MODELS OF LIGHT REFLECTION FOR COMPUTER SYNTHESIZED PICTURES

James F. Blinn

University of Utah

ABSTRACT

In the production of computer generated pictures of three dimensional objects, one stage of the calculation is the determination of the intensity of a given object once its visibility has been established. This is typically done by modelling the surface as a perfect diffuser, sometimes with a specular component added for the simulation of hilights. This paper presents a more accurate function for the generation of hilights which is based on some experimental measurements of how light reflects from real surfaces. It differs from previous models in that the intensity of the hilight changes with the direction of the light source. Also the position and shape of the hilights is somewhat different from that generated by simpler models. Finally, the hilight function generates different results when simulating metallic vs. nonmetallic surfaces. Many of the effects so generated are somewhat subtle and are apparent only during movie sequences. Some representative still frames from such movies are included.

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

CR Categories: 3.17, 5.12, 8.2

INTRODUCTION

In producing computer generated pictures of three dimensional objects, two types of calculation must be performed. The first, and most popularly discussed, is the hidden surface problem; determining which object is visible where on the screen and what is the normal vector to the object at that point. The second is the intensity calculation; given the normal vector and the position of the light sources, what is the proper intensity for the corresponding spot on the picture. Very simple models are typically used which simulate ideal diffuse reflectors. This uses the, so called, Lambert's law which states that the surface will diffuse incident light equally in all directions. Differences in intensity are then caused by the different amounts of incident light per unit area intercepted by portions of the surface at various angles to the light source. This will be proportional to the cosine of the angle between the normal to the surface, N, and the vector to the light source, L. This cosine is evaluated by computing the dot product of the two vectors after normalizing them to a length of 1. If this dot product is negative it indicates that the viewer is on the opposite side of the surface from the light source. The intensity should then be set to zero.

In addition, some constant value is usually added to the intensity to simulate the effects of ambient light on the surface. This assumes that a small amount of light falls on the surface uniformly from all directions in addition to the main point light source. The integral of this ambient light from all directions yields a constant value for any normal direction. The net function is:

$$d = \max(0, N \cdot L)$$

$$i = p_a + d\, p_d$$

where

i = percieved intensity

p_a = proportion of ambient reflection

p_d = proportion of diffuse reflection

d = amount of diffuse reflection

N = Normal vector to surface

L = Light direction vector

This model is simple to compute and quite adequate for many applications.

SIMPLE HILIGHT MODELS

A more realistic lighting model was introduces by Phong [2] as part of a technique for improving the appearance of images of curved surfaces. The function makes use of the fact that, for any real surface, more light is reflected in a direction making an equal angle of incidence with reflectance. The additional light reflected in this direction is referred to as the specular component. If the surface was a perfect mirror light would only reach the eye if the surface normal, N, pointed halfway between the source direction, L, and the eye direction, E. We will name this direction of maximum hilights H, where

$$H = \frac{L+E}{\mathrm{len}(L+E)}$$

Siggraph '77, July 20-22 San Jose, California

For less than perfect mirrors, the specular component falls off slowly as the normal direction moves away from the specular direction. The cosine of the angle between H and N is used as a measure of the distance a particular surface is away from the maximum specular direction. The degree of sharpnesssof the highlights is adjusted by taking this cosine to some power, typically 50 or 60. The net Phong shading function is then:

$$d = \max(0, N \cdot L)$$

$$s = (N \cdot H)^{c_1}$$

$$i = p_a + d\,p_d + s\,p_s$$

where

i = percieved intensity

p_s = proportion of specular reflection

s = amount of specular reflection

c_1 = measure of shininess of surface

other values as defined above

In addition, when simulating colored surfaces, there is a different intensity value for each primary. These. should be calculated by scaling only the diffuse and ambient components by the color of the object. The highlights then appear desaturated or white.

TORRANCE- SPARROW MODEL

The reflection of light from real surfaces has been the subject of much theoretical and experimental work by physicists Is], [6] and illumination engineers [4]. The experimental results generally match the Phong shading function but some differences do arise. The main one is the fact that the the specular bump, represented by the parameter p_s above, varies with the direction of the light source. Also the direction of peak specular reflection is not always exactly along H. In 1967 Torrance and Sparrow [7] derived a theoretical model to explain these effects. The match between their theoretically predicted functions and experimentally measured data is auite imoressive. In this section we derive the Torrance-Sparrow highlight function in terms of the vectors N, L, H and E, all of which are assumed to be nonnallized.

The surface being simulated is assumed to be composed of a collection of mirror like micro facets. These are oriented in random directions all over the surface. The specular component of the reflected light is assumed to come *'from* reflection from those facets oriented in the direction of H. The diffuse component comes from multiple reflections between facets and from internal scattering. The specular reflection is then a combination of four factors:

$$s = \frac{DGF}{(N \cdot E)}$$

D is the distribution function of the directions of the micro facets on the surface. G is the amount by which the facets shadow and mask each other. F is the Fresnel reflection law. Each of these factors will now be examined in ***turn.***

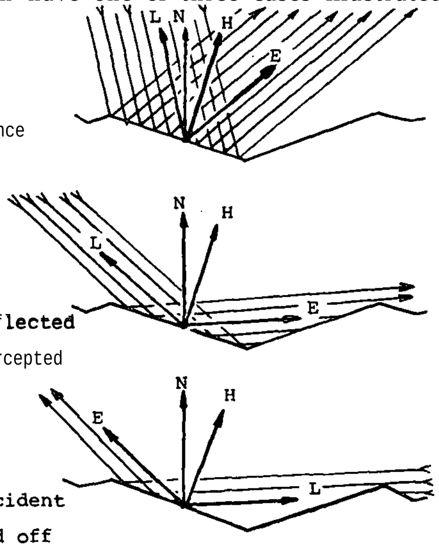
The light reflected specularly in any given direction can come only from the facets oriented to reflect the light in that direction. That is, the facets whose local normal vectors point in the direction of H. The first term in the specular reflectance is the evaluation of the distribution of the number of facets pointing in that.direction. The distribution used by Torrance and Sparrow was a simple Gaussian:

$$D_2 = e^{-(\alpha c_2)^2}$$

D_2 is the proportionate number of facets oriented at an angle α from the average normal to the surface. The factor c_2 is the standard deviation for the distribution and is a property of the surface being modelled. Large values yield dull surfaces and small values yield shiny surfaces. We are interested in the number of facets pointing in the direction of H so the angle α here is $\cos^{1}(N \cdot H)$.

Since the intensity is proportional to the number of facets pointing in the H direction, we must take into account the observer sees more of the surface area when the surface is tilted. The increase in area is inversely proportional to the cosine of the angle of tilt. The tilt angle is is the angle between the average surface normal, N, and the eye, E. This explains the division by (N.E).

Counteracting this effect is the fact that some of the facets shadow each other. The degree to which this shadowing occurs is called the "geometrical attenuation factor", G. It is a value from 0 to.1 representing the proportionate amount of light remaining after the masking or shadowing has taken place. Calculation of G assumes that the micro facets exist in the form of V shaped grooves with the sides at equal but opposite angles to the average surface normal. We are interested only in grooves where one of the sides points in the specular direction H. For differing positions of the light source and eye oosition we can have one of three cases illustrated in Figure 1.



Case a
No interference

Case b
Some of the reflected light is intercepted

Case c
Some of the incident light is masked off

Figure 1

Note that the vectors L and E do not necessarily lie in the plane of the figure (i.e. the plane containing N and H). We can see this by considering a top view as in Figure 2.

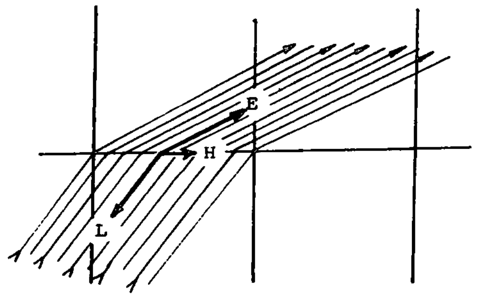


Figure 2 - Top view of reflection from a micro-facet

The value of G for case a of Figure 1 is 1.0, signifying no attenuation.

To compute G for case b we need to compute the ratio $1-(m/\ell)$ which is the proportionate amount of the facet contributing to the reflected light. See Figure 3.

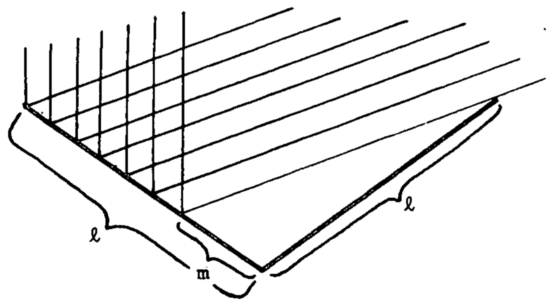


Figure 3 - Light which escapes is $1-(m/\ell)$

We can reduce the problem to two dimensions if we project E onto the plane containing N and H (the plane of the diagram). Calling this projection E_P and labeling relevant angles we have Figure 4.

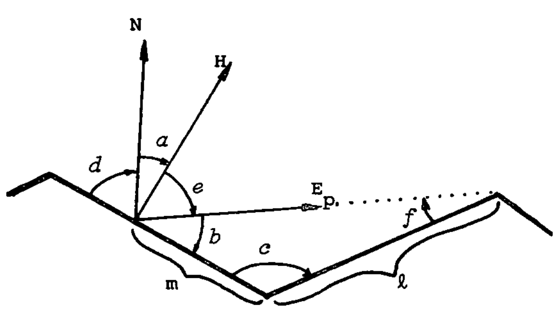


Figure 4 - Measurement of m/ℓ

Applying the law of sines we have

$$m/\ell = \sin f/\sin b$$

Then we note that

$$\sin b = \cos e$$
$$\cos b = \sin e$$

Since the angles of the triangle must sum to 2π we have

$$\sin f = \sin(b+c)$$
$$= \sin b \cos c + \cos b \sin c$$

Due to the symmetry of the groove and the complementarity of d and a

$$c = 2d$$
$$\cos c = 1-2\sin^2 d = 1-2\cos^2 a$$
$$\sin c = 2\cos d \sin d = 2\sin a \cos a$$

Plugging these into the expression for sinf

$$\sin f = \cos e(1-2\cos^2 a) + 2\sin e \cos a \sin a$$
$$= \cos e - 2\cos a(\cos e \cos a - \sin e \sin a)$$
$$= \cos e - 2\cos a \cos(e+a)$$
$$= (H\cdot E_p)-2(N\cdot H)(N\cdot E_p)$$

Since E_p is the projection of B onto the N,H plane then $N\cdot E_p = N\cdot E$ and $H\cdot E_p = H\cdot E$ so that

$$G_b = 1 - \frac{m}{\ell} = 2(N\cdot H)(N\cdot E)/(E\cdot H)$$

Examining the diagram for Gc we see that it is the same as that for G_b but with the roles of L and E exchanged. Thus

$$G_c = 2(N\cdot H)(N\cdot L)/(H\cdot L) = 2(N\cdot H)(N\cdot L)/(E\cdot H)$$

For a particular situation, the effective value of G will be the minimum of G_a, G_b and G_c.

The final factor in the specular reflection is the Fresnel reflection. This gives the fraction of the light incident on a facet which is actually reflected as opposed to being absorbed. This is a function of the angle of incidence on the micro facet and the index of refraction on the substance. It is given by

$$F = \frac{1}{2}\left(\frac{\sin^2(\phi-\theta)}{\sin^2(\phi+\theta)} + \frac{\tan^2(\phi-\theta)}{\tan^2(\phi+\theta)}\right)$$

where $\sin\theta = \sin\phi/n$
ϕ = angle of incidence
n = index of refraction

In our case, the angle of incidence is $\phi = \cos^{-1}(L\cdot H) = \cos^{-1}(E\cdot H)$. The interesting thing about this function is that it has a substantially different form for metallic vs. nonmetallic substances. For metals, corresponding to large values of n, $F(\phi,n)$ is nearly constant at 1. For nonmetals, corresponding to small values of n, it has a more exponential appearance, starting out near zero for $\phi=0$ and going to 1 at $\phi=\pi/2$.

FACET DISTRIBUTION FUNCTIONS

One thing in the above model can be improved upon. This is the facet distribution function. This function takes an angle, a, and a measure of the shininess of the surface and computes the proportionate area of facets pointing in that direction. The angle a is the angle between H and N; we can evaluate its cosine as (N•H).

The Phong model effectively uses the distribution function of the cosine raised to a power.

$$D_1 = \cos^{c_1}\alpha$$

The Torrance Sparrow model uses the standard Gaussian distribution already mentioned.

$$D_2 = e^{-(\alpha c_2)^2}$$

A third function has been proposed by Trowbridge and Reitz [8]. They showed that a very general class of surface properties could be generated by modelling the microfacets as ellipsoids of revolution. This leads to the distribution function

$$D_3 = \left(\frac{c_3^2}{\cos^2\alpha(c_3^2-1)+1}\right)^2$$

Where c_3 is the eccentricity of the ellipsoids and is 0 for very shiny surfaces and 1 for very diffuse surfaces.

Each of these functions has a peak value of 1 at $\alpha=0$ (for facets pointing along the average surface normal) and falls off as a increases or decreases. The rate of fall off is controlled by the values c_1, c_2 and c_3. In comparing the functions it is necessary to specify this rate in a uniform unit. A convenient such unit is the angle at which the distribution falls to one half. In terms of this angle, β, the three coefficients are:

$$c_1 = -\frac{\ln 2}{\ln \cos\beta}$$

$$c_2 = \frac{\sqrt{\ln 2}}{\beta}$$

$$c_3 = \left(\frac{\cos^2\beta-1}{\cos^2\beta-\sqrt{2}}\right)^{1/2}$$

If these three functions are plotted with equal values of β it can be seen that they are very similar in shape. However, since there is some experimental as well as theoretical justification for D_3 and since it is the easiest to compute, it is the one we shall choose.

COMPUTATIONAL CONSIDERATIONS

There are several observations which can be made to speed up the computation of the hilight function.

If β does not change within a frame the function D_3 can be calculated using the intermediate values (calculated once per frame):

$$k_1 = 1/(c_3^2-1)$$

$$k_2 = k_1+1$$

whereupon

$$D_3 = \left(\frac{k_2}{\cos^2\alpha+k_1}\right)^2$$

A simplification which is often made is to assume that the light source is at infinity. Thus the vector L is a constant for each point of the picture. We may also model the eye as being far away from the object so that E = (0 0 -1). This allows the calculation of the direction of H to be done once per change in light direction.

It is possible to avoid a potential division by zero when computing G by combining it with the term 1/(N•E) and finding the minimum of Ga, Gb and Gc before doing the divisions:

```
if(N•E)<(N•L) then
  if 2(N•E)(N•H)<(E•H) then G:=2(N•H)/(E•H)
                       else G:=1/(N•E)
else
  if 2(N•L)(N•H)<(E•H) then G:=2(N•H)(N•L)/(E•H)(N•E)
                       else G:=1/(N•E)
```

The Fresnel reflection is a function only of the index of refraction and the dot product (E•H). If E is assumed constant at (0 0 -1) then this calculation needs to be made only once per change in light source direction. In addition, by some trigonometric identities it can be shown that the Fresnel formula can be calculated by:

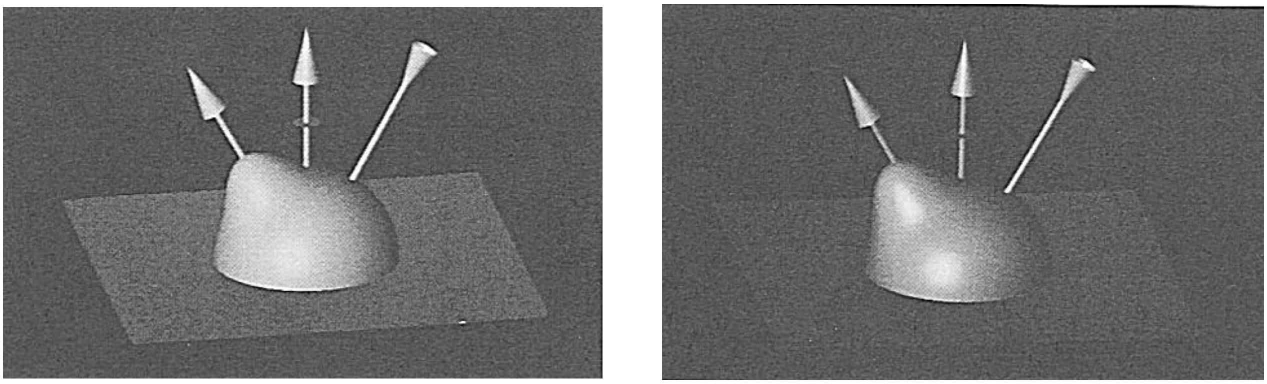
$$F = \frac{(g-c)^2}{(g+c)^2}\left(1 + \frac{(c(g+c)-1)^2}{(c(g-c)+1)^2}\right)$$

where $c = (E\cdot H)$

$$g = \sqrt{n^2+c^2-1}$$

COMPARISON WITH PHONG SHADING

Now that we have derived this hilight function we should compare it with the Phong function to see where and by how much they differ. Figure 5 shows a plot of the amount of light reflected from a surface as a result of an incident ray at 30 degrees from the surface normal. The distance of the surface in a particular direction from the center represents the amount of light reflected in that direction. The incoming ray is from the right. A vector pointing to the left at the specular direction is shown for reference. The hemispherical portion of the function is the diffuse reflection; equal amounts in each direction. The bump is the specular reflection.. For this angle of incidence the functions.are almost identical. Figure 6 shows the same function for an incident ray at 70 degrees. Note that the specular bump is much larger for the Torrance



Phong Model | Torrance-Sparrow Model

Figure 5

Comparison of Phong and Torrance-Sparrow reflection distributions for incident light at 30° from normal

Sparrow function and not in quite the same direction. This indicates that the new function will be materially different only for shallow angles of incident light and that the specular reflection will be much higher there. This may be verified by the simple experiment of holding a matte sheet of paper edge on to a light and noting that it looks quite shiny.

Figure 7 shows images of an object made using the two hilight functions with both an edge-on lighting direction and a front-on direction. Figure 7a simulates an aluminum metallic surface using the experimentally measured parameters :

$$p_s = .4$$
$$p_d = .6$$
$$n = 200$$
$$c_3 = .35$$

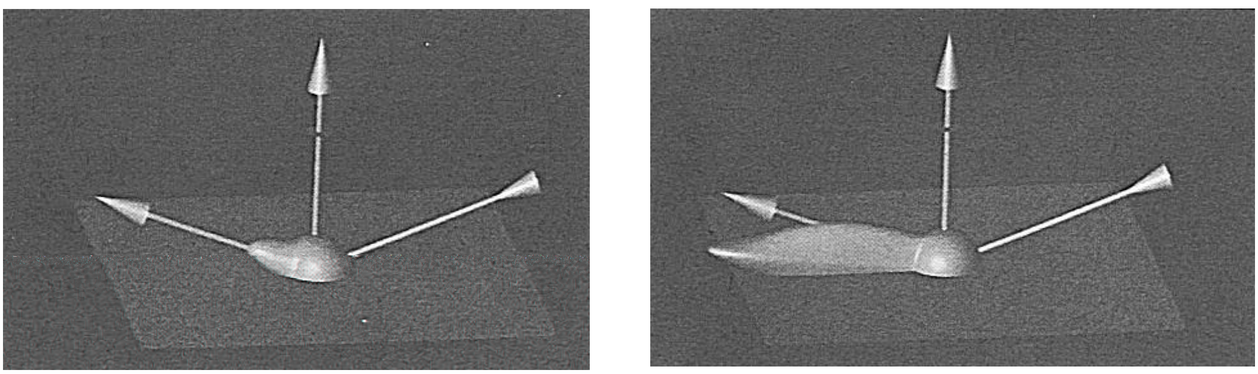
Figure 7b simulates a Magnesium Oxide ceramic (a standard diffuse reflector) using the experimental parameters:

$$p_s = .667$$
$$p_d = .333$$
$$n = 1.8$$
$$c_3 = .35$$

Note that the ceramic looks quite diffuse for light hitting it almost perpendicularly and very specular (even more. so than. the aluminum) for light hitting it almost tangentially.

VARYING SURFACE SHININESS

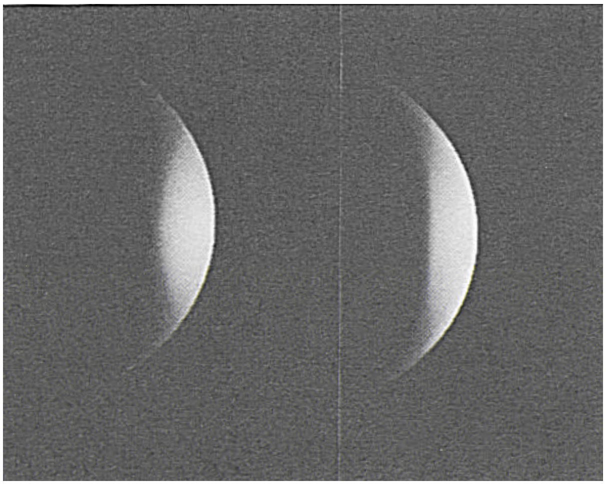
In [1] and [3] a technique for mapping texture patterns onto. bicubic surfaces was described. The object-was defined as a biparameteric surface and the parameter values were

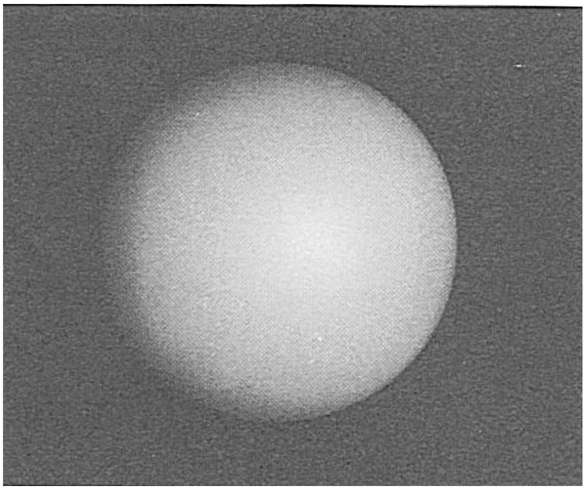


Phong Model | Torrance-Sparrow Model

Figure 6

Comparison of Phong and Torrance-Sparrow reflection distributions for incident light at 70° from normal





Phong Torrance-Sparrow

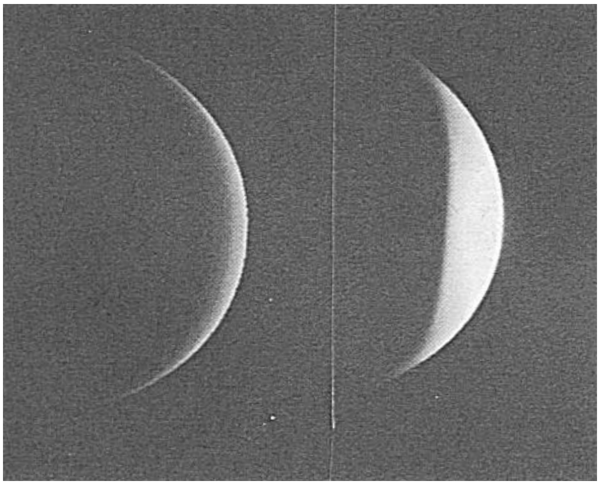
Edge Lit

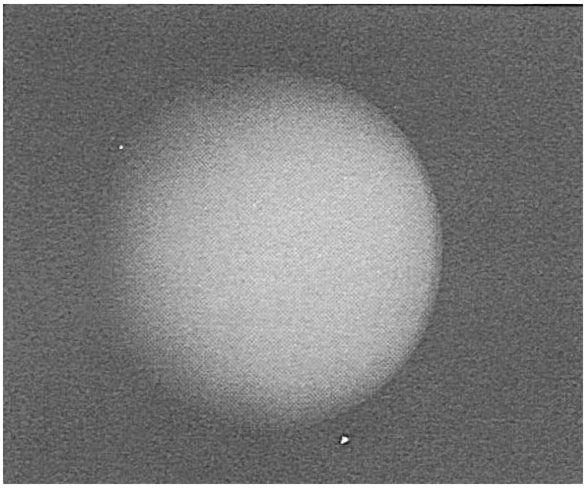
Both Models Essentially Same

Front Lit

Figure ?a

Simulation of Aluminum Surface





Phong Torrance-Sparrow

Edge Lit

Both Models Essentially Same

Front Lit

Figure 7b

Simulation of Magnesium Oxide Surface

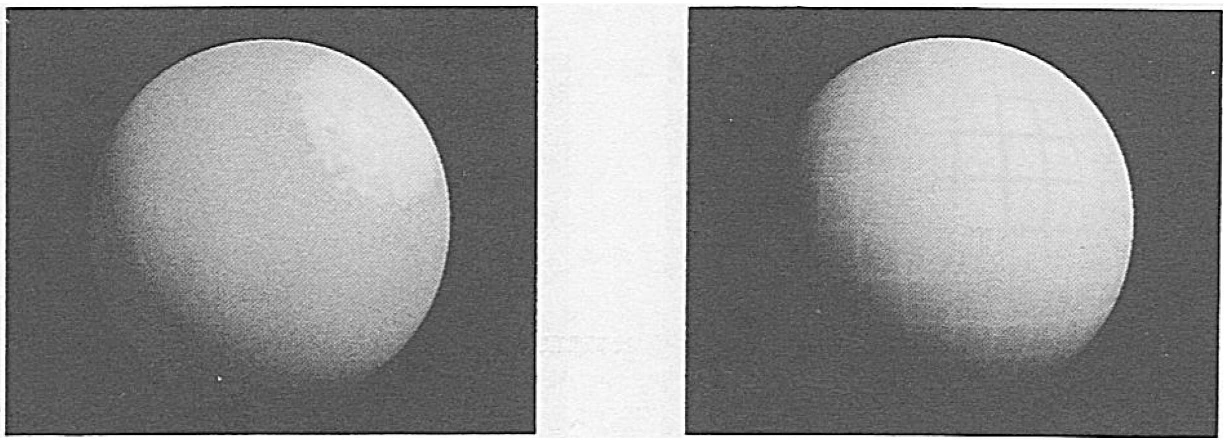


Figure 8

Surface shininess varying as a function of two different texture patterns

used as input to a texture function which scaled the diffuse component of the reflection. This form of mapping is good for simulating patterns painted on the surface but attempts to simulate bumpy surfaces were disappointing. This effect can, however, be better approximated by using the same texture mapping approach applied to the local surface roughness c_3.

If c_3 is going to change from place to place on the surface we must worry about nonnallization of the D_3 function. In its original derivation in [8] D_3 differed from that shown here by a factor of c_3^2. This additional factor was included here as a normallizing constant to make $D_3(0)=1$. Since, now, c_3 is varying across the surface, we wish to use a constant normallizing factor based on its minimum value over the surface. The texture modulated distribution function should then be:

$$c_3 = c_{min} + (1-c_{min})\, t(u,v)$$

$$D_3 = \left(\frac{c_{min}\, c_3}{\cos^2\alpha(c_3^2-1)+1}\right)^2$$

where $t(u,v)$=texture value

Figure 8 shows some images made with various texturing functions.

CONCLUSIONS

The Torrance-Sparrow reflection model differs from the Phong model in the inclusion of the G, F and $1/(N \cdot E)$ terms. This has a noticable effect primarily for non-metallic and edge lit objects. The use of the D_3 micro facet distribution function provides a better match to experimental data and is, happily, easier to compute than D_1 or D_2. This savings effectively offsets the extra computation time for G and F yielding a hilight generation function having a high degree of realism for no increase in computation time.

REFERENCES

1. Blinn, J. F. and Newell, M. E. Texture and reflection incomputer generated images. Comm ACM 19, 10(Oct 1976), 542-547

2. Bui-Tuong Phong. Illumination for computer generated images. Comm ACM 18, 6(June 1975) 311-317

3. Catmull, E. A. Computer display of curved surfaces. Proc. Conf. on Comptr. Graphics. May 1975 (IEEE Cat. No. 75CH0981-1C)11-17

4. Gilpin, F. H. Effect of the variation of the incident angle on the coefficient of diffused reflection. Trans. Illum. Eng. Soc. Vol 5, 1910 854-873

5. Middleton, W. E. K. and Mungall, A. G. The luminous directional reflectance of snow. J. Opt. Soc. Am. 42, 8(Aug 1952) 572-579

6. Torrance, K. E. and Sparrow, E. M. Polarization, directional distribution, and off-specular peak phenomena in light reflected from roughened surfaces. J. Opt. Soc. Am. 56, 7(Jul 1966) 916-925

7. Torrance, K. E. and Sparrow, E. M. Theory for off-specular reflection from roughened surfaces. J. Opt. Soc. Am.. 57, 9(Sep 1967) 1105-1114

8. Trowbridge, T. S. and Reitz, K. P. Average irregularity representation of a roughened surface for ray reflection. J. Opt. Soc. Am. 65, 1975) 531-536