Diffuse and specular reflectance from rough surfaces

Bram van Ginneken, Marigo Stavridi, and Jan J. Koenderink

We present a reflection model for isotropic rough surfaces that have both specular and diffuse components. The surface is assumed to have a normal distribution of heights. Parameters of the model are the surface roughness given by the rms slope, the albedo, and the balance between diffuse and specular reflection. The effect of roughness on diffuse reflection is taken into account, instead of our modeling this component as a constant Lambertian term. The model includes geometrical effects such as masking and shadowing. The model is compared with experimental data obtained from goniophotometric measurements on samples of tiles and bricks. The model fits well to samples with very different reflection properties. Measurements of the sample profiles performed with a laser profilometer to determine the rms slope show that the assumed surface model is realistic. The model could therefore be used in machine vision and computer graphics to approximate reflection characteristics of surfaces. It could also be used to predict the texture of surfaces as a function of illumination and viewing angles.

OCIS codes: 120.5700, 240.5700, 240.0240.

1. Introduction

Reflection models have many applications, for instance, in the generation of realistic images of scenes (computer graphics) and for extraction of information such as shape or physical properties of materials from photographs (computational vision, photometric stereo) or from satellite images (remote sensing, astrophysics). Our model is based on the view that two key factors determine the reflection properties of materials: the reflection properties of points on the surface relative to their local surface normal, and the macroscopic structure of the surface (roughness).

Regarding the first factor, local reflection properties, two limiting cases arise. The light can be reflected in the perfect specular direction or it can be scattered diffusely, uniformly in all directions of the hemisphere, independent of the incident direction. Specular reflection predominates in polished metallic surfaces, whereas materials such as chalk or paper reflect nearly all incident light diffusely. In most surfaces, however, specular and diffuse reflection coexist.

The second factor, macroscopic surface roughness, implies local deviations from the mean surface normal. This has a marked effect on both the specular and the diffuse components. In the case of specular reflection a smooth surface acts as a mirror, reflecting all incident light in the specular direction, whereas a rough surface reflects the light into a lobe around the specular direction. With increasing roughness the visual appearance of the surface will change from shiny to glossy. In the case of diffuse reflection a surface with no macroscopic roughness will appear equally bright from all directions, as was first observed by Lambert.[1] For a rough surface this is not the case. Parts of the surface with normals oriented toward the illumination receive more light, since these parts are illuminated head-on. Likewise, a sensor receives more light from surface parts oriented toward it. Therefore the surface will appear brighter as the direction of observation approaches the illumination direction.

The effect of roughness on the specular component is incorporated into many reflection models. Of these, the Torrance–Sparrow model[2] is among the most popular in the computational vision and computer graphics community; the surface is modeled as a collection of specular V cavities whose surface normals are assumed to be normally distributed. The calculation of reflected radiation is based on geometrical optics. Geometrical optics is applicable when the surface irregularities are much larger than the wavelength of incident radiation. Nayar *et al.*[3] showed that in such a case the Torrance–Sparrow model approximates the physical optics model developed by Beckmann and Spizzichino.[4] A number of

The authors are with the Helmholtz Institute, Utrecht University, Princetonplein 5, Utrecht, The Netherlands.

Received 2 December 1996; revised manuscript received 9 July 1997.

0003-6935/98/010130-10$10.00/0

alternative physical-optics approaches to the problem of specular reflection from rough surfaces are found in the literature. Barrick[5] obtained an expression for the average number of specular points on a Gaussian surface, the same surface model we use in this paper. Leader[6] used a similar model for surface roughness to derive an elaborate reflection model. He *et al.*[7] combined results from several specular reflection models.

But all these models ignore the effect of roughness on the diffuse component. This term is omitted or treated as a constant Lambertian term. However, this effect is incorporated into a model recently put forward by Oren and Nayar.[8] Their model is based on the same surface model as the Torrance–Sparrow model but treats the V cavities as diffuse reflectors. Both models take the geometrical effects of masking and shadowing into account. To permit this, the V cavities have to be very long (so that effects at the end of the cavities can be ignored), and surface isotropy requires the cavities to have no preferred direction. Obviously, such surfaces cannot exist, because the geometry is inconsistent. This is a major drawback of both the Torrance–Sparrow model and the Oren–Nayar model.

In this paper we develop a model that can be used to predict reflectance from macroscopic rough surfaces that have both a specular and a diffuse reflection component. We use the Gaussian surface model,[9] a realistic statistical approach to model roughness, that has been widely used in theoretical work before. The distribution of heights is assumed to be normal. The standard deviation of heights and the autocorrelation function determine the rms slope of the surface, which is a parameter for surface roughness. We derive radiance expressions as a function of illumination and viewing angles for both specular and diffuse local reflection. The use of geometrical optics enables us to treat both these components on an equal footing. Similar expressions for the specular reflection component from a Gaussian surface have been presented before.[5–7,10] But the effect on the diffuse component, which can be seen as a generalization of Lambertian reflection for Gaussian surface roughness, has not been investigated. To take shadowing and masking into account, we extend the shadowing theory of Smith[11] and propose a new expression for a general bidirectional shadowing function.

We start with a derivation of the model. Then we fit the present model and a combination of the Torrance–Sparrow and Oren–Nayar models to experimental data from several samples. Measurements of the sample profiles are used to determine the roughness and to check the accuracy of the surface model. It is shown that the distribution of heights of our samples is normal and that the shadowing–masking function is accurate. Therefore it is argued that our theory may also be used to predict the texture of surfaces as a function of illumination and viewing angles.

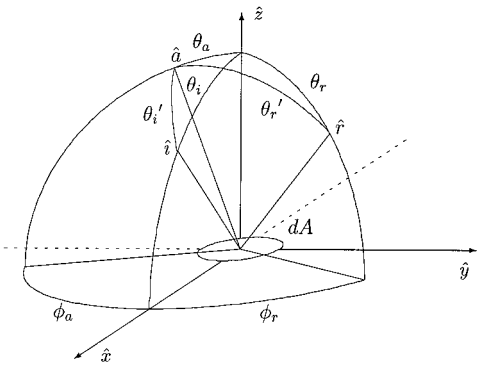

Fig. 1. Coordinate system.

2. Derivation of the Model

Figure 1 shows the coordinate system used in this derivation. We consider reflection from a surface patch dA whose mean surface normal coincides with the z axis. The surface is assumed to be isotropic, and therefore we can choose to locate the direction of illumination $\hat{\imath}$ in the xz plane and specify it with θ_i. The viewing direction $\hat{r}$ is given by (θ_r, ϕ_r), where ϕ_r can take values between 0 and π (owing to surface isotropy, negative values of ϕ_r are equivalent to their absolute values). Because the surface is rough, the local surface normal $\hat{a} = (\theta_a, \phi_a) \neq \hat{z}$. The local angles of incidence and reflection, θ_i' and θ_r', respectively, are determined by

$$\cos\theta_i' = (\hat{\imath}, \hat{a}) = \cos\phi_a \sin\theta_i \sin\theta_a + \cos\theta_i \cos\theta_a, \quad (1)$$

$$\cos\theta_r' = (\hat{r}, \hat{a}) = \cos(\phi_a - \phi_r)\sin\theta_r \sin\theta_a + \cos\theta_r \cos\theta_a. \quad (2)$$

A. Modeling Surface Roughness

We assume that the distribution of the heights $P_z\mathrm{d}z$ of the points on the surface is normal. Since we have chosen the mean height to be $z = 0$, we have

$$P_z(z, \sigma)\mathrm{d}z = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-z^2}{2\sigma^2}\right)\mathrm{d}z, \quad (3)$$

where σ is the standard deviation of height. This distribution specifies only the vertical scale of the relief. The horizontal scale of the relief is contained within the autocorrelation function $R(\tau)$, defined as the average product of the heights of two points at a distance τ. We assume the surface to be isotropic, so the value of the autocorrelation function does not depend on the direction of τ, and the distribution of ϕ_a is uniform:

$$P_{\phi_a}\mathrm{d}\phi_a = \frac{1}{2\pi}\mathrm{d}\phi_a. \quad (4)$$

The autocorrelation function and the distribution of heights [Eq. (3)] completely specify the surface and can be used to derive other distributions needed in the subsequent derivation (Chaps. 7 and 8 of Ref. 12, and Ref. 13). First we need the distribution of slopes in a given direction. We write the derivatives of

height in the x and y directions as $z_x = \partial z/\partial x$ and $z_y = \partial z/\partial y$. For z_x we have

$$P_{z_x}(z_x, r)\mathrm{d}z_x = \frac{1}{\sqrt{2\pi}r}\exp\left(\frac{-z_x^2}{2r^2}\right)\mathrm{d}z_x, \tag{5}$$

where $r = [-R''(0)]^{1/2}$, the rms slope of the surface. The distribution $P_{z_y}\mathrm{d}z_y$ is similar. The distributions of z, z_x, and z_y are all statistically independent and Gaussian. The rms slope r is an indicator for the roughness of the surface and is used as a parameter in the model. Note that by slope we mean the derivative of height in a given direction (e.g. the x axis) and not the slope in the direction of maximum increase in height. That slope, indicated as z_m, is given by $[z_x^2 + z_y^2]^{1/2}$. It can be calculated from the distributions $P_{z_x}\mathrm{d}z_x$ and $P_{z_y}\mathrm{d}z_y$ and is given by

$$P_{z_m}(z_m, r)\mathrm{d}z_m = \frac{z_m}{r^2}\exp\left(\frac{-z_m^2}{2r^2}\right)\mathrm{d}z_m. \tag{6}$$

Note that $P_{z_m}\mathrm{d}z_m$ is not a normal distribution, and its standard deviation is $\sqrt{2}r$. Now, using $\tan\theta_a = z_m$, we obtain the distribution of θ_a:

$$P_{\theta_a}(\theta_a, r)\mathrm{d}\theta_a = \frac{\sin\theta_a}{r^2\cos^3\theta_a}\exp\left(\frac{-\tan^2\theta_a}{2r^2}\right)\mathrm{d}\theta_a. \tag{7}$$

Finally, we require an expression for the probability $P_{\mathrm{d}\omega_a}\mathrm{d}\omega_a$ that a surface normal (θ, ϕ) lies within a solid angle $\mathrm{d}\omega_a$ with its center at (θ_a, ϕ_a). Assuming that $\mathrm{d}\omega_a$ is small, we may approximate this by the probability that θ lies in a small interval around θ_a, and ϕ in a small interval around ϕ_a. The size of the allowed interval for θ is proportional to the diameter of $\mathrm{d}\omega_a$. The size of the allowed interval around ϕ is proportional to the diameter of $\mathrm{d}\omega_a$ divided by the circumference of the circle that is obtained by our taking the projection of (θ_a, ϕ_a) on the xy plane for fixed value of θ_a and letting ϕ_a go from 0 to 2π. The circumference of this circle is $2\pi\sin\theta_a$. Thus we can express the probability that a surface normal lies within $\mathrm{d}\omega_a$ as

$$P_{\mathrm{d}\omega_a}(\theta_a, r)\mathrm{d}\omega_a = \frac{1}{\sqrt{\pi}U(-1/2, 0, 1/(2r^2))r^2\cos^3\theta_a} \times \exp\left(\frac{-\tan^2\theta_a}{2r^2}\right)\mathrm{d}\omega_a, \tag{8}$$

where the term U in the normalization constant is the confluent hypergeometric function $U(a, b, z)$.

B. Masking and Shadowing

On any rough surface it is likely that some points will not receive light (shadowing) and some will not be visible (masking). Only those parts of the surface that are both illuminated and visible contribute directly to the reflected flux. We can incorporate this effect into the reflection model by deriving an expression for the bistatic shadowing function, i.e., the probability that a point on the surface is both visible and illuminated. We have not been able to find such an expression in the literature, although the problem is of considerable importance in many applications.

A rigorous derivation of the bistatic shadowing function of a surface with the features of the surface model described above leads to very complicated expressions, and therefore some reasonable approximations must be made. In fact, approximations are needed even for the case of shadowing or masking alone, and several researchers[11,14,15] have used them. We follow Smith[11,16] because we found that his approach agrees very well with computer simulations of the actual bistatic shadowing function. Others[17,18] also performed simulations and arrived at similar conclusions. We extend Smith's shadowing theory to the bistatic case.

First we discern two types of shadowing and masking, which we treat separately: (1) self-shadowing and self-masking, which means that the point is turned away from the illumination ray and viewing ray, respectively; and (2) shadowing and masking by intersection, which occurs when the illumination or viewing ray cannot reach a point because it is intersected by the surface. These types are quite distinct, because the former is local (and thus tractable), whereas the latter is global.

1. *Self-Shadowing and Self-Masking*

Self-shadowing occurs when the slope of the point, z_x, is larger than the (always positive) slope of the illumination ray, $\cot\theta_i$. Using $z_x = \cos\phi_a\tan\theta_a$, we can express the fact that a point is not self-shadowed in terms of a condition on ϕ_a:

$$\begin{aligned} &\text{if} \quad \theta_a \le \frac{\pi}{2} - \theta_i, \quad \text{then} \quad -\pi < \phi_a \le \pi, \\ &\qquad\qquad\quad \text{else} \quad a < \phi_a < b, \\ &\qquad\qquad\quad \text{where} \quad a = -\pi + \arccos(\cot\theta_i\cot\theta_a), \\ &\qquad\qquad\qquad\quad\;\; b = \pi - \arccos(\cot\theta_i\cot\theta_a). \end{aligned} \tag{9}$$

The slope in the viewing direction is given by $\cos(\phi_a - \phi_r)\tan\theta_a$. By requiring that the point not be self-masked, we obtain another condition on ϕ_a:

$$\begin{aligned} &\text{if} \quad \theta_a \le \frac{\pi}{2} - \theta_r, \quad \text{then} \quad -\pi < \phi_a \le \pi, \\ &\qquad\qquad\quad \text{else} \quad a < \phi_a < b, \\ &\qquad\qquad\quad \text{where} \quad a = \phi_r - \pi + \arccos(\cot\theta_r\cot\theta_a), \\ &\qquad\qquad\qquad\quad\;\; b = \phi_r + \pi - \arccos(\cot\theta_r\cot\theta_a). \end{aligned} \tag{10}$$

The allowed range of ϕ_a in the bistatic case can be found by our combining both conditions. Note that the range need not be connected.

2. *Shadowing and Masking by Intersection*

The next step is to derive an approximation for the probability that a point is not shadowed and not

masked by intersection, given that it is not self-shadowed and not self-masked. We indicate this probability by $P_{\text{ill+vis}}$. We have

$$P_{\text{ill+vis}} = P_{\text{ill}} P_{\text{vis|ill}}, \tag{11}$$

where P_{ill} is the probability that the point is illuminated. $P_{\text{vis|ill}}$ is the probability that the point is visible given that it is illuminated, conditional on the point's being neither self-shadowed nor self-masked.

An expression for P_{ill} is derived by Smith.[11,16] From this derivation we repeat some steps that we need for the derivation of $P_{\text{vis|ill}}$ and, subsequently, $P_{\text{ill+vis}}$. (For details see Ref. 11 or 16.) P_{ill} can be written as

$$P_{\text{ill}} = \exp\left[-\int_0^\infty g(\tau)\mathrm{d}\tau\right], \tag{12}$$

where $g(\tau)\mathrm{d}\tau$ is the probability that the surface intersects the illumination ray in $x \in [\tau, \tau + \mathrm{d}\tau]$ given that it does not do so in $x \in [0, \tau]$. Smith approximates this with the probability that the surface intersects the illumination ray in $x \in [\tau, \tau + \mathrm{d}\tau]$ given that it does not at $x = \tau$. Thus he finds for $g(\tau)$

$$g(\tau) = \frac{\displaystyle\int_{\cot\theta_i}^\infty \mathrm{d}z_x'(z_x' - \cot\theta_i)[P_\tau]_{z'=z+\tau\cot\theta_i}}{\displaystyle\int_{-\infty}^\infty \mathrm{d}z_x' \int_{-\infty}^{z+\tau\cot\theta_i} \mathrm{d}z' P_\tau}, \tag{13}$$

where P_τ is the joint distribution of the height z' and slope z_x' at $x = \tau$, conditional on z and z_x at $x = 0$. But Smith ignores the correlation between height and slope at $x = \tau$ and $x = 0$. As a result, P_{ill} does not depend on the orientation (θ_a, ϕ_a) and the integrations in Eqs. (13) and (12) can be performed analytically, leading to

$$P_{\text{ill}}(\theta_i, r, z, \sigma)\mathrm{d}z = \left[1 - \frac{1}{2}\operatorname{erfc}\left(\frac{z}{\sqrt{2}\sigma}\right)\right]^{\Lambda(r,\,\theta_i)} \mathrm{d}z, \tag{14}$$

where erfc is the error function complement and

$$\Lambda(r, \theta_i) = \frac{r}{\sqrt{2\pi}\cot|\theta_i|} \exp\left(\frac{-\cot^2\theta_i}{2r^2}\right) - \frac{1}{2}\operatorname{erfc}\left(\frac{\cot|\theta_i|}{\sqrt{2}r}\right). \tag{15}$$

By our multiplying $P_{\text{ill}}\mathrm{d}z$ by $P_z\mathrm{d}z$ and integrating over all heights, the expression no longer depends on z and σ, leading to

$$P_{\text{ill}}(\theta_i, r) = \int_{-\infty}^\infty P_{\text{ill}}(\theta_i, r, z, \sigma) P_z(z, \sigma)\mathrm{d}z = \frac{1}{1 + \Lambda(r, \theta_i)}. \tag{16}$$

The corresponding expression for masking, indicated as $P_{\text{vis}}(\theta_r, r)$, is obtained by our replacing θ_i with θ_r.

To derive $P_{\text{vis|ill}}$, we need to find an appropriate form for P_τ in Eq. (12). In this case P_τ is the joint probability of height and slope at $(x, y) = (\tau\cos\phi_r, \tau\sin\phi_r)$ conditional on z, z_x, and on the fact that the point is illuminated; that is, all points on the positive x axis lie below the illumination ray. For small values of ϕ_r it is clearly not possible to ignore this correlation. Consider the extreme case $\phi_r = 0$, when both the illumination and the viewing ray come from the positive x axis. Now there is full correlation between each point along the viewing direction and the illumination direction at τ. If $\theta_i > \theta_r$, we find that $g = 0$, and therefore $P_{\text{vis|ill}} = 1$, which means that an illuminated point is certain to be visible. The argument can be reversed for $\theta_r > \theta_i$, so we get

$$P_{\text{ill+vis}}(\theta_i, \theta_r, \phi_r = 0, r) = \frac{1}{1 + \Lambda(r, \max(\theta_i, \theta_r))}. \tag{17}$$

On the other hand, for large ϕ_r the correlation will be small. If we ignore the correlation, we can evaluate the integrations in Eqs. (13) and (12) and arrive at

$$P_{\text{ill+vis}}(\theta_i, \theta_r, \phi_r \to \pi, r) \to \frac{1}{1 + \Lambda(r, \theta_i) + \Lambda(r, \theta_r)}. \tag{18}$$

Relations (17) and (18) provide an upper and lower limit for the actual value of $P_{\text{ill+vis}}$. So far all these results can be found in the literature. Equation (17) was derived by Smith,[11] and relation (18) has been used in many instances as an approximation of the bistatic shadowing function for $\phi_r > 0$. The need for a better approximation was pointed out by Sancer,[10] and we derive an expression below.

Using the full expression for P_τ is cumbersome, since we would need to include the correlation between the point at τ and the given upper limit on the height for the infinity of points on the positive x axis. We choose to approximate this by the correlation between the height at τ in the viewing direction and the upper limit $z + \tau\cot\theta_i$ for the height z' at τ in the illumination direction. Like Smith, we ignore the correlations between slopes. This makes the integrations over z_x' in Eq. (13) easy. We find the distribution of heights at τ by integrating the bivariate distribution[12] from $-\infty$ to $z + \tau\cot\theta_i$:

$$P_\tau = C_\tau \exp\left(\frac{-z'^2}{2}\right)\left\{1 + \operatorname{erf}\left(\frac{z + \tau\cot\theta_i - \rho z}{[\sqrt{2(1-\rho^2)}]^{1/2}}\right)\right\}, \tag{19}$$

$$\rho = R(d) = \exp\left(\frac{-d^2}{T^2}\right), \tag{20}$$

$$d = 2\tau\sin\frac{\phi_r}{2}, \tag{21}$$

where we have assumed a Gaussian autocorrelation function. The normalization constant C_τ cancels when P_τ is inserted into Eq. (13).

We lose dependancy on z' by integrating over the

distribution of heights of illuminated points $P_{z|\mathrm{ill}}\mathrm{d}z$, which follows from Eqs. (14) and (16):

$$P_{z|\mathrm{ill}}(\theta_i, z, \sigma, r)\mathrm{d}z = [1 + \Lambda(r, \theta_i)] \times \left(1 - \frac{1}{2}\,\mathrm{erfc}\,\frac{z}{\sqrt{2}\sigma}\right)^{\Lambda(r,\,\theta_i)} P_z\mathrm{d}z. \quad (22)$$

To construct a convenient expression for $P_{\mathrm{ill+vis}}$, we calculated the integrations described above numerically for a large set of values $(\theta_i, \theta_r, \phi_r, r)$ for $0 < r < 1$ and determined an approximation in the min–max sense of the form

$$P_{\mathrm{ill+vis}}(\theta_i, \theta_r, \phi_r, r) \rightarrow \frac{1}{1 + \Lambda[r, \max(\theta_i, \theta_r)] + \alpha\Lambda[r, \min(\theta_i, \theta_r)]}. \quad (23)$$

This form was chosen because it automatically obeys the upper and lower limits given in Eqs. (17) and (18), provided that $\alpha = 0$ when $\phi_r = 0$ and $\alpha \rightarrow 1$ for increasing ϕ_r. We found that setting

$$\alpha = \frac{4.41\phi_r}{4.41\phi_r + 1}, \quad (24)$$

yields an error in $P_{\mathrm{ill+vis}}$ that never exceeds 0.03 (for $r < 1$) and is usually much smaller. Using more complicated expressions for α, in which not only ϕ_r but also θ_i, θ_r, and r appear, led to only minor improvements. Thus we suggest that relations (23) and (24) are a convenient and accurate approximation of the actual bistatic shadowing function of a Gaussian surface. We specifically mention this result, since it is of interest by itself and may prove useful in many other applications.

C. Radiance due to Specular Reflection

A point on the surface dA will reflect light specularly from source to detector only if its normal lies within the solid angle $d\omega_a$ whose center (θ_a, ϕ_a) is located in the plane formed by $\hat{\imath}$ and $\hat{r}$ at the position where $\theta_i' = \theta_r'$. From the geometry we find that θ_a (now indicated as $\theta_{a\ \mathrm{spec}}$) is given by

$$\theta_{a\ \mathrm{spec}} = \arccos\{(\cos\theta_i + \cos\theta_r)[(\cos\phi_r \sin\theta_r + \sin\theta_i)^2 + \sin^2\phi_r \sin^2\theta_r + (\cos\theta_i + \cos\theta_r)^2]^{-1/2}\}. \quad (25)$$

The size of $\mathrm{d}\omega_a$ is related to the size of $\mathrm{d}\omega_r$:

$$\mathrm{d}\omega_a = \frac{\mathrm{d}\omega_r}{4\cos\theta_i'}. \quad (26)$$

This relation follows from basic geometric considerations, assuming that the incident direction is perfectly parallel and that the surface is not perfectly smooth, so that the specularly reflected rays are scattered in a cone around the specular direction. A derivation can be found in App. D of Nayar *et al.*[3]

The flux reflected by any point is proportional to its irradiance. The irradiance of a point that is illuminated head-on is denoted E_0. Then the irradiance of a point with normal $\hat{a}$ is equal to $E_0 \cos\theta_i'$. The radiance L_r of the surface is defined as the flux reflected by the surface per unit solid angle per unit foreshortened area.[19]

Thus the contribution of a specularly reflecting part of the radiance can be expressed as

$$L_{rs}(\theta_i, \theta_r, \phi_r) = \frac{E_0 \cos\theta_i'}{(\cos\theta_a)\mathrm{d}A\ (\cos\theta_r)\mathrm{d}\omega_r}. \quad (27)$$

The factor $\cos\theta_a$ is included because the projected area of a small part (θ_a, ϕ_a) of the surface is its total area over $\cos\theta_a$. The total projected area of specularly reflecting parts is given by $P_{d\omega_a}(\theta_{a\ \mathrm{spec}}, r)\mathrm{d}\omega_a\mathrm{d}A$. Inserting this into Eq. (27) and using Eqs. (26) and (8), we obtain the following expression for the total radiance due to specular reflection:

$$L_{rs}(\theta_i, \theta_r, \phi_r, r) = \frac{C_s P_{\mathrm{ill+vis}}(\theta_i, \theta_r, \phi_r, r)}{\cos\theta_r \cos^4\theta_{a\ \mathrm{spec}}} \times \exp\left(\frac{-\tan^2\theta_{a\ \mathrm{spec}}}{2r^2}\right), \quad (28)$$

where $\theta_{a\ \mathrm{spec}}$ is given by Eq. (25). C_s is given by

$$C_s = \frac{E_0}{4\sqrt{\pi}U(-1/2, 0, 1/(2r^2))}. \quad (29)$$

The factor $P_{\mathrm{ill+vis}}(\theta_i, \theta_r, \phi_r, r)$ accounts for masking and shadowing by intersection. It can easily be seen that specularly reflecting points are never self-shadowed or self-masked.

D. Radiance due to Diffuse Reflection

In the case of diffuse reflection any point that is illuminated and visible contributes to the reflected flux. The relative contribution of each point is given by Lambert's model[1] and is $(\rho/\pi)E_0 \cos\theta_i' \cos\theta_r'$. Here, again, $E_0 \cos\theta_i'$ is the irradiance of the point, and ρ is the albedo, a constant between 0 and 1. Thus the contribution of the diffusely reflecting parts to the radiance can be expressed as

$$L_{rd}(\theta_i, \theta_r, \phi_r, \theta_a, \phi_a) = \frac{\rho}{\pi} E_0 \frac{\cos\theta_i' \cos\theta_r'}{(\cos\theta_a)\mathrm{d}A \cos\theta_r}, \quad (30)$$

where θ_i' and θ_r' are given by Eqs. (1) and (2). Again, the factor $\cos\theta_a$ in the denominator is included because we consider the projected area of each diffusely reflecting part. The total radiance that is due to diffuse reflection, corrected for masking and shadowing, is now given by integration over dA

$$L_{rd}(\theta_i, \theta_r, \phi_r, r) = P_{\mathrm{vis+ill}}(\theta_i, \theta_r, \phi_r, r) \times \int_0^{\pi/2} \left[\int_a^b L_{rd}(\theta_i, \theta_r, \phi_r, \theta_a, \phi_a) P_{\phi_a}\mathrm{d}\phi_a\right] P_{\theta_a}(\theta_a, r)\mathrm{d}\theta_a. \quad (31)$$

The integration over ϕ_a yields

$$\int_a^b L_{rd}(\theta_i, \theta_r, \phi_r, \theta_a, \phi_a) P_{\phi_a} \mathrm{d}\phi_a = C_d \left[\left(\frac{c_1}{2} + c_5 \right)(b - a) + \frac{c_1}{4}(\sin 2b - \sin 2a) + \frac{c_2}{4}(\cos 2a - \cos 2b) + c_3(\sin b - \sin a) + c_4(\cos a - \cos b) \right], \quad (32)$$

where

$$C_d = \frac{\rho E_0}{2\pi^2 \cos \theta_r \cos \theta_a},$$

$$c_1 = \sin \theta_i \sin^2 \theta_a \cos \phi_r \sin \theta_r,$$

$$c_2 = \sin \theta_i \sin^2 \theta_a \sin \phi_r \sin \theta_r,$$

$$c_3 = (\sin \theta_a \cos \theta_a)(\sin \theta_i \cos \theta_r + \cos \theta_i \cos \phi_r \sin \theta_r),$$

$$c_4 = \cos \theta_i \cos \theta_a \sin \phi_r \sin \theta_r \sin \theta_a,$$

$$c_5 = \cos \theta_i \cos \theta_r \cos^2 \theta_a, \quad (33)$$

and a and b are the lower and the upper limits for the range of allowed values of ϕ_a for which points are not self-shadowed and not self-masked.

The resulting integral over θ_a could not be evaluated analytically, and thus we calculated it numerically. Note that for $r = 0$ we have, by definition, $\theta_a = 0$, and L_{rd} reduces to the Lambertian model.

E. Total Radiance

We can write the total radiance as the sum of the contributions of specular reflection and diffuse reflection:

$$L_r(\theta_i, \theta_r, \phi_r, r, g, C) = C[gL_{rs}(\theta_i, \theta_r, \phi_r, r) + (1 - g)L_{rd}(\theta_i, \theta_r, \phi_r, r)] \quad (34)$$

The model contains three parameters. The rms slope r is a measure of the roughness of the sample. The parameter g has a value between 0 and 1 and indicates the balance between diffuse and specular reflection at the surface. For $g = 0$ there is only a diffuse reflection component; for $g = 1$ there is only a specular reflection component. The parameter C is proportional to the incident light flux and the overall albedo. The constants E_0 and ρ are absorbed in C and are effectively set to 1. If E_0 is known, the bidirectional reflection distribution function, defined as the ratio of radiance to irradiance, can be evaluated. In that expression C represents the overall albedo.

The assumption that g does not depend on θ_i is a good approximation for some surfaces, but not if the surface is a dielectric that is smooth at wavelength scale. In that case we can incorporate the Fresnel reflection coefficient into the model by writing the total radiance as

$$L_r(\theta_i, \theta_r, \phi_r, r, C, n, \rho) = C\{F(n, \theta_i) \times L_{rs}(\theta_i, \theta_r, \phi_r, r) + [1 - F(n, \theta_i)] \times L_{rd}(\theta_i, \theta_r, \phi_r, r, \rho)\}. \quad (35)$$

For unpolarized light the Fresnel coefficient F as a function of θ_i and the index of refraction n are given in Ref. 20 as

$$F(n, \theta_i) = \frac{1}{2}\left[\frac{\sin^2(\theta_i - \theta_t)}{\sin^2(\theta_i + \theta_t)} + \frac{\tan^2(\theta_i - \theta_t)}{\tan^2(\theta_i + \theta_t)} \right],$$

$$\text{with} \quad \theta_t = \arcsin\left(\frac{\sin \theta_i}{n} \right). \quad (36)$$

The albedo ρ is now also a parameter. Therefore C in Eq. (35) is not directly comparable with C in Eq. (34).

3. Experiments

To test the performance of the model, we did goniophotometric measurements on samples of real materials: bricks, concrete, and ceramic tiles. The samples, measuring approximately 5 cm × 5 cm, were placed in a sample holder that could be oriented automatically in any direction. Samples were illuminated with a 150-W halogen lamp, mounted at the end of an 80-cm-long bar. A mirror was mounted behind the lamp to increase the light intensity, and a lens system was installed in front of it to yield a parallel beam. Radiance was measured with a spectroradiometer (PR-704/PR-714, Photo Research) positioned at the end of another bar. Both bars were attached to a rotary stage and could be moved manually. This enabled us to measure radiance for any combination (θ_i, θ_r, ϕ_r). Radiance was measured from an area of 2 mm × 2 mm (when $\theta_r = 0$). Because the detector had a fixed aperture, the area seen changed proportionally to $1/\cos \theta_r$. All measurements were repeated for five different spots on the sample and averaged.

In Figs. 2–5 we give the results for four typical samples: a rather smooth colored tile, a rougher tile, a grey concrete brick, and a red brick that was thinly painted with matte white latex. We show experimental data and model fits. All measurements were made in the plane of incidence (negative values of θ_r correspond to $\phi_r = \pi$). Radiance is shown as a function of θ_r for various values of θ_i. The roughness values (r parameter) were calculated from surface profiles of the samples, measured with a laser profilometer (Perthometer S5P, Mahr GmbH, with a resolution of 0.5 μm), at the Foundation for Research and Technology–Hellas, Institute of Electronic Structure and Laser, in Greece. The other model parameters were estimated with an iterative technique to minimize the sum of squared deviations between model predictions and experimental data. In our implementation we used routine E04FDF from the

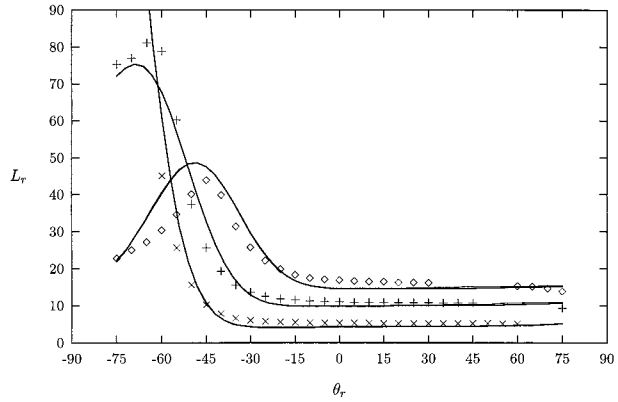

Fig. 2. Measured radiance from sample 1 (a fairly smooth tile) with $r = 0.12$, determined by profilometer measurements. Radiance is given as a function of θ_r, for $\theta_i = 45°$ ($\diamond$), $\theta_i = 60°$ (+), and $\theta_i = 75°$ ($\times$). The thick curve is the best fit for the present model when Eq. (35) is used with $C = 15.1$, $\rho = 0.484$, and $n = 2.17$. The thin curve (not visible because it coincides with the thick curve) is the best fit for the TSON model with $C = 15.2$, $\rho = 0.485$, and $n = 2.18$.

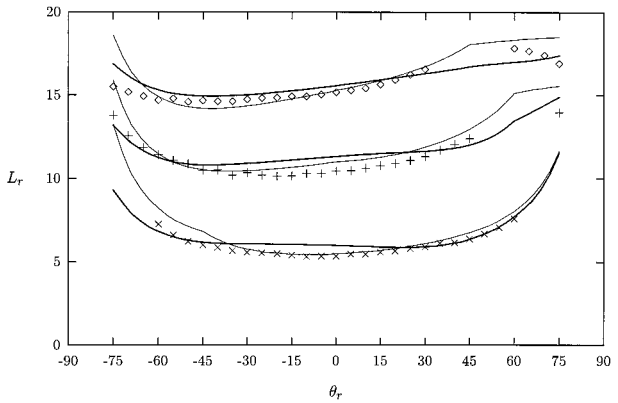

Fig. 4. Measured radiance from sample 3 (a grey concrete brick) with $r = 0.49$, determined by profilometer measurements. Radiance is given as a function of θ_r, for $\theta_i = 45°$ ($\diamond$), $\theta_i = 60°$ (+), and $\theta_i = 75°$ ($\times$). The thick curve is the best fit for the present model when Eq. (34) is used with $C = 68.5$ and $g = 0.0200$. The thin curve is the best fit for the TSON model with $C = 75.07$ and $g = 0.0385$.

NAG (Numerical Algorithms Group, Oxford, UK) FORTRAN library; this routine is based on the work of Gill and Murray.[21]

We also computed the best fit, using the same method, for a combination of the Torrance–Sparrow model[2] and the Oren–Nayar model.[8] We refer to this combined model as the TSON model. The Torrance–Sparrow model and the Oren–Nayar model are both derived from a surface model that differs from the statistical model adopted here; the surface is thought to consist of numerous long, symmetric, V-shaped cavities running in all directions. Torrance and Sparrow assume the cavities to be specular reflectors, and this leads to a model comparable with Eq. (28). Oren and Nayar treat the cavities as diffuse reflectors, and this produces a model comparable with Eq. (31). The distribution of the surface normals of the cavities, expressed in terms of their polar angles θ, can have any form in the Torrance–Sparrow model, and various distributions have been proposed.[22] Oren and Nayar used what they refer to as a slope-area distribution, and to make a consistent combination of both models we used this specific distribution for the Torrance–Sparrow model as well. From this distribution a rms slope can be computed. We combined both models in the way described in Section 2 for the present theory.

Figure 2 shows results obtained for a rather smooth colored tile (sample 1). The sample shows both a specular reflection component, noticeable in the peaks around the specular positions, and a diffuse reflection component, since the radiance far away from the specular peak is approximately constant. Although the sample is not very rough, the effect of

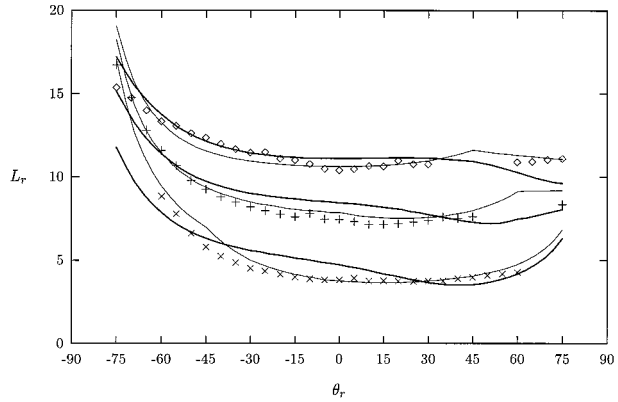

Fig. 3. Measured radiance from sample 2 (a rough pink tile) with $r = 0.51$, determined by profilometer measurements. Radiance is given as a function of θ_r, for $\theta_i = 45°$ ($\diamond$), $\theta_i = 60°$ (+), and $\theta_i = 75°$ ($\times$). The thick curve is the best fit for the present model when Eq. (34) is used with $C = 39.8$ and $g = 0.00501$. The thin curve (not visible because it coincides with the thick curve) is the best fit for the TSON model with $C = 45.6$ and $g = 0.0903$.

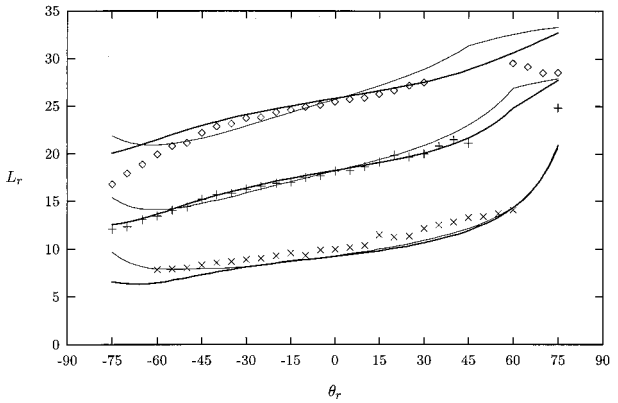

Fig. 5. Measured radiance from sample 4 (a red brick thinly painted with matte white latex paint) with $r = 0.43$, determined by profilometer measurements. Radiance is given as a function of θ_r, for $\theta_i = 45°$ ($\diamond$), $\theta_i = 60°$ (+), and $\theta_i = 75°$ ($\times$). The thick curve is the best fit for the present model when Eq. (34) is used with $C = 126$ and $g = 0.00527$. The thin curve is the best fit for the TSON model with $C = 134$ and $g = 0.0102$.

Table 1. Statistical Measures of Sample Surfaces

Sample	r	σ (μm)	R_a (μm)	R_a/σ
1	0.12	9.54	7.21	0.76
2	0.51	36.6	29.5	0.81
3	0.49	28.6	22.5	0.79
4	0.43	28.1	22.2	0.79

the surface roughness on the specular peaks is already clearly visible: the peaks are not sharp. In the diffuse component no noticeable effect of roughness is seen (in the sense of an increase when the detector approaches the light source). In the case of this sample, significantly better results were obtained when the experimental data were fitted to Eq. (35), which includes the Fresnel reflection coefficient. This was not the case for the other three samples, so for those we used Eq. (34). There is good agreement between data and model fits.

Figure 3 shows results for a rougher tile (sample 2). The greater roughness clearly has a greater effect on the reflection characteristics of this tile than for sample 1. The roughness broadens the specular peaks so much that only a general increase in the specular direction remains. The effect of roughness on the diffuse term is still not noticeable. Again, both models fit the data reasonably well.

Figure 4 shows results for a grey concrete brick (sample 3). We measured different types of brick, and apart from the overall albedo (yellow bricks appear brighter than red bricks) we obtained similar results for all these samples. The reflection properties of the bricks are different from those of sample 2, although the roughness values are comparable. The bricks clearly show that roughness affects the diffuse component: radiance increases when the detector approaches the light source. In the model fits we find lower values for g; that is, the local reflection properties are shifted from specular to diffuse.

Finally we performed measurements on a red brick sprayed with a thin layer of white matte latex paint (sample 4). The idea was to remove the specular component without changing the surface structure too much. Indeed, the paint did not affect the roughness value much. Figure 5 shows that the specular component has disappeared while the effect of roughness on the diffuse component remains. The fitting procedures come up with g values that are lower than those of samples 2 and 3. The C values are higher, as expected, because this sample was painted white and has a high albedo.

For each sample we obtained between 10–17 profiles of 4 mm length with a resolution of 0.5 μm. The profiles included both horizontal and vertical directions. We found no systematic difference between these directions, which supports the assumption of surface isotropy. From these profiles we calculated σ, the standard deviation of height; r, the rms slope; and R_a, the arithmetic average defined as the mean absolute deviation from the mean height. These values are given in Table 1, together with the ratio of

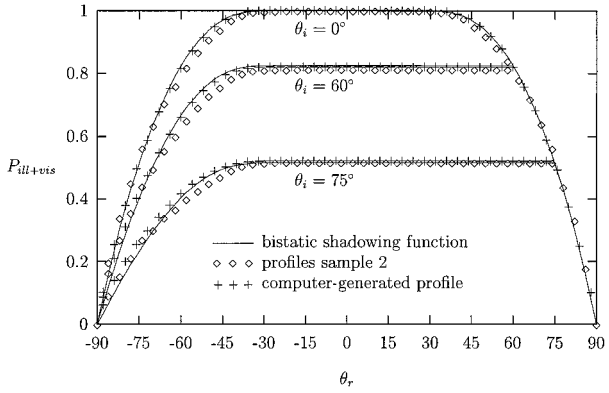

Fig. 6. Illuminated and visible portion of the surface as a function of θ_r for various values of θ_i. The values computed from the profiles of sample 2, measured by a profilometer, are denoted $\diamond$. For these profiles a rms slope of $r = 0.51$ was computed. The bistatic shadowing function for this r is plotted as a solid curve. Computed values for computer-generated profiles with $r = 0.51$ are denoted $+$.

R_a and σ. If the surface is Gaussian, this ratio must have the value $(2/\pi)^{1/2} \approx 0.80$, as can be seen when a normal distribution is inserted into the definition of R_a:

$$R_a = \int_{-\infty}^{\infty} |z| P_z(z, \sigma) \mathrm{d}z = \int_{-\infty}^{\infty} \frac{|z|}{\sqrt{2\pi}\sigma} \exp\left(\frac{-z^2}{2\sigma^2}\right) \mathrm{d}z$$
$$= \sqrt{\frac{2}{\pi}}\,\sigma \approx 0.80\sigma. \tag{37}$$

R_a/σ is therefore a straightforward test for checking the hypothesis of a Gaussian height distribution (page 92 of Ref. 13, and Ref. 23).

To verify the accuracy of the masking–shadowing function, we performed many computer simulations in which we used as input both the digitized sample profiles measured by the profilometer and computer-generated surfaces. Then we calculated the visible and illuminated portion of the surface as a function of θ_i, θ_r, and ϕ_r. This portion can be compared with the bistatic shadowing function for the same r value, with both self-shadowing and self-masking and shadowing and masking by intersection taken into account. In all cases we found good agreement between the masking–shadowing theory and computer experiments. A typical example is given in Fig. 6.

Our theory can be used to generate texture patterns. In Fig. 7 we show an example of a generated texture, together with an actual image of sample 4, the white painted brick.

The textures look very similar. Differences can be attributed to the difference in surface structure. The brick seems to consist of more or less spherical parts that are not seen in the Gaussian surface. Furthermore, the shadowed parts in the generated image are completely black, which is because interreflections have been ignored.

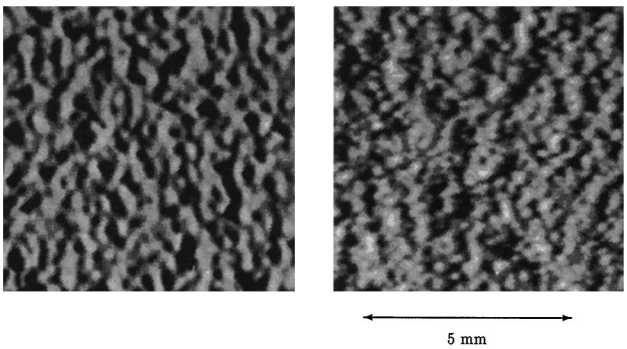

Fig. 7. Left, texture pattern generated by use of the present theory ($\theta_i = 75°$, $\theta_r = 0°$, $r = 0.50$, $g = 0$). Right, image taken with a CCD camera from the white painted brick (sample 4) for $\theta_i = 75°$, $\theta_r = 0°$. For this sample $r = 0.43$, determined by profilometer measurements; and $g = 0.0143$, determined by model fitting.

4. Discussion

From Figs. 2–5 it is clear that there are large differences between the reflection characteristics of the four samples. Although the models have only three or four parameters, they are able to capture all these variations. This flexibility supports the starting point of this paper, namely, that the reflection properties of materials are determined mainly by surface roughness and the balance between specular and diffuse reflection components. If the theory were to be simplified, the results would be much poorer. For instance, the omission of the diffuse component leads to huge errors for all the samples presented and is in practice allowed only when surfaces closely resemble a perfect mirror. Likewise, the removal of the specular component is inadmissable unless diffuse reflection is dominant, as in sample 4. The effect of roughness on the diffuse component, which is ignored in many reflection models, is also essential for an accurate description of the reflection properties of rougher surfaces, as is clear from the results for the bricks (samples 3 and 4). This has also been demonstrated in experiments by Oren and Nayar,[8] involving samples with a large rms slope and a dominating diffuse component. Finally, in cases for which both the roughness and one of the polar angles are large, it is necessary to include a bistatic shadowing function to avoid large deviations between theory and experiment.

There is no major difference in the performance of the present model and the TSON model. It turns out that both models show similar behavior for values of $r < 0.2$ except at grazing viewing angles. [When $\theta_r \rightarrow \infty$, L_r is the product of a very large number ($1/\cos \theta_r$) and a very small number, namely, the masking correction; since this correction is different for both models, the deviation between the two will become large.] For larger roughness values differences show up because of the different surface models used and, as a consequence, the different ways in which the correction for masking and shadowing is handled. Clearly, the surface model used here is preferable, because that of the TSON model is inconsistent.

Note that one of the model parameters, the rms slope r, was not determined by fitting but was obtained from an independent measurement. Since C is only a scaling parameter, this means that for samples 2 to 4 just one free parameter (g) was fitted, and for sample 1 only two (ρ and n) were fitted. This makes the performance of the model even more impressive. If we include r as a parameter in the fitting procedure, we find roughness values comparable with the measured r and only slightly better fits.

From the experimental data presented here, it seems that for many natural materials a simple linear combination of diffuse and specular reflection components, as given in Eq. (34), will be adequate. The linear combination reflects the idea that the samples are inhomogeneous mixtures of which some parts reflect the irradiation specularly and others scatter it diffusely. There exists also a class of materials for which the combination with a Fresnel coefficient [Eq. (35)] is preferable. Sample 1 is a representative of this class. Without the Fresnel coefficient, the model is not able to explain the increase in the specular peak for increasing θ_i. Note that in fitting Eq. (35) to goniophotometric data, one obtains an estimation of the refractive index n of the material. The value we found for sample 1 ($n = 2.18$) is probably too high, since n ranges from 1.3 to 2.0 for most dielectrics. But this is hardly an indication that the physical assumptions behind the theory are wrong, because forcing n to a lower value results in a fit that is only slightly worse. Strictly speaking, to calculate F, we should use θ_i' instead of θ_i, and we could also, following the ideas of Wolff,[24] include a term to correct for the possibility that (diffuse) rays that reappear at the surface are refracted back into the dielectric. But contrary to what one would expect, this only reduced the agreement between model fit and the experimental data for sample 1. This suggests that the complete separation of the specular and the diffuse component, in the way described here, is an oversimplification of the actual physical process of light interacting with materials.

Another possible explanation for the deviations between theory and experiment is that we have ignored the contributions of interreflections to the reflected radiation. The treatment of specular and diffuse interreflections from a Gaussian rough surface is a complicated problem that is beyond the scope of this paper.

Although the work of Oren and Nayar,[8] who calculated a correction term for diffuse interreflections, shows that this contribution is relatively small, even for surfaces with a high albedo, it would still be interesting to include the effect for at least two reasons. First, interreflections have a large influence on the spectrum of the reflected radiation (for higher-order reflections the spectrum will shift more toward the diffuse spectrum). Making the albedo ρ wavelength dependent and including this effect could be used to extend the present model to a color reflection model. Second, interreflections are important in texture

analysis and generation, since the brightness of shadowed parts is determined solely by interreflections.

There is a natural link between reflection models and texture analysis. Applications using reflection models will often also use texture analysis. In computer graphics or machine vision the reflection model deals with all structure smaller than the resolution of the images used, whereas structure at higher scales appears as texture. Since most reflection models used in computer vision applications do not take the surface structure into account, it is perhaps not surprising that the reflection properties of surfaces and their textural appearance are usually treated as separate matters.

Now we see one of the key benefits of the present theory. It is based on an accurate and realistic surface model, as is shown from the profilometer data analysis in Table 1 and from the agreement between the masking–shadowing function calculated from surface profiles and the bistatic shadowing theory developed here. Moreover, many authors have reported the applicability of the normal height distribution to surfaces as diverse as the moon,[25] the sea, and many man-made and natural rough surfaces.[13,23] Therefore, the theory can be used to generate realistic-looking textures for given angles (θ_i, θ_r, ϕ_r) and parameters (r, g, C). This distinguishes the present theory from the TSON model. The latter is based on a mathematically inconsistent surface model: long V-cavities, running in all directions. Therefore the surface structure cannot be visualized as texture.

This research was supported by the European Strategic Programme for Research and Development in Information Technology Program Realise of the European Community. The authors thank L. Boonen for helping with the experimental set up.

References

1. J. H. Lambert, *Photometria sive de mensura et gradibus luminus, colorum et umbrae* (Eberhard Klett, Augsburg, Germany, 1760).
2. K. Torrance and E. Sparrow, "Theory for off-specular reflection from roughened surfaces," J. Opt. Soc. Am. **57,** 1105–1114 (1967).
3. S. K. Nayar, K. Ikeuchi, and T. Kanade, "Surface reflection: physical and geometrical perspectives," IEEE Trans. Pattern Anal. Mach. Intell. **13,** 611–634 (1991), Appendix D.
4. P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces* (Pergamon, New York, 1963).
5. D. E. Barrick, "Rough surface scattering based on the specular point theory," IEEE Trans. Antennas Propag. **16,** 449–454 (1968).
6. J. C. Leader, "Analysis and prediction of laser scattering from rough-surface materials," J. Opt. Soc. Am. **69,** 610–628 (1979).
7. X. D. He, K. E. Torrance, F. X. Sillion, and D. P. Greenberg, "A comprehensive physical model for light reflection," ACM Comput. Graphics **25,** 175–186 (1991).
8. M. Oren and S. K. Nayar, "Generalization of the Lambertian model and implications for machine vision," Int. J. Comput. Vision **14,** 227–251 (1995).
9. M. S. Longuet-Higgins, "Reflection and refraction at a random moving surface. II: Number of specular points in a Gaussian surface," J. Opt. Soc. Am. **50,** 845–850 (1960).
10. M. I. Sancer, "Shadow-corrected electromagnetic scattering from a randomly rough surface," IEEE Trans. Antennas Propag. **17,** 577–585 (1969).
11. B. G. Smith, "Geometrical shadowing of a random rough surface," IEEE Trans. Antennas Propag. **15,** 668–671 (1967).
12. D. Middleton, *Introduction to Statistical Communications Theory* (McGraw-Hill, New York, 1960).
13. T. R. Thomas, *Rough Surfaces* (Longmans, London, 1982).
14. P. Beckmann, "Shadowing of random rough surfaces," IEEE Trans. Antennas Propag. **13,** 384–388 (1965).
15. R. J. Wagner, "Shadowing of randomly rough surfaces," J. Acoust. Soc. Am. **41,** 138–147 (1966).
16. B. G. Smith, "Lunar surface roughness: Shadowing and thermal emission," J. Geophys. Res. **72,** 4059–4067 (1967).
17. R. A. Brockelman and T. Hagfors, "Note on the effect of shadowing on the backscattering of waves from a random rough surface," IEEE Trans. Antennas Propag. **14,** 621–629 (1966).
18. D. J. Schertler and N. George, "Backscattering cross section of a roughened sphere," J. Opt. Soc. Am. A **11,** 2286–2297 (1994).
19. F. E. Nicodemus, J. C. Richmond, and J. J. Hsia, *Geometrical Considerations and Nomenclature for Reflectance*, Monogr. 160 (National Bureau of Standards, Gaithersburg, Md., 1977).
20. G. Kortüm, *Reflectance Spectroscopy* (Springer-Verlag, Berlin, 1969).
21. P. E. Gill and W. Murray, "Algorithms for the solution of the nonlinear least-squares problem," SIAM J. Numeri. Anal. **15,** 977–992 (1978).
22. T. S. Trowbridge and K. P. Reitz, "Average irregularity representation of a rough surface for ray reflection," J. Opt. Soc. Am. **65,** 531–536 (1975).
23. F. Abdellani, G. Rasigni, M. Rasigni, and A. Lleberia, "Distributions of zero crossings for the profile of random rough surfaces," Appl. Opt. **31,** 4534–4539 (1992).
24. L. B. Wolff, "Diffuse-reflectance model for smooth dielectric surfaces," J. Opt. Soc. Am. A **11,** 2956–2968 (1994).
25. K. Lumme, J. Peltoniemi, and W. M. Irvine, "Some photometric techniques for atmosphereless solar system bodies," Adv. Space Res. **10,** 1187–1193 (1990).