Surface Reflection II

Lecture #5:	Wednesday, 25 October 2000
Lecturer:	Pat Hanrahan
Scribe:	Yi-Ren Ng
Reviewer:	Tim Purcell, Pat Hanrahan

1 Introduction

This lecture is a continuation of the previous lecture on surface reflection. Three topics are covered: correct normalization of microfacet distribution functions; shadowing on a Gaussian rough surface; and a model for surface reflection based on thoroughly pitted surfaces.

2 Normalization of Microfacet Distributions

We will explore three important factors in creating correctly normalized, physically plausible BRDFs: area normalization, energy conservation, and self-shadowing.

2.1 Area Normalization

As we saw in the previous lecture, we were able to derive a microfacet distribution function, given by $P(z')dz'$ where $z' = \frac{\partial z}{\partial x}$, directly from the height distribution function of a rough surface. In particular, last week we derived the microfacet distribution function for a surface with Gaussian statistics. Since these distribution functions are probability distributions, integrating over all slopes gives unity.

$$\int P(z')dz' = 1$$

This equation imposes a normalization condition on the system. Geometrically, the condition means that if we were to project a rough surface with such a microfacet distribution onto the surface base plane perpendicular to the (global) surface normal, we would get unity area. This makes sense because each point comprising a height field projects to unique point on the base plane.

In the Torrance-Sparrow model that we derived last lecture, we saw that the microfacet distribution $D(\omega_f)dA$ represented the area of all the microfacets oriented in the direction ω_f. Applying the area normalization condition above to this microfacet distribution model, we require that for a height field

$$\int D(\omega_f)\cos\theta_f d\omega_f = 1$$

Here θ_f represents the angle between the microfacet normal and the surface normal. The cosine term gives the projected area of the microfacet on the surface base plane.

A consequence of this requirement is that if we want to relate any the Torrance-Sparrow microfacet distribution $D(\omega_f)$ to the slope distribution function $P(\omega_f)$ we must divide through by $\cos\theta_f$ to normalize it correctly.

$$D(\omega_f) = \frac{P(\omega_f)}{\cos\theta_f}$$

2.2 Energy normalization

The integral of the BRDF f_r over its hemisphere is called the directional-hemispherical reflectance distribution function, and is denoted ρ_{dh}. We can interpret this function as the proportion of the incoming light that is reflected for an incoming ray direction. By conservation of energy, we know that this distribution must be less than or equal to one, otherwise the surface would emit more energy than it received.

$$\rho_{dh}(\omega_i) = \int f_r(\omega_i \to \omega_r)\cos\theta_r d\omega_r \le 1 \tag{1}$$

To relate this to $D(\omega_f)$, we recall from last lecture that, ignoring the Fresnel effect and self-shadowing effects,

$$\begin{aligned}
f_r(\omega_i \to \omega_r) &= \frac{D(\omega_f)}{4\cos\theta_i\cos\theta_r} \\
\rho_{dh}(\omega_i) &= \frac{1}{4\cos\theta_i}\int D(\omega_f)d\omega_r \\
&= \frac{1}{4\cos\theta_i}\int D(\omega_f)4\cos\theta' d\omega_f \\
&= \frac{1}{\cos\theta_i}\int D(\omega_f)\cos\theta' d\omega_f
\end{aligned}$$

where θ' is the angle between the incoming ray and the microfacet normal. Substituting into Equation 1 we get

$$\frac{1}{\cos\theta_i}\int D(\omega_f)\cos\theta' d\omega_f \le 1$$

or

$$\int D(\omega_f)\cos\theta' d\omega_f \le \cos\theta_i$$

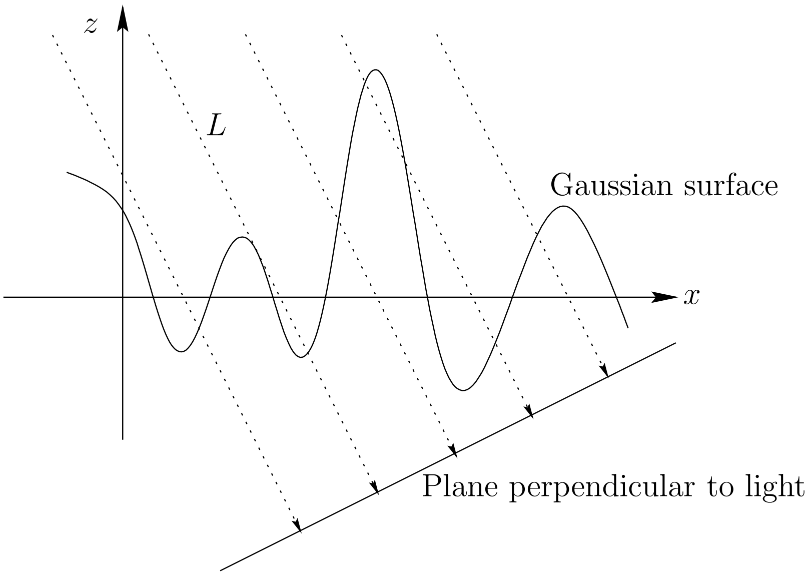

Figure 1: Projection of a height field surface onto the plane of incident light to estimate the area of visible microfacets. Note that if we ignore occlusion, the area of microfacets in the valleys of the height field may be counted even though they are not visible.

This last inequality, called the Self-Consistency Condition, is necessary for energy balance.

What is the physical significance of the Self-Consistency Condition? The integrand in the integral is the projected area of a microfacet in the direction of the incoming ray (Figure 1). The integral itself represents the total projected area of all the microfacets, and hence must be $\cos\theta_i$. Moreover, this condition must hold for all incoming direction.

The projection of a height field immediately exposes some problems. The projected area of each microfacet may be positive or negative depending on whether it points towards or away from the light. If we eliminate all the negative terms, since they are locally occluded, that does not solve the problem. Even in this case, some parts of the height field oriented towards the light will be occluded by other parts of the rough surface. Thus the calculated projected area will likely be greater than $\cos\theta_i$. The only way the projected area will equal $\cos\theta_i$ is if we only count microfacets *visible* to the light.

2.3 Self-Shadowing

We proceed by defining a self-shadowing term in terms of ρ_{dh} and scaling our BRDF by this factor to ensure that the system is self-consistent: Let

$$g(\omega_i) = \int D(\omega_f)\cos\theta' d\omega_f$$

Now define shadowing functions

$$\begin{aligned} S(\omega_i) &= \frac{\cos\theta_i}{g(\omega_i)} \\ S(\omega_r) &= \frac{\cos\theta_r}{g(\omega_r)} \end{aligned}$$

The resulting BRDF is

$$\begin{aligned} f_r &= S(\omega_i)S(\omega_r)\frac{D(\omega_f)}{4\cos\theta_i\cos\theta_r} \\ &= \frac{D(\omega_f)}{4g(\omega_i)g(\omega_r)} \end{aligned}$$

Note that in general analytic solutions for $g(\omega_i)$ and $g(\omega_r)$ may not be available, but they can be computed numerically. As a result, it is possible for any microfacet distribution to be energy normalized. This is the approach presented in the paper by Ashikhmin, Premoze and Shirley in SIGGRAPH 2000.

3 Shadowing on a Gaussian Rough Surface

Shadowing on a Gaussian rough surface is interesting for a number of reasons. First, it allows for consistent self-shadowing. That is, the self-shadowing function is consistent with the microfacet distribution function since they were both derived from the same height distribution function. Second, the resulting shadowing function may be used for the stochastic generation of texture (Illuminance Textures). Such textures, were discussed briefly in the last lecture. Finally, the derivation of the self-shadowing equations is an interesting application of statistical techniques that may have application to other rendering problems.

The discussion here focuses on the 1D case, but can be extended to 2D isotropic and anisotropic surfaces.

3.1 Random Height Field

Given the geometry in Figure 2, where the height of the surface is normally distributed, we would like to calculate the statistics of occlusion and self-shadowing. The surface is described by a height-field function $z(x)$. We consider a ray leaving the surface at $(0, z(0))$, at an angle of θ from the global surface normal. The question that we are trying to answer is: What is the probability that the ray escapes the surface without intersecting a peak as shown in Figure 2?

Define

$$s(x) = \Pr[\text{ray has not intersected the surface at x}]$$

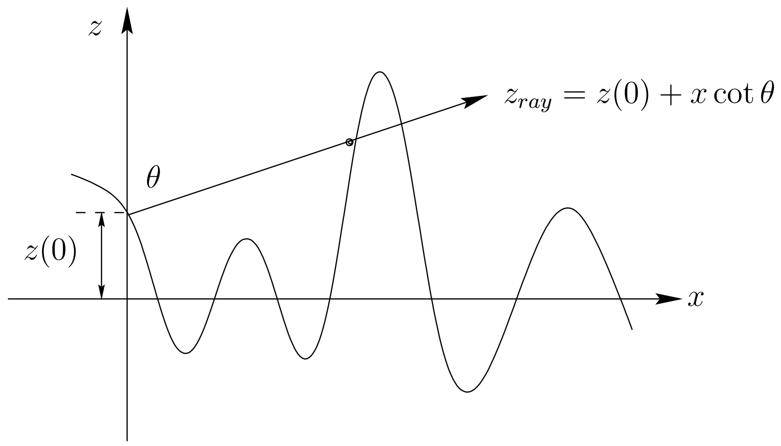

Figure 2: A ray of light leaving a Gaussian rough surface. We are interested in the probability that a ray which originates at $x = 0$ will escape the surface. Note that $z'(0) > \cot\theta$ so that the ray exits away from the surface.

We think of the probability that the ray hits the surface as an opacity, leading to a formulation of the problem as light attenuation.

$$\frac{\partial s}{\partial x} = -g(x)s(x)$$

Here $g(x)$ is the conditional probability that the ray intersects the surface between x and $x + dx$, given that it does not at x. g acts like an opacity, attenuating the probabilistic shadowing function. Integrating, we get

$$s(x) = s(0)e^{-\int_0^x g(x')dx'}$$

Here, $s(0) = H(\cot\theta - p_0)$. H is the heaviside step function, which has value 0 left of the origin and 1 right of the origin.

Next, we need to consider the possible values of $z(x)$ and $z'(x)$ at an intersection point. Clearly, we need the surface to be taller than the height of the ray: $z(x) \geq z(0) + x\cot\theta$ at an intersection. Second, we only want to consider intersections points that are visible to the ray (see Figure 3), so we need $z'(x) \geq \cot\theta$ for an intersection at x.

A technical consideration that we will mention but not elaborate on is the relationship between dz and dx. From Figure 4 we see that $dz = (z' - \cot\theta)dx$.

These considerations give

$$g(x) = p(z_{ray}) \int_{\cot\theta}^{\infty} (z' - \cot\theta)p(z')dz'$$

Our first assumption is the independence of the position and slope distributions for the surface. Let $p(z)$ be the probability that the surface is of height z, and $p(z')$ be the probability that it has slope z'. We assume that these two distributions are independent;

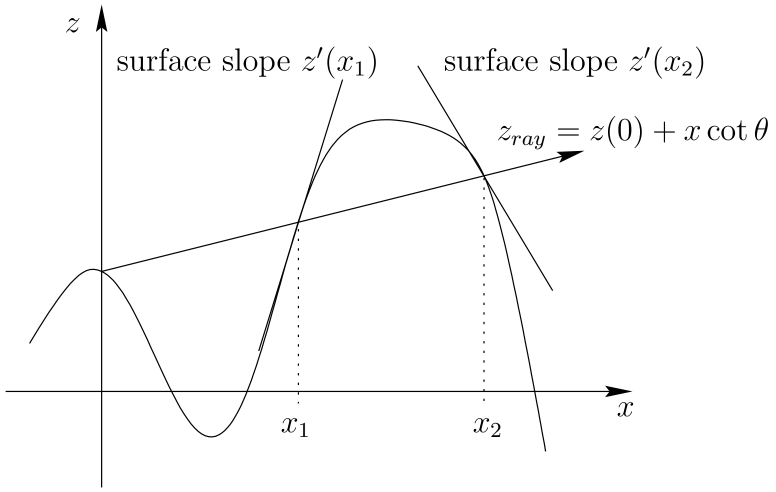

Figure 3: Details about the intersection point of a ray with the surface. Notice that the ray intersects the surface in two locations, but only the intersection on the left is visible to the ray. $z'(x)$ gives the slope of the surface at the intersection point; $\cot\theta$ is the slope of the ray. We require $z'(x) > \cot\theta$ so that the ray intersects the surface from the front.

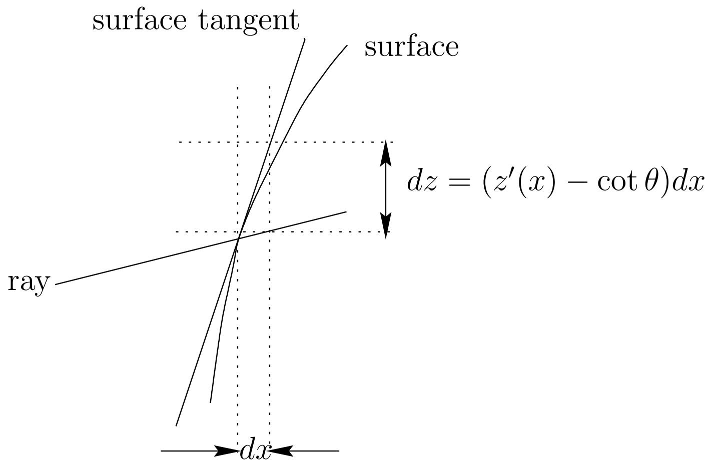

Figure 4: The relationship between dz and dx at an intersection point between the ray and the surface.

that is, $p(z, z') = p(z)p(z')$ We note that the two distributions $p(z_{ray})$ and $p(z')$ are Gaussian. Solving the integral is difficult but can be done analytically. The solution is:

$$S(\theta, z(0), z'(0)) = H(\cot\theta - z'(0)) \left[1 - \frac{1}{2} erfc\left(\frac{z(0)}{\sqrt{2}\sigma}\right)\right]^{\Lambda(\cot\theta)}$$

where $erfc(z) = \int_{-z}^{z} \frac{1}{\sqrt{\pi}} e^{-z^2} dz$

This solution makes intuitive sense because if $z(0)$ is large and positive (the ray originates high on the surface), then $1 - \frac{1}{2} erfc\left(\frac{z(0)}{\sqrt{2}\sigma}\right)$ is large, so the ray is likely to escape the surface. Conversely, if $z(0)$ is large and negative (the ray originates deep in a valley), then $1 - \frac{1}{2}erfc\left(\frac{z(0)}{\sqrt{2}\sigma}\right)$ is small and the ray is likely to intersect the surface.

By integrating over the appropriate Gaussian probability distributions, we can eliminate the dependence on z and z' to obtain:

$$S(\theta) \quad = \quad \frac{1 - \frac{1}{2}erfc\left(\frac{\cot\theta}{\sqrt{2}\sigma}\right)}{1 + \Lambda(\cot\theta)}$$

Having taken into account the statistical self-shadowing of the surface, we would expect $S(\theta)$ to be self-consistent as defined in Section 1. Indeed, Smith proves that

$$\int S(\theta)D(\omega_f)\cos\theta' d\omega_f = \cos\theta$$

3.1.1 Extensions

Here are several possible extensions to the model presented above:

Use in Raytracing
When BRDFs are used, we usually treat the surface as smooth (in this case we would use $S(\theta)$ as the self-shadowing term). However, as we saw above, we actually have more information in the explicit form for $S(z, z', \theta)$. Thus, in a raytracer when we intersect a gaussian surface, we could plug not only the incident angle of the ray into the self-shadowing expression, but also the height and slope of the surface. This technique may be applied to create statistical *illuminance textures*.

Geometry as Opacity
We saw how the statistical distribution of the surface heights and slopes lent itself to interpretation of the surface as a participating media in the calculation of self-shadowing. This interpretation may lend itself to a more general framework of geometry surfaces as participating media, in which the statistics of the geometry are used to calculate an opacity for rays traveling through the environment.

From 1D to 2D
Although the derivation above deals only with the 1D case, it can be extended to 2D. For real surfaces, the distributions of the Gaussian heights in each dimension can be uncorrelated, allowing for anisotropic surfaces.

4 Thoroughly Pitted Surfaces

We have seen that without considering self-shadowing, a micro-geometrical model may create macroscopic noticable discrepancies: e.g. the surface may not conserve light energy. Similarly, without considering inter-reflections within the microgeometry of the surface, we may not account for all the energy reflected from the surface.

Surfaces in general consist of convex and concave pieces. Locally, interreflection does not occur in convex regions, since all incoming or outgoing light is not locally occluded (although it may interact with more distant parts of the surface). Interreflection is important for concave parts of the surface, since light rays will intercept neighboring parts of the surface. Most models of interreflection assume locally concave surface features. Such interdependencies between neighboring parts of the surface greatly complicate the calculation of analytic solutions.

The first model to use concave microscale elements was the v-groove model of Oren-Nayar. In this model, rough surfaces are modeled with distributions of infinitely long v-grooves. Within a v-groove, self-shadowing is solved analytically, and interreflections involving one or two bounces of light are considered.

The "Thoroughly Pitted Surface" model of Koenderink, van Doorn, Dana and Nayar uses spherical pits for microscale features. This model improves on Oren-Nayar model in two ways. First, interreflections are solved completely using well-known radiosity results, taking advantage of the simplicity of interreflections within a sphere. Second, the surface-elements are more physically plausible since they are finite in extent rather than infinitely long as in the Oren-Nayar model.

In the Koenderink et al. paper, the spherical pits are in the shape of spherical caps, with arbitrary depth (up to the radius of the sphere). In this lecture we derive their results for the simplest case of hemispherical pits. Our general approach consists of three steps. First, we find the direct illumination within a pit. Second, we solve for interreflections. Finally, we sum the direct and indirect components to calculate the exiting radiance.

4.1 Direct Illumination

When illuminated by a distant light source, the shadow cast inside a hemispherical pit divides the pit into illuminated and shadowed regions. The dividing line between these two regions is an arc of a great circle. To see this, we first look at a cross-section of the pit in the plane of incidence. The plane of incidence contains both the surface normal and the direction to the light source. Consider the plane of incidence that passes through the center of the hemispherical pit as illustrated in the top image of Figure 5. In this hemicircular cross-section the shadowed region subtends an angle 2α and the illuminated region subtends an angle of $(2\alpha - \pi)$. Now consider any other plane parallel to the plane of incidence that intersects the hemispherical pit. The resulting cross-section will still be a hemicicle, but the radius will be smaller. However, the illuminated region still subtends the same angle, $(2\alpha - \pi)$. Therefore, the locus of points swept out by the family of cross-sectional planes will all subtend the same angle with respect to the center of the sphere. Thus all the points on the boundary between the illuminated and shadowed regions must lie on a plane through the center of the sphere at the angle $(2\alpha - \pi)$. The profile of this plane is indicated by the dotted line in the top of Figure 5. The intersection of this plane and the hemispherical pit is thus a great circle.

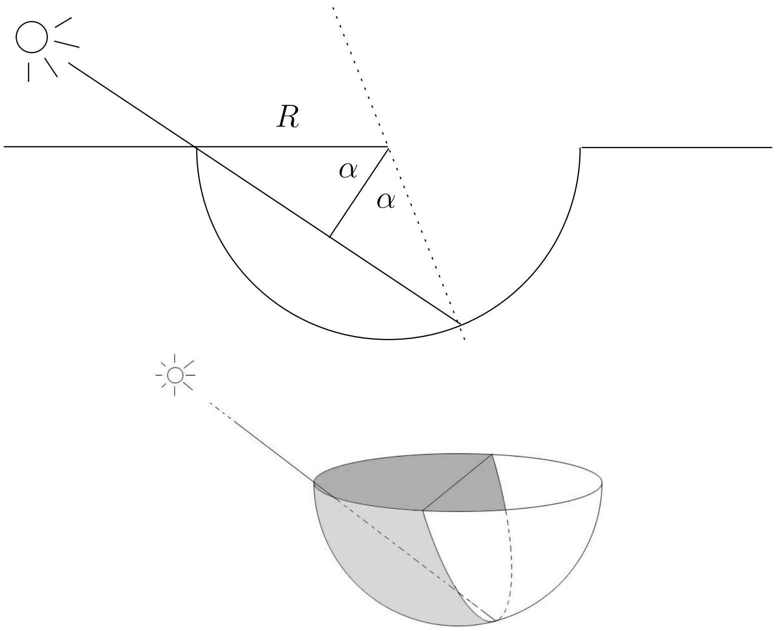

Figure 5: Shadow and illuminated regions in a hemispherical pit. A cross-section of the pit (top) in the plane of incidence. A 3D image of the pit (bottom) showing illuminated and shadowed regions.

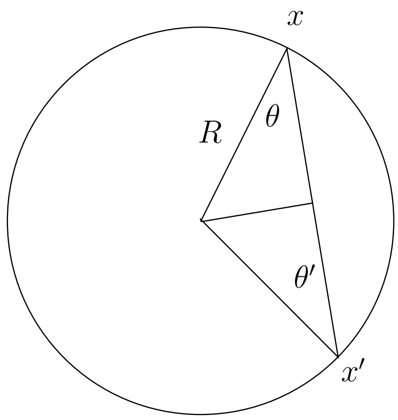

Figure 6: Geometry for calculation of the form factor between x and x' on the surface of a spherical chamber.

4.2 Indirect Illumination

We solve for the indirect illumination using the radiosity equation:

$$B(x) = B_D(x) + \rho \int B(x')G(x,x')dA'$$

First we compute the form factor between two points x and x' on a sphere. Consulting Figure 6, we see that θ and θ' are equal. Further, the distance between the x and x' is

$2R\cos\theta$. The form factor G between x and x' is given by

$$
\begin{aligned}
G(x,x') &= \frac{\cos\theta\cos\theta'}{\pi|x-x'|^2} \\
&= \frac{\cos^2\theta}{\pi(2R\cos\theta)^2} \\
&= \frac{1}{4\pi R^2}
\end{aligned}
$$

and is constant.

Since the form factor is constant, after the light bounces the energy becomes equally distributed independent of the initial distribution. Thus, we need only find the total energy coming through the circular operning of the pit to calculate the total indirect illumination. The distribution of the direct illumination is unimportant.

The power on the circular opening is $\Phi_i = \pi R^2 E_0 \cos\theta$.
The irradiance on the hemisphere is $E_D = \frac{\Phi_i}{2\pi R^2} = \frac{1}{2}E_0\cos\theta$.
And the average radiosity is $B_D = \rho E_D = \frac{\rho}{2}E_0\cos\theta$.

Returning to the radiosity equation for the hemispherical pit, after the first bounce we can drop the dependence of radiance on position

$$
\begin{aligned}
B &= B_D + \frac{\rho B}{4\pi R^2}\int dA' \\
&= B_D + \frac{\rho B}{2} \\
&= \frac{B_D}{1-\frac{\rho}{2}}
\end{aligned}
$$

We can thus solve for the indirect component.

$$
\begin{aligned}
B_I &= B - B_D \\
&= \frac{\rho}{2}\frac{(1-(1-\rho/2))}{(1-\rho/2)}E_0\cos\theta \\
&= \frac{(\rho/2)^2}{1-\rho/2}E_0\cos\theta
\end{aligned}
$$

4.3 Exiting Radiance

The exiting radiance is the average radiance leaving the cavity. To calculate this value we sum the contributions from two areas: the visible and directly irradiated area of the pit; and the indirectly illuminated, yet visible area. This is equivalent to summing the indirect diffuse irradiation that is visible and the direct irradiation that is visible.

The indirect component is easy, since we know it is uniformly distributed inside the hemispherical cavity. Since the inside surface is diffuse, the indirect radiance on the inside surface of the pit is also constant. No matter from which direction we look at the pit, we will see a circular opening with a constant indirect radiance.

$$L_o = \frac{B_I}{\pi}$$

The direct illumination component is more difficult. However, if the viewing direction is in the plane of incidence analytic solutions may be derived. Four cases are distinguished: light scattered forward, with illuminated region visible; light scattered forward, with only shadow region visible; light scattered backwards with shadow region visible; and finally, light scattered backwards with only illuminated region visible. The component of the exiting radiance due to direct illumination is difficult to calculate because it depends on a complex region of integration: the intersection of the region that is visible to the light and the region that is visible to the eye. When the viewing direction is out of the plane of incidence, the calculations become significantly even more complicated. Solutions in terms of elementary functions are not known, but values can be computed numerically to arbitrary precision.

One interesting note is the prediction of a retroreflection lobe in the BRDF. Geometrically, this results because most of the bright illuminated region is visible to the eye when it views the pit in almost the same direction as the light vector.

The geometry of this intersection region – the bright, visible region of the pit – is characteristic of the viewing direction. These regions are called textons, and could be the basis of a vocabulary for surface textures (see Figure 7). The lay out of the Figure is as follows: Each panel is a picture of the surface from the viewing direction β (co-elevation), ϵ (azimuth) and directino of incidence α (co-elevation). The subpanels correspond to the triples (α, β, ϵ) (the value of ϵ is omitted when irrelevant):

$(0°, 0°)$	$(0°, 30°)$	$(0°, 60°)$
$(30°, 0°)$	$(30°, 30°, 0°)$	$(30°, 60°, 0°)$
	$(30°, 30°, 180°)$	$(30°, 60°, 180°)$
$(60°, 0°)$	$(60°, 30°, 0°)$	$(60°, 60°, 0°)$
	$((60°, 30°, 180°)$	$(60°, 60°, 180°)$

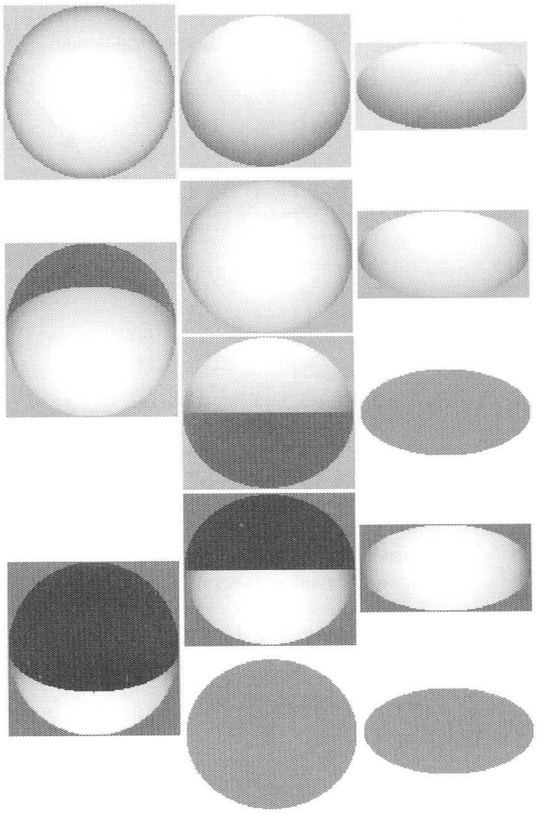

Figure 7: "Textons" for the hemispherical pit, that are views of the pit for various viewing and source geometries. For illustrative purposes the maximum radiance is represented by the paper white in each picture. The layout of the panel is explained in the text. Source: Koenderink, et al.*Bidirectional Reflection Distribution Funtion of Thoroughly Pitted Surface.* IJCV, April 1999.