{L}, \mathbf{V})}{2}. \tag{52}$$

where the $m(\mathbf{L}, \mathbf{V})$ metric is described by formula (29) belonging to the retro-reflective model.

Let us define angle θ' in the following way:

$$\theta' = \arcsin\left(\frac{m(\mathbf{L}, \mathbf{V})}{2}\right). \tag{53}$$

In conclusion, the proposed new model is:

$$f_{r,\text{metallic}}(\mathbf{L}, \mathbf{V}) = \frac{s^2}{2\pi} \cdot e^{-s \cdot |\mathbf{h} - (\mathbf{N} \cdot \mathbf{h})\mathbf{N}|} \cdot \text{Fresnel}_\lambda(\theta'). \tag{54}$$

where $\theta' = \arcsin(|\mathbf{g} - (\mathbf{N} \cdot \mathbf{g})\mathbf{N}|/2)$.

The model is the product of a reflective and a retro-reflective model. Importance sampling can be performed efficiently on the reflective factor, the retro-reflective factor modifies the weights of the elementary rays as a multiplier.

7.3. Plastics and polishing

Using a real refraction index in the Fresnel term instead of the complex index of metals, the proposed model can also be used for modelling the specular reflection of plastics. Since the diffuse reflection of plastics is also significant, the BRDF function is defined as a sum of the diffuse and specular terms. To meet the requirement of energy conservation, the sum of the
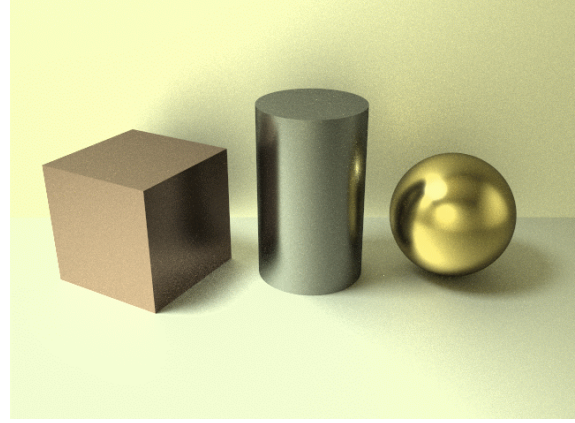
Figure 19: *A simple scene with a cube (s = 2) cylinder (s = 4) and sphere (s = 8)*

albedos of the diffuse and specular part must be less than 1. The BRDF is then

$$f_{r,\text{plastics}} = \frac{a_{\text{diffuse}}}{\pi} + (1 - a_{\text{diffuse}}) \cdot C_{\max} \cdot \xi(m) \cdot \text{Fresnel}(\theta'). \tag{55}$$

Certain materials, as for example lacquered objects, reflect sharp mirror images for large viewing angles [13]. To simulate this phenomenon, any of the proposed BRDFs can be supplemented by an ideal mirror in the following way:

$$f_{r,\text{mirrorextension}} = f_r(\mathbf{L}, \mathbf{V}) + (1 - a(\mathbf{L})) \cdot f_{r,\text{idealmirror}}(\mathbf{L}, \mathbf{V}). \tag{56}$$

Note that the extension is limited by the "missing albedo" of the original BRDF. The BRDF of the ideal mirror is a Dirac function:

$$f_{r,\text{idealmirror}}(\mathbf{L}, \mathbf{V}) = \frac{\text{Fresnel}(\theta') \cdot \delta(\mathbf{L}' - \mathbf{V})}{\cos \theta}. \tag{57}$$

7.4. Generalised Lambertian models

The $\xi(m)$ function can be monotone increasing as well. In this case the reflected radiance will have smaller slope as the incident angle changes than the cosine function obtained from the Lambertian model, which conforms to the measurements using a certain class of materials, including ceramic, wall, foam, cloth, etc. Due to the local interreflections - especially for great incident angles - these materials reflect more light backwards, than forwards in the mirror direction.

Oren et. al.[11] also proposed a model for these types of materials. Here a different approach is presented. To obtain such a model, let us assume that the light source is placed in the viewpoint. An appropriate function of the reflected radiance that has a "smoother" slope than the Lambertian model is $\cos^{1-d} \theta$ where θ
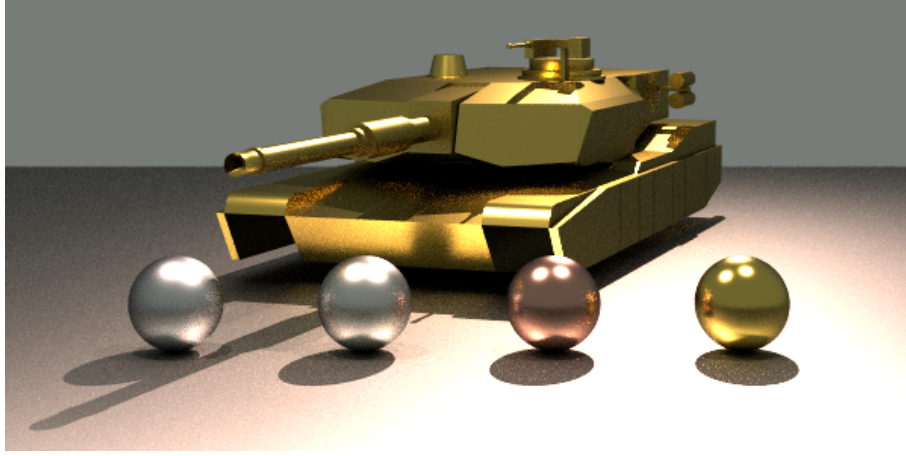
Figure 18: *Scene rendered using Russian roulette including a golden tank (s = 10), silver base plane (s = 4), aluminium (s = 6), silver (s = 8), copper (s = 15) and golden (s = 20) spheres*

is the angle of the light and of the normal vectors and d is the smoothness parameter. In the limit case of $d = 0$ the diffuse case is obtained, for the $d = 1$ limit case we get a disk of constant radiance, which is like the full moon. This selection leads to the following ξ function:

$$\xi(m) = \left(1 - \left(\frac{m}{2}\right)^2\right)^{-d/2}, \qquad 0 \le d \le 1. \quad (58)$$

Unfortunately, these models do not satisfy the principle of the conservation of energy, so we have to modify definition (58). The simplest way of correcting this is to limit the BRDF with a constant. Another possible approach is to consider the MacLaurin's series of the function, retaining only a few terms, such that the truncated series will be less than the original function.

As an example, let us consider the case when the function is clipped by a constant:

$$f_r(\mathbf{L}, \mathbf{V}) = C_{\max} \cdot \min\left(\left(1 - \left(\frac{m}{2}\right)^2\right)^{-d/2}, \xi_{\max}\right), \quad (59)$$

where $0 \le d \le 1$.

Table 3 has been generated using a clipping where the function was taken to be constant above $m = 1.9$, that is $\xi_{\max} = \xi(1.9)$. The $C_{\max}$ value was determined such that the maximum value of the albedo is 1. For the above model, this is reached always for $m = 2$. The minimum of the albedo is always in $m = 0$. So, contrary to the case of monotone decreasing functions, the pure diffuse component is defined by the case corresponding to $m = 0$.

The model defined by formula (59) can always be
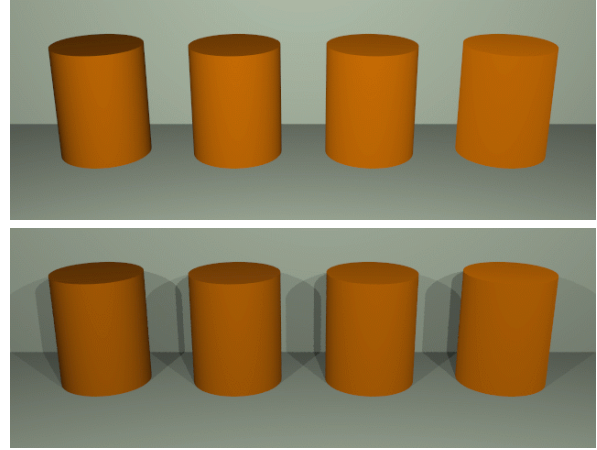
Figure 20: *Objects (d = 1, d = 0.75, d = 0.5, d = 0.25) rendered using generalised Lambertian BRDFs computing only the direct illumination effects. In the upper image the light source is at the view position, in the lower image there are two point light sources*

described as the sum of a Lambertian model with an albedo (reflectivity) of a_{Diffuse} and of a complementary model. The albedo of the latter is zero for $m = 0$ and it is a monotone increasing function of m. Formula (59) has the same behaviour as the model defined by equation (58), except for the $m > 1.9$ domain. In the case of $1.9 \le m \le 2$, the reflected radiance for the above mentioned sphere will have a decrease proportional to the cosine function at $[0, 1]$.

The "full moon" limit case means that for a sphere

d	C_{max}	f_{max}	$a_{diffuse}$	
0	$1/\pi$	1.0	1.0	LAMBERTIAN
0.1	0.30884	1.123440	0.97689	
0.2	0.29939	1.262117	0.95352	
0.3	0.28998	1.417912	0.92989	
0.4	0.28061	1.592939	0.90602	
0.5	0.27128	1.789571	0.88193	
0.6	0.26200	2.010475	0.85764	
0.7	0.25278	2.258647	0.83318	
0.8	0.24363	2.537454	0.80856	
0.9	0.23456	2.850676	0.78381	
1.0	0.22556	3.202563	0.75895	"FULL MOON"

Table 3: *Parameters obtained for the generalised Lambertian model*

illuminated from the viewpoint the radiance will be constant, only at the margins will decrease with a cosine character. The width of the "cosine-ring" can be controlled by modifying ξ_{max}. It is advisable to choose the value of d in formula (59) to be less than 1.

For the above model, a good approximation of the importance sampling is to generate directions (with a "cosine" distribution) like in the case of the Lambertian model and to use an appropriate multiplier correction factor for each ray.

8. Conclusions and future work

The paper presents a new, simple and unified definition for obtaining a good visual approximation for many materials. The new model class can be used easily in the majority of the commercial programs. By allowing the interactive definition of the $\xi(m)$ highlight profile functions, BRDF characteristics can be specified in a new manner.

Due to their simplicity, the computation cost of the new models is low. According to our running time measurements using Heckbert's BRDF viewer[22], the computational time of the new model is practically equal to the time of the Phong-Blinn model and half of the time of the Ward model. Importance sampling, on the other hand, is simpler and faster for the new model than for previous models.

One of the problems to be solved by further research is to investigate which of the physically plausible BRDFs (fulfilling reciprocity, energy conservation and non-negativity) can be defined by suitable

$$(T(\mathbf{L}), m(\cdot, \cdot), p(R))$$

triplets (see section 5). Another important question is which materials can be modelled as the sum, product

and other generalisations of the introduced reflective and retro-reflective models.

References

1. R. Lewis. Making shaders more physically plausible. In *Rendering Techniques '93*, pages 47–62, 1993.

2. P. Beckmann and A. Spizzichino. *The Scattering of Electromagnetic Waves from Rough Surfaces.* MacMillan, 1963.

3. K. Torrance and M. Sparrow. Off-specular peaks in the directional distribution of reflected thermal distribution. *Journal of Heat Transfer — Transactions of the ASME*, pages 223–230, May 1966.

4. R. Cook and K. Torrance. A reflectance model for computer graphics. *Computer Graphics*, 15(3), 1981.

5. X. He, K. Torrance, F. Sillion, and D. Greenberg. A comprehensive physical model for light reflection. *Computer Graphics*, 25(4):175–186, 1991.

6. J. T. Kajiya. Anisotropic reflection models. In *Computer Graphics (SIGGRAPH '85 Proceedings)*, pages 15–21, 1985.

7. G. Ward. Measuring and modeling anisotropic reflection. *Computer Graphics*, 26(2):265–272, 1992.

8. B. T. Phong. Illumination for computer generated images. *Communications of the ACM*, 18:311–317, 1975.

9. J. F. Blinn. Models of light reflection for computer synthesized pictures. In *Computer Graphics (SIGGRAPH '77 Proceedings)*, pages 192–198, 1977.

10. T. Whitted. An improved illumination model for shaded display. *Communications of the ACM*, 23(6):343–349, 1980.

11. M. Oren and S. Nayar. Generalization of Lambert's reflectance model. *Computer Graphics (SIGGRAPH '94 Proceedings)*, pages 239–246, 1994.

12. J. Beard and J. R. Maxwell. Bidirectional reflectance model validation and utilization. Technical Report AFAL-TR-73-303, 1973.

13. E. Lafortune, S. Foo, K. Torrance, and D. Greenberg. Non-linear approximation of reflectance functions. *Computer Graphics (SIGGRAPH '97 Proceedings)*, pages 117–126, 1997.

14. L. Neumann and A. Neumann. A new class of brdf models with fast importance sampling. Technical Report TR-186-2-96-24, Institute of Computer Graphics, Vienna University of Technology, 1996. www.cg.tuwien.ac.at/.

15. R. F. Tobler, L. Neumann, M. Sbert, and W. Purgathofer. A new form factor analogy and its application to stochastic global illumination algorithms. In *Rendering Techniques '98*, 1998.

16. D. S. Immel, M. F. Cohen, and D. P. Greenberg. A radiosity method for non-diffuse environments. In *Computer Graphics (SIGGRAPH '86 Proceedings)*, pages 133–142, 1986.

17. E. Lafortune and Y. D. Willems. Using the modified phong reflectance model for physically based rendering. Technical Report RP-CW-197, Department of Computing Science, K.U. Leuven, 1994.

18. L. Neumann and A. Neumann. Radiosity and hybrid methods. *ACM Transactions on Graphics*, 14(3):233–265, 1995.

19. L. Szirmay-Kalos. *Theory of Three Dimensional Computer Graphics*. Akadémia Kiadó, Budapest, 1995. http: //www.iit.bme.hu/~szirmay.

20. Ch. Schlick. A customizable reflectance model

for everyday rendering. In *Fourth Eurographics Workshop on Rendering*, pages 73–83, Paris, France, 1993.

21. L. Neumann and A. Neumann. Photosimulation: Interreflection with arbitrary reflectance models and illumination. *Computer Graphics Forum*, (8):21–34, 1989.

22. P. Heckbert. Brdf viewer. 1997. http: //www.cs.cmu.edu/afs/cs.cmu.edu/user/ph/ www/src/illum.

23. P. Poulin and A. Fournier. A model for anisotropic reflection. *Computer Graphics*, 24(4):273–281, 1990.

24. M. Minnaert. The reciprocity principle in lunar photometry. *Astrophysical Journal*, 93:403–410, 1941.

25. E. Ken. Reflectance phenomenology and modeling tutorial. 1994. http://www.erim.org.

26. B. Hapke. A theoretical photometric function for the lunar surface. *Journal of Geophysical Research*, 68(15), 1963.

27. R. Hall. *Illumination and Color in Computer Generated Imagery*. Springer-Verlag, New York, 1989.

28. I. Deák. *Random Number Generators and Simulation*. Akadémia Kiadó, Budapest, 1989.

29. L. Neumann, A. Neumann, and L. Szirmay-Kalos. New simple reflectance models for metals and other specular materials. Technical Report TR-186-2-98-17, Institute of Computer Graphics, Vienna University of Technology, 1998. www.cg.tuwien.ac.at/.

30. L. Neumann, A. Neumann, and L. Szirmay-Kalos. Analysis and pumping up the albedo function. Technical Report TR-186-2-98-20, Institute of Computer Graphics, Vienna University of Technology, 1998. www.cg.tuwien.ac.at/.

31. L. Neumann, A. Neumann, and L. Szirmay-Kalos. Compact metallic reflectance models. *Computer Graphics Forum (Eurographics'99)*, 18(3):161–172, 1999.

32. L. Neumann, A. Neumann, and L. Szirmay-Kalos. Reflectance models by pumping up the albedo function. *Machine Graphics and Vision*, 8(1):3–18, 1999.

A Microfacet-based BRDF Generator

Michael Ashikhmin   Simon Premože   Peter Shirley

University of Utah   www.cs.utah.edu

Abstract

A method is presented that takes as an input a 2D microfacet orientation distribution and produces a 4D bidirectional reflectance distribution function (BRDF). This method differs from previous microfacet-based BRDF models in that it uses a simple shadowing term which allows it to handle very general microfacet distributions while maintaining reciprocity and energy conservation. The generator is shown on a variety of material types.

CR Categories: I.3.7 [Computing Methodologies]: Computer Graphics—3D Graphics

Keywords: Reflectance & Shading Models, Rendering

1 Introduction

Physically-based rendering systems describe reflection behavior using the *bidirectional reflectance distribution function* (BRDF) [7]. At a given point on a surface the BRDF is a function of two directions, one toward the light and one toward the viewer. The characteristics of the BRDF will determine what "type" of material the viewer thinks the displayed object is composed of, so the choice of BRDF model and its parameters is important. There are a variety of basic strategies for modeling BRDFs that we categorize as follows.

Direct measurement. BRDFs can be measured directly using *gonioreflectometers* which mechanically vary the direction to a small light source and a spectral sensor and thus collect a large number of point samples for the BRDF [7]. Simpler and less accurate devices can also be constructed using CCD imaging devices [26]. More complex CCD devices can also be used which gather data quickly with accuracy almost that of full gonioreflectometry [12]. If enough is known about the microstructure of a material, a BRDF can be simulated by using a *virtual gonioreflectometer*, where statistical ray tracing followed by density estimation is used to create BRDF data [3, 5, 27].

Empirical methods. There exist a variety of purely empirical reflection models, the most familiar being the models introduced by Gouraud [6] and Phong [15]. These two initial models were meant to be used with hand-chosen parameters, and thus these parameters are intuitive. A variety of more complex methods have been introduced to improve characteristics of the Phong model for efficiency [19], to include anisotropy [26], and enforce physical constraints such as reciprocity [9]. Other models have been developed to fit measurement data as opposed to being intuitive [10].
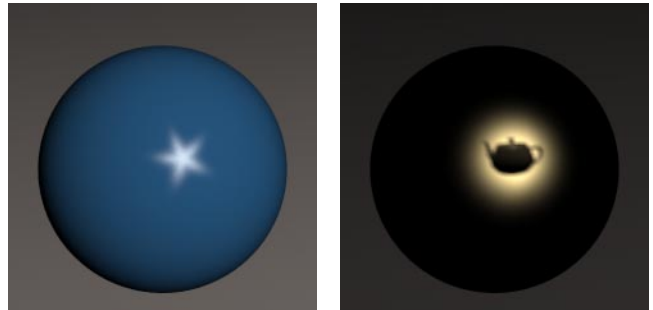
Figure 1: *Images generated using the new BRDF model with unusual microfacet distributions. The BRDFs used to create these images are both reciprocal and energy-conserving. The only illumination is a small distant source, and the highlights will stay unchanged if the spheres rotate about the axes through their north and south poles.*

Height correlation methods. In these methods a random rough surface is a realization of some Gaussian random process. Such a process can be described by its correlation function which is directly related to surface height correlations. This is the most complete surface representation used in computer graphics. Some of the most detailed descriptions of light scattering by a surface, including wave optics effects, were obtained using this approach [8, 22].

Microfacet methods. Somewhere between the height correlation methods and empirical methods lie models based on microfacet theory [2, 4]. Microfacet models assume the surface consists of a large number of small flat "micromirrors" (facets) each of which reflect light only in the specular direction. By computing the number of visible microfacets at the appropriate orientation to specularly reflect light from the source to the viewer, one can determine the BRDF.

All of these methods have their place. In applications where little is known about the low-level properties of the surface, measurement is essential. Where physical optics effects are important, height correlation methods should be used. Our interest is in visual computer graphics applications which do not have obvious physical optics effects (e.g. metal with relatively large scratches, fabric). The lesson from empirical models is that in many cases viewers are not particularly sensitive to the fine details of light scattering as long as the main character of the reflection is conveyed correctly. This paper uses this aspect of human sensitivity to suggest a new microfacet model specifically intended to capture the main character of reflection.

Microfacet models are able to capture the main character of reflection for surfaces whose appearance is dominated by surface scattering. Although microfacet models lack the precision of height correlation methods, they tend to be more intuitive with simpler expressions. However, to date there has been no microfacet model that is reasonably general in its assumptions, maintains a simple formulation, and conserves energy. In this paper we develop a model with all of these characteristics by introducing assumptions about surfaces that we believe are reasonable. These assumptions allow

Figure 2: *Geometry of reflection. Note that $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{h}$ share a plane, which usually does not include $\mathbf{n}$. On the left, the microfacet can "see" in directions $\mathbf{k}_1$ and $\mathbf{k}_2$ so it contributes to the BRDF. On the right, direction $\mathbf{k}_2$ is blocked and the microfacet does not contribute. Note that the microfacet distribution is not restricted to height fields.*

$(\mathbf{ab})$	scalar (dot) product of vectors $\mathbf{a}$ and $\mathbf{b}$
$\mathbf{k}_1$	normalized vector to light
$\mathbf{k}_2$	normalized vector to viewer
$\mathbf{n}$	surface normal to macroscopic surface
$\rho(\mathbf{k}_1, \mathbf{k}_2)$	BRDF
$\mathbf{h}$	normalized half-vector between $\mathbf{k}_1$ and $\mathbf{k}_2$
$p(\mathbf{h})$	probability density function of microfacet normals
$F(\cos\theta)$	Fresnel reflectance for incident angle θ
$P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{h})$	Probability that light from $\mathbf{k}_1$ reflecting in direction $\mathbf{k}_2$ is *not* shadowed
$\langle f \rangle$	average of function f over distribution $p(\mathbf{h})$ (see Equation 9)
$\Omega_+(\mathbf{k})$	set of directions $\mathbf{h}$ where $(\mathbf{hk}) > 0$ (see Figure 4)
$g(\mathbf{k})$	average of positive $(\mathbf{hk})$ (see Equation 18)

Table 1: Important terms used in the paper

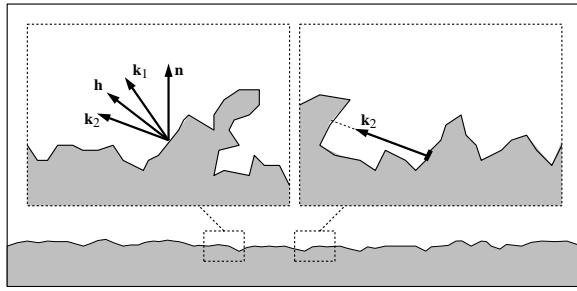us to create a relatively simple formula for the probability that a microfacet at a certain orientation is visible to the light/viewer. The BRDF produced by this process is compact, reciprocal and energy-conserving with only mild restrictions on the distribution of microfacet orientation (e.g., the very general distributions in Figure 1).

Our assumptions and guiding principles in relation to microfacet theory are given in Section 2. Formalisms are developed in Section 3. The key development of the paper, a simplified shadowing term, is introduced in Section 4, and the resulting BRDF is derived. Section 5 shows that this BRDF model conserves energy, and derives a diffuse term to account for secondary and subsurface reflection. The model is applied to a variety of surfaces in Section 6. This last section serves as a set of case-studies which both show how the model can be applied, and that it is more general than previous microfacet approaches. We believe the only other method which is able to handle such a diverse set of surface microgeometries is the "virtual gonireflectometer" approach involving explicit modeling of the surface structure and statistical averaging the results of light scattering simulations.

2 Overview

The strategy behind our model is in balancing issues of practicality and accuracy to produce a simple formulation that is still expressive, reciprocal, and conserves energy. In this section we discuss the basic ideas of microfacet models, as well as our strategy for using this theory to produce BRDFs. Important symbols used in the paper are listed in Table 1.

Microfacet models assume that the surface consists of a large number of small flat "micromirrors" (facets) each of which reflects light only in the specular direction with respect to its own normal $\mathbf{h}$ (Figure 2) and the overall appearance of the surface is governed by two assumptions:

- the microfacet normals have an underlying probability density function $p(\mathbf{h})$.

- a microfacet contributes to BRDF for a given pair of directions if and only if it is visible (not shadowed) relative to the lighting direction $\mathbf{k}_1$ and the viewing direction $\mathbf{k}_2$.

The BRDF for a given direction pair $(\mathbf{k}_1, \mathbf{k}_2)$ is determined entirely by the Fresnel reflectance for that angle, the fraction of microfacets with normal vector $\mathbf{h}$ exactly between $\mathbf{k}_1$ and $\mathbf{k}_2$, and the *shadowing term*: the fraction of those microfacets which are visible to both eye and light (Figure 2). Microfacet theory's only knowledge of the

surface configuration is $p(\mathbf{h})$, and this alone does not uniquely determine the shadowing term. However, the shadowing term is still heavily constrained by energy conservation.

The shadowing term is the most complex part of most microfacet-based models, even if additional $p(\mathbf{h})$-specific information about the surface geometry is used. Because there are many possible surface geometries that are consistent with a given $p(\mathbf{h})$, it is the case that no specific shadowing function is "right". We believe that in most cases the shape of $p(\mathbf{h})$ function itself has a much greater impact on the appearance than the shadowing. This suggests the key idea in this paper: the shadowing term should be made as simple as possible while remaining physically plausible. Such a shadowing term is developed in Section 4. This key extension of the standard microfacet theories allows us to construct a general procedure to create a BRDF for a statistical surface starting from $p(\mathbf{h})$.

Note that surface description in the language of $p(\mathbf{h})$ is less detailed than that of using height correlation functions. Nevertheless, we believe that the microfacet normal distribution is more intuitive to deal with than the correlation functions. As we will emphasize in Section 6, enough useful information about $p(\mathbf{h})$ can be obtained from general notion of surface structure obtained through visual examination of the surface and the specular reflection highlight. Moreover, attempting to obtain more detailed information about the distribution might not be worth the effort. As in any other model, we make simplification in our approach which affect the final result, but what we are trying to do is generate a physically plausible BRDF having the general character of the surface reflection while restricting the range of allowed surface microstructures as little as possible. This is in contrast to most other physics-based approaches which concentrate on a particular type of surface, usually Gaussian height field, and emphasize the need for precise knowledge of surface characteristics.

Some care should be exercised when specifying $p(\mathbf{h})$. In particular, because we do not make the common assumption of a surface being a height field, in this general case $p(\mathbf{h})$ should refer only to the distribution of "visually important" or "surface" part of the microfacets. For example, a homogeneous porous substance thought of as a collection of microfacets will have an overall "volume" distribution of microfacets $p_v(\mathbf{h}) = const$ over the whole sphere of directions. However, that most of these microfacets will be completely hidden and will not be of any significance for the scattering process which occurs on the surface. In this case it is rather difficult to separate surface from the rest of the substance and judge the exact shape of $p(\mathbf{h})$. Fortunately, because we are not trying to reproduce all the details of the reflection function, a reasonable guess for $p(\mathbf{h})$

is all we need and for this surface; it might be $p(\mathbf{h}) = const$ in the upper hemisphere and $p(\mathbf{h}) = 0$ in the lower one. Note, that by making this particular choice for $p(\mathbf{h})$ the surface is restricted to be a height field. The initial choice can be refined later if necessary but in this particular case it the surface will be mostly diffuse and small refinements will not dramatically change the appearance.

We are concerned with single-bounce reflections from the microfacets and stay within the limits of geometric optics and Fresnel reflection. The result is a new form of the specular component of the BRDF which constitutes the main contribution of the paper. The complete BRDF can also have a diffuse term which accounts for multiple bounces and subsurface scattering. This issue along with other important properties of the BRDFs produced with the generator are briefly discussed in Section 5. Our framework is modular and allows the user to choose the form of the final BRDF most appropriate for the particular application.

3 Microfacet Theory

We now review the main results of microfacet theory as developed by Torrance and Sparrow [23] and later introduced to computer graphics by Cook and Torrance [4]. We follow their approach of considering a collection of microfacets of small but finite size, and we derive the basic formula for BRDF in terms of quantities convenient for our model.

The quantity we wish to derive an expression for is the BRDF $\rho(\mathbf{k}_1, \mathbf{k}_2)$ which gives the ratio of radiance observed by a viewer in the direction $\mathbf{k}_2$ to irradiance from infinitesimal solid angle about $\mathbf{k}_1$. Throughout the paper, all vectors are shown in **bold**. They are assumed to be normalized, and all quantities with subscript 1 refer to incident direction while those with subscript 2 belong to the outgoing direction. Both $\mathbf{k}_1$ and $\mathbf{k}_2$ and all normals point outward from the surface. If we expose the surface to a uniform radiance of L_1 coming from a small solid angle $\delta\omega_1$ around $\mathbf{k}_1$, the outgoing radiance in direction $\mathbf{k}_2$ will be

$$L_2 = \rho(\mathbf{k}_1, \mathbf{k}_2) \ L_1 \ (\mathbf{k}_1\mathbf{n})\delta\omega_1, \tag{1}$$

where $\mathbf{n}$ is the surface geometric normal and two vectors written next to each other in parenthesis denotes their scalar product, i.e., the cosine of the angle between them. The use of δ is not standard notation, but is used to make the algebra less cluttered without losing the gist of the argument. By the definition of radiance, if $(\mathbf{k}_2\mathbf{n})A$ is the projected surface element area in the direction $\mathbf{k}_2$ and $\delta E(\mathbf{k}_1 \rightarrow \mathbf{k}_2)$ is the power reflected by the surface in the direction $\mathbf{k}_2$, then

$$L_2 = \frac{\delta E(\mathbf{k}_1 \rightarrow \mathbf{k}_2)}{A(\mathbf{k}_2\mathbf{n})\delta\omega_2}, \tag{2}$$

and BRDF can be written as

$$\rho(\mathbf{k}_1, \mathbf{k}_2) = \frac{\delta E(\mathbf{k}_1 \rightarrow \mathbf{k}_2)}{AL_1(\mathbf{k}_2\mathbf{n})(\mathbf{k}_1\mathbf{n})\delta\omega_1\delta\omega_2}. \tag{3}$$

Only a fraction of all microfacets will participate in scattering the energy from $\mathbf{k}_1$ to $\mathbf{k}_2$. If the number of these active microfacets is N_{active} and all microfacets have the same area A_{mf}, their total projected area in the direction of $\mathbf{k}_1$ is $N_{active}A_{mf}(\mathbf{kh})$ and the total scattered power is

$$\delta E(\mathbf{k}_1 \rightarrow \mathbf{k}_2) = L_1\delta\omega_1 N_{active}A_{mf}(\mathbf{kh})F((\mathbf{kh})), \tag{4}$$

where $\mathbf{h}$ is the normalized half-vector between $\mathbf{k}_1$ and $\mathbf{k}_2$ and $F((\mathbf{kh}))$ is Fresnel coefficient giving the fraction of incoming light which is specularly reflected by a microfacet. Note that we will drop subscripts in our notations if either of incoming and outgoing direction can be used in an expression (e.g., $(\mathbf{kh})$).

Out of the total of N surface microfacets, only $Np(\mathbf{h})\delta\omega_h$ will have their normals oriented in the appropriate direction. The density $p(\mathbf{h})$ does not specify all surface properties uniquely, but in our simplified approach this is the only characteristic of the surface we will use in our analysis. Note that this function operates in the domain of microfacet normals which is different from the space of incoming and outgoing light directions. In particular, for the case of specularly reflecting microfacets, the relationship between elementary solid angles [23] can be shown to be

$$\delta\omega_2 = 4(\mathbf{k}_1\mathbf{h})\delta\omega_h. \tag{5}$$

Even if a microfacet has the required orientation, it might still not contribute to the single-bounce highlight if it is shadowed by other microfacets for either incoming or outgoing direction. Introducing the probability for a microfacet *not* to be shadowed in either incoming or outgoing directions as $0 \leq P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{h}) \leq 1$ we will have $N_{active} = Np(\mathbf{h})P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{h})\delta\omega_h$ and BRDF in the form

$$\rho(\mathbf{k}_1, \mathbf{k}_2) = \frac{NA_{mf}p(\mathbf{h})P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{h})F((\mathbf{kh}))}{4A(\mathbf{k}_1\mathbf{n})(\mathbf{k}_2\mathbf{n})}. \tag{6}$$

Equation 6 is a somewhat modified version of the original result of Torrance and Sparrow who present its more detailed derivation [23].

The area A of the surface element can be written as a sum of the projected areas of all microfacets:

$$A = \sum_{facets} A_{mf}(\mathbf{hn})P(\mathbf{n}, \mathbf{h}), \tag{7}$$

where we introduce probability $P(\mathbf{n}, \mathbf{h})$ for a microfacet not to be "shadowed" in the surface normal direction $\mathbf{n}$ by other microfacets. If the surface is a height field, $P(\mathbf{n}, \mathbf{h}) = 1$ but in the general case some microfacets may not contribute to the area A of the projection. This question is related to the general shadowing term $P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{h})$ and we postpone its discussion until the next section. The "P" is used with a variable number of arguments that depend on what assumptions are in play for that equation.

Given a large number of microfacets, Equation 7 can be rewritten using the average over the ensemble of microfacets as

$$A = NA_{mf}\langle(\mathbf{hn})P(\mathbf{n}, \mathbf{h})\rangle_{ens}, \tag{8}$$

where $\langle...\rangle_{ens}$ denotes the averaging procedure. One of the most fundamental results in statistics states that as the size of the ensemble increases, for a certain function f of a random variable its average over ensemble $\langle f\rangle_{ens}$ converges with probability one to its average $\langle f\rangle$ over the distribution of the random variable. In our case we can write for any quantity $f(\mathbf{h})$:

$$\langle f(\mathbf{h})\rangle_{ens} = \langle f(\mathbf{h})\rangle = \int_\Omega f(\mathbf{h})p(\mathbf{h})d\omega_h, \tag{9}$$

where the integration is done over the unit sphere Ω of microfacet normal directions (Gaussian sphere). So, for the BRDF we finally have

$$\rho(\mathbf{k}_1, \mathbf{k}_2) = \frac{p(\mathbf{h})P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{h})F((\mathbf{kh}))}{4(\mathbf{k}_1\mathbf{n})(\mathbf{k}_2\mathbf{n})\langle(\mathbf{nh})P(\mathbf{n}, \mathbf{h})\rangle}, \tag{10}$$

and in the important special case of surface being a height field,

$$\rho(\mathbf{k}_1, \mathbf{k}_2) = \frac{p(\mathbf{h})P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{h})F((\mathbf{kh}))}{4(\mathbf{k}_1\mathbf{n})(\mathbf{k}_2\mathbf{n})\langle(\mathbf{nh})\rangle}. \tag{11}$$

Although we have assumed that all microfacets have equal area A_{mf} the result does not change if there is an arbitrary distribution of microfacet areas so long as this distribution is not correlated with $p(\mathbf{h})$, the distribution of normals.

Given a density $p(\mathbf{h})$, all terms in Equation 11 are straightforward to compute except for the shadowing term $P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{h})$. We now turn to the discussion of this shadowing term which is necessary to complete our formulation of the specular part of BRDF.

4 Shadowing Term

Most of the complexity of microfacet-based models arise from the shadowing function $P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{h})$. In this section we describe how previous models deal with this term and introduce a new simplified shadowing term.

4.1 Previous Shadowing Terms

On any rough surface it is likely that some microfacets will either not receive light, or light reflected by them will be blocked by other microfacets. The first situation is referred to by many authors as *shadowing* and the second as *masking*. However, these events are symmetrical and for simplicity we will refer to both of them as *shadowing*. A rigorous derivation of the probability that a point on the surface is both visible and illuminated (also known as the bistatic shadowing function) leads to very complicated expressions and a set of approximations is made to make the problem tractable. Several forms of the shadowing term have been derived in different fields [1, 18, 21, 23, 25] and some of them (usually after further simplification) were later introduced to computer graphics reflection models [4, 8, 22].

The most popular shadowing functions currently used are modifications of those of Smith [21], Sancer [18] and the original Torrance and Sparrow shadowing term [23]. The first two formulations are rather complex and are designed only for Gaussian height fields. Smith, in addition, assumes an isotropic surface. The shadowing function by Torrance and Sparrow is simple, but assumes an inconsistent model of an isotropic surface exclusively made by very long V-cavities. None of the existing functions is flexible enough to accommodate a sufficiently general distribution of microfacets. Also, most of the formulations operate with height distributions, not the more intuitive normal distribution $p(\mathbf{h})$. In addition to space limitations, this is the reason we do not present the expressions of previously derived shadowing functions here.

The reason most authors deal with height distribution functions is that shadowing is clearly a non-local event intimately related to the height distribution of the surface and this information is necessary for rigorous treatment of shadowing. In the next subsection we will, however, make several assumptions which allows us to derive a very general form of the shadowing term $P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{h})$ sufficient for our purposes.

4.2 New Shadowing Term

As indicated by the preceding discussion, we cannot treat shadowing rigorously if we assume a general form for the microfacet normal density function. Therefore, our generator is most appropriate in cases where the effects of shadowing are secondary compared with the influence of normal distribution shape. Even in these cases, however, we cannot ignore the shadowing term $P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{h})$. As can be seen from Equation 10, at the very least shadowing should take care of the divergence at grazing angles where the denominator terms disappear: $(\mathbf{k}_1\mathbf{n})(\mathbf{k}_2\mathbf{n}) \rightarrow 0$.

The shadowing term can be written as

$$P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{h}) = P(\mathbf{k}_1, \mathbf{h})P(\mathbf{k}_2, \mathbf{h} \,|\, \mathbf{k}_1), \tag{12}$$

where $P(\mathbf{k}_1, \mathbf{h})$ is the probability of not being shadowed in the direction $\mathbf{k}_1$ and $P(\mathbf{k}_2, \mathbf{h} \,|\, \mathbf{k}_1)$ is conditional probability of not being shadowed in the direction $\mathbf{k}_2$ given that the facet is not shadowed in direction $\mathbf{k}_1$. In general, $P(\mathbf{k}_2, \mathbf{h} \,|\, \mathbf{k}_1) \neq P(\mathbf{k}_2, \mathbf{h})$. For example, it is easy to see that in the extreme case where $\mathbf{k}_1 = \mathbf{k}_2$ we have $P(\mathbf{k}_2, \mathbf{h} \,|\, \mathbf{k}_1) = 1$. This shows that visibilities in the incoming and outgoing directions are correlated. Most of shadowing functions, however, are derived under the assumption of uncorrelated

visibilities. Van Ginneken et al. [24] considered how this correlation affects Smith's shadowing function, and found that its effect can be accounted for by modifying the uncorrelated expression.

In most of this paper we will use the uncorrelated form of the shadowing term written as a product of the two independent factors for each of the two directions:

$$P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{h}) = P(\mathbf{k}_1, \mathbf{h})P(\mathbf{k}_2, \mathbf{h}). \tag{13}$$

This leads to some underestimation of the BRDF if directions $\mathbf{k}_1$ and $\mathbf{k}_2$ are close to each other. If the viewing conditions are such that this arrangement is of particular importance (in a night driving simulator, for example) or if retroreflection is one of the pronounced features of surface appearance (see Section 6.4) we propose using a different form of the shadowing term:

$$P(\mathbf{k}_1, \mathbf{k}_2, \mathbf{h}) = (1 - t(\phi))P(\mathbf{k}_1, \mathbf{h})P(\mathbf{k}_2, \mathbf{h}) + t(\phi)min(P(\mathbf{k}_1, \mathbf{h}), P(\mathbf{k}_2, \mathbf{h})), \tag{14}$$

where $-\pi < \phi < \pi$ is the angle between the projections of vectors $\mathbf{k}_1$ and $\mathbf{k}_2$ onto the tangent plane and $t(\phi)$ is a correlation factor with values between 0 and 1. The case $t(\phi) = 0$ corresponds to the completely uncorrelated case. This form of correlated shadow term was chosen because it is simple and the resulting BRDF will still conserve energy with arbitrary $t(\phi)$, as will be shown in Section 5.3. We have not done extensive experimentation with the particular form of $t(\phi)$ but we do not believe it makes a large difference as long as $t(0) = 1$ and $t(\phi)$ monotonically decreases to almost zero as $|\phi|$ increases. The range of correlation effects was found in [24] to be on the order of 15-25 degrees, so we use a Gaussian in ϕ with the width of 15 degrees.

All we need now is an expression for $P(\mathbf{k}, \mathbf{h})$, the probability for a microfacet to be visible in a given direction $\mathbf{k}$. Note that $P(\mathbf{n}, \mathbf{h})$ in Equations 7, 8 and 10 of the previous section is just a special case of this probability with $\mathbf{k} = \mathbf{n}$. The key assumption we make is that probability for a microfacet to be visible in direction $\mathbf{k}$ does not depend on the microfacet's orientation $\mathbf{h}$ as long as it is not turned away from $\mathbf{k}$ (not self-shadowed), namely

$$P(\mathbf{k}, \mathbf{h}) = \begin{cases} P(\mathbf{k}) & \text{if } (\mathbf{kh}) > 0 \\ 0 & \text{if } (\mathbf{kh}) \leq 0 \end{cases} \tag{15}$$

This assumption is equivalent to the absence of correlation between the microfacet orientation and its position. This "distant shadower" assumption has been invoked before to simplify complicated shadowing expressions obtained in other fields [1, 21, 25] but we will use it in a different way - as a basis for deriving a simple and general shadowing function. Intuitively, it corresponds to rather rough surfaces and does not hold if the microfacets with certain orientation are more likely to be found at a certain height. For example, a surface made of cylinders as shown in Figure 3a will not obey this assumption while a very similar surface in Figure 3b might. In general, the more correlated the surface microfacets are, the less likely $P(\mathbf{k}, \mathbf{h})$ is to obey Equation 15.

The two surfaces in Figure 3 may still have the same distribution $p(\mathbf{h})$ and there is no way for us to distinguish between the two cases. Similarly, we will not be able to distinguish, for example, between "positive" and "negative" cylinders of Poulin and Fournier [16] but from their images it is clear that the differences in appearance due to microfacet visibility issues and not to the distribution of microfacets are minor in this case. If finer details of microfacet arrangement not captured by $p(\mathbf{h})$ are expected to substantially affect the appearance, some different framework should be used (see also Section 6.4).

The total projected area of a surface element onto direction $\mathbf{k}$ is $A(\mathbf{kn})$. It can also can be written in a way similar to Equation 7:

$$A(\mathbf{kn}) = \sum_{facets} A_{mf}(\mathbf{hk})_+ P(\mathbf{k}). \tag{16}$$
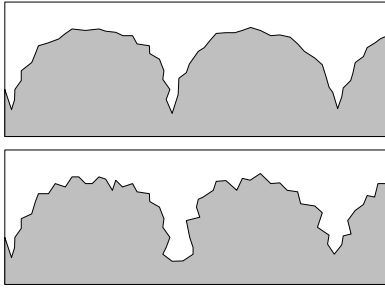
Figure 3: *Examples of surface microgeometry. Top: microfacets with almost vertical orientation are more likely to be found near the "bottom" of the surface and, therefore, are more likely to be shadowed. Bottom: orientation and height are largely uncorrelated.*

Here the subscript '+' refers to the fact that the summation is performed only over microfacets turned towards $\mathbf{k}$, namely the ones with $(\mathbf{hk}) > 0$. Introducing averaging over microfacets and, as before, replacing it by averaging over distribution, we get

$$A(\mathbf{kn}) = N A_{mf} P(\mathbf{k}) \langle (\mathbf{hk})_+ \rangle. \tag{17}$$

We are able to take $P(\mathbf{k})$ out of the averaging integral because of our assumption that it does not depend on $\mathbf{h}$. Because of the great importance of quantity $\langle (\mathbf{hk})_+ \rangle$ we introduce a new notation

$$g(\mathbf{k}) = \langle (\mathbf{hk})_+ \rangle = \int_{\Omega_+(\mathbf{k})} (\mathbf{hk})_+ p(\mathbf{h}) d\omega_h, \tag{18}$$

where the integration is done in $\mathbf{h}$-space over the hemisphere $\Omega_+(\mathbf{k})$ of directions $(\mathbf{hk}) > 0$ (Figure 4). Note that if the surface is a height field, $P(\mathbf{n}) = 1$ and Equations 8 and 17 immediately give a useful expression for $P(\mathbf{k})$:

$$P(\mathbf{k}) = \frac{(\mathbf{kn}) g(\mathbf{n})}{g(\mathbf{k})}. \tag{19}$$

In this special case $p(\mathbf{h}) = 0$ in the lower hemisphere and the averaging in $g(\mathbf{n})$ is effectively done over the complete distribution.

To handle a more general case, we note that each microfacet turned away from the direction $\mathbf{k}$ will have a shadow with area $A_{mf}(\mathbf{hk})$. This area must be subtracted from the contribution of microfacets turned towards $\mathbf{k}$. Again replacing sums by averages over ensemble and then over distribution, we write the projected area on the right-hand side of Equation 17 as

$$N A_{mf} P(\mathbf{k}) \langle (\mathbf{hk})_+ \rangle =$$
$$N A_{mf} \langle (\mathbf{hk})_+ \rangle + N A_{mf} \langle (\mathbf{hk})_- \rangle, \tag{20}$$

or

$$P(\mathbf{k}) = 1 + \frac{\langle (\mathbf{hk})_- \rangle}{g(\mathbf{k})}. \tag{21}$$

The second term is negative and the integration in it is done over the part $\Omega_-(\mathbf{k})$ of distribution complimentary to $\Omega_+(\mathbf{k})$ (Figure 4). It is clear from this equation that $P(\mathbf{k}) \leq 1$ as it should be. For a distribution of microfacet normals $p(\mathbf{h})$ to represent a valid surface, at the very least the average normal vector over the entire distribution must lie in the direction of the geometric normal $\mathbf{n}$ of the surface:

$$\int_{\Omega_+(\mathbf{k})} \mathbf{h} p(\mathbf{h}) d\omega_h + \int_{\Omega_-(\mathbf{k})} \mathbf{h} p(\mathbf{h}) d\omega_h =$$
$$\int_{\Omega} \mathbf{h} p(\mathbf{h}) d\omega_h = \mathbf{n}(\langle \mathbf{h} \rangle \mathbf{n}) \tag{22}$$
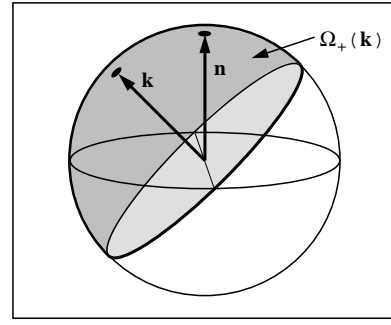
Figure 4: *Integration domain for $g(\mathbf{k})$*

Multiplying both sides of this equation by scalar $\mathbf{k}$ we have

$$\langle (\mathbf{hk})_+ \rangle + \langle (\mathbf{hk})_- \rangle = (\mathbf{kn}) \langle (\mathbf{hn}) \rangle, \tag{23}$$

or

$$\langle (\mathbf{hk})_- \rangle = (\mathbf{kn}) \langle (\mathbf{hn}) \rangle - g(\mathbf{k}). \tag{24}$$

Substituting this into Equation 21 we obtain an expression for $P(\mathbf{k})$:

$$P(\mathbf{k}) = \frac{(\mathbf{kn}) \langle (\mathbf{hn}) \rangle}{g(\mathbf{k})}. \tag{25}$$

Averaging in the numerator is done over the complete sphere Ω of directions. Note that Equation 19 is now just a special case of Equation 25 and that Equations 21 and 25 show that for any physically valid distribution $p(\mathbf{h})$ our probability of being visible will indeed lie between 0 and 1.

The combination of Equations 10, 13 (or 14) and 25 completely describes the specular part of BRDF. Using the uncorrelated form of shadowing term of Equation 13, we get

$$\rho(\mathbf{k}_1, \mathbf{k}_2) = \frac{p(\mathbf{h}) \langle (\mathbf{hn}) \rangle F((\mathbf{kh}))}{4 g(\mathbf{k}_1) g(\mathbf{k}_2)}. \tag{26}$$

Note the interesting fact that p.d.f. $p(\mathbf{h})$ does not even have to be normalized to be used in this equation. The above formula is well-suited to evaluation. Given $p(\mathbf{h})$, it is straightforward to evaluate the BRDF. Equation 26 is the main contribution of this paper. For the rest of the paper we will discuss implications and applications of this formula.

5 Extensions and Discussion

In this section we discuss several issues related to the specular-only single bounce BRDF model derived in the last section. In particular, we discuss an energy-conserving diffuse term, implementation issues, extension to non-Fresnel microfacets, and prove energy conservation.

5.1 Diffuse Term

Equation 26 describes the part of scattering process due to single-bounce reflections from microfacets. In addition to this specular part there will be other scattering events, such as multiple bounces and subsurface scattering. A complete description of these processes is rarely attempted in a general-purpose BRDF model and

their combined contribution is usually represented by adding a diffuse component to the specular BRDF. The most common form of the diffuse term is Lambertian:

$$\rho(\mathbf{k}_1, \mathbf{k}_2) = \frac{k_d \rho_d}{\pi} + k_s \rho_s(\mathbf{k}_1, \mathbf{k}_2), \qquad (27)$$

where $0 \leq \rho_d \leq 1$ is diffuse albedo of the surface while k_d and k_s are user-specified constants controlling the relative importance of specular and diffuse reflections. This is a perfectly valid option in our case as well. We can simply use Equation 26 for ρ_s and ensure that $k_d + k_s \leq 1$ to preserve the energy conservation achieved for the specular part (Section 5.3).

However, this simple form of diffuse term has problems. First of all, it is not obvious how to choose weights k_d and k_s. Second, it is clear that as more light is being reflected specularly, less of it is available for diffuse scattering, so the relative weights k_d and k_s of diffuse and specular reflections should not be constants. If Fresnel effects can cause k_s to approach one for grazing angles, k_d must be set to zero for all angles (since it is a constant). To take this effect into account in a way preserving reciprocity, we use a method of Shirley et al. [20] and write for k_d

$$k_d(\mathbf{k}_1, \mathbf{k}_2) = c(1 - R(\mathbf{k}_1))(1 - R(\mathbf{k}_2)), \qquad (28)$$

where

$$R(\mathbf{k}) = \int \rho_s(\mathbf{k}, \mathbf{k}')(\mathbf{k}'\mathbf{n}) d\omega_{k'} \qquad (29)$$

is the directional hemispherical reflectance of the specular term, where $\mathbf{k}'$ is the mirrored direction of $\mathbf{k}$. We also completely dispose of k_s by allowing the specular reflection to "have its way" and adjust the diffuse term so that it consistently follows the specular reflection. The normalization constant c is computed such that for $\rho_d = 1$ the total incident and reflected energies are the same. A complete BRDF will have the form

$$\rho(\mathbf{k}_1, \mathbf{k}_2) = c(1 - R(\mathbf{k}_1))(1 - R(\mathbf{k}_2))\rho_d + \rho_s(\mathbf{k}_1, \mathbf{k}_2). \quad (30)$$

This form of diffuse term implicitly assumes that there is no absorption on the surface and all the energy which is not reflected specularly is available for diffuse scattering. The situation is different in case of metals. First, if f_0 is the normal reflectance of the metal, only approximately f_0 fraction of incoming light is not absorbed by a flat metal surface. Second, diffuse scattering here is exclusively due to multiple bounces and thus the diffusely scattered light has a more saturated color of the metal than the primary reflection does. We attempt to take both of these effects into account by replacing ones in Equation 30 by f_0 and assigning ρ_d for a metal (which otherwise does not have any physical sense) to be f_0. Because the true fraction of non-absorbed light is greater than f_0, factor $(f_0 - R(\mathbf{k}))$ can become negative for some surfaces due to our approximation. We simply set the diffuse term to zero in such cases.

5.2 Implementation Issues

Implementation of our model in a rendering system is straightforward. For the Fresnel coefficient we use Schlick's approximate formula [19]

$$F((\mathbf{kn})) = f_0 + (1 - f_0)(1 - (\mathbf{kn}))^5 \qquad (31)$$

where again f_0 is the Fresnel factor at normal incidence. Note that we could also use the full Fresnel equations, but we use Schlick's formula only for convenience. This should not lead to significant accuracy problems as for the error introduced by Schlick's formula is smaller than one percent compared with the full Fresnel expression [19]. To generate a BRDF for a new distribution $p(\mathbf{h})$ all we

need, in addition to the implementation of $p(\mathbf{h})$ itself, are values for $g(\mathbf{k})$ and $R(\mathbf{k})$. Unfortunately, because of the non-standard integration domain of $g(\mathbf{k})$, analytical expressions for this function can be obtained only for the most trivial $p(\mathbf{h})$'s and we need to resort to numerical integration.

However, the integrals are well-behaved and the results are smooth functions for non-singular $p(\mathbf{h})$. This allows us to compute values of both $g(\mathbf{k})$ and $R(\mathbf{k})$ on a very coarse grid using available numerical packages, store the results in a table and use bilinear interpolation during the rendering process. We have used a total of 200 grid points (for many distributions an even coarser grid should be sufficient). Integration was done using both Matlab and a simple home-built Monte Carlo routine. Two sets of computed $R(\mathbf{k})$ (one with $f_0 = 1$ and one with $f_0 = 0$) are sufficient to compute $R(\mathbf{k})$ for a material with arbitrary f_0 for a given microfacet distribution.

In the BRDF generation phase we start from $p(\mathbf{h})$ and output a compact numerical representation of three two-dimensional functions: $g(\mathbf{k})$, $R(\mathbf{k})$ with $f_0 = 0$ and $R(\mathbf{k})$ with $f_0 = 1$. The last two functions are only used for the diffuse term and are not required for its simpler form in Equation 27. During rendering we use these data to compute the full four dimensional BRDF for arbitrary $\mathbf{k}_1$ and $\mathbf{k}_2$. At this stage we also use data for normal reflectance f_0 and diffuse albedo ρ_d. Wavelength dependence of these quantities controls the color of the surface. We have not done a careful performance analysis but from our experience for a non-trivial $p(\mathbf{h})$ most of the BRDF computation time is due to evaluating this normal distribution function.

Note that most distributions have some symmetry which can be exploited to further reduce the amount of data and/or generation time. Data for an anisotropic Gaussian distribution of normals, for example, need be computed only over a quarter of the hemisphere and for any isotropic distribution functions $g(\mathbf{k})$ and $R(\mathbf{k})$ become one dimensional.

Finally, if a particular type of parameterized distribution (Gaussian, for example) is used often it should be possible to approximate $g(\mathbf{k})$ with a simple function of $\mathbf{k}$ and distribution parameters as is commonly done to increase the efficiency of reflection models. The same is true for $R(\mathbf{k})$ but these functions usually have more complex shapes.

5.3 Energy Conservation

By inspection of the formulas, it is clear that generated BRDFs are reciprocal. We now prove now that they also conserve energy for any physically plausible $p(\mathbf{h})$. To do this, we assume the worst-case scenario of $F((\mathbf{kh})) = 1$ and shadowing term in Equation 14 with $t(\phi) = 1$ (because $P(\mathbf{k}) \leq 1$ this corresponds to the largest possible shadowing term for our model). The BRDF in this case will be

$$\rho(\mathbf{k}_1, \mathbf{k}_2) = \frac{p(\mathbf{h}) min(P(\mathbf{k}_1), P(\mathbf{k}_2))}{4\langle(\mathbf{hn})\rangle(\mathbf{k}_1\mathbf{n})(\mathbf{k}_2\mathbf{n})} \leq$$

$$\frac{p(\mathbf{h})P(\mathbf{k}_1)}{4\langle(\mathbf{hn})\rangle(\mathbf{k}_1\mathbf{n})(\mathbf{k}_2\mathbf{n})}$$

Hemispherical reflectance for a given incoming direction is

$$R(\mathbf{k}_1) = \int \rho_s(\mathbf{k}_1, \mathbf{k}_2)(\mathbf{k}_2\mathbf{n}) d\omega_2 \leq$$

$$\frac{P(\mathbf{k}_1)}{4\langle(\mathbf{hn})\rangle(\mathbf{k}_1\mathbf{n})} \int p(\mathbf{h}) d\omega_2 =$$

$$\frac{P(\mathbf{k}_1)}{4\langle(\mathbf{hn})\rangle(\mathbf{k}_1\mathbf{n})} \int p(\mathbf{h}) 4(\mathbf{k}_1\mathbf{h}) d\omega_h$$

The last transition is done using Equation 5. The integration is done over a complex region of h-space which is in any case contained

in the hemisphere $\Omega_+(\mathbf{k}_1)$. Extending the integral over the whole $\Omega_+(\mathbf{k}_1)$ and using definitions 18 of $g(\mathbf{k})$ and 25 of $P(\mathbf{k})$ we complete the proof:

$$R(\mathbf{k}_1) \leq \frac{P(\mathbf{k}_1)}{\langle(\mathbf{hn})\rangle(\mathbf{k}_1\mathbf{n})} \int_{\Omega_+(\mathbf{k}_1)} (\mathbf{k}_1\mathbf{h})p(\mathbf{h})d\omega_h =$$

$$\frac{P(\mathbf{k}_1)g(\mathbf{k}_1)}{\langle(\mathbf{hn})\rangle(\mathbf{k}_1\mathbf{n})} = 1 \qquad (32)$$

The only fact we used in our proof is that $P(\mathbf{k}) \leq 1$ for any $\mathbf{k}$. In Section 4, in turn, this was shown to be the case for any $p(\mathbf{h})$ whose average normal vector $\langle\mathbf{h}\rangle$ is parallel to the geometric normal of the surface. This is the only restriction on microfacet distribution $p(\mathbf{h})$. If it is satisfied, the generated BRDF will conserve energy.

5.4 Non-Fresnel Microfacets

Our model is not restricted to perfectly specular microfacets. In general, microfacets with many orientations will contribute to surface BRDF for given incoming and outgoing directions and integration of their contribution is necessary.

Let all microfacets have elementary BRDF β. Then we can repeat with some modifications the derivation from Sections 3 and 4 to arrive at the result

$$\rho(\mathbf{k}_1, \mathbf{k}_2) = \frac{P(\mathbf{k}_1, \mathbf{k}_2)}{(\mathbf{k}_1\mathbf{n})(\mathbf{k}_2\mathbf{n})\langle(\mathbf{nh})\rangle}$$

$$\int \beta(\mathbf{k}_1, \mathbf{k}_2)(\mathbf{k}_1\mathbf{h})_+(\mathbf{k}_2\mathbf{h})_+p(\mathbf{h})d\omega_h \qquad (33)$$

The integration is done over the sector where both $(\mathbf{k}_1\mathbf{h})$ and $(\mathbf{k}_2\mathbf{h})$ are positive and any of shadowing terms $P(\mathbf{k}_1, \mathbf{k}_2)$ from Section 4 can be used. Note that $\beta(\mathbf{k}_1, \mathbf{k}_2)$ is usually specified with respect to microfacet's local coordinate system and a coordinate transformation is necessary to obtain its value for the integral in Equation 33.

Although this extension considerably broadens the range of surfaces our model is applicable to, we also lose one of the main advantages of our approach: compactness. Before, we could represent a general four dimensional BRDF using only two dimensional functions. The integral in Equation 33, however, is a four dimensional function by itself and does not, in general, allow lower dimensional representation. For some special cases, such as Lambertian elementary BRDF coupled with isotropic $p(\mathbf{h})$ the integral becomes three dimensional and, therefore, feasible to compute, store and use in a way similar to that described in Section 5.2. For an isotropic Gaussian distribution of Lambertian microfacets the general behavior of the generated BRDF is similar to that of Oren-Nayar's model [14], namely, retroreflection is increased compared to a Lambertian surface (Figure 8).

6 Applications

In this section we apply our model to a variety of surface types. Although we have implemented our model in a Monte Carlo ray tracer capable of handling complex geometries and illumination effects, our images in this section intentionally show very simple objects and lighting conditions. In particular, illumination is coming from a single small light source far from the scene and indirect lighting is not included. This is done to emphasize effects due to BRDF of the material and to make the comparison with previous results easier.

Reflectance data of gold are used as f_0 (see Section 5.2) for all metal objects while for non-metals f_0 is set to 5% across the visible spectrum.

Figure 5: *Anisotropic Gaussian golden spheres with $\sigma_x = 0.1$, $\sigma_y = 1.0$. Left: Ward. Right: new model.*
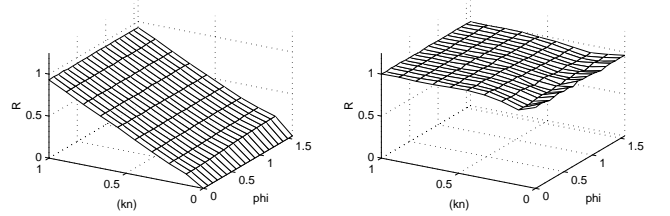
Figure 6: *Directional hemispherical reflectance as a function of incoming angle for perfectly reflecting microfacets with Gaussian distribution $\sigma_x = 0.1$, $\sigma_y = 0.2$. For an ideal flat surface R should be 1.0 everywhere. Left: Ward. Right: new model.*
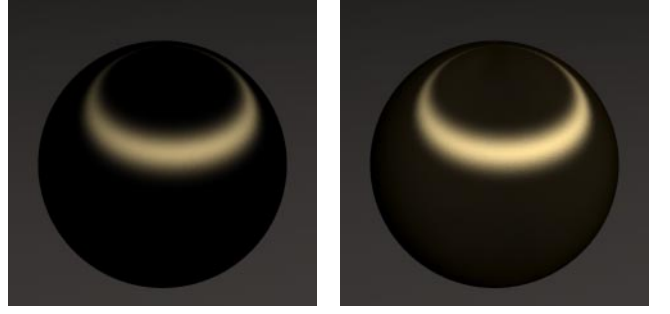

Figure 7: *Anisotropic Gaussian golden painted plastic spheres with $\sigma_x = 0.1$, $\sigma_y = 0.2$. Left: Ward. Right: new model.*
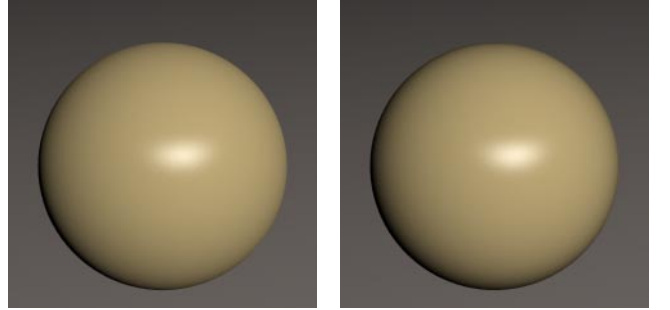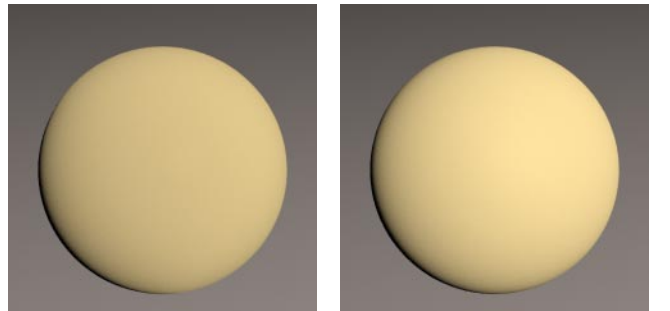

Figure 8: *Gaussian spheres with Lambertian microfacets. Right: new model with $\sigma_x = \sigma_y = 1.0$. Left: Oren-Nayar with compatible parameters.*

6.1 Gaussian Surfaces

By far the most popular distribution used in BRDF research literature is Gaussian. This is due to both its practical importance and nice mathematical properties. Gaussians are used in all four major categories of BRDF models outlined in the introduction. While some of this work is closer to our approach in its theoretical foundations, we feel that from the practical point of view our model is closest to that of Ward [26]. Ward's BRDF is simple, handles anisotropic distributions and seeks to reproduce the main character of the material's reflectance behavior without attempting an overly detailed description. Other previous models do not simultaneously possess all these properties.

To create an anisotropic Gaussian BRDF, we use the distribution

$$p(\mathbf{h}) = c * \exp(-\tan^2\theta(\cos^2\phi/\sigma_x^2 + \sin^2\phi/\sigma_y^2)) \qquad (34)$$

where θ is the angle between the half vector $\mathbf{h}$ and the surface normal, ϕ is the azimuth angle of $\mathbf{h}$ and c is a normalization constant.

Two side-by-side comparisons of our model with Ward's are shown in Figures 5 and 7. Note that the shape of highlight is nearly identical while there are some differences in the diffuse part of images which is due to Ward effectively using a simpler form (Equation 27) of the diffuse component. In particular, for our metal sphere on Figure 5 the diffuse component appears automatically when there is enough energy left after single-bounce scattering. To achieve the same effect in Ward's model (and any other using the popular Lambertian diffuse term) it would be necessary to manually adjust the diffuse reflectance parameter.

This figure also shows that the highlight is brighter for our BRDF. The general reason for this is clear from Figure 6 where the hemispherical reflectance R is plotted versus the incoming light direction. To make the plots directly comparable, we show data for most reflecting specular BRDF in both cases ($f_0 = 1$ for our model and $\rho_s = 1$ in Equation 5 of Ward's paper [26]) and do not include the diffuse term. For the values of parameters shown, the surface is quite close to being flat, so one would expect that R should be close to that of flat surface, 1.0 in this case. One can see from the plots that our model behaves as expected while Ward's does not. Note also that the true value for R at the grazing angle ($(\mathbf{kn}) = 0$) is infinite for Ward's model [13] and we simply extrapolate previous behavior to get the data point at the grazing angle.

While our approach does require an extra generation step, computation time during the rendering process of our BRDF is close to that of Ward's and our model is a viable alternative where energy conservation is of great importance for a particular application.

Figure 8 compares a BRDF generated for an isotropic Gaussian distribution of Lambertian microfacets with an extension of our process (Section 5.4) and Oren-Nayar model with compatible parameters. Both BRDFs have the tendency to make objects appear "flatter" than the Lambertian BRDF due to increased retroreflection.

6.2 Grooved Surface

A surface consisting of ideal V-grooves all running in a given direction will have its $p(\mathbf{h})$ proportional to the sum of two delta functions, each accounting for microfacets forming one side of a groove. Replacing these delta functions with narrow Gaussians ($\sigma = 0.1$) to account for imperfections and going through our generation process, we create a BRDF which correctly shows the main feature of a grooved surface's reflectance, double reflections. Figure 9 shows a piece of grooved metal illuminated by a *single* light source. The orientation of the grooves on the left is perpendicular to the viewing direction while on the right they are parallel.
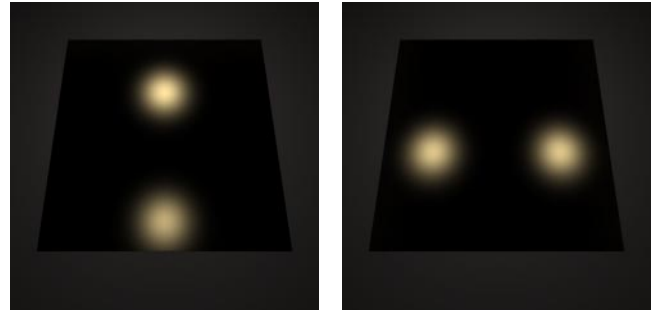

Figure 9: *Double highlights from a single light source for the same metallic grooved surface at two orientations of the grooves. Grooves are symmetrical with the angle of 40 degrees*

6.3 Satin

The microstructure of woven cloth is usually thought of as a symmetric pattern of interwoven cylindrical fibers running in perpendicular directions. While it would be possible to generate a BRDF corresponding to this structure with our approach, the surface of particular fabric we studied had a different microstructure shown in Figure 10. It is created almost exclusively by fibers running in one direction with about 70% of the fiber length lying in the relatively flat part of the fiber while the other 30% corresponding to the bent parts at the ends. We model the distribution of microfacets as a linear combination of two terms corresponding to these flat and bent parts of the cylindrical fiber: $p(\mathbf{h}) = 0.7 * p_{flats}(\mathbf{h}) + 0.3 * p_{ends}(\mathbf{h})$. The coefficients reflect mutual area contributions of the two parts to the complete distribution. Both $p_{flats}(\mathbf{h})$ and $p_{ends}(\mathbf{h})$ were chosen to be "cylindrical" Gaussian heightfields ($\sigma_y = \infty$, $p(\mathbf{h}) = 0$ for $(\mathbf{hn}) < 0$) with different widths. Values $\sigma_x = 0.1$ for $p_{flats}(\mathbf{h})$ and $\sigma_x = 0.3$ for $p_{ends}(\mathbf{h})$ were used. Strictly speaking, the shape of real $p_{ends}(\mathbf{h})$ would probably be more accurately modeled by a distribution with flatter top and faster drop-offs than that of a Gaussian. This was attempted but the results were almost identical visually, so a simpler Gaussian distribution was used for the final image. This is consistent with our belief that the very precise characterization of the microfacet distribution is not needed for visual applications. Note that because $g(\mathbf{k})$ is linear in $p(\mathbf{h})$, no new integration is necessary to compute $g(\mathbf{k})$ if g's corresponding to $p_{flats}(\mathbf{h})$ and $p_{ends}(\mathbf{h})$ are already computed. This suggests an efficient way of creating new distributions as a linear combination of ones for which $g(\mathbf{k})$ has been previously computed. For example, small contribution due to perpendicular fibers can be added in this manner if necessary.

Because the appearance of real cloth is dramatically affected by the presence of characteristic wrinkles, we used a dynamic simulation method [17] to create cloth geometry. The left side of Figure 11 shows a satin tablecloth rendered with generated BRDF. It is interesting to contrast this image with the image on the right using the same geometric model with the BRDF described in the next section.

6.4 Velvet

Velvet is another example of a material with interesting reflectance properties not easily conveyed by conventional BRDFs. In their virtual gonioreflectometer, Westin et al. [27], model velvet microstructure as a forest of narrow cylinders (fibers) with the orientation of each cylinder perturbed randomly. While it is difficult to write an exact $p(\mathbf{h})$ corresponding to such "surface" for the reasons outlined in Section 2, a simple intuitive form of this function written as an "inverse Gaussian" heightfield is enough to capture the main char-
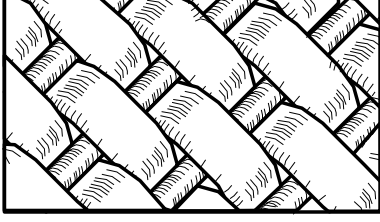
Figure 10: *Microgeometry of our sample of satin.*

Figure 11: *Synthetic satin (left) and velvet (right) tablecloths. The geometries are identical.*
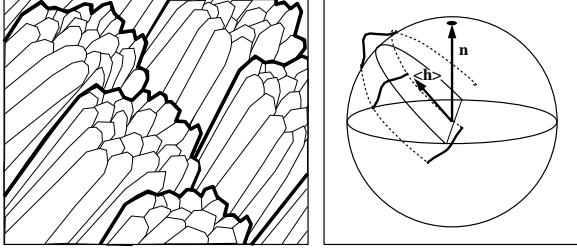

Figure 12: *Microgeometry of velvet (left) and $p(\mathbf{h})$ used to model it (right).*

Figure 13: *A tablecloth made of two different colors of slanted fiber velvets.*

acter of the distribution:

$$p(\mathbf{h}) = c * \exp(-\cot^2 \theta / \sigma^2), \qquad (35)$$

with $\sigma = 0.5$ for the image on the right of Figure 11 which shows a material with distinct velvet-like reflectance properties. Because retroreflection is one of the most pronounced reflection properties of velvet [11], we used the correlated form of shadowing term (Equation 14) to generate both this and slanted fiber (see below) velvet BRDFs. Contrary to Westin et al. we ignore the tips of the fibers due to their very small area. If there were any specular highlights due to the tips, their contribution can be easily added by forming a linear combination of an inverse Gaussian with a regular Gaussian distribution.

Although this approach produced good results, a symmetric forest of fibers was not what we saw when we examined a piece of real velvet. More realistic structure is shown on the left of Figure 12. The fabric consists of rows of tightly woven bundles of filament. Each bundle is slanted with the angle of about 40 degrees with respect to the geometric normal of the cloth surface. We can call this arrangement milliscale geometry in contrast with microgeometry formed by the thin fibers themselves. Similar geometry was credited as the major reason for velvet anisotropic reflection behavior by Lu et al. [11]. Strictly speaking, our model does not take into account visibility issues due to this higher-order arrangement of microfacets. The most consistent approach therefore would be to model this structure explicitly, for example as a collection of slanted cylinders applying two different BRDFs (both of which can be generated by our process) to the tops and to the sides of these cylinders. An easier alternative would be an attempt to create a simple distribution of microfacets $p(\mathbf{h})$ which, although potentially non-physical, can account for the milliscale visibility and produce a BRDF with necessary reflection properties.

Looking carefully at the velvet highlight structure we saw that it is the sides of the bundles and not the tops which contribute the most to the reflection. This suggests that we can try to reproduce most of the behavior with a specular BRDF based exclusively on the $p(\mathbf{h})$ accounting for the microfacets on the sides of the bundles. A "slanted" version of cylindrical Gaussian distribution ($\sigma_y = \infty$, $\sigma_x = 0.5$) schematically shown on the right of Figure 12 was used. The only place where we used the part of distribution due to the tops of the bundles is the computation of $\langle(\mathbf{h}\mathbf{n})\rangle$ when we double this value due to the tops contribution. Note two facts about this distribution: it is not a height field and its average vector $\langle\mathbf{h}\rangle$ does not point in the direction of geometric normal. While the first feature does not present any problem in our approach, the second one shows that this distribution is not physically realizable and, as a result, the energy conservation of the generated BRDF is not guaranteed. Computations of $R(\mathbf{k})$ show that this quantity indeed exceeds one for 14 out of our 200 directional data points in the hypothetical case of perfectly reflecting ($f_0 = 1.0$) fiber material but was never a problem for our $f_0 = 0.05$ synthetic fibers.

Figure 13 shows the results of this process. The illumination and viewing directions are almost parallel but due to the slant of the fibers the left side of the tablecloth is substantially brighter than the right one. This is in good agreement with the behavior of real velvet we observed. The right image of Figure 13 shows some limitations of our approach. Because we do not handle the details of multiple-bounce scattering and simply introduce a diffuse term to account for them, the right side of the red tablecloth does not look as it does for the real velvet. In the real material, light experiences multiple bounces among the red fibers for this viewing geometry acquiring a deep dark (almost black) color in the process. This is not captured by our simple diffuse term.

6.5 Unusual Distributions

We can take to extreme the use of the desired reflection properties as the only guidance in creating the distribution $p(\mathbf{h})$ regardless of whether a material described by this function exists or is even physically possible. For example, we can modulate a Gaussian $p(\mathbf{h})$ with an arbitrary function or even an image to create the unusual highlights shown in Figure 1. As long as the modulation is symmetric enough to keep the average vector $\langle\mathbf{h}\rangle$ in the normal direction (such as the distribution used for the image on the left of Figure 1), the BRDF will be energy conserving. A more general modulation may result in $\langle\mathbf{h}\rangle$ no longer parallel to $\mathbf{n}$ but in practice we notice that as long as this effect is not very strong, the energy conservation is not affected. For example, image on the right of Figure 1 was created with an energy conserving BRDF. While such unusual distributions are not of great value in realistic image synthesis, they clearly demonstrate the generality of our approach and can potentially find applications in the special effects industry.

7 Conclusion

The new BRDF model presented in this paper is well-suited to surfaces whose primary characteristic is the shape of the specular highlight. We have found it reasonably straightforward to design new BRDFs for surfaces because the diffuse term and energy conservation are handled in a natural manner that does not require substantial user intervention, and the parameters used in the model are intuitive. However, for surfaces whose appearance is not dominated by the specular highlight, our model is not well-suited.

We have found that using our model does not require much hand-tuning of parameters; the images in the last section were generated with very few iterations on parameter values. We speculate that a model for subsurface effects in a similar spirit to our model is possible. The user would specify some simple parameters analogous to $p(\mathbf{h})$ and a BRDF would be generated. We also believe that there should ultimately be separate terms for the components of the BRDF accounted for by primary specular reflection, multiple-bound specular reflection, and subsurface scattering.

Acknowledgements

References

[1] BECKMANN, P. Shadowing of random rough surfaces. *IEEE Transactions on Antennas and Propagation 13* (1965), 384–388.

[2] BLINN, J. F. Models of light reflection for computer synthesized pictures. *Computer Graphics (Proceedings of SIGGRAPH 77) 11*, 2 (July 1977), 192–198.

[3] CABRAL, B., MAX, N., AND SPRINGMEYER, R. Bidirectional reflectance functions from surface bump maps. *Computer Graphics 21*, 4 (July 1987), 273–282. ACM Siggraph '87 Conference Proceedings.

[4] COOK, R. L., AND TORRANCE, K. E. A reflectance model for computer graphics. *Computer Graphics 15*, 3 (August 1981), 307–316. ACM Siggraph '81 Conference Proceedings.

[5] GONDEK, J. S., MEYER, G. W., AND NEWMAN, J. G. Wavelength dependent reflectance functions. In *Proceedings of SIGGRAPH '94 (Orlando, Florida, July 24–29, 1994)* (July 1994), A. Glassner, Ed., Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH, ACM Press, pp. 213–220.

[6] GOURAUD, H. Continuous shading of curved surfaces. *Communications of the ACM 18*, 6 (June 1971), 623–629.

[7] GREENBERG, D. P., TORRANCE, K. E., SHIRLEY, P., ARVO, J., FERWERDA, J. A., PATTANAIK, S., LAFORTUNE, E. P. F., WALTER, B., FOO, S.-C., AND TRUMBORE, B. A framework for realistic image synthesis. *Proceedings of SIGGRAPH 97* (August 1997), 477–494.

[8] HE, X. D., TORRENCE, K. E., SILLION, F. X., AND GREENBERG, D. P. A comprehensive physical model for light reflection. *Computer Graphics 25*, 4 (July 1991), 175–186. ACM Siggraph '91 Conference Proceedings.

[9] LAFORTUNE, E. P., AND WILLEMS, Y. D. Using the modified phong BRDF for physically based rendering. Tech. Rep. CW197, Computer Science Department, K.U.Leuven, November 1994.

[10] LAFORTUNE, E. P. F., FOO, S.-C., TORRANCE, K. E., AND GREENBERG, D. P. Non-linear approximation of reflectance functions. *Proceedings of SIGGRAPH 97* (August 1997), 117–126.

[11] LU, R., KOENDERINK, J. J., AND KAPPERS, A. M. L. Optical properties (bidirectional reflection distribution functions) of velvet. *Applied Optics 37*, 25 (1998), 5974–5984.

[12] MARSCHNER, S. R., WESTIN, S. H., LAFORTUNE, E. P. F., TORRANCE, K. E., AND GREENBERG, D. P. Image-based BRDF measurement including human skin. *Eurographics Rendering Workshop 1999* (June 1999).

[13] NEUMANN, L., NEUMANN, A., AND SZIRMAY-KALOS, L. Compact metallic reflectance models. *Computer Graphics Forum 18*, 13 (1999).

[14] OREN, M., AND NAYAR, S. K. Generalization of lambert's reflectance model. In *Proceedings of SIGGRAPH '94 (Orlando, Florida, July 24–29, 1994)* (July 1994), A. Glassner, Ed., Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH, ACM Press, pp. 239–246.

[15] PHONG, B.-T. Illumination for computer generated images. *Communications of the ACM 18*, 6 (June 1975), 311–317.

[16] POULIN, P., AND FOURNIER, A. A model for anisotropic reflection. *Computer Graphics 24*, 3 (August 1990), 267–282. ACM Siggraph '90 Conference Proceedings.

[17] PROVOT, X. Deformation constraints in a mass-spring model to describe rigid cloth behavior. In *Proceedings of Graphics Interface '95* (1995), pp. 147–154.

[18] SANCER, M. I. Shadow corrected electromagnetic scattering from randomly rough surfaces. *IEEE Transactions on Antennas and Propagation AP-17*, 5 (September 1969), 577–585.

[19] SCHLICK, C. An inexpensive BRDF model for physically-based rendering. *Computer Graphics Forum 13*, 3 (1994), 233—246.

[20] SHIRLEY, P., HU, H., SMITS, B., AND LAFORTUNE, E. A practitioners' assessment of light reflection models. In *Pacific Graphics* (October 1997), pp. 40–49.

[21] SMITH, B. G. Geometrical shadowing of a random rough surface. *IEEE Transactions on Antennas and Propagation 15* (1967), 668–671.

[22] STAM, J. Diffraction shaders. *Proceedings of SIGGRAPH 99* (August 1999), 101–110.

[23] TORRANCE, K. E., AND SPARROW, E. M. Theory for off-specular reflection from roughened surfaces. *Journal of Optical Society of America 57*, 9 (1967).

[24] VAN GINNEKEN, B., STAVRIDI, M., AND KOENDERINK, J. J. Diffuse and specular reflectance from rough surfaces. *Applied Optics 37*, 1 (1998), 130–139.

[25] WAGNER, R. J. Shadowing of randomly rough surfaces. *Journal of Acoustic Society of America 41* (1967), 138–147.

[26] WARD, G. J. Measuring and modeling anisotropic reflection. *Computer Graphics 26*, 4 (July 1992), 265–272. ACM Siggraph '92 Conference Proceedings.

[27] WESTIN, S. H., ARVO, J. R., AND TORRANCE, K. E. Predicting reflectance functions from complex surfaces. *Computer Graphics 26*, 2 (July 1992), 255–264. ACM Siggraph '92 Conference Proceedings.

An Anisotropic Phong Light Reflection Model

Michael Ashikhmin   Peter Shirley

University of Utah
www.cs.utah.edu

Abstract.
We present a new BRDF model that attempts to combine the advantages of
the various empirical models currently in use. In particular, it has intuitive pa-
rameters, is anisotropic, energy-conserving, reciprocal, has an appropriate non-
Lambertian diffuse term, and is easy to use in a Monte Carlo framework.

1 Introduction

Physically-based rendering systems describe reflection behavior using the *bidirectional
reflectance distribution function* (BRDF) [3]. At a given point on a surface the BRDF
is a function of two directions, one toward the light and one toward the viewer. The
characteristics of the BRDF will determine what "type" of material the viewer thinks
the displayed object is composed of, so the choice of BRDF model and its parameters
is important.

We would like to have a BRDF model that works for "common" surfaces such as
metal and plastic, and has the following characteristics:

1. **Plausible:** as defined by Lewis [5], this refers to the BRDF obeying energy con-
   servation and reciprocity.
2. **Anisotropy:** the material should model simple anisotropy such as seen on brushed
   metals.
3. **Intuitive parameters:** for material such as plastics there should be parameters
   such as R_d for the substrate and R_s for the normal specular reflectance as well as
   two roughness parameters n_u and n_v.
4. **Fresnel behavior:** specularity should increase as the incident angle goes down.
5. **Non-Lambertian diffuse term:** The material should allow for a diffuse term,
   but the component should be non-Lambertian to assure energy conservation in
   the presence of Fresnel behavior.
6. **Monte Carlo friendliness:** there should be some reasonable probability den-
   sity function that allows straightforward Monte Carlo sample generation for the
   BRDF.

Neumann et al's metallic model [6] captures items 1, 3, 4, and 6. Schlick's model [8]
captures items. Ward's model [10] captures items 2, and 3. It only violates 1 for energy
conservation at grazing angles. It also approximates Monte Carlo friendliness by giving
a sample generation method but does not specify what the underlying density function
is.

Our goal is to find a BRDF with all the properties outlined. Our basic strategy is
to make a Fresnel-weighted Phong-style cosine lobe model that is anisotropic. This
strategy borrows pieces from Ward's model [10] and from Neumann and Neumann's
model [6]. In addition, we add some correction terms that are crucial to keep the di-
rectional hemispherical reflection near the desired level. For the diffuse term we use

1
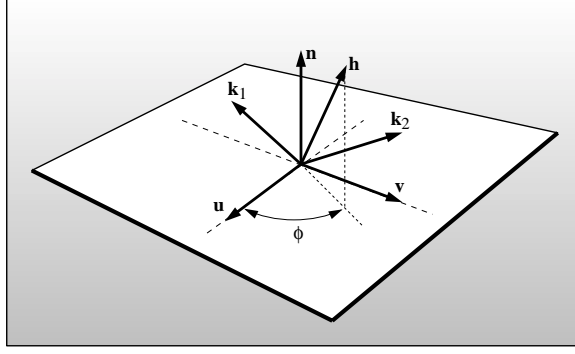
Fig. 1. *Geometry of reflection. Note that* $\mathbf{k}_1$, $\mathbf{k}_2$, *and* $\mathbf{h}$ *share a plane, which usually does not include* $\mathbf{n}$.

$(\mathbf{ab})$	scalar (dot) product of vectors $\mathbf{a}$ and $\mathbf{b}$
$\mathbf{k}_1$	normaized vector to light
$\mathbf{k}_2$	normaized vector to viewer
$\mathbf{n}$	surface normal to macroscopic surface
$\rho(\mathbf{k}_1, \mathbf{k}_2)$	BRDF
$\mathbf{h}$	normalized half-vector between $\mathbf{k}_1$ and $\mathbf{k}_2$
$p_h(\mathbf{h})$	probability density function for half-vector
$p(\mathbf{v})$	probability density function for reflection sampling rays
$F(\cos\theta)$	Fresnel reflectance for incident angle θ

Table 1. Important terms used in the paper

the basic method of Shirley et al. [9] to allow the diffuse-specular tradeoff to conserve energy.

We decompose the BRDF into a specular component and a diffuse component. Accordingly, we write our BRDF as the classical sum of two parts:

$$\rho(\mathbf{k}_1, \mathbf{k}_2) = \rho_s(\mathbf{k}_1, \mathbf{k}_2) + \rho_d(\mathbf{k}_1, \mathbf{k}_2), \tag{1}$$

where the first term accounts for the specular reflection and will be presented in the next section. While it is possible to use the Lambertian BRDF as diffuse term $\rho_d(\mathbf{k}_1, \mathbf{k}_2)$ in our model, we will discuss a better solution in section 3. We discuss how to implement the model in Section 4. Readers who just want to implement the model should skip to that section.

2 Anisotropic specular BRDF

Several shapes for the specular lobe have been proposed in the literature with Phong power-of-cosine lobe being by far the most popular. This is primarily due to its simplicity. The original form of the Phong shader [7] has several problems which triggered at creating a more physically plausible Phong-style BRDF [4, 5, 6]. We will also use a Phong-style specular lobe in our model but will make this lobe anisotropic and incorporate Fresnel behavior while attempting to preserve the simplicity of the initial model and physical plausibility achieved earlier for the Phong BRDF by other researchers.

As our starting point we will choose recent result of Neumann and Neumann [6] who improved energy conservation properties of Phong model and made the BRDF well-suited for importance sampling in a Monte-Carlo framework. Their main result in our notation is:

$$\rho(\mathbf{k}_1, \mathbf{k}_2) = c * \frac{(\mathbf{r}_1 \mathbf{k}_2)^n}{\max((\mathbf{n}\mathbf{k}_1), (\mathbf{n}\mathbf{k}_2))} F(\cos\theta), \tag{2}$$

where n is Phong exponent, $\mathbf{r}_1$ is the unit vector in the direction of mirror reflection of vector $\mathbf{k}_1$ around the surface normal, c is a normalization constant and $F(\cos\theta)$ is the Fresnel fraction. Several choices of argument θ are discussed by the authors. The division by $\max((\mathbf{n}\mathbf{k}_1), (\mathbf{n}\mathbf{k}_2))$ "pumps up" the total hemispherical reflectance $R(\mathbf{k})$ of the surface and in the limit $n \to \infty$ (Phong representation of a perfect mirror) gives $R(\mathbf{k}) = 1$ for any $\mathbf{k}$ not exactly at grazing incidence. While there are several ways to achieve this behavior, this particular form preserves reciprocity and avoids the divergence near the grazing angle frequently observed for other simple models.

To extend this model to anisotropic surfaces we use an approach similar to Ward's [10] who made the parameters of his gaussian lobe model depend on the azimuthal angle of the unit half vector with respect to a system of coordinates attached to the surface. Instead of single Phong parameter n in Equation 2 we introduce two parameters n_u and n_v and write the exponent as $n_u \cos^2 \phi + n_v \sin^2 \phi$ where ϕ is the azimuthal angle of half-vector $\mathbf{h}$. To get a better intuition about the model and, more importantly, to allow more efficient importance sampling of the specular lobe in a way discussed below, we also replace the Phong cosine $(\mathbf{r}_{\mathbf{k}1} \mathbf{k}_2)$ by $(\mathbf{n}\mathbf{h})$, a transformation originally proposed by Blinn [1]. Our BRDF is now

$$\rho(\mathbf{k}_1, \mathbf{k}_2) = c * \frac{(\mathbf{n}\mathbf{h})^{n_u \cos^2 \phi + n_v \sin^2 \phi}}{\max((\mathbf{n}\mathbf{k}_1), (\mathbf{n}\mathbf{k}_2))} F(\cos\theta). \tag{3}$$

Although our model is mostly empirical, to proceed further it is useful to interpret certain parts of Equation 3 in terms of physics-based microfacet models [2]. These models treat a surface as a collection of small mirror-like facets. Reflection from these facets is is governed by Fresnel laws. At a high level, a BRDF obtained with such models have the form

$$\rho(\mathbf{k}_1, \mathbf{k}_2) = c * p_h(\mathbf{h}) Vis(\mathbf{k}_1, \mathbf{k}_2, \mathbf{h}) F((\mathbf{k}\mathbf{h})), \tag{4}$$

where $p_h(\mathbf{h})$ is the microfacet probability density function, F is the Fresnel fraction and Vis is the microfacet visibility function which gives the probability for a given microfacet to be visible from both directions $\mathbf{k_1}$ and $\mathbf{k_2}$ and accounts for most of the complexity of a given microfacet model. Visibility function is also responsible for ensuring the energy conservation. We will not attempt to find a direct analog of this complicated Vis function in our empirical model and will be simply concerned with providing the means to conserve energy. However, other terms of equation 4 do have direct counterparts in equation 3. For example, it is immediately clear that the appropriate choice for the argument of the Fresnel fraction F is $(\mathbf{k}\mathbf{h})$. Note that throughout the paper we will drop the subscript of vector $\mathbf{k}$ if either $\mathbf{k}_1$ or $\mathbf{k}_2$ can be used. We will also introduce notation

$$p_h(\mathbf{h}) = \frac{\sqrt{(n_u + 1)(n_v + 1)}}{2\pi} (\mathbf{n}\mathbf{h})^{n_u \cos^2 \phi + n_v \sin^2 \phi} \tag{5}$$

3

where the normalization constant is chosen so that $p_{(}\mathbf{h})$ is a true probability density function (integrates to one over the hemisphere of possible $\mathbf{h}$ directions). Energy conservation requirement can be written as

$$R(\mathbf{k_1}) = \int_{\mathbf{k}_2} \rho(\mathbf{k_1}, \mathbf{k_2})(\mathbf{k_2 n}) d\omega_{k_2} \leq 1 \tag{6}$$

for any $\mathbf{k_1}$. Division by $\max((\mathbf{nk}_1), (\mathbf{nk}_2))$ in our model will be cancelled (or replaced by a number less than 1) by $(\mathbf{k_2 n})$ factor and we obtain the condition

$$c' * \int_{\mathbf{k}_2} p_h(\mathbf{h}) F((\mathbf{kh})) d\omega_{k_2} \leq 1 \tag{7}$$

The assumption of mirror reflection from microfacets gives an important relationship between differential solid andles in the space of reflected rays and the $\mathbf{h}$-space of microfacet normal directions [**?**]:

$$d\omega_{k_2} = 4(\mathbf{k_1 h}) d\omega_h. \tag{8}$$

Using this formula and the fact that $F \leq 1$ we obtain

$$c' * \int_{\mathbf{h}} p_h(\mathbf{h}) 4(\mathbf{k_1 h}) d\omega_h \leq 1 \tag{9}$$

The integration is now done over a complex subregion of $\mathbf{h}$-space. However, being conservative, we can extend the integral over the whole hemisphere of directions. This formula shows that if we divide our BRDF by $4(\mathbf{kh})$ and set $c' = 1$ we will guarantee that our model will conserve energy since $p_h(\mathbf{h})$ integrates to one over the hemisphere. Putting all this together, we arrive at the final form of our anisotropic specular BRDF:

$$\rho(\mathbf{k_1}, \mathbf{k_2}) = \frac{\sqrt{(n_u + 1)(n_v + 1)}}{8\pi} \frac{(\mathbf{nh})^{n_u \cos^2 \phi + n_v \sin^2 \phi}}{(\mathbf{hk}) \max((\mathbf{nk}_1), (\mathbf{nk}_2))} F((\mathbf{kh})) \tag{10}$$

In our implementation we use Schlick's approximation to Fresnel fraction [8]:

$$F((\mathbf{kh})) = R_s + (1 - R_s)(1 - (\mathbf{kh}))^5, \tag{11}$$

where R_s is material's reflectance for the normal incidence.

As a visualization of the energy normalization of the model, we rendered a variety of spheres with different parameters shown in Figure 2. The spheres are in a "furnace" with radiance one in all directions. Perfectly reflecting spheres, regardless of BRDF, would also be white. Essentially it is a visualization of the directional hemispherical reflectance (directional albedo) for a variety of input angles.

The specular BRDF 10 described in this section is useful for representing metallic surfaces where the diffuse component of reflection is very small. Figure 3 shows a set of golden spheres on a texture-mapped Lambertian plane. As the values of parameters n_u and n_v change, the appearance of the spheres shift from rough metal to almost perfect mirror, and from highly anisotropic to the more familiar phong-like behavior.

$n_v = 10000$

$n_v = 1000$

$n_v = 100$

$n_v = 10$

$n_u = 10$ $n_u = 100$ $n_u = 1000$ $n_u = 10000$
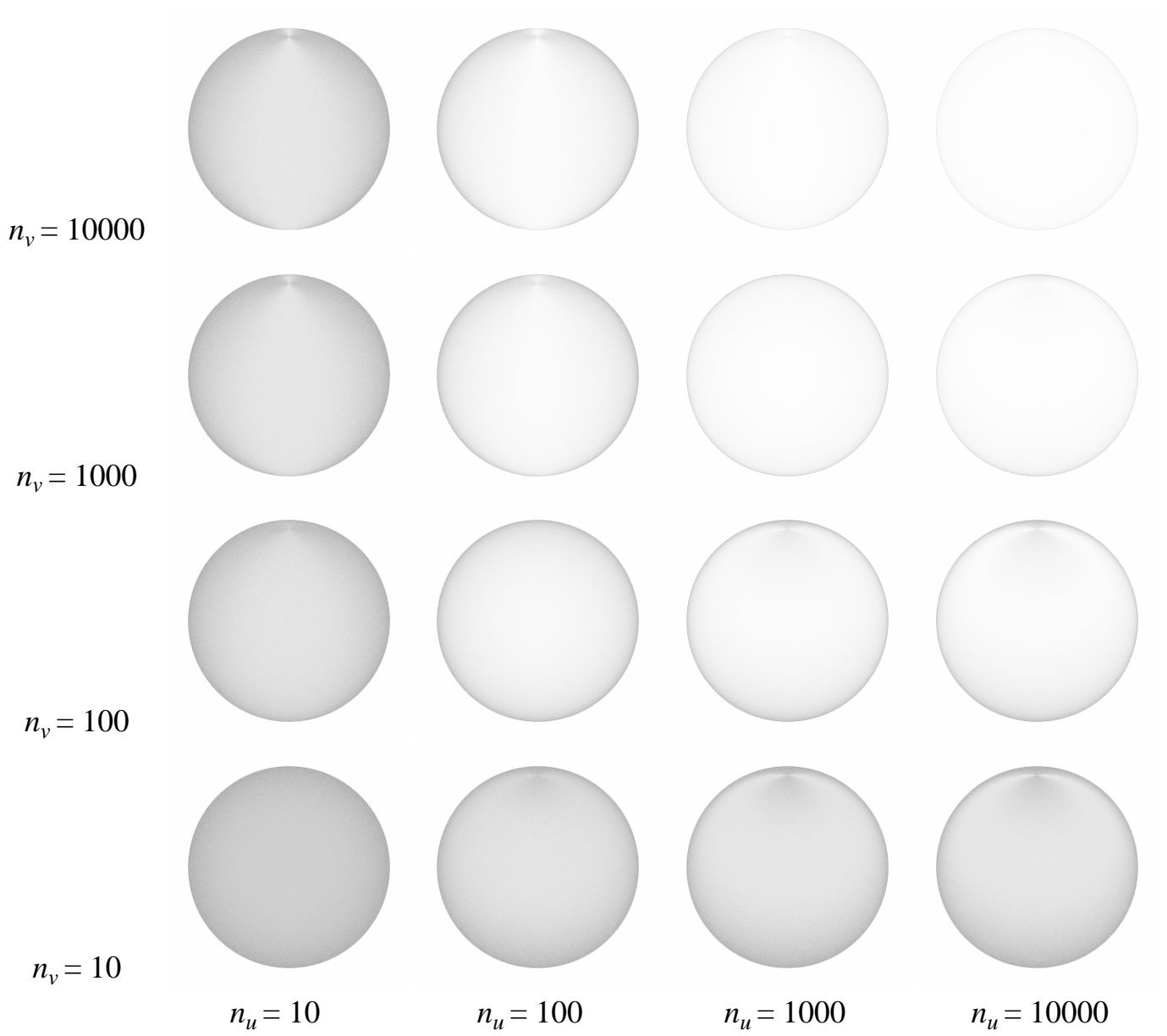
Fig. 2. *Spheres in a furnace. As the exponents get larger, less energy is "lost". For the center of the darkest sphere, $n_u = n_v = 10$, the luminance is about 68% of the background luminance.*
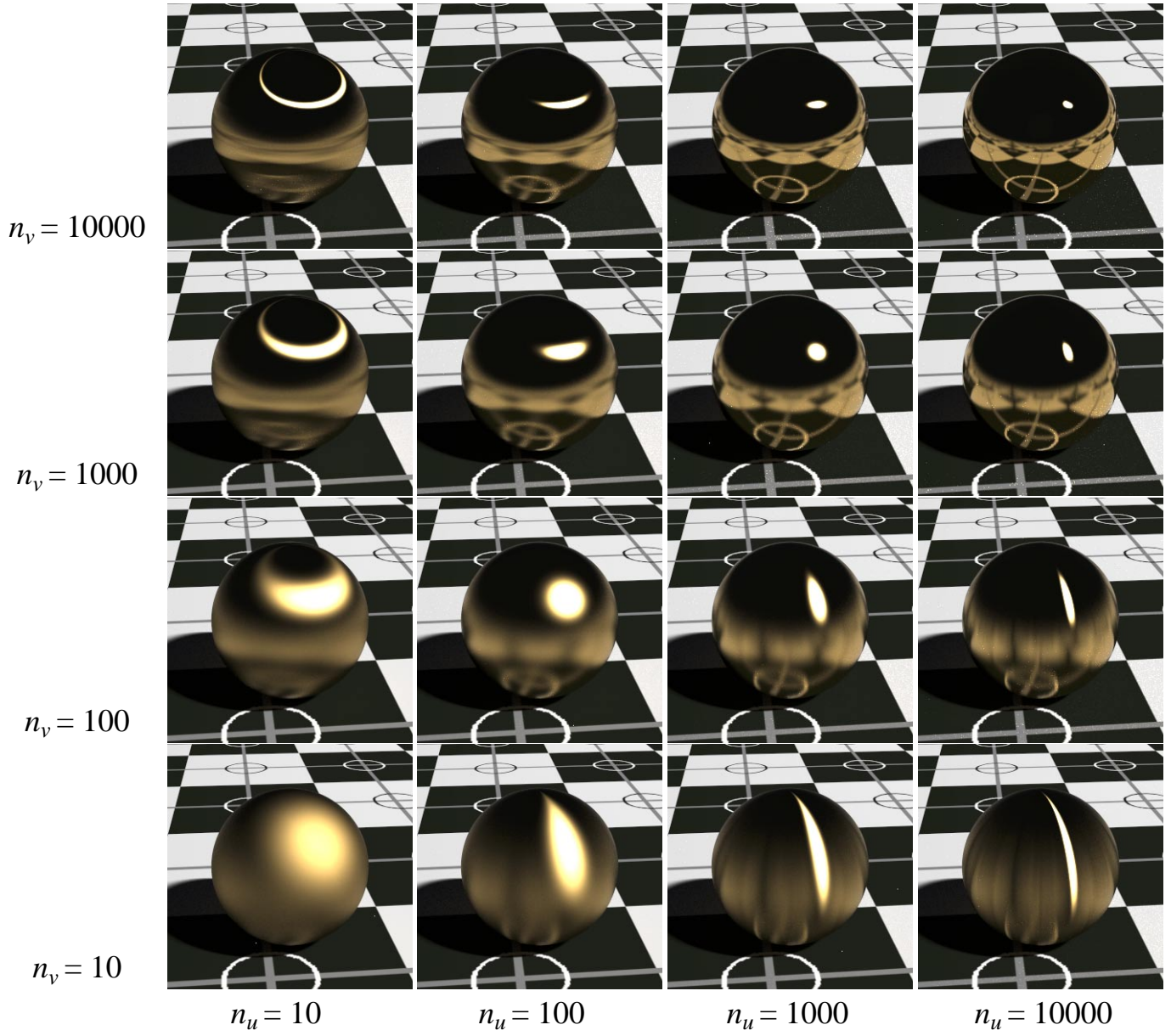
5

$n_v = 10000$

$n_v = 1000$

$n_v = 100$

$n_v = 10$

$n_u = 10$ $n_u = 100$ $n_u = 1000$ $n_u = 10000$

Fig. 3. *Metallic spheres for various exponents.*

6

3 Diffuse term

It is certainly possible to use a Lambertian BRDF together with our specular term in a way this is usually done for most models [8, 10]. However, in this section we will derive a simple angle-dependent form of the diffuse component which takes into account the fact that the amount of energy available for diffuse scattering varies due to the dependence of specular term's total reflectance on the incident angle. In particular, diffuse color of a surface disappears near the grazing angle because the total specular reflectance is close to one in this case. This well-known effect can not be reproduced with a Lambertian diffuse term and is therefore missed by most reflection models. Another, perhaps more important, limitation of the Lambertian diffuse term is that it must be set to zero to ensure energy conservation in the presence of a Fresnal-weighted term.

Shirley et al. [9] proposed a simple form of a non-Lambertian diffuse BRDF which takes this issue into account while preserving overall energy conservation and reciprocity. We use this result in the following form:

$$\rho_d(\mathbf{k}_1, \mathbf{k}_2) = c * R_d(1 - R(\mathbf{k}_1))(1 - R(\mathbf{k}_2)), \tag{12}$$

where $R(\mathbf{k})$ is the total hemispherical reflectance of the specular term as defined by equation 6, $0 < R_d < 1$ is the diffuse albedo of the surface and c is a normalization constant computed such that for $R_d = 1$ the total incident and reflected energies are the same.

For this form to be directly used in our model, we need a closed-form expression for $R(\mathbf{k})$. Unfortunately, specular BRDF 10 does not allow for analytical integration of Equation 6. It is possible, however, to find an approximation to $R(\mathbf{k})$ which will be sufficient for our purposes. To ensure overall energy conservation we will be looking for a simple function $r(\mathbf{k})$ which is bounded by $R(\mathbf{k})$ from below, i.e. $R(\mathbf{k}) \leq r(\mathbf{k})$ for any $\mathbf{k}$. First of all, we will ignore the loss of energy by the specular component due to the specular lobe going below horizon. This effect is hard to account for for an arbitrary n and it becomes negligible for large n, so we will approximate $R(\mathbf{k})$ as 1 in the absence of Fresnel effects ($R_s = 1$ in equation 11). This allows us to write

$$R(\mathbf{k_1}) = \int_{\mathbf{k}_2} f(\mathbf{k_1}, \mathbf{k}_2)(\mathbf{k_2 n})F((\mathbf{kh}))d\omega_{k_2} \leq R_s + (1 - R_s)\int_{\mathbf{k}_2} f(\mathbf{k_1}, \mathbf{k}_2)(\mathbf{k_2 n})(1 - (\mathbf{kh}))^5 d\omega_{k_2},$$

$$\tag{13}$$

where $f(\mathbf{k_1}, \mathbf{k_2})$ is the part of the specular BRDF without the Fresnel fraction and our approximation says that $\int_{\mathbf{k}_2} f(\mathbf{k_1}, \mathbf{k_2})(\mathbf{k_2 n})d\omega_{k_2} = 1$. For a given incident vector $\mathbf{k_1}$ scalar product $(\mathbf{kh})$ is minimal if $\mathbf{h}$ lies in the plane of incidence and bisects the angle between $\mathbf{k_1}$ and a vector in $\mathbf{uv}$ coordinate plane farthest from $\mathbf{k_1}$. In this case $(\mathbf{kh})_{min} = \sqrt{\frac{1-\sqrt{1-(\mathbf{k_1 n})^2}}{2}}$ and we can choose $r(\mathbf{k}) = R_s + (1 - R_s)(1 - (\mathbf{kh})_{min})^5$. We will further simplify this expression by replacing $(\mathbf{kh})_{min}$ with approximation $(\mathbf{kh})_{min} \geq (\mathbf{kn})/2$. Our approximate hemispherical reflectance becomes $r(\mathbf{k}) = R_s + (1 - R_s)(1 - (\mathbf{kn})/2)^5$. We can now substitute this as $R(\mathbf{k})$ into equation 12 and perform integration to obtain the normalization constant c. The diffuse term becomes

$$\rho_d(\mathbf{k}_1, \mathbf{k}_2) = \frac{28R_d}{23\pi}(1 - R_s)\left(1 - \left(1 - \frac{(\mathbf{nk}_1)}{2}\right)^5\right)\left(1 - \left(1 - \frac{(\mathbf{nk}_2)}{2}\right)^5\right) \tag{14}$$

Note that our diffuse BRDF does not depend on n_u and n_v, so we can judge the quality of approximations we made in its derivation by looking at a single image on figure 2

Fig. 4. Half of a diffusely illuminated sphere with $R_s = 0.05$ and $R_d = 1$.

created in a setting identical to the "furnace" of Figure 4. For large n there is little loss of energy by the specular term, so any darkening of the sphere is due to the diffuse component.

A set of polished red spheres with different phong exponents n_u, n_v is shown in Figure 6. For all spheres R_s is set to 0.05 across the visible spectrum which is a typical value for plastics. In addition to anisotropic highlights and blurred reflections we can observe strengthening of the specular reflection near the silhouette of the sphere along with simultaneous decrease in the intensity of the red color. This effect is more prominent in Figure 5 where three different views of the same scene are shown.

4 Implementing the model

Recall the BRDF is a combination of diffuse and specular components:

$$\rho(\mathbf{k}_1, \mathbf{k}_2) = \rho_s(\mathbf{k}_1, \mathbf{k}_2) + \rho_d(\mathbf{k}_1, \mathbf{k}_2), \tag{15}$$

The diffuse component is given in Equation 14. The specular component is given in Equation 10. It is not necessary to call trigonometric functions to compute the exponent, so the specular BRDF is:

$$\rho(\mathbf{k}_1, \mathbf{k}_2) = \frac{\sqrt{(n_u + 1)(n_v + 1)}}{8\pi} \frac{(\mathbf{nh})^{\frac{(n_u(\mathbf{hu})^2 + n_v(\mathbf{hv})^2)}{(1-(\mathbf{hn})^2)}}}{(\mathbf{hk})\max((\mathbf{nk}_1), (\mathbf{nk}_2))} F((\mathbf{kh})) \tag{16}$$

In a Monte-Carlo setting we are also interested in the following problem: given $\mathbf{k}_1$, generate samples of $\mathbf{k}_2$ with a distribution which shape is similar to the cosine weighted BRDF. The key part of our thinking on this is inspired by discussion by Zimmerman [**?**] and by Lafortune [**?**] who point out that greatly undersampling a large value of the integrand is a serious error while greatly oversampling a small value is acceptable in practice. The reader can verify that the densities suggested below have this property.

We can just use the probability density function $p_h(\mathbf{h})$ of Equation 5 to generate a random $\mathbf{h}$. However, to evaluate the rendering equation we need both a reflected vector $\mathbf{k}_2$ and a probability density function $p(\mathbf{k}_2)$. It is important to note that if you generate $\mathbf{h}$ according to $p_h(\mathbf{h})$ and then transform to the resulting $\mathbf{k}_2$:

$$\mathbf{k}_2 = -\mathbf{k}_1 + 2(\mathbf{k}_1 \mathbf{h})\mathbf{h}, \tag{17}$$
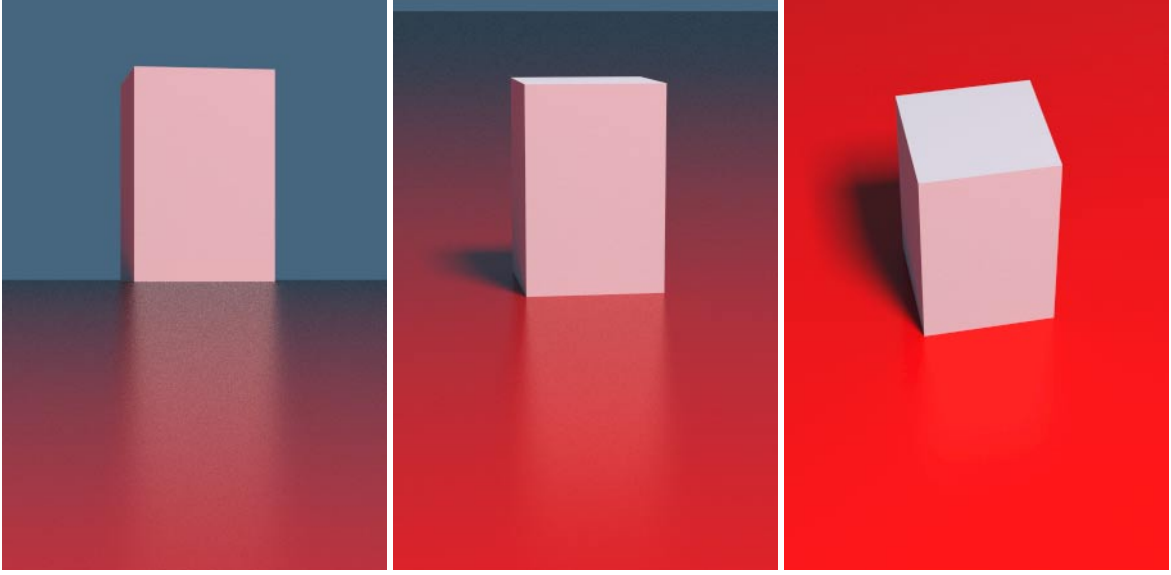
Fig. 5. Three views for $n_u = n_v = 400$ and a red substrate.

the density of the resulting $\mathbf{k}_2$ is **not** $p_h(\mathbf{k}_2)$. This is because of the difference in measures in $\mathbf{h}$ and $\mathbf{v}_2$ space described in Equation 8. So the actual density $p(\mathbf{k}_2)$ is:

$$p(\mathbf{k}_2) = \frac{4(\mathbf{k}_1\mathbf{h})}{p_h(\mathbf{h})} \tag{18}$$

Note that it is possible to generate an $\mathbf{h}$ vector whose corresponding vector $\mathbf{k}_2$ will point inside the surface, i.e. $\mathbf{k_2 n} < 0$. The weight of such a sample should be set to zero. This situation corresponds to the specular lobe going below the horizon and is the main source of energy loss in the model. Clearly, this problem becomes progressively less severe as n_u, n_v become larger.

The only thing left now is to describe how to generate $\mathbf{h}$ vectors with pdf of Equation 5. We will start by generating $\mathbf{h}$ with its spherical angles in the range $(\theta, \phi) \in [0, \frac{\pi}{2}] \times [0, \frac{\pi}{2}]$. Note that this is only the first quadrant of the hemisphere. Given two random numbers (ξ_1, ξ_2) uniformly distributed in $[0, 1]$, we can choose

$$\phi = \arctan\left(\sqrt{\frac{n_u + 1}{n_v + 1}} \tan\left(\frac{\pi\xi_1}{2}\right)\right) \tag{19}$$

and then use this value of ϕ to obtain θ according to

$$\cos\theta = (1 - \xi_2)^{\frac{1}{n_u \cos^2\phi + n_v \sin^2\phi + 1}} \tag{20}$$

To sample the entire hemisphere, the standard manipulation where ξ_1 is mapped to one of four possible functions depending one whether it is in $[0, 0.25)$, $[0.25, 0.5)$, $[0.5, 0.75)$, or $[0.75, 1.0)$. For example for $\xi_1 \in [0.25, 0.5)$, find $\phi(1 - 4(0.5 - \xi_1))$ via Equation 19, and then "flip" it about the $\phi = \pi/2$ axis. This ensures full coverage and stratification.

$n_v = 10000$

$n_v = 1000$

$n_v = 100$

$n_v = 10$

$n_u = 10$ $n_u = 100$ $n_u = 1000$ $n_u = 10000$
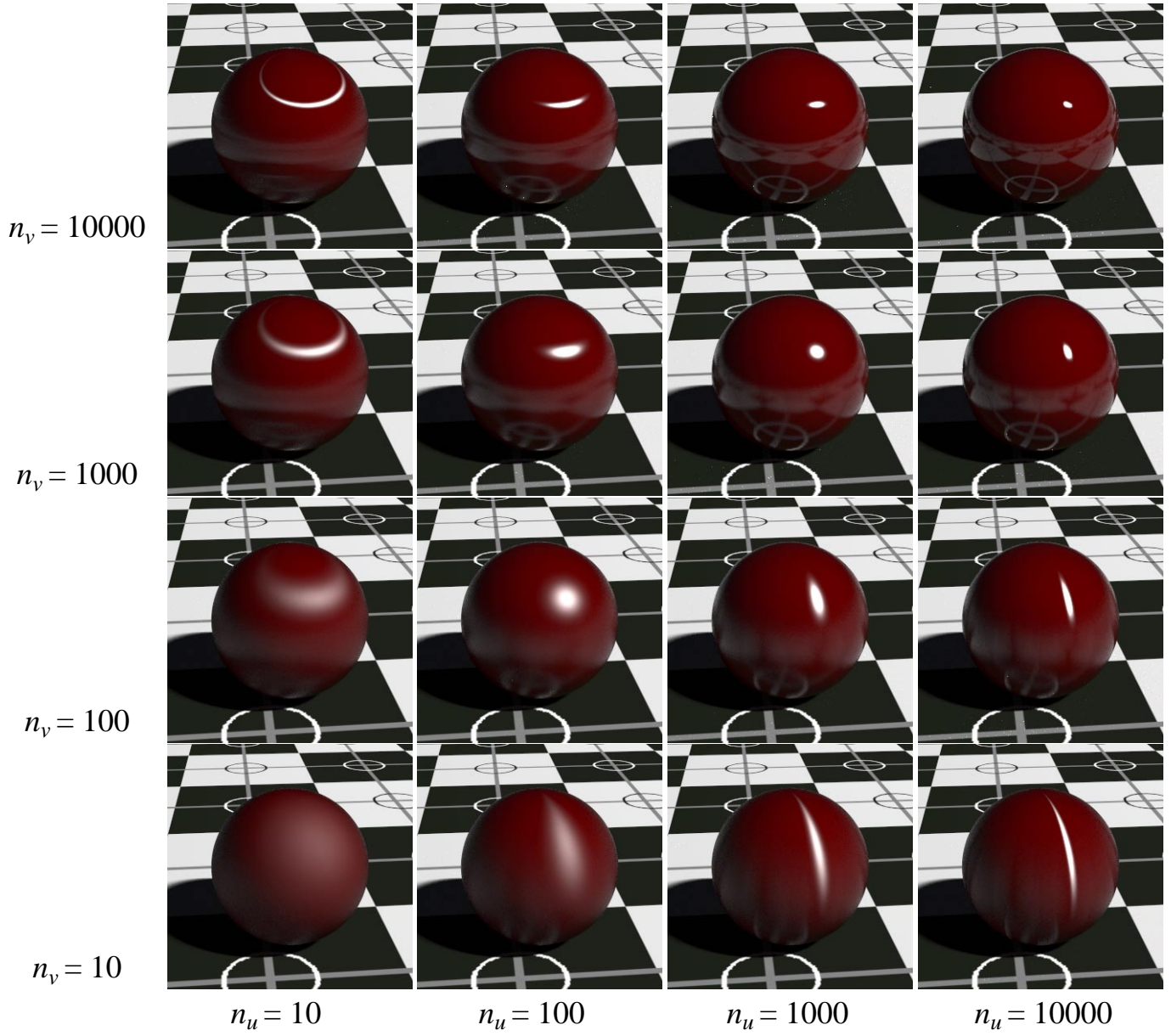
Fig. 6. *Diffuse spheres for various exponents.*

10

Fig. 7. *A closeup of the model implemented in a path tracer with 9, 26, and 100 samples.*

Fig. 8. *Fancy image.*

It would be possible to do importance sample with a density close to cosine-weighted BRDF 14 in a way similar to that described by Shirley et al [9], but we use a simpler approach and generate samples according to cosine distribution. This is sufficiently close to the complete diffuse BRDF to substantially reduce variance of the Monte-Carlo estimation.

References

1. James F. Blinn. Models of light reflection for computer synthesized pictures. *Computer Graphics (Proceedings of SIGGRAPH 77)*, 11(2):192–198, July 1977.
2. Robert L. Cook and Kennneth E. Torrance. A reflectance model for computer graphics. *Computer Graphics*, 15(3):307–316, August 1981. ACM Siggraph '81 Conference Proceedings.
3. Donald P. Greenberg, Kenneth E. Torrance, Peter Shirley, James Arvo, James A. Ferwerda, Sumanta Pattanaik, Eric P. F. Lafortune, Bruce Walter, Sing-Choong Foo, and Ben Trumbore. A framework for realistic image synthesis. *Proceedings of SIGGRAPH 97*, pages 477–494, August 1997.
4. Eric P. Lafortune and Yves D. Willems. Using the modified phong BRDF for physically based rendering. Technical Report CW197, Computer Science Department, K.U.Leuven, November 1994.
5. Robert Lewis. Making shaders more physically plausible. In Michael F. Cohen, Claude Puech, and Francois Sillion, editors, *Fourth Eurographics Workshop on Rendering*, pages 47–62. Eurographics, June 1993. held in Paris, France, 14–16 June 1993.
6. László Neumann, Attila Neumann, and László Szirmay-Kalos. Compact metallic reflectance models. *Computer Graphics Forum*, 18(13), 1999.
7. Bui-Tuong Phong. Illumination for computer generated images. *Communications of the ACM*, 18(6):311–317, June 1975.
8. Christophe Schlick. An inexpensive BRDF model for physically-based rendering. *Computer Graphics Forum*, 13(3):233—246, 1994.
9. Peter Shirley, Helen Hu, Brian Smits, and Eric Lafortune. A practitioners' assessment of light reflection models. In *Pacific Graphics*, pages 40–49, October 1997.
10. Gregory J. Ward. Measuring and modeling anisotropic reflection. *Computer Graphics*, 26(4):265–272, July 1992. ACM Siggraph '92 Conference Proceedings.

An Illumination Model for a Skin Layer Bounded by Rough Surfaces

Jos Stam

Alias $|$ wavefront

1218 Third Ave, 8th Floor,

Seattle, WA 98101

Abstract

In this paper we present a novel illumination model that takes into account multiple anisotropic scattering in a layer bounded by two rough surfaces. We compute the model by a discrete-ordinate solution of the equation of radiative transfer. This approach is orders of magnitude faster than a Monte Carlo simulation and does not suffer from any noisy artifacts. By fitting low order splines to our results we are able to build analytical shaders. This is highly desirable since animators typically want to texture map the parameters of such a shader for higher realism. We apply our model to the important problem of rendering human skin. Our model does not seem to have appeared before in the optics literature. Most previous models did not handle rough surfaces at the skin's boundary. Also we introduce a novel analytical bidirectional transmittance distribution function (BTDF) for an isotropic rough surface by generalizing the Cook-Torrance model. We believe our work to be both of practical and theoretical importance.

Keywords:
Illumination. Illumination Effects. Reflectance & Shading Models.

1 Introduction

The work described in this paper was motivated by the desire to model the appearance of human skin under various lighting conditions. A good model for the reflection of light from skin has many obvious applications in the entertainment industry, where there is a keen interest in making virtual actors appear more life-like. However, despite the importance of this problem there are very few analytical models that convincingly model the appearance of skin. This is probably because the interaction of light with human skin is a very complicated physical phenomenon. Skin appearance depends not only on the skin's surface but also on the layer directly below it: incoming light is not only reflected specularly (as in oily shiny skin) but is also scattered diffusely within the flesh. This explains why traditional Phong-like surface-based reflection models fail to capture the subtle appearance of skin. Most traditional models approximate the contributions due to the subsurface layer inadequately using a Lambertian cosine term. A good model for subsurface scattering is also important to model substances other than human skin such as paints and tissues. An effective skin reflection model should ideally depend analytically on a set of meaningful parameters, such as the

skin's surface roughness and amount of melanin. Animators typically want to texture map these parameters to add visual detail, such as freckles, pores or a shiny forehead.

In this paper we propose a new model for subsurface reflections based on linear transport theory. This theory has matured well and has been applied to a wide range of disciplines, including nuclear physics, the atmospheric sciences, astrophysics and computer graphics. It seems that the only transport theoretical model for subsurface scattering in computer graphics is the one proposed by Hanrahan and Krueger [5]. They introduce an analytical model for layers that only scatter weakly using the single-scattering approximation. In addition, they assume a perfectly smooth reflecting and refracting surface at the top of the layers. These approximations are, however, ill suited for skin, since the skin's surface is rarely perfectly smooth and multiple scattering is very important. To address the latter problem, Hanrahan and Krueger ran a Monte Carlo simulation to precompute reflection maps for different configurations of sub-layers. Unfortunately, the Monte Carlo method converges very slowly, so handling a wide range of interesting skin parameters requires huge amounts of data to be computed. Because the model of Hanrahan and Krueger did not satisfy our needs, we decided to derive our own model in a more general setting.

We first turned to the abundant literature on the subject from other fields. Most relevant to our problem is the literature in the medical sciences studying the optics of skin for such applications as non-invasive surgery. We first consulted the review articles [23] and [24]. The latter article discusses an approximation which is too coarse for our purposes. The first reference from a Russian journal only mentions Monte Carlo methods and the so-called adding-doubling method. We have already drawn attention to the drawbacks of the Monte Carlo approach. The adding-doubling technique was used by Prahl and co-workers to extract parameters from skin measurement [18] and is based on earlier work in astronomy [6]. This technique, however, is iterative in nature and does not handle rough surface boundaries. At this point we turned to the atmospheric sciences. The problem addressed there is how to compute the global interchange of radiation between the atmosphere and the ocean. The ocean is very much like skin since it has a rough boundary and light is scattered below it. We found a very attractive model based on the discrete-ordinate approximation of radiative transfer [3]. In particular, Stamnes and co-workers developed a general solution framework for the atmosphere-ocean radiative problem [8, 21]. Unfortunately their model does not handle rough surfaces. It seems that the "state of the art" model in this area resorts to a Monte Carlo simulation to determine the effect of a rough surface [14].

After this review, we decided to extend the discrete-ordinate approach of Jin and Stamnes [8] to include rough surfaces. We first required a good analytical model for the Bidirectional Transmittance Distribution Function (BTDF). We are aware of only one such model in computer graphics based on the wave theory of light [7]. Unfortunately this model is fairly complex, so we derive a simpler one in this paper for the first time. Our model is an extension of the BRDFs of Cook-Torrance [4] and of van Ginneken et al. [25]. Extending a reflectance model to include transmittance may seem straightforward at first, but we encountered some subtle issues. The first contribution of this paper is to resolve these issues. The second contribution is a general discrete-ordinate solution for a scattering layer bounded by rough surfaces. Our model is, therefore, of interest to computer graphics and potentially to other fields. We show how to efficiently solve the problem by a suitable "diagonalization" of the "transfer matrix." We use the machinery of Fourier transforms and eigenanalysis to perform this task. To build practical reflection models for computer graphics, we fit low-order splines to our discretized functions. Our approach is orders of magnitude faster than Monte Carlo methods, requires less memory, and does not suffer from any noisy artifacts. The parameters of our model can also be texture mapped without the need for
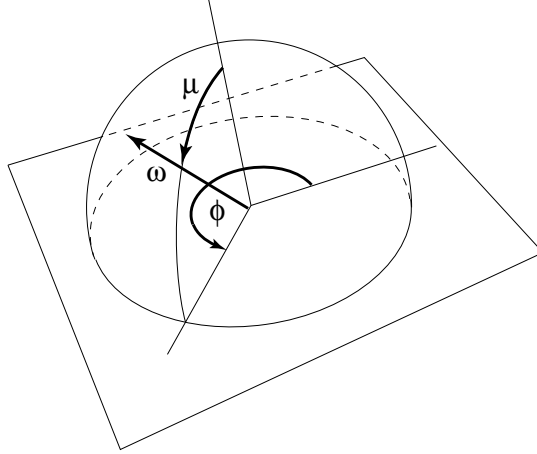
Figure 1: Definition of a direction $\omega = (\mu, \phi)$.

any recomputations. It is therefore ideally suited for an implementation as a shader in a standard rendering package.

We briefly mention here that the discrete-ordinate method has been used before in a different context in computer graphics. Both Max and Languenou et al. used this technique to compute the scattering in non-constant densities such as clouds [11, 13]. Their methods, however, do not lead to analytical reflection models since they consider arbitrary densities. They also do not address the problem of rough surfaces refracting and reflecting light at the boundaries. The same comments apply to the method of Kajiya and Von Herzen [9] and the radiosity-based approach of Rushmeier et al. [20].

The rest of the paper is organized as follows. The next section details the physics involved and introduces the equivalent discrete problem. In Section 3 we show how to solve the discrete problem efficiently. Section 4 presents a derivation of our new BTDF. Section 5 clarifies many implementation issues and discusses the corresponding "skin shader". In Section 6 we present some results and compare them to experimental data, while in Section 7 we conclude and discuss future research. Material of a rather technical nature is addressed in the appendices.

1.1 Notational Preliminaries

Much of the material in this paper can be presented more elegantly using a "matrix operator approach" [17]. Many relations are expressed more compactly without indices in vector and matrix form. In this paper all vectors are denoted by bold lowercase characters: $\mathbf{v}$. The elements of $\mathbf{v}$ are denoted by the corresponding italicized letter: v_i is the i-th component of $\mathbf{v}$. An element of a vector should not be confused with an indexed vector such as $\mathbf{v}_k$. A matrix is denoted by a bold upper case character such as $\mathbf{M}$ and its elements are denoted by $M_{i,j}$. The transpose of $\mathbf{M}$ is written $\mathbf{M}^T$.

2 Discretizing the Physics

2.1 Physical parameters

The physical quantity corresponding to a visual stimulus is the radiance u. This quantity has units of power per unit area per unit solid angle and gives the amount of radiant power flowing from a particular position in a particular direction. Following Hanrahan and Krueger [5] we assume that the skin depth is along the z-direction and that the skin's properties are uniform in each xy-plane. In this setting it is more convenient to use the dimensionless optical depth τ rather than the depth z, where

$$\tau = z/L$$

and L is the mean free path of a photon in the medium. Consequently, the radiance is a function of optical depth and direction. We represent a direction by an ordered pair $\pm\omega = (\pm\mu, \phi)$ where $\mu = \cos\theta$ is the cosine of the elevation angle θ and ϕ is the azimuthal angle (see Figure 1). In the following we always assume that the cosine $\mu \geq 0$. We therefore denote a downward direction by $-\omega = (-\mu, \phi)$. The use of the minus sign is purely notational in this context.

The optical properties of the skin are modeled by two parameters that describe how light scatters at each point. They are the albedo Ω and the anisotropy factor g. The albedo gives the fraction of light that is scattered versus absorbed and is typically close to one for skin. The distribution of scattered light is defined by a phase function p. This function gives the probability that a photon travelling in a direction ω' is scattered in another direction ω. We rely on the *Henyey-Greenstein phase function*, a useful model frequently seen in the optics literature:

$$p(\omega', \omega) = \frac{1 - g^2}{\left(1 + g^2 - 2g\cos\gamma\right)^{3/2}}.$$

Here γ is the angle between the directions ω' and ω. The anisotropy factor $g \in [0, 1]$ models how much light is scattered forward. For $g = 0$ the medium scatters isotropically, while for $g = 1$ the scattering is in the forward direction only. Scattering in skin is typically highly anisotropic with values of g in the range $[0.7, 0.95]$. Both the albedo and the anisotropy factor of skin depend on depth and wavelength. It is interesting that the popularity of the Henyey-Greenstein phase function stems from the fact that it has a very simple expansion in terms of the associated Legendre functions (see Appendix C). This is rarely mentioned in the computer graphics literature. The expansion coefficients are the powers of g. The higher the anisotropy, the more terms are required in the expansion. We associate with the phase function a linear scattering functional:

$$\mathcal{S}\{u\}(\tau, \omega) = \frac{\Omega}{4\pi} \int_{4\pi} p(\omega', \omega) u(\tau, \omega') \, d\omega', \tag{1}$$

where the integration is over all possible directions.

The reflection and refraction at the skin's boundaries are modeled as isotropic rough surfaces. In our model we assume that the skin has a uniform index of refraction n_2 and is bounded above and below by media having indices equal to n_1 and n_3, respectively. To model the reflection, we use a variant of the Cook-Torrance model [4] proposed by van Ginneken et al. [25]. For the transmission we derive a new model in Section 4 since we could not find a satisfactory one in the optics literature. Both the transmission and reflection models depend on a roughness parameter σ. The BRDF and BTDF are denoted by $r_{ij}(\omega, \omega')$ and $t_{ij}(\omega, \omega')$, respectively, for light coming from
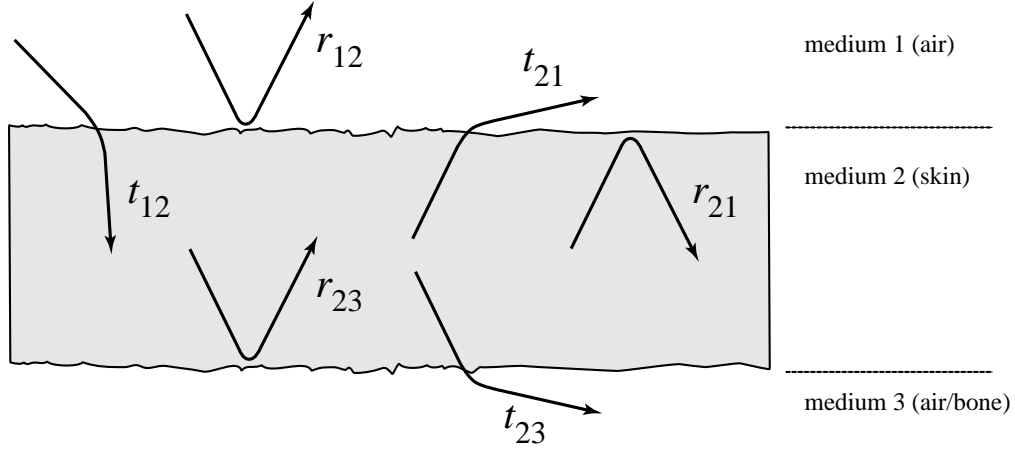
Figure 2: Nomenclature for the BRDFs and BTDFs.

material $i \in 1, 2, 3$ arriving at material $j \in 1, 2, 3$, where, for example, r_{12} models the reflection off of the top surface. This nomenclature is clarified in Figure 2. We associate with these distributions linear reflection and transmission operators:

$$\mathcal{R}_{ij}\{u\}(\tau, \pm\omega) = \int_{2\pi} r_{ij}(\mp\omega', \pm\omega)u(\tau, \mp\omega')\mu' \, d\omega', \quad (2)$$

$$\mathcal{T}_{ij}\{u\}(\tau, \pm\omega) = \int_{2\pi} t_{ij}(\pm\omega', \pm\omega)u(\tau, \pm\omega')\mu' \, d\omega', \quad (3)$$

where the integration is over the positive hemisphere and the signs depend on the BRDF or BTDF considered, e.g., r_{12} has opposite signs from r_{21} as is evident from Figure 2.

2.2 Equation of Transfer

An equation for the radiance within the skin is obtained by considering its variation in an infinitesimal cylinder aligned with the direction ω. The change is equal to the amount of light scattered into this direction minus the light absorbed and scattered out of this direction:

$$-\mu\frac{du}{d\tau} = -u + \mathcal{S}\{u\}, \quad (4)$$

where $\mathcal{S}$ is the scattering operator defined in Equation 1. To completely specify the problem, this equation requires boundary conditions at the top and the bottom of the skin layer. At the skin's surface ($\tau = 0$) the downwelling radiance is equal to the transmitted radiance plus the internal reflections of the radiance coming from the internal layer:

$$u(0, -\omega) = t_{12}(-\omega_0, -\omega) + \mathcal{R}_{21}\{u\}(0, -\omega). \quad (5)$$

Similarly, if we assume there are no sources below the skin, the upwelling radiance at the bottom of the layer ($\tau = \tau_b$) is given by

$$u(\tau_b, \omega) = \mathcal{R}_{23}\{u\}(\tau_b, \omega). \quad (6)$$

Once Equation 4 is solved using these boundary conditions, the BRDF and the BTDF due to scattering in the skin's layer are equal to

$$r_s(-\omega_0, \omega) = \mathcal{T}_{21}\{u\}(0, \omega)/\mu_0 \quad \text{and} \quad t_s(-\omega_0, -\omega) = \mathcal{T}_{23}\{u\}(\tau_b, -\omega)/\mu_0,$$

5

respectively. In addition, the reflection due to an ambient light source of radiance is modeled by integrating the skin's BRDF over all incident directions $-\omega_0$:

$$r_a(\omega) = \int_{2\pi} r_s(-\omega_0, \omega)\, \mu_0\, d\mu_0.$$

The total amount of light reflected off the skin is the sum of the part directly reflected by the surface, the ambient term and the radiance leaving the subsurface layer:

$$r_{tot}(\omega) = r_{12}(-\omega_0, \omega) + r_a(\omega) + r_s(-\omega_0, \omega).$$

In Section 4 we provide a model for r_{12} (and the other r_{ij} and t_{ij}) while the next section describes a method of solution for r_s and t_s.

2.3 Angular Discretization

We discretize the angular part of Equation 4 in two steps. Because we assume that the surface roughness is isotropic and that the skin is horizontally uniform, we can decompose the azimuthal dependence of the radiance into a cosine series:

$$u(\tau, \omega) = \sum_{k=0}^{N} u_k(\tau, \mu) \cos k(\phi - \phi_0). \tag{7}$$

Next we discretize the cosines μ into $2M$ discrete samples (see Appendix A for how they are chosen):

$$\mu_1, \mu_2, \cdots, \mu_M, -\mu_1, -\mu_2, \cdots, -\mu_M. \tag{8}$$

These values are also known as "ordinates," hence the name "discrete-ordinates" to refer to this type of discretization. The corresponding values of the discretized radiances are stored in a $2M$ vector

$$\mathbf{u}_k(\tau) = (u_k(\tau, \mu_1), \cdots, u_k(\tau, -\mu_M))^T \qquad k = 0, \cdots, N.$$

As shown in Appendix A, the scattering operator in Equation 4 is discretized into a collection of $N + 1$ matrices $\mathbf{S}_k$ ($k = 0, \cdots, N$), each of size $2M \times 2M$. These discretizations convert the transfer equation into $N + 1$ decoupled linear ordinary differential vector equations:

$$-\mathbf{W}\frac{d\mathbf{u}_k(\tau)}{d\tau} = -\mathbf{u}_k(\tau) + \mathbf{S}_k\mathbf{u}_k(\tau),$$

where $\mathbf{W}$ is a diagonal matrix containing the samples of Equation 8. The last equation can be written more compactly as

$$\frac{d\mathbf{u}_k(\tau)}{d\tau} = \mathbf{M}_k\mathbf{u}_k(\tau), \tag{9}$$

where $\mathbf{M}_k = \mathbf{W}^{-1}(\mathbf{I} - \mathbf{S}_k)$ and $\mathbf{I}$ is the identity matrix. Equation 9 is the main equation of this paper. In the next section we show how to solve it efficiently.

3 Direct Solution of the Discrete Problem

This section is inspired by the work of Jin and Stamnes [8]. However, our compact vector/matrix notation greatly simplifies the presentation. Our approach is also more general, since we consider surfaces of arbitrary roughness at the boundaries.

3.1 Diagonalization

We assume that the skin is composed of a layer with constant optical properties sandwiched between two isotropic rough surfaces. In order to simplify the notation in this section, we will drop the dependence of all quantities on the index "k". This is justified because the equations for different terms in the cosine expansion are entirely decoupled. In the skin the radiance satisfies the following equation:

$$\frac{d\mathbf{u}(\tau)}{d\tau} = \mathbf{M}\mathbf{u}(\tau).$$

Ignoring the boundary conditions for the moment, we see that this is a homogeneous vector ordinary differential equation. Such an equation is solved efficiently by putting the matrix $\mathbf{M}$ into diagonal form. Indeed, in diagonal form the equations are decoupled and can be solved analytically. Diagonalizing $\mathbf{M}$ is equivalent to computing its eigenvalues and eigenvectors:

$$\mathbf{M} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}.$$

Here $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\mathbf{M}$:

$$\mathbf{\Lambda} = \mathrm{diag}\left(\lambda_1, \cdots, \lambda_M, -\lambda_1, \cdots, -\lambda_M\right)$$

where $\lambda_i > 0$ for $i = 1, \cdots, M$ (see Appendix E) and $\mathbf{V}$ contains the eigenvectors stored column-wise. If we let $\mathbf{w}(\tau)$ be the transformed radiance $\mathbf{w} = \mathbf{V}^{-1}\mathbf{u}$, then

$$\frac{d\mathbf{w}(\tau)}{d\tau} = \mathbf{\Lambda}\mathbf{w}(\tau).$$

The exact solution to this differential equation is given by:

$$\mathbf{w}(\tau) = e^{\mathbf{\Lambda}\tau}\mathbf{u}_0, \tag{10}$$

where the exponential is simply the diagonal matrix whose elements are the exponential of the elements of $\mathbf{\Lambda}\tau$. The vector $\mathbf{u}_0$ in Equation 10 is to be determined from the boundary conditions. The radiance in the layer is then obtained by inverting our earlier transformation:

$$\mathbf{u}(\tau) = \mathbf{V}\mathbf{w}(\tau) = \mathbf{V}e^{\mathbf{\Lambda}\tau}\mathbf{u}_0. \tag{11}$$

Our next step is to find a vector $\mathbf{u}_0$ satisfying the boundary conditions.

3.2 Solving the Discrete Problem: Boundary conditions

We have just shown that the radiance in each layer can be solved for directly in terms of the eigenvectors and eigenvalues of the transfer matrix. We can rewrite Equation 11 separating the parts corresponding to upward and downward directions:

$$\begin{pmatrix} \mathbf{u}^+(\tau) \\ \mathbf{u}^-(\tau) \end{pmatrix} = \begin{pmatrix} \mathbf{V}^+ & \mathbf{V}^- \\ \mathbf{V}^- & \mathbf{V}^+ \end{pmatrix} \begin{pmatrix} \mathbf{E}(\tau)\mathbf{u}_0^+ \\ \mathbf{E}(-\tau)\mathbf{u}_0^- \end{pmatrix}, \tag{12}$$

where each of the $\mathbf{E}$ matrices contains half of the exponentials:

$$\mathbf{E}(t) = e^{\mathbf{\Lambda}^+ t}.$$

The goal in this section is to compute the unknown vectors $\mathbf{u}_0^+$ and $\mathbf{u}_0^-$ given by Equations 5 and 6. First, let $\mathbf{R}_{ij}$ and $\mathbf{T}_{ij}$ denote the discrete versions of $\mathcal{R}_{ij}$ and $\mathcal{T}_{ij}$ respectively. Since they are defined only over the positive hemisphere they are of size $M \times M$. The top and bottom boundary conditions in terms of these matrices are

$$\mathbf{u}^-(0) = \mathbf{T}_{12}\mathbf{d}_0 + \mathbf{R}_{21}\mathbf{u}^+(0) \quad \text{and} \tag{13}$$

$$\mathbf{u}^+(\tau_b) = \mathbf{R}_{23}\mathbf{u}^-(\tau_b). \tag{14}$$

The vector $\mathbf{d}_0$ represents the incident radiances, and for a directional light source is zero for each entry except for the entry corresponding to $-\mu_0$ where it is equal to one. By substituting Equations 13 and 14 into Equation 12 and rearranging,

$$\begin{pmatrix} \mathbf{V}^- - \mathbf{R}_{21}\mathbf{V}^+ & \mathbf{V}^+ - \mathbf{R}_{21}\mathbf{V}^- \\ (\mathbf{V}^+ - \mathbf{R}_{23}\mathbf{V}^-)\mathbf{E}(\tau_b) & (\mathbf{V}^- - \mathbf{R}_{23}\mathbf{V}^+)\mathbf{E}(-\tau_b) \end{pmatrix} \begin{pmatrix} \mathbf{u}_0^+ \\ \mathbf{u}_0^- \end{pmatrix} = \begin{pmatrix} \mathbf{T}_{12}\mathbf{d}_0 \\ \mathbf{0} \end{pmatrix}.$$

This system, however, is ill-conditioned because the matrix $\mathbf{E}(\tau_b)$ has entries that grow exponentially with τ_b. Fortunately, we can easily fix this problem by setting $\mathbf{u}_0^+ = \mathbf{E}(-\tau_b)\tilde{\mathbf{u}}_0^+$ and solving for $(\tilde{\mathbf{u}}_0^+, \mathbf{u}_0^-)$ instead [21]. The new system becomes:

$$\begin{pmatrix} (\mathbf{V}^- - \mathbf{R}_{21}\mathbf{V}^+)\mathbf{E}(-\tau_b) & \mathbf{V}^+ - \mathbf{R}_{21}\mathbf{V}^- \\ \mathbf{V}^+ - \mathbf{R}_{23}\mathbf{V}^- & (\mathbf{V}^- - \mathbf{R}_{23}\mathbf{V}^+)\mathbf{E}(-\tau_b) \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{u}}_0^+ \\ \mathbf{u}_0^- \end{pmatrix} = \begin{pmatrix} \mathbf{T}_{12}\mathbf{d}_0 \\ \mathbf{0} \end{pmatrix}. \tag{15}$$

This linear system is well behaved and can be solved using any standard linear solver. Once the solution is obtained, the upward radiance at the top and the downward radiance at the bottom of the layer are given by:

$$\mathbf{u}^+(0) = \mathbf{V}^+\mathbf{E}(-\tau_b)\tilde{\mathbf{u}}_0^+ + \mathbf{V}^-\mathbf{u}_0^- \quad \text{and}$$

$$\mathbf{u}^-(\tau_b) = \mathbf{V}^-\tilde{\mathbf{u}}_0^+ + \mathbf{V}^+\mathbf{E}(-\tau_b)\mathbf{u}_0^-,$$

respectively. These are the radiances just inside the rough surfaces of the skin layer. To compute the radiances exiting the surface, we have to multiply these radiances by the transmission matrices $\mathbf{T}_{21}$ and $\mathbf{T}_{23}$, respectively:

$$\mathbf{u}_r = \mathbf{T}_{21}\mathbf{u}^+(0) \quad \text{and} \quad \mathbf{u}_t = \mathbf{T}_{23}\mathbf{u}^-(\tau_b). \tag{16}$$

3.3 Summary

First we restore the subscript "k" to indicate that the radiances of Equation 16 correspond to a single term in the cosine series. Consequently, the complete description of the radiances is given by the following vectors

$$\mathbf{u}_{r,0}, \cdots, \mathbf{u}_{r,N} \quad \text{and} \quad \mathbf{u}_{t,0}, \cdots, \mathbf{u}_{t,N}.$$

These radiances are for a particular incoming direction $-\omega_0 = (-\mu_0, 0)$. To get a discrete description of the BRDF r_s of the skin layer we sample the incoming directions at the ordinates $\mu_1, \cdots, \mu_M$. The discrete representation of the BRDF is, therefore, a collection of $N+1$ matrices $\mathbf{R}_k$ of size $M \times M$ ($k = 0, \cdots, N$). The i-th column of this matrix consists of the vector $\mathbf{u}_{r,k}$ computed for the incident direction $(-\mu_i, 0)$, $i = 1, \cdots, M$. In a similar fashion we build a set of matrices $\mathbf{T}_k$ for the BTDF t_s of the skin layer ($k = 0, \cdots, N$). A high level description of the algorithm that computes these matrices is given in Figure 3.

ComputeRT:
    For $k = 0, \cdots, N$ do
        Compute the scattering matrix $\mathbf{M}_k$ (Appendix A)
        Compute the reflection and transmission matrices $\mathbf{R}_{ij}$ and $\mathbf{T}_{ij}$ (Appendix D)
        Compute eigenstructure of $\mathbf{M}_k$ (see Appendix E)
        For $i = 1, \cdots, M$
            Solve linear system for incoming direction μ_i (Equation 15)
            Transmit radiances out of the layer (Equation 16)
            Set the i-th columns of $\mathbf{R}_k$ and $\mathbf{T}_k$
        next
    next

Figure 3: Summary of our algorithm.

parameter	num. of samples	samples									
T	10	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
σ	9	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Ω	7	0.1	0.5	0.75	0.8	0.9	0.999	1.0			
g	10	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9

Table 1: Samples used for each of our parameters.

We have precomputed these matrices for different values of the parameters that model the skin layer. These parameters are the transparency $T = e^{-\tau_b}$, the albedo Ω, the anisotropy factor g of the phase function, and the roughness σ of the surfaces bounding the skin layer. Each parameter is dimensionless and takes on values between zero and one. The precomputations were done for all possible combinations of the parameter values listed in Table 1. The ratio of the indices of refraction is kept constant throughout: it is set to 1.4, roughly that of human skin. The number of ordinates M was determined from the discretizations of the BRDF and the BTDF of the skin's surface (derived in the next section). For a roughness $\sigma = 0.1$ we nedeeded $M = 30$ ordinates while for other values $M = 24$ was sufficient. The number of cosine series is always set to twice the number of ordinates: $N = 2M$ [8].

Because the scattering matrix depends only on Ω and g, we first computed the eigenstructures for all $7 \times 10 = 70$ possible values of the parameters. We used the RG routine from EISPACK [16] to compute the eigenvectors and eigenvalues. We encountered no numerical problems except when the albedo was exactly one. An easy fix is simply to set the albedo to a value almost equal to one, i.e., $\Omega = 0.999999$. Once the eigenstructures were available we used them to precompute the reflection r_s and transmission t_s of the skin layer for all possible combinations of the samples listed in Table 1. We used the routine DGESL from LINPACK to solve the linear system of Equation 15.

The precomputation generates a huge data set. Our next task was to compress the data using well chosen approximations. We first experimented with the elegant non-linear representation of Lafortune et al. [10]. We did get some good matches using three cosine lobes. However, in many cases the non-linear least square solver got stuck in local minima. For these reasons we adopted a less efficient but more straightforward compression scheme. First, not all cosine terms need to be included. For the reflection at the top of the layer we found that in general, 5
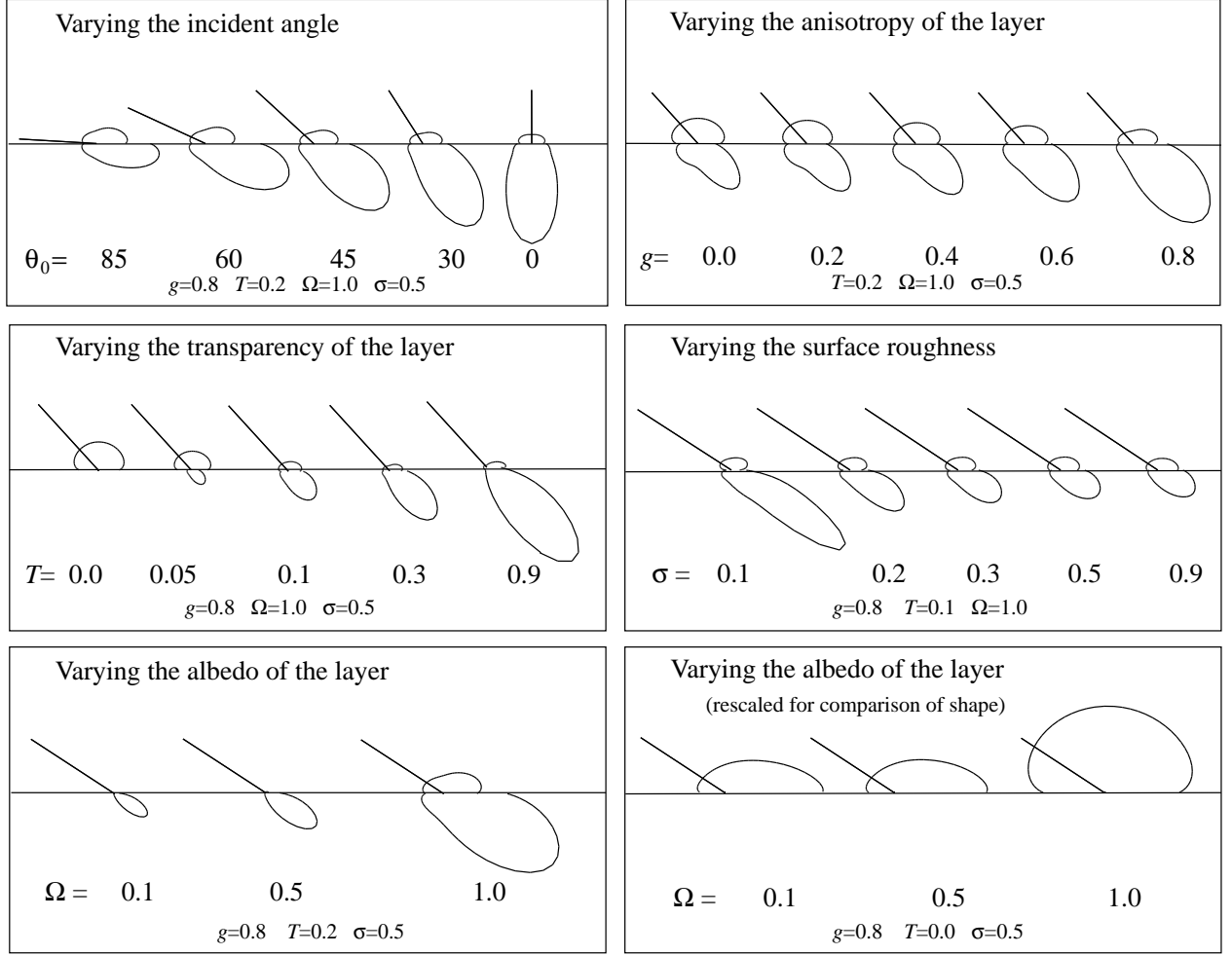
9

Figure 4: Cross-sections of the reflection and transmission functions of the skin layer for different values of the parameters.

terms ($N = 4$) were sufficient, while for the transmission at the bottom 15 ($N = 14$) terms were required. These numbers were obtained by visually comparing the data to the approximation. We further compressed the data by fitting a cubic Bezier surface to the data stored in the reflection (resp. transmission) matrix $\mathbf{R}_k$ (resp. $\mathbf{T}_k$). We constrained the control vertices to respect the symmetry of these matrices (Helmholtz reciprocity).

In Figure 4 we demonstrate the effect of our parameters on the reflection and transmission functions. The simple shapes of the lobes first led us to believe that they might be modeled by simple analytical expressions. The variation with each parameter is, however, quite subtle and none of our analytical estimations could handle all variations at the same time. Analytical solutions are rare when multiple scattering is included. Even the simplest case of a semi-infinite constant medium with isotropic scattering does not admit an analytical solution [3]. The distributions are clearly different from a simple constant Lambert term. The main difference is that the reflected lobe is flatter and has a bias towards the forward direction. This is consistent with some of the experimental data in [12]. Our distributions also vary from those of Hanrahan and Krueger [5]. Because they assumed a smooth surface, their distributions tend to zero for glancing angles. We get the same behavior asymptotically when the surface roughness tends to zero (see Figure 4). The

shape of their distributions corresponds to the plots with a low albedo in Figure 4.

To further validate our model we wrote a simple Monte Carlo simulator, compared the results for a set of different parameters and found good agreement. Of course we could not verify this for all possible values listed in Table 1 because of the inefficiency of the Monte Carlo technique.

4   Reflection and Refraction from Rough Surfaces

In this section we derive BRDF and BTDF models for an isotropic random rough surface. As in van Ginneken et al. [25] we derive these models directly from a statistical model of the height field. Indeed, our BRDF is essentially the same as theirs. Our main contribution is of course the new BTDF model. To the best of our knowledge, a similar model has not appeared before in the optics literature. We have chosen to derive the BRDF here as well for two reasons. First, it makes the derivation easier to follow. Second, since we need an explicit expression for the BRDF, it makes the paper self-contained.

We assume that our surface is an isotropic gaussian random height field [1]. The probability that a normal ω_a lies within an infinitesimal solid angle $d\omega_a = (d(\cos\theta_a), d\phi_a)$ is given by the Beckmann function [1]:

$$P(\omega_a)\, d\omega_a = \frac{1}{2\pi\sigma^2 \cos^3\theta_a} \exp\left(-\frac{\tan^2\theta_a}{2\sigma^2}\right)\, d\omega_a, \tag{17}$$

where σ is the RMS slope of the surface. Let the surface be illuminated by a directional source of irradiance E_0 of direction $\omega_0 = (\cos\theta_0, \phi_0)$. For each direction ω_r, resp. ω_t, there is a unique normal ω_a that will reflect (resp. refract) the incoming light in the direction ω_r (resp. ω_t). In the case of reflection, this vector is simply the vector halfway between ω_0 and ω_r. For refraction it is the (normalized) vector equal to the sum of ω_0 and $\eta\omega_t$, where η is the ratio of the indices of refraction above and below the surface. Notice that η can be smaller than 1 when computing t_{21}, for example. In cases where $\eta > 1$ it's possible that no normal exists that refracts the incoming light in the direction ω_t. This happens whenever ω_t lies outside of the cone of refraction. In this case the BTDF is simply zero for that direction.

The incoming power at a surface element dA with normal ω_a is equal to the incoming irradiance times the projected area:

$$\Phi_0 = E_0\ \cos\theta_0'\ dA,$$

where $\cos\theta_0'$ is the cosine of the angle between the normal and the incoming direction. The amount of power that is reflected and refracted is determined by the Fresnel factor $F(\cos\theta_0', \eta)$ [2]. Indeed, a fraction F of the power is reflected while a portion $(1 - F)$ is refracted. The radiance reflected is by definition the power reflected per solid angle and foreshortened area. To get the total radiance reflected into direction ω_r, we multiply the radiance reflected by a point of the surface with normal ω_a by the Beckmann probability function defined in Equation 17:

$$u_r = \frac{F\ E_0\ \cos\theta_0'\ P(\omega_a)\ d\omega_a}{\cos\theta_r\, d\omega_r}.$$

The solid angles $d\omega_r$ and $d\omega_a$ are not independent. This can be understood intuitively: by varying the normal in the cone $d\omega_a$ we get a corresponding variation around the reflected direction ω_r. The

size of this variation is exactly the factor which relates the two solid angles. The precise relation between them was cited by Torrance and Sparrow [22]:

$$d\omega_r = 4\cos\theta'_0 \ d\omega_a.$$

Nayar provides an elegant geometric proof of this result [15]. In Appendix B we give an alternative proof which easily generalizes to the case of refraction to be discussed below. Consequently, the BRDF for a surface of roughness σ and with ratio of indices of refraction η is:

$$r = \frac{F(\cos\theta'_0, \eta) \ P(\omega_a)}{4 \ \cos\theta_r \ \cos\theta_0}.$$

This result, when multiplied by a shadowing function, is essentially the Cook-Torrance BRDF.

We now derive the BTDF in a similar fashion. As in the reflected case, the total radiance refracted into a direction ω_t is given by:

$$u_t = \frac{(1-F) \ E_0 \ \cos\theta'_0 \ P(\omega_a) \ d\omega_a}{\cos\theta_t \ d\omega_t}.$$

The relation between the solid angles $d\omega_t$ and $d\omega_a$ is, however, very different. At first we did not pay too much attention to this relationship and simply assumed $d\omega_a = d\omega_t$. But when we compared our analytical model with a Monte Carlo simulation for validation, we found large discrepancies. Finally, after a careful analysis of other BRDF derivations [4, 25] we realized the importance of this relation. In Appendix B we prove that:

$$d\omega_t = \frac{\left(\cos\theta'_0 - \sqrt{\cos^2\theta'_0 + \eta^2 - 1}\right)^2}{\eta\sqrt{\cos^2\theta'_0 + \eta^2 - 1}} \ d\omega_a = G(\cos\theta'_0, \eta) \ \cos\theta'_0 \ d\omega_a.$$

It is interesting to note that for $\eta = 1$ this factor is zero, i.e., when there is no surface, light travels unperturbed in a straight line. With this factor the BTDF is equal to:

$$t = \frac{(1 - F(\cos\theta'_0, \eta)) \ P(\omega_a)}{\cos\theta_t \cos\theta_0 \ G(\cos\theta'_0, \eta)}.$$

This last expression is the main result of this section: a new BTDF model for an isotropic rough surface. We also multiply this function by the shadowing function proposed by van Ginneken at al. [25]. We prefer this shadowing function over the one used by Cook and Torrance [4] since it is consistent with the underlying model for the surface.

The BRDF and the BTDF are shown for different ratios of indices of refraction η and roughness values σ in Figure 5. The top figure corresponds to a ratio $\eta = 1.4$ which is that of skin. These plots correspond to the functions r_{12}, t_{12}, r_{21} and t_{21} of our skin model. As mentioned above we have validated our derivation using a Monte Carlo simulation. Whether they are a good model for rough surfaces is another matter to be settled by experiment. At least, Cook and Torrance reported good agreement with experiment for the function r_{12} [4].

5 The Skin Shader

The main motivation behind our work was to create good skin shaders. It is clear that our illumination model has many other applications. For example, Hanrahan and Krueger used their model

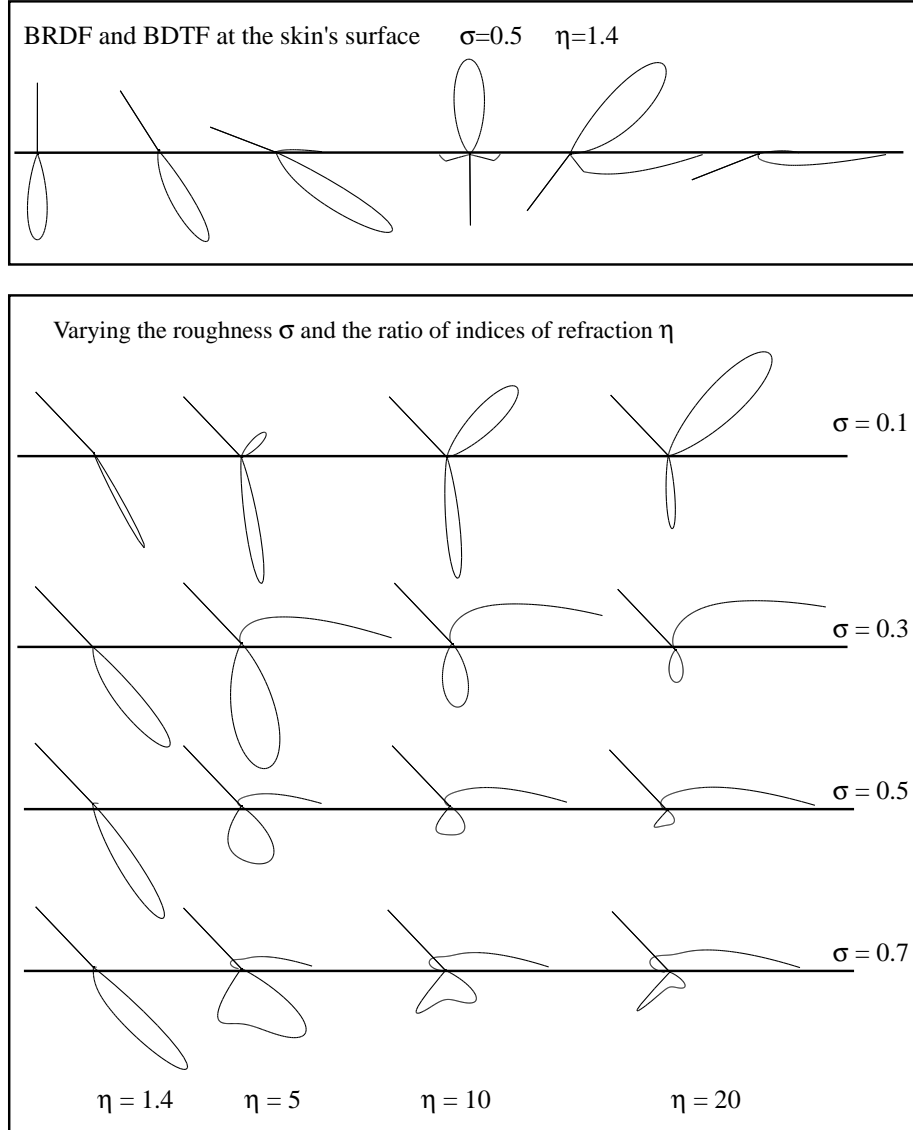Figure 5: Our new BRDFs and BTDFs for a rough surface. The distributions are rescaled to fit in the figure.

parameter	type	typical values
epidermis	transparency (RGB)	depends the race
dermis albedo	color (RGB)	$(0.993, 0.979, 0.943)$
dermis anisotropy	color (RGB)	$(0.860, 0.854, 0.823)$
surface roughness	scalar	$0.1 - 0.9$

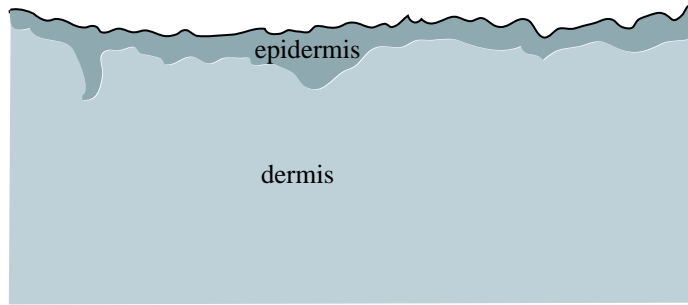Table 2: Parameters of the skin shader.

Figure 6: The skin is composed of a rough surface, the epidermis and the dermis. Most of the scattering occurs in the dermis.

to render leaves [5]. As shown in Figure 6, human skin consists of a rough surface and two layers below it: the epidermis and the dermis. The epidermis is a thin layer just below the skin's surface which only scatters weakly and mainly absorbs light. This layer determines the skin's general tone and corresponds to someone's "race". Below the epidermis is the dermis where all the scattering occurs. This layer is almost completely opaque ($T = 0$). The scattering there is characterized by a high albedo ($\Omega \approx 1$) and high forward scattering ($g \approx 0.8$). The shape of the distribution of reflected light depends only on the scattering in the dermis. The effect of the epidermis is simply to scale this distribution. Table 2 lists the parameters of our model. We used the precomputed data mentioned in Section 3.3 to evaluate our shader. We employed a simple quadri-linear interpolation for parameter values different from the ones listed in Table 1. Consequently, all of these parameters can be smoothly texture mapped to achieve many different effects.

We have implemented our reflection model as a shader plugin in our animation software MAYA. The plugin is available for free on our company's web page[1]. The web page also provides more information on the parameters of our skin shader. Several of our customers have recently started to use our shader in production with good results.

Figure 7 shows several examples of human heads rendered using our new skin shader. Figure 7.(a) compares our model (right) to a Lambert shader (left) and the Hanrahan-Krueger (HK) model (center). Our model seems to be a blend between these two models, which is consistent with the plots in Figure 4. Unfortunately, the comparison is necessarily very vague. Indeed, we manually tried to find a set of parameters for both the Lambert shader and the HK model which was as close as possible to our results. In particular, we had to "brighten up" the HK model since it assumes single-scattering. Figure 7.(b) shows our model illuminated by different area light sources. Notice also that we texture mapped both the albedo and the roughness of the lips. Figure 7.(c) is similar for a male head. Finally 7.(d) demonstrates a non-photorealistic application of our shader (notice that the surfaces have been bump mapped).

6   Conclusions and Future Work

In this paper, for the first time, we compute the reflection and the transmission of light through a skin layer bounded by rough surfaces. We achieved this through a discretization of the equation of radiative transfer. We were able to efficiently solve the discrete problem using Fourier transforms

---

[1] `http://www.aliaswavefront.com` by following "community" and "Download".
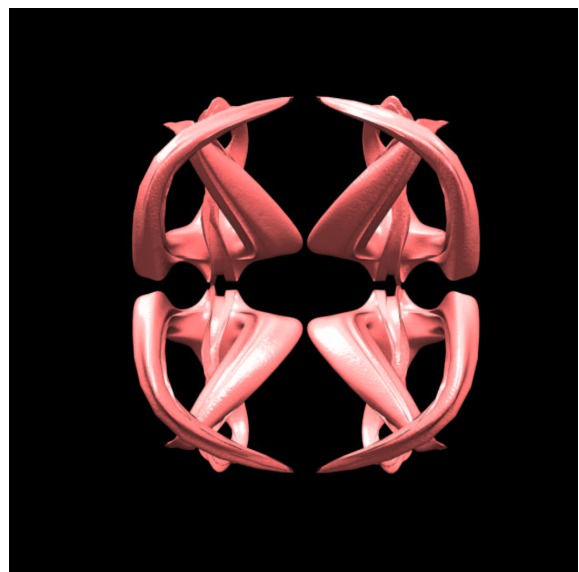
(a) Comparison of our shader (right) with a Lambertian shader(left) and the Hanrahan-Krueger model (center)
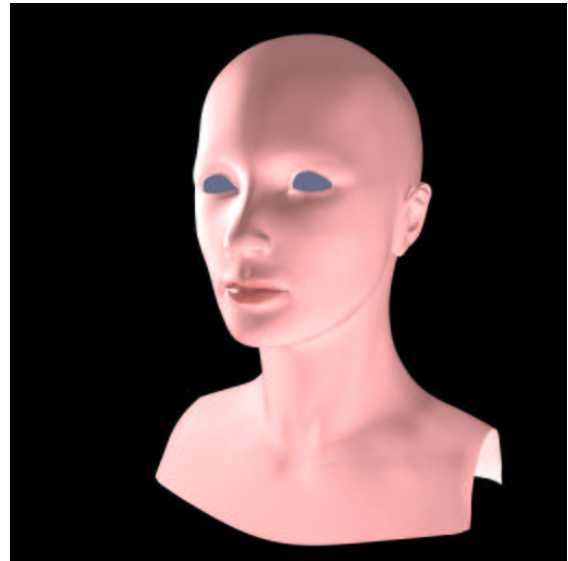
(b) Head under different lighting conditions. Flash-like area source (left) and two area light sources (right)

(c) Another head model with lips and freckles texture mapped.

(d) Non-photorealistic application of our model.

Figure 7: Renderings created using our new skin shader.

and eigenanalysis of the scattering matrix. Our model takes into account multiple anisotropic scattering and also handles reflections and refractions at the rough boundaries. To the best of our knowledge a model of this generality has not appeared before in computer graphics or in any of the related fields such as the atmospheric sciences. Our model is therefore of potential interest to these other fields as well. We have precomputed the distributions of reflected and transmitted light for different parameters and used an approximation of that data to build a skin shader. Also, for the first time we derive an analytical model for the Bidirectional Transmittance Distribution Functions (BTDFs) for a rough surface following earlier derivations for the Bidirectional Reflectance Distribution Function.

We have compared our solutions for a subset of our parameters with the output of more expensive Monte Carlo simulations. There was good agreement in each case. We are therefore confident that we have solved the physical formulation of the problem. The question of whether our models match reality has to be settled by comparing them with experimental data. We found a general agreement with the Cornell data [12], however, we feel that more comparisons are needed. Assuming our solver to be bug-free, it would be interesting if there were large discrepancies between our model and experiment. This would suggest that the linear transport theory is perhaps inadequate for this problem. On the practical side, our animators appreciated our new skin shader as it gave them an effect previously possible only with texture mapped Phong-like models. Possible future effects they have requested include "glowing ears", better skin bump-maps, tiny hairs, etc, which are of course hard to model using a shader alone. We are currently investigating how to achieve these effects.

There are many obvious extensions to our model. First, it is easy to include different scattering layers in our model, and our implementation can handle different layers. However, we found that one layer was good enough for the skin shader. To keep this paper as readable as possible, we decided not to add another index referring to the layers. The model can also be extended to handle anisotropic surfaces: simply use Fourier series instead of Cosine series for the azimuthal dependence of the radiances.

We intend to make our data publicly available in the hope that it might lead someone to find a better approximation scheme. Ideally, we would like to have a simple analytical model that fits the data. This would be of great practical and theoretical interest to many applied fields.

A   Details: Angular Discretization

In this appendix we provide the missing details of Section 2.3 that lead to explicit expressions for the matrices $\mathbf{M}_k$.

First, we expand the phase function into a cosine series as well:

$$p(\omega', \omega) = \sum_{k=0}^{N} p_k(\mu', \mu) \cos k(\phi - \phi'), \tag{18}$$

where the p_k are functions of the anisotropy factor g and the associated Legendre functions as shown in Appendix C. If we substitute Equations 7 and 18 into Eq. 4 we get the following $N + 1$ equations ($k = 0, \cdots, N$):

$$\mu \frac{d}{d\tau} u_k(\mu) = u_k(\mu) - \Omega_k \int_{-1}^{+1} p_k(\mu', \mu) \, u_k(\mu') \, d\mu', \tag{19}$$

where

$$\Omega_k = \Omega \frac{(1 + \delta_{0,k})}{4}$$

and $\delta_{i,j}$ is the Kronecker symbol. We now discretize the problem further by approximating the integrals in Equations 19 using a quadrature:

$$\int_{-1}^{+1} f(\mu') \, d\mu' \approx \sum_{m=1}^{M} w_m \left\{ f(-\mu_m) + f(\mu_m) \right\},$$

where w_m are the weights and $\mu_m \geq 0$ are the ordinates of the quadrature. With this approximation Equation 19 becomes a set of linear equations:

$$\mu_n \frac{d}{d\tau} u_k(\mu_n) = u_k(\mu_n) - \Omega_k \sum_{m=1}^{M} w_m \left\{ p_k(-\mu_m, \mu_n) u_k(-\mu_m) + p_k(\mu_m, \mu_n) u_k(\mu_m) \right\}$$

$$-\mu_n \frac{d}{d\tau} u_k(-\mu_n) = u_k(-\mu_n) - \Omega_k \sum_{m=1}^{M} w_m \left\{ p_k(-\mu_m, -\mu_n) u_k(-\mu_m) + p_k(\mu_m, -\mu_n) u_k(\mu_m) \right\}.$$

Since $p_k(-\mu_m, \mu_n) = p_k(\mu_m, -\mu_n)$ and $p_k(\mu_m, \mu_n) = p_k(-\mu_m, -\mu_n)$ we introduce the following two $M \times M$ matrices:

$$
\begin{aligned}
(A_k)_{n,m} &= \left(\Omega_k w_m p_k(\mu_m, \mu_n) - \delta_{n,m} \right) / \mu_n \quad \text{and} \\
(B_k)_{n,m} &= \Omega_k w_m p_k(-\mu_m, \mu_n) / \mu_n.
\end{aligned}
$$

Consequently, recalling the vector notations introduced in Section 2.3:

$$\mathbf{M}_k = \begin{pmatrix} -\mathbf{A}_k & -\mathbf{B}_k \\ \mathbf{B}_k & \mathbf{A}_k \end{pmatrix}.$$

B Relating $d\omega$ to $d\omega_a$

Computing the relationship between the two differentials $d\omega_a$ and $d\omega$ is mathematically equivalent to computing the Jacobian of the change of coordinates $\omega_a \to \omega$. In this appendix we compute the Jacobians for both the reflected and the refracted solid angles. We compute the change of coordinates in three steps:

$$\omega_a \longrightarrow (x_a, y_a) \longrightarrow (x, y) \longrightarrow \omega.$$

The relation between the spherical and cartesian coordinates is well known and given by

$$dx_a \, dy_a = \mu_a \, d\omega_a \quad \text{and} \quad dx \, dy = \mu \, d\omega.$$

Following the approach of Nayar et al. [15] we assume without loss of generality that the solid angle of the normal $d\omega_a$ is centered along the normal $(0, 0, 1)$. We also assume that the source is coming from the direction $(\mu_0, 0)$, or in cartesian coordinates $\mathbf{v}_0 = (\sqrt{1 - \mu_0^2}, 0, \mu_0)$. See Figure 8. Let

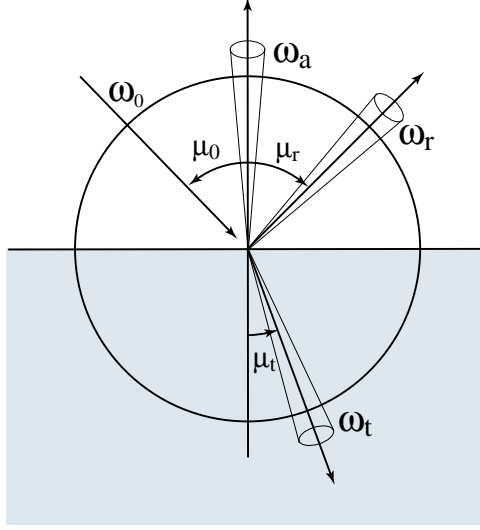$$\mathbf{n}(x_a, y_a) = (x_a, y_a, \sqrt{1 - x_a^2 - y_a^2})$$

Figure 8: Solid angles involved.

be a normal in $dx_a\ dy_a$. Then the reflected and refracted directions are equal to

$$\mathbf{r} = 2(\mathbf{n} \cdot \mathbf{v}_0)\mathbf{n} - \mathbf{v}_0 \quad \text{and}$$
$$\eta\mathbf{t} = \left(\mathbf{n} \cdot \mathbf{v}_0 - \sqrt{(\mathbf{n} \cdot \mathbf{v}_0)^2 + \eta^2 - 1}\right)\mathbf{n} - \mathbf{v}_0,$$

respectively. The transformation from the normal to the reflected vector corresponds to a change of coordinates $(x_a, y_a) \to (x, y)$, where

$$x = 2(\sqrt{1 - \mu_0^2}x_a + \mu_0\sqrt{1 - x_a^2 - y_a^2})x_a - \sqrt{1 - \mu_0^2}$$
$$y = 2(\sqrt{1 - \mu_0^2}x_a + \mu_0\sqrt{1 - x_a^2 - y_a^2})y_a.$$

The Jacobian of this change of coordinates at $(x_a, y_a) = (0, 0)$ is easily calculated to be equal to:

$$J = \begin{vmatrix} \frac{\partial x}{\partial x_a} & \frac{\partial x}{\partial y_a} \\ \frac{\partial y}{\partial x_a} & \frac{\partial y}{\partial y_a} \end{vmatrix} = \begin{vmatrix} 2\,\mu_0 & 0 \\ 0 & 2\,\mu_0 \end{vmatrix} = 4\,\mu_0^2.$$

Therefore,

$$\mu_0 d\omega = dx dy = 4\mu_0^2 dx_a dy_a = 4\mu_0^2 d\omega_a.$$

In other words,

$$d\omega = 4\mu_0\ d\omega_a,$$

which agrees with Nayar's result [15]. Our derivation might seem unnecessarily complicated compared to that of Nayar [15]. However, our derivation has the merit that it can easily be applied to the refraction problem as well. The Jacobian of the change of coordinates corresponding to the refraction at $(0, 0)$ is equal to

$$\eta^2 J = \begin{vmatrix} \mu_0 - \mu_t & 0 \\ 0 & \mu_0 - \mu_t \end{vmatrix} = (\mu_0 - \mu_t)^2,$$

18

where

$$\mu_t = \sqrt{\mu_0^2 + \eta^2 - 1}$$

is the cosine of the refracted direction. Again we have the chain of relations

$$\frac{1}{\eta}\mu_t \, d\omega = dxdy = \frac{1}{\eta^2}\left(\mu_0 - \mu_t\right)^2 \, dx_a dy_a = \frac{1}{\eta^2}\left(\mu_0 - \mu_t\right)^2 \, d\omega_a.$$

Therefore,

$$d\omega = \frac{\left(\mu_0^2 - \mu_t\right)^2}{\eta\mu_t} \, d\omega_a,$$

as advertised in Section 4.

C Representation of the Phase Function

The Henyey-Greenstein Phase function has the nice property that it can be expanded explicitly in a cosine series given by Equation 18. The coefficients in the expansion are expressed in terms of the associated Legendre functions [3]. This explains why this phase function is so popular in the radiative transfer literature. The expansion follows from the following result [6]:

$$p(\cos\gamma) = \sum_{k=0}^{N}(2k+1)g^k P_k(\cos\gamma),$$

where P_k is the Legendre polynomial of degree k. From a well known relation between the Legendre polynomials and the associated ones[2], we see that the coefficients in the expansion of Equation 18 are given by:

$$p_k(\mu', \mu) = \left(2 - \delta_{0,k}\right)\sum_{n=k}^{N}(2n+1)g^n \frac{(n-k)!}{(n+k)!} P_n^k(\mu')P_n^k(\mu),$$

where $P_n^k(x)$ are the *associated Legendre functions* [19].

D Discrete Representation of the BRDF and BTDF

Let $\rho(\omega', \omega)$ be one of our BRDFs or BTDFs. We then want to compute the coefficients ρ_k in the cosine series:

$$\rho(\omega', \omega) = \sum_{k=0}^{N}\rho_k(\mu', \mu)\cos k(\phi - \phi').$$

Unlike the phase function in Appendix C, we cannot express these coefficients analytically for the BRDF and BTDF derived in Section 4. For the given set of ordinates $\mu_1, \cdots, \mu_M$ we approximate the integrals:

$$I_k(\mu_n, \mu_m) = \int_0^{2\pi} \rho(\pm\mu_n, 0; \pm\mu_m, \phi) \, \cos k\phi \, d\phi,$$

for $k = 0, \cdots, N$ and $n, m = 1, \cdots, M$. The signs in the integrand depend on the BRDF/BTDF being computed. The discrete representation of the linear operators associated with ρ are then given by matrices $\mathbf{L}_k$ whose elements are

$$(L_k)_{n,m} = I_k(\mu_n, \mu_m).$$

---

[2]This relation was used in [9], for example.

E   Computation of Eigenvalues and Eigenvectors of Eq. 9

We seek a $2M$-dimensional vector $(\mathbf{a}, \mathbf{b})^T$ and a scalar λ that satisfy

$$\lambda \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} = \begin{pmatrix} -\mathbf{A} & -\mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}.$$

Equivalently,

$$\begin{aligned} \lambda(\mathbf{a} + \mathbf{b}) &= (\mathbf{A} - \mathbf{B})(\mathbf{b} - \mathbf{a}) \\ \lambda(\mathbf{b} - \mathbf{a}) &= (\mathbf{A} + \mathbf{B})(\mathbf{a} + \mathbf{b}) \end{aligned}$$

hence

$$\lambda^2(\mathbf{a} + \mathbf{b}) = (\mathbf{A} - \mathbf{B})(\mathbf{A} + \mathbf{B})(\mathbf{a} + \mathbf{b}).$$

Thus we have reduced the problem to size M. We can solve this problem using standard numerical methods to get M eigenvalues $\alpha_1, \cdots, \alpha_M > 0$ and M eigenvectors $\mathbf{e}_1, \cdots, \mathbf{e}_M$. Now, let $\lambda_n = \sqrt{\alpha_n}$ and $\lambda_n \mathbf{f}_n = -(\mathbf{B} + \mathbf{A})\mathbf{e}_n$, and

$$\mathbf{v}_n^+ = \frac{1}{2}(\mathbf{e}_n + \mathbf{f}_n) \quad \text{and} \quad \mathbf{v}_n^- = \frac{1}{2}(\mathbf{e}_n - \mathbf{f}_n).$$

One can verify that these definitions provide eigenvalues and eigenvectors for our original problem:

$$\lambda_n \begin{pmatrix} \mathbf{v}_n^+ \\ \mathbf{v}_n^- \end{pmatrix} = \begin{pmatrix} -\mathbf{A} & -\mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{v}_n^+ \\ \mathbf{v}_n^- \end{pmatrix} \quad \text{and} \quad -\lambda_n \begin{pmatrix} \mathbf{v}_n^- \\ \mathbf{v}_n^+ \end{pmatrix} = \begin{pmatrix} -\mathbf{A} & -\mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{v}_n^- \\ \mathbf{v}_n^+ \end{pmatrix}.$$

These results can be written more compactly using the following matrices:

$$\mathbf{V}^+ = \begin{pmatrix} \mathbf{v}_1^+ \cdots \mathbf{v}_M^+ \end{pmatrix}, \quad \mathbf{V}^- = \begin{pmatrix} \mathbf{v}_1^- \cdots \mathbf{v}_M^- \end{pmatrix} \quad \text{and} \quad \mathbf{\Lambda}^+ = \mathrm{diag}(\lambda_1, \ldots, \lambda_M).$$

We obtain the following matrices of eigenvectors and eigenvalues:

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}^+ & \mathbf{V}^- \\ \mathbf{V}^- & \mathbf{V}^+ \end{pmatrix} \quad \text{and} \quad \mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}^+ & \mathbf{0} \\ \mathbf{0} & -\mathbf{\Lambda}^+ \end{pmatrix}.$$

References

[1] P. Beckmann and A. Spizzichino. *The Scattering of Electromagnetic Waves from Rough Surfaces*. Pergamon, New York, 1963.

[2] M. Born and E. Wolf. *Principles of Optics. Sixth (corrected) Edition*. Cambridge University Press, Cambridge, U.K., 1997.

[3] S. Chandrasekhar. *Radiative Transfer*. Dover, New York, 1960.

[4] R. L. Cook and K. E. Torrance. A Reflectance Model for Computer Graphics. *ACM Computer Graphics (SIGGRAPH '81)*, 15(3):307–316, August 1981.

[5] P. Hanrahan and W. Krueger. Reflection from Layered Surfaces due to Subsurface Scattering. In *Proceedings of SIGGRAPH '93*, pages 165–174. Addison-Wesley Publishing Company, August 1993.

[6] J. E. Hansen and L. D. Travis. Light Scattering in Planetary Atmospheres. *Space Science Reviews*, 16:527–610, 1974.

[7] X. D. He. *Physically-Based Models for the Reflection, Transmission and Subsurface Scattering of Light by Smooth and Rough Surfaces, with Applications to Realistic Image Synthesis*. PhD thesis, Cornell University, Ithaca, New York, 1993.

[8] Z. Jin and K. Stamnes. Radiative transfer in nonuniformly refracting layered media: atmosphere-ocean system. *Applied Optics*, 33(3):431–442, January 1994.

[9] J. T. Kajiya and B. P. von Herzen. Ray Tracing Volume Densities. *ACM Computer Graphics (SIGGRAPH '84)*, 18(3):165–174, July 1984.

[10] E. P. F. Lafortune, S-C. Foo, K. E. Torrance, and D. P. Greenberg. Non-Linear Approximation of Reflectance Functions. In *Computer Graphics Proceedings, Annual Conference Series, 1997*, pages 117–126, August 1997.

[11] E. Languénou, K.Bouatouch, and M.Chelle. Global illumination in presence of participating media with general properties. In *Proceedings of the 5th Eurographics Workshop on Rendering*, pages 69–85, Darmstadt, Germany, June 1994.

[12] S. R. Marschner, S. H. Westin, E. P. F. Lafortune, K. E. Torrance, and D. P. Greenberg. Image-based brdf measurement including human skin. *Eurographics Workshop on Rendering*, 1999.

[13] N. Max. Efficient light propagation for multiple anisotropic volume scattering. In *Proceedings of the 5th Eurographics Workshop on Rendering*, pages 87–104, Darmstadt, Germany, June 1994.

[14] C. D. Mobley. A numerical model for the computation of radiance distributions in natural waters with wind-roughened surfaces. *Limnology and Oceanography*, 34(8):1473–1483, 1989.

[15] S. K. Nayar, K. Ikeuchi, and T. Kanade. Surface Reflection: Physical and Geometrical Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):611–634, July 1991.

[16] NETLIB. The code is publicly available from `http://netlib.org`.

[17] G. N. Plass, G. W. Kattawar, and F. E. Catchings. Matrix operator theory of radiative transfer. 1: Rayleigh scattering. *Applied Optics*, 12(2):314–329, February 1973.

[18] A. A. Prahl, M. J. C. van Gemert, , and A. J. Welch. Determining the optical properties of turbid media by using the adding-doubling method. *Applied Optics*, 32:559–568, 1993.

[19] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C. The Art of Scientific Computing*. Cambridge University Press, Cambridge, 1988.

[20] H. E. Rushmeier and K. E. Torrance. The Zonal Method for Calculating Light Intensities in the Presence of a Participating Medium. *ACM Computer Graphics (SIGGRAPH '87)*, 21(4):293–302, July 1987.

[21] K. Stamnes and P. Conklin. A New Multi-Layer Discrete Ordinate Approach to Radiative Transfer in Vertically Inhomogeneous Atmospheres. *Journal of Quantum Spectroscopy and Radiative Transfer*, 31(3):273–282, 1984.

[22] K. E. Torrance and E. M. Sparrow. Theory for Off-Specular Reflection From Roughened Surfaces. *Journal of the Optical Society of America*, 57(9):1105–1114, September 1967.

[23] V. V. Tuchin. Light scattering study of tissue. *Physics - Uspekhi*, 40(5):495–515, 1997.

[24] M. J. C. van Gemert, S. L. Jacques, H. J. C. M. Sterenborg, and W. M. Star. Skin optics. *IEEE Transactions on Biomedical Engineering*, 36(12):1146–1154, December 1989.

[25] B. van Ginneken, M. Stavridi, and J. J. Koenderink. Diffuse and specular reflectance from rough surfaces. *Applied Optics*, 37(1):130–139, January 1998.