A COMPARISON OF PHYSICALLY AND EMPIRICALLY BASED ILLUMINATION MODELS

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF MASTER OF SCIENCE
IN THE FACULTY OF SCIENCE

By

Dean Gammage

Department of Computer Science

October 1993

Contents

List of figures 6

List of tables 9

Abstract 10

Declaration 11

Dedication 12

Acknowledgements 13

1 Introduction 14

2 Local empirically based illumination models 18

2.1 Wylie model 18

2.2 Lambert model 19

2.3 Bouknight model 20

2.4 Phong model 21

2.5 PHIGS 24

2.5.1 Light source types 24

2.5.2 Illumination model 26

3 Physically based illumination models 28

3.1 Beckmann model 29

3.2 Stogryn model 31

3.3 Torrance and Sparrow model 31

3.4 Blinn model 33

3.5 Whitted model 34

3.6 Moravec model 36

3.7 Cook and Torrance model 36

4 He, Torrance, Sillion and Greenberg illumination model 40

4.1 Details of the new model 40

4.2 Colour 43

4.2.1 RGB 44

4.2.2 The Perception of Colour 45
4.2.3 CIEXYZ 47

5 Implementation of the new illumination model 50

5.1 Ray tracing 50
5.1.1 Overview 50
5.1.2 Sphere intersection 54
5.2 Scene description 56
5.3 Spectral samples 56
5.4 Conversion of spectral samples to an RGB triplet 58
5.5 Gamma correction 61
5.6 Fresnel term 62
5.6.1 Theory 62
5.6.2 Implementation 64
5.7 Facet distribution function 66
5.8 Effective surface roughness 67
5.9 Specular term 69
5.10 Ambient term 70

5.11 Uniform diffuse term 72

6 Results and conclusions 74

6.1 Results 74

6.1.1 Ambient images 77

6.1.2 Diffuse images 77

6.1.3 Specular images 81

6.1.4 Complete Images 83

6.1.5 Computation times 86

6.2 Conclusion 86

6.3 Further work 86

A Mathematics of the New Illumination Model 89

B An MSDL data file 93

References 95

List of Figures

1.1 Rendering pipeline 15

2.1 Uniform diffuse reflection from a surface 20

2.2 A simple scene 21

2.3 Vectors used in the Phong model 23

2.4 Specular intensity as a function of **V** and n 23

2.5 Light source types in PHIGS 24

2.6 Cone of influence of a spot light in PHIGS 25

3.1 Vectors used in Kirchhoff integral 30

3.2 Tangent plane approximation 30

3.3 A surface composed of microfacets 32

3.4 Shadowing and masking 33

3.5 Vectors used in the Whitted model 35

4.1 Reflection by a Real Surface 41

4.2 RGB Colour Space 44

4.3 The Human Retina (courtesy of Chris Lilley) 46

4.4 Luminous efficiency (courtesy of Chris Lilley) 47

4.5 Cone sensitivities on a log scale (courtesy of Chris Lilley) 48

4.6 CIEXYZ matching functions (courtesy of Chris Lilley) 49

5.1 Forward ray tracing 51

5.2 Backward ray tracing 52

5.3 Ray tracing a scene 52

5.4 Pattern of rays cast through each pixel 53

5.5 Pseudocode for a simple ray tracer 54

5.6 Geometry of sphere intersection 55

5.7 The Effects of Different Numbers of Spectral Samples 57

5.8 Wavelength selection (courtesy of Chris Lilley) 58

5.9 Computing CIE tristimulus values (courtesy of Chris Lilley) . . . 59

5.10 The Need for Gamma Correction 62

5.11 Approximated Fresnel Reflectance for Copper using the measured values of n = 0.617 and k = 2.630 65

5.12 Variation of Reflectance with Angle 65

5.13 Spectral Reflectance Curves: (1) galvanized steel; (2) aluminium; (3) stainless steel; (4) brass; (5) copper; (6) zinc; (7)lead; (8) magnesium alloy 66

5.14 Pseudocode of iterative procedure 69

5.15 Specular Reflection from a Surface 70

5.16 The incident hemisphere 71

5.17 Reflection at a surface 72

6.1 Specular, directional diffuse, uniform diffuse and ambient terms for Copper Sphere 75

6.2 Sum of specular, directional diffuse, uniform diffuse and ambient terms for copper sphere 76

6.3 Specular, directional diffuse and ambient terms for aluminium sphere 76

6.4 Sum of specular, directional diffuse and ambient terms for aluminium sphere 77

A.1 Vectors used by Physical Model 90

List of tables

5.1 Selected wavelengths 57

5.2 Details of Monitor Phosphors 59

5.3 Parameters used in $erfc(\beta)$ approximation 68

Abstract

UNIVERSITY OF MANCHESTER

ABSTRACT OF THESIS submitted by **Dean Gammage** for the Degree of Master of Science and entitled **A Comparison of Physically and Empirically Based Illumination Models**

Month and Year of Submission: October 1993

In order to produce highly realistic computer generated images it is necessary to accurately simulate the natural illumination process rather than generating only the illusion of realism. This thesis presents the results of a comparison between an existing model used in PHIGS PLUS and a new general reflectance model based on physical optics.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning.

Dedication

To my mum and dad.

Acknowledgements

Special thanks to my supervisor Terry Hewitt for his help and advice, Neil Gatenby for his modifications to MSDL and generally being a star, Chris Lilley for his help on colour, Tony Arnold for the use of his camera, Fenquiang Lin for his help with PHIGS and Martin Preston for his proof reading and putting up with my endless questions.

I would also like to mention Dave Hutchinson, Mary McDerby, Paul Lever, Alex Knowles and Helen Morphet for their help and comments throughout.

Finally, personal thanks to Emily Blanchet, Martin Samm, Phil Hyman and Jude Mott for their encouragement, without which this thesis would probably never have been completed.

Chapter 1

Introduction

The process of converting a database representation of a three dimensional scene into a shaded two dimensional image is known as *rendering*. The rendering pipeline involves a number of separate processes, as can be seen in figure 1.1. A database contains information about objects and lights. Once the objects have been defined, transformations may be applied which position the objects within the scene and our view of it. When the three dimensional world has been completely defined, the surfaces can be rasterized, which is the conversion of the surfaces into sets of pixels. The hidden surfaces are then be removed from the scene and each pixel in the image shaded by the application of an illumination model.

An illumination model is a set of rules which describe, in terms of the nature of the incident light and the properties of the surface, the way in which light interacts with the surface of a material. The illumination model is usually applied at red, green and blue wavelengths, and the resulting values passed to the display

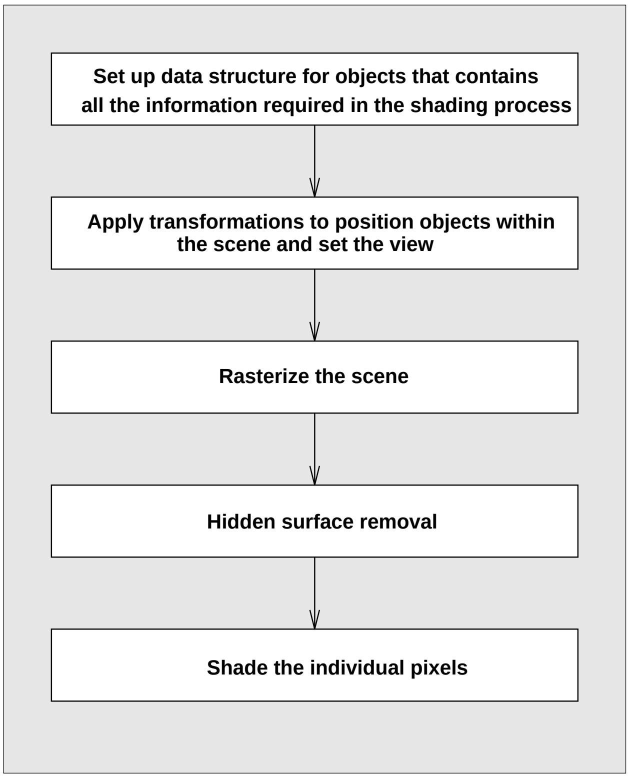

Figure 1.1: Rendering pipeline

medium as an RGB triplet. The realism required of the image determines the complexity of the illumination model and hence the computational expense of image generation. It is only those applications which require the utmost realism that would benefit by the use of a physically correct illumination model, since the computational expense would prohibit its use where such realism is of lesser importance.

Photorealism is not always of importance in image generation. A builder, for example, may only require the simplest of images in order to visualize a building. In some instances, the image is made deliberately unrealistic in order to convey information not normally available, false colour images of heat flow are a good example.

In order to produce highly realistic computer generated images it is necessary to accurately simulate the natural illumination process rather than generating only the illusion of realism. The simulation of natural illumination requires an illumination model which has a solid physical foundation.

The quest for photorealism is an attempt to try and produce an image on a two dimensional screen which has the same shading and colour as would be expected if the screen were in fact a photograph of a real three dimensional world. The illumination model plays an important role in the conversion of the geometric model of a scene into a realistic two dimensional image [16].

Illumination models fall into one of two classes:

- *Local illumination models* [46, 33, 5, 39], which simulate the interaction of the surface with light coming directly from the light sources and do not consider light from other objects within the scene. Therefore, these models do not take object-object interactions within the scene into account and so do not produce shadows or other important effects, such as object inter-reflections and colour bleeding. Unless stated otherwise, the illumination models discussed in the following chapters should be assumed to be of this type.

- *Global illumination models* [45, 30], which simulate the interaction of the surface with all light in the scene, irrespective of whether it comes directly from a light source or from another object in the scene. Models of this type allow complicated object-object interactions to be simulated, and therefore complex shadowing effects and colour bleeding may be generated.

This thesis compares the images produced by a typical empirically based local illumination model and those produced by a new general model based on physical optics. It will consider the image quality and rendering time of the two models.

For a typical empirically based model, I will use the model contained within the Programmers Hierarchical Interactive Graphics System (PHIGS). The physically based model will be a new model proposed by He, Torrance, Sillion and Greenberg [20].

Chapters 2 and 3 discuss the evolution of empirical and physical illumination models. The new illumination model is described in chapter 4, with particular attention paid to the colour model used and its conversion into a displayable form. The implementation of the new model will be discussed in chapters 5 and a comparison of the images generated will be made in chapters 6.

Chapter 2

Local empirically based illumination models

In order to compare the PHIGS illumination model to the new model, it is necessary to understand the principles upon which the PHIGS model is based. This chapter discusses the development of empirically based illumination models.

2.1 Wylie model

Empirically based illumination models are fairly simple models which generate only the illusion of realism. One of the first models used in the shading of computer images was proposed by Wylie [46] in 1967. It stated that the intensity of a point on the surface of an object was inversely proportional to its distance from the light source, and its form was:

$$I_r = I_i \beta d$$

I_r is the reflected light intensity. I_i is the incident light intensity.

d is the distance to the light source. β is a constant

Parallel planes at different distances from a light source were then displayed at different intensities, but no allowance was made for individual surface properties such as reflectivity or roughness. Models such as this are only used in special applications which simulate atmospheric effects such as fog or mist.

2.2 Lambert model

In order to improve on the Wylie model, the orientation of the reflecting surface with respect to a light source had to be considered. This is known as Lambertian or uniform diffuse reflection [6], and can be seen in figure 2.1. Assuming that incident light is scattered equally in every direction (isotropic reflection), the brightness at a given position on the surface is proportional to the cosine of the angle between the surface normal **N**, and the light vector **L**. The Lambert model is therefore given by:

$$I_r = I_i K_d (\mathbf{N} \cdot \mathbf{L})$$

Where K_d is a diffuse reflection coefficient.

Although this model produced reasonable results, all surfaces appeared matte and dull. Most real materials are imperfect diffuse reflectors and so could not be represented. However, an imperfectly diffuse surface can be split into a number of patches, each of which are perfectly diffuse and can therefore be represented using this model.

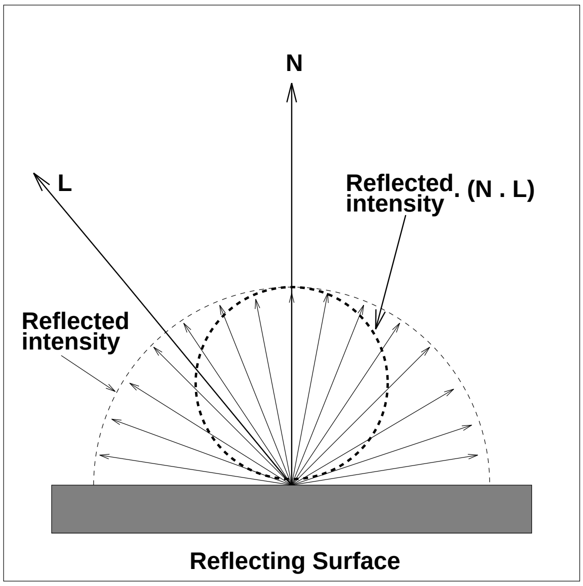

Figure 2.1: Uniform diffuse reflection from a surface

2.3 Bouknight model

If the scene shown in figure 2.2 were rendered using a local illumination model, only the top face of the cube would be visible, with the remainder rendered black. However, in the real world, reflections from other surfaces would illuminate the object from every direction, making all faces visible. In order to overcome this problem, Bouknight [7] used *ambient* and Lambertian terms to develop a new model in 1970. The ambient term in this model is an approximation to the background illumination in the scene. Bouknight's model can be described by:

$$I_r = I_a K_a + K_d(\mathbf{N} \cdot \mathbf{L})$$

Where I_a is the ambient light intensity and K_a is an ambient reflection coefficient.

The inclusion of an ambient term in the Bouknight model meant that scenes illuminated by only a point light source exhibited a higher degree of realism than

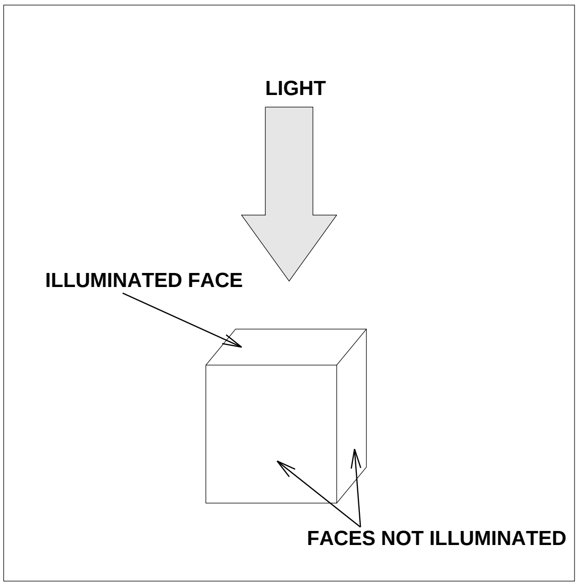

Figure 2.2: A simple scene

with previous models. However, this model still failed to address the problem represented by shiny surfaces, which is an obvious limitation.

2.4 Phong model

In 1975, Bui Tuong Phong [33] introduced a model which was designed to represent reflections from shiny surfaces. The model contained a linear combination of a diffuse component, an ambient component and an empirically based specular component. Phong assumed that light travelling from the $\mathbf{L}$ direction would be reflected along the reflection vector $\mathbf{R}$, as can be seen in figure 2.3. If the view vector $\mathbf{V}$, is equal to $\mathbf{R}$, a bright specular highlight would be observed, and as $\mathbf{V}$ moves away from the $\mathbf{R}$, the intensity of the specular highlight falls off by an amount influenced by the smoothness of the surface. Phong modelled this fall off as $(\mathbf{R} \cdot \mathbf{V})^n$, where n is the *specular index*, which would be infinite for a perfect reflector, the effect of increasing n is shown in figure 2.4. The Phong model is described by:

$$I_r = I_a k_a + I_i(k_d(\mathbf{L} \cdot \mathbf{N}) + k_s(\mathbf{R} \cdot \mathbf{V})^n)$$

Where K_s is a specular reflection coefficient.

It should be noted that the specular reflection coefficient is wavelength independent, meaning that specular highlights are always the same colour as the light source [44, 11]. This is inaccurate but produces pleasing results.

One of the main faults with the Phong model is that it does not consider the variation of the specular peak with the position of the light source. Phong recognized this problem, but did not recommend a solution to it. All geometry except the surface normals was ignored, which meant that the light sources and the viewer were located at infinity. Any intensity distribution of the light source was also ignored.

This model is still widely used in computer graphics as it produces reasonable results and requires relatively little computation when compared to many more complicated models. The Phong model can also be encoded into hardware, which reduces the computation time considerably.

So far, numerous illumination models have been presented, each producing good results for certain surfaces under specific lighting conditions. The Bouknight model, and all later models which included an ambient term, were good at representing surfaces illuminated by only a point light source. However, the Phong model produced good results for shiny surfaces illuminated by a light at an infinite distance. It became clear that a model was needed that was capable of representing many types of surface and numerous light source types.

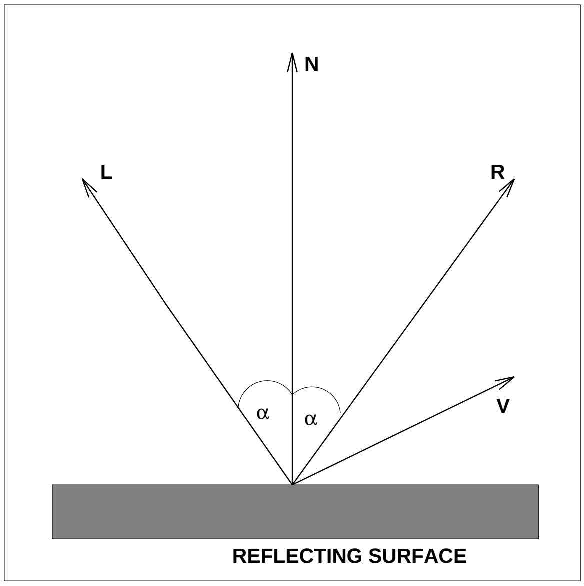

Figure 2.3: Vectors used in the Phong model

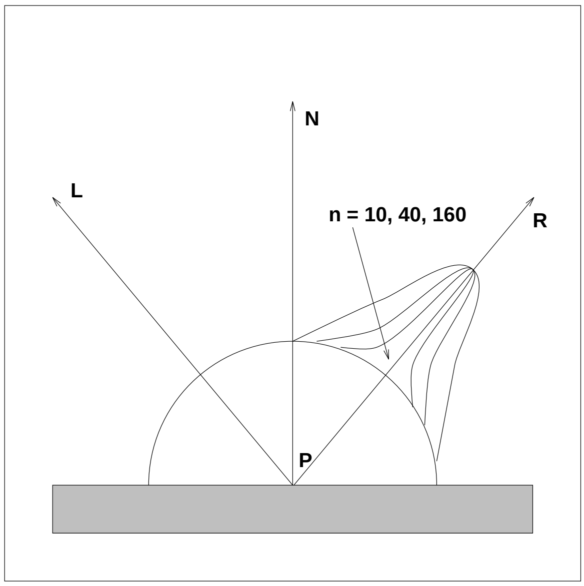

Figure 2.4: Specular intensity as a function of **V** and n

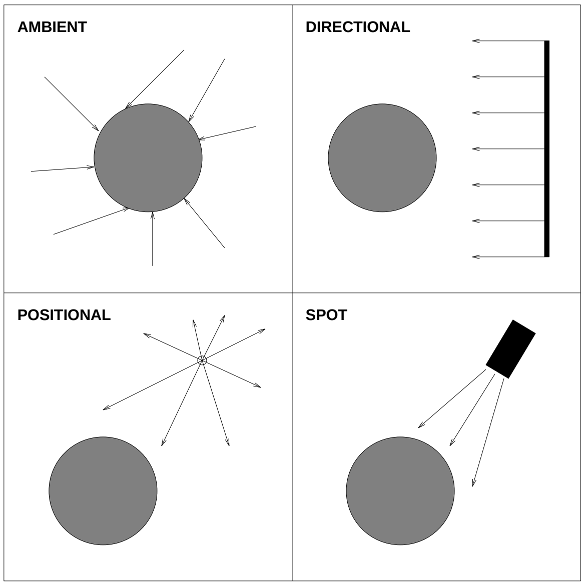

Figure 2.5: Light source types in PHIGS

2.5 PHIGS

2.5.1 Light source types

During the 1980's, the Programmer's Hierarchical Interactive Graphics System (PHIGS) [25, 24, 26] was developed. For the first time, different types of light source were included in the model [23], these were ambient, spot, positional and directional light sources as shown in figure 2.5.

In the real world, light is scattered from an objects surroundings, causing the object to be illuminated independently of its position and orientation. In order to approximate this, PHIGS models ambient light as a constant, diffuse reflection from all sides of the object. Ambient light sources have an intensity defined by an RGB triplet and can therefore be of any colour.

Directional light sources have a colour and a direction. Light originates from the specified direction, and since the light is located at infinity, this direction is the

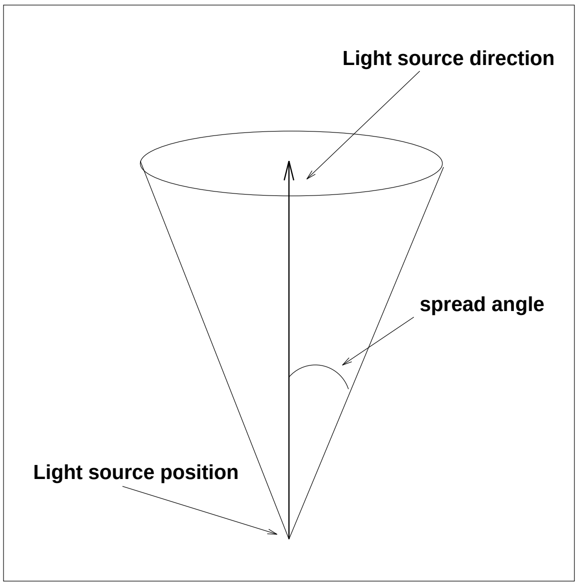

Figure 2.6: Cone of influence of a spot light in PHIGS

same for all objects, regardless of their position.

The third light source type included in PHIGS is the positional light. This is a point source and light rays travel in all directions from its specified position. The rays from this type of light source are not parallel, and so the position of an object relative to the light source affects the angle at which the light hits the surface. In order to specify a point light source, the user must supply a position, an intensity, and two attenuation coefficients. The attenuation coefficients describe how the brightness of the light falls off with distance from the source, and can therefore be used to model effects such as fog or mist.

The final light source type available in PHIGS is the spot light source. This is a point light source with a cone of influence, outside of which no light is emitted, as can be seen in figure 2.6.

2.5.2 Illumination model

In order to standardise existing illumination models, PHIGS combined them into one set of equations [34], it is this which will be compared to the new physical model. The illumination model used by PHIGS can be expressed as:

$$I_r = \sum_{i=1}^{n}(I_{a_i} + I_{d_i} + I_{s_i})$$

Where:

I_a is the ambient light source contribution.

I_d is the diffuse light source contribution.

I_s is the specular light source contribution.

n is the total number of active light sources.

The ambient, diffuse and specular contributions from these lights are:

$$I_a = K_a L_c O_d$$

$$I_d = K_d L_c O_d (1 + L_a((\mathbf{L}_d \cdot -\mathbf{L})^{L_e}))(\mathbf{N} \cdot \mathbf{L})$$

$$I_s = K_s L_c O_s (1 + L_a((\mathbf{L}_d \cdot -\mathbf{L})^{L_e}))(\mathbf{V} \cdot \mathbf{R})^{O_e}$$

Where:

$\mathbf{L}_d$ is the light vector.

L_c is the light source colour.

$\mathbf{L}$ is the light position.

L_e is the light concentration exponent.

O_d is the diffuse colour of the object.

O_s is the specular colour of the object.

O_e is the specular exponent.

Although this model produces reasonable results, it has many limitations with respect to realistic image generation, these will be discussed in chapter 6.

Chapter 3

Physically based illumination models

Radiosity [17, 11] and ray-tracing [14] were developed to obtain realistic images for diffuse and specular environments respectively. The application of these techniques is limited by the lack of a comprehensive physical model for light reflection. A general physically based model is needed to produce realistic pictures for such environments.

Physically based illumination models are far more complicated than the empirically based models discussed in the previous section. Although these models produce more realistic images, they are computationally expensive and therefore are only used in applications which require a high degree of realism. For an explanation of the physics upon which the following models are based, refer to [28, 35, 21].

3.1 Beckmann model

An early physical model proposed by Beckmann [5] used:

- The *Kirchhoff theory of scalar diffraction* to calculate the electromagnetic field at an arbitrary point above a surface using the scalar field E_r and its normal derivative dE_r/dn on any enclosing surface Γ (for a single surface, Γ reduces to the area of the reflecting surface), figure 3.1. The Kirchhoff theory has the following form:

$$E(R) = \frac{1}{4\pi} \int_\Gamma \left(E(r) \frac{dZ(R,r)}{dn} - Z(R,r) \frac{dE(r)}{dn} \right) d\Gamma$$

  Where $Z(R,r)$ is the free space Green's function and has the form:

$$Z(R,r) = \frac{e^{ik|R-r|}}{|R-r|}$$

- A *tangent plane approximation* to calculate the electric field at a point on the surface. This approximation states that the field at a point on the surface is the field that would exist if the surface were replaced by its local tangent plane, as can be seen in figure 3.2. This is a valid assumption if the radius of curvature of the surface is much greater than the wavelength of the incident light.

- A *statistical surface* to supply information, at length scales the order of the incident light wavelength, necessary for the Kirchhoff theory. The surface height distribution was assumed to be gaussian and spatially isotropic.

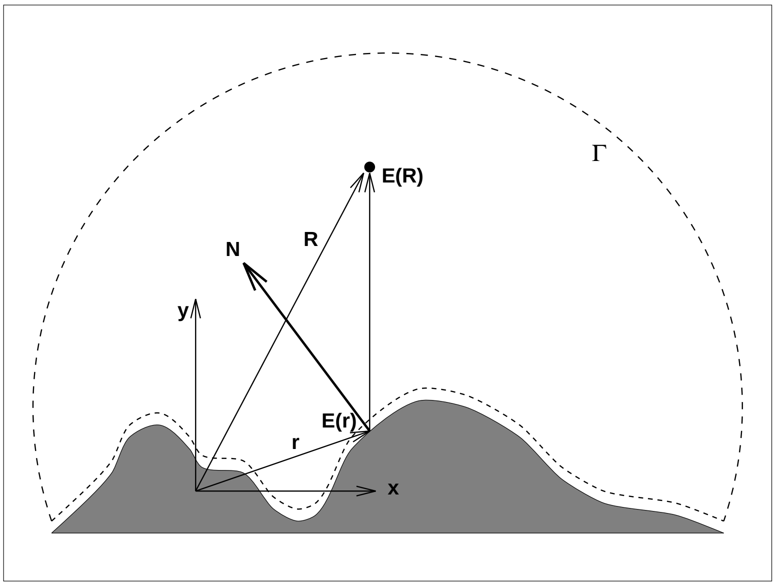

Figure 3.1: Vectors used in Kirchhoff integral

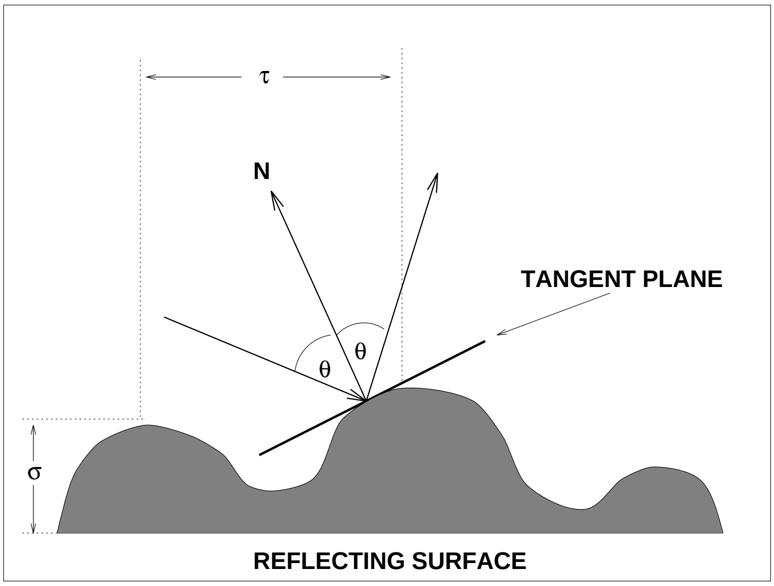

Figure 3.2: Tangent plane approximation

3.2 Stogryn model

A more general physical model was later proposed by Stogryn [39]. It was based on the Beckmann model and included:

- A *vector* form of Kirchhoff's theory, which included effects such as Fresnel reflectivity.
- Statistical averages over surface height *and slope*.

However, this model did not consider shadowing/masking effects [4, 8] and was only suitable for special incident polarizations.

3.3 Torrance and Sparrow model

Following work by Torrance, a new model was proposed by Torrance and Sparrow [41] in 1967, the aim being to improve the appearance of curved surfaces. The following assumptions were made during its derivation:

- Surfaces are composed of collections of randomly orientated *microfacets*, as can be seen in figure 3.3.
- The specular component comes from reflection off facets whose surface normal points in the direction of the average of the light vector and view vector.
- The diffuse component comes from multiple surface reflection and internal scattering.

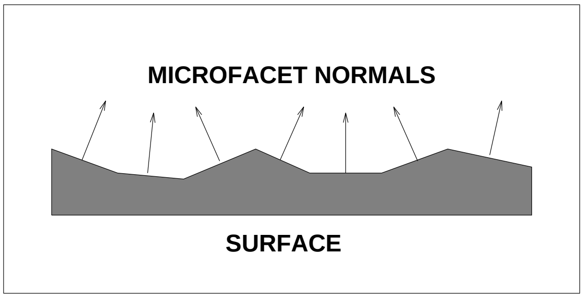

Figure 3.3: A surface composed of microfacets

The form of the specular term was:

$$I_s = \frac{D(\phi)S(\mathbf{L},\mathbf{N},\mathbf{V})F(\theta)}{\mathbf{N}\cdot\mathbf{V}}$$

Where:

$D(\phi)$ is a *facet distribution function*, which is a measure of the number of facets facing in a given direction and is thus an indication of how shiny the surface is, a simple gaussian distribution was chosen by Torrance and Sparrow. This term depends only on the viewing angle, ϕ.

$S(\mathbf{L},\mathbf{N},\mathbf{V})$ is a *shadowing/masking* factor, which describes the shadowing of incident light and the masking of reflected light by the surface, as can be seen in figure 3.4. This term depends on the light vector, the view vector, the surface normal and the surface roughness.

$F(\theta)$ is the *Fresnel Reflectance*, which is the fraction of incident light reflected rather than absorbed. The wavelength dependent nature of this term is not considered in this model.

$1/(\mathbf{N}\cdot\mathbf{V})$ accounts for the fact that the observed surface area is inversely proportional to the angle between the vectors $\mathbf{N}$ and $\mathbf{V}$.

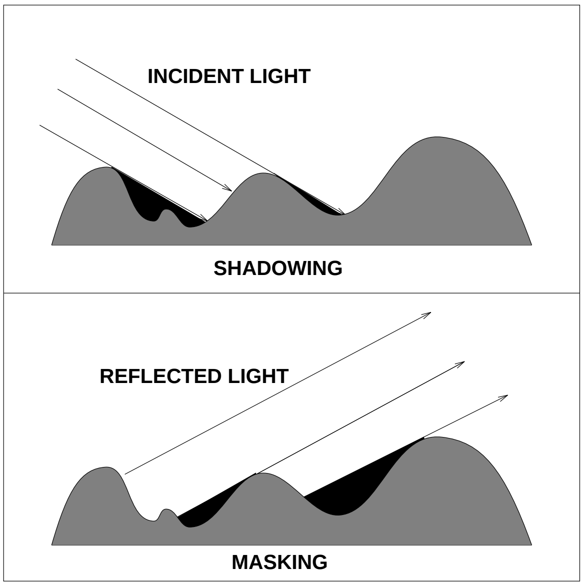

Figure 3.4: Shadowing and masking

The Torrance-Sparrow model correctly predicted that the specular component is not always along the reflection vector **R** and that *off-specular peaks* [40] arise at shallow angles of incidence, these effects are mainly noticeable for non-metallic and edge lit objects. This model also predicted that the size of the specular peak varies with the angle of incidence. Although it required considerably more computation than existing models, which could be reduced by the use of a faster distribution function (such as that proposed by Trowbridge and Reitz [43]), the match between the Torrance-Sparrow model and experiment is impressive.

3.4 Blinn model

Blinn [6] proposed a model that contained a more accurate function for the generation of specular highlights. The model was based on the work of Torrance and Sparrow, whose model provided a good scientific basis for the simulation of reflection. The Blinn model differed from previous models in that the magnitude of the intensity and the position of the specular highlight varied with the angle of incidence. The model accurately predicted *off-specular peaks* at grazing angles of incidence, and provided a better match to experimental data than the Torrance-Sparrow model [16]. The Blinn model produced essentially the same images as the Phong model, except at grazing incidence, but was computationally more expensive and therefore not an attractive alternative [3].

3.5 Whitted model

A model proposed by Whitted [45] in 1980 used global information to calculate reflected intensities. The new model was based on the Blinn model, but included a term for ideal specular reflection from perfectly smooth surfaces. Unlike local illumination models, the Whitted model considers light sources within the scene together with object-object interactions, and therefore produces far more realistic specular components. A simple model for the reflection and transmission from a perfectly smooth surface is provided by ray optics, as can be seen in figure 3.5. The direction of the reflection vector **R**, is given by

$$\theta_i = \theta_r$$

and the direction of the transmission vector **T**, is given by Snell's law, which states that

$$\frac{n_i}{n_t} = \frac{\sin\theta_t}{\sin\theta_r}$$

Where n_i and n_t are the refractive indices of the two materials. If the first substance is air, the equation is simplified by setting n_i to one.

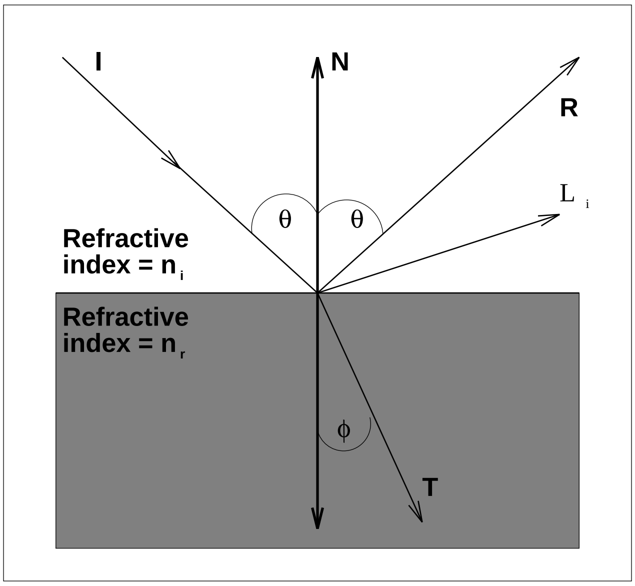

Figure 3.5: Vectors used in the Whitted model

The intensity detected by the viewer consists of specular reflection and transmission, the diffuse term from the Phong model was also included. The new model is:

$$I_r = I_a + K_d \sum_{i=1}^{n} (\mathbf{N} \cdot \mathbf{L}_i) + K_s S + K_t T$$

Where:

S is the intensity of light incident *from* the **R** direction.

T is the intensity of light incident *from* the **T** direction.

$\mathbf{L}_i$ is the ith light vector.

k_t is a transmission coefficient.

k_s and k_t were taken as constants, but for greater accuracy they should include an approximation of the Fresnel reflection law. This model did not spread the specular component like the Phong model, so a random perturbation was added in order to improve realism [45].

3.6 Moravec model

In 1981 Moravec [30] proposed a global illumination model based on the wave model of light . It was claimed that the wave model had important computational advantages for complex scenes. Light was represented by wavefronts, which were stored as two dimensional arrays of complex numbers, and therefore this model was well suited to array processors.

This method eliminated aliasing problems and produced good results for extended light sources and multiple scattering. However, problems of diffraction effects and depth of focus limitations arise, as well as the fact that even the simplest of calculations are computationally very expensive.

3.7 Cook and Torrance model

During the same year, Cook and Torrance [10] proposed an extension of the Torrance-Sparrow model in order to correctly render polished metallic surfaces and to produce the correct specular component at grazing incidence. This model describes the distribution of the reflected light and a colour shift that occurs as the reflectance changes with incident angle. A method was proposed for determining the *spectral energy distribution* of the light reflected from a real material and deriving a colour from this distribution.

This model differs from the Phong model in that:

- It considers the energy, not the intensity, of the incident light and thus the brightness of an object could be related to the intensity and size of the light source [16].

- The specular term is determined from a physically based microfacet model, which used a facet distribution function derived from wave optics and originally developed by Beckmann [5]. The formulation allowed surfaces with multiple roughness levels to be simulated.

- The use of Fresnel's law and measured surface properties mean that the colour variation of the specular component can be determined.

Previous models assumed that reflection consisted of ambient, diffuse and specular components. Ambient and diffuse reflections were the same colour as the surface, and specular reflection was the same colour as the light source. In these models, Fresnel equations may have been used to obtain the angular variation of intensity, but not the colour of the specular component, the resulting models produced realistic images for certain types of materials only.

Existing models were applied at red, green and blue wavelengths, since these are the colours of the monitor phosphors. The RGB values could then be passed directly to the monitor for display. The Cook-Torrance model used wavelength dependent spectral energy distributions and reflectances. This provided a more accurate simulation than one that only considers red, green and blue components. A more detailed discussion of this topic can be found in chapter 4.

The reflected light intensity was determined by the intensity of the light source together with reflectance and surface properties of the material. The *spectral composition* of the reflected light was determined by the spectral composition of the light source and the wavelength dependent reflection from the surface.

Cook and Torrance defined *bidirectional reflectance*, R, to be the ratio of reflected intensity in a given direction to incident energy from another direction. The reflected intensity reaching the viewer from each light source is given by:

$$I_r = RI_i(\mathbf{N} \cdot \mathbf{L})d\omega_i$$

Where $d\omega_i$ is the solid angle of the incident light.

The bidirectional reflectance can be split into specular R_s, and diffuse R_d components, which will have different colours if the material is non-uniform. The bidirectional reflectance is now given by:

$$R = sR_s + dR_d$$

where s + d = 1

The ambient reflectance R_a, was assumed to be constant over the hemisphere above the surface. Therefore, the ambient light reflected in a given direction was the integral of the ambient light intensity times R_a. The ambient light intensity was also assumed to be constant and thus a linear combination of R_s and R_d. In real scenes the ambient intensity depends on the closeness of objects to each other, in order to approximate this a *blocking factor, f*, was introduced, which was the fraction of the hemisphere not blocked by neighbouring objects. If R_a was assumed to be independent of viewing angle, reflected intensity due to ambient light is given by:

$$I_r = I_a R_a f$$

The overall reflection model proposed by Cook and Torrance was therefore:

$$I_r = I_a R_a + \sum_{l=1}^{n} I_{il}(\mathbf{N} \cdot \mathbf{L}_l)(sR_s + dR_d)d\omega_{il}$$

Where n is the number of light sources in the scene.

The specular reflection coefficient depends on the position of the observer, and is determined by assuming that the surface consists of microfacets which each reflect specularly. The form of the specular term is thus very similar to the term introduced by Torrance and Sparrow. All three components of this model exhibit wavelength dependence and several conclusions can be drawn:

- The specular component of reflected light is usually the same colour as the material, not the colour of the light source as assumed in previous models.
- Bidirectional reflectances are needed to simulate different light sources and materials in a scene.
- Plastics and painted objects have diffuse and specular components of different colours.
- Metals have a specular component determined from the light source characteristics and the reflectance of the surface. The diffuse component is often negligible.
- The Fresnel term correctly predicts a colour shift of the specular component at grazing angles of incidence.
- The brightness of an object depends on the intensity and size of the light source.

Chapter 4

He, Torrance, Sillion and Greenberg illumination model

4.1 Details of the new model

In 1991 He, Torrance, Sillion and Greenberg [20] proposed a model which used physical optics to describe specular, directional diffuse and uniform diffuse reflection by a real surface, as can be seen in figure 4.1, this is the second of the two models which are to be compared in this thesis. The mathematics of this model is very complicated and therefore only a brief description of it is given in this chapter, for more information refer to appendix A.

The model is based on Beckmann's [5] model but has the following improvements:

- A vector form of Kirchhoff's theory is used. This allows correct treatment of polarization and Fresnel effects.

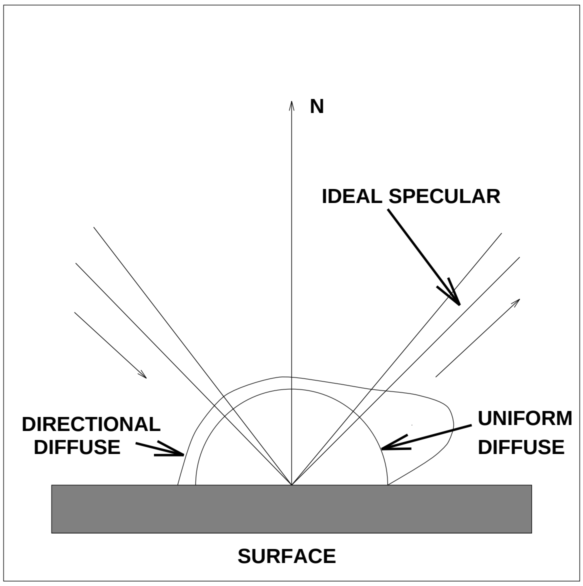

Figure 4.1: Reflection by a Real Surface

- The surface averaging scheme of Stogryn [39] is modified to average only over unshadowed/unmasked portions of the surface. The probability function uses an effective roughness σ, which depends on incident and reflected angles and can be much less than the actual surface roughness σ_0.

- A multiplicative factor proposed by Smith [37] is used to represent shadowing/masking.

The model is represented in terms of a *Bidirectional Reflectance Distribution Function* (BRDF) ρ_{bd}, which is the ratio of incident intensity and reflected intensity in a given direction. The BRDF was proposed to consist of three components, specular, directional diffuse and uniform diffuse:

$$\rho_{bd} = \rho_{bd,sp} + \rho_{bd,dd} + \rho_{bd,ud}$$

Where:

$\rho_{bd,sp}$ is the specular component from the first surface reflection process (reflection by mean surface) and has the form:

$$\rho_{bd,sp} = \frac{\rho_s \Delta}{cos\theta_i d\omega_i}$$

Where ρ_s is the specular reflectivity of the surface, which contains a surface roughness function, a shadowing/masking factor and a Fresnel term. For rough surfaces, the specular term is reduced by the surface roughness and shadowing/masking components. The delta function, Δ, is unity inside the specular cone and zero outside. The solid angle of the incident light ray $d\omega_i$ is included, this is because an incident ray with a large solid angle will produce a small specular highlight. Finally, the $cos\theta_i$ term is included to account for the fact that the projected surface area is reduced as the angle increases.

$\rho_{bd,dd}$ is the directional diffuse component, which represents diffraction scattering from a rough surface, and has the form:

$$\rho_{bd,dd} = \frac{|F|^2 GSD}{\pi cos\theta_i cos\theta_r}$$

Where $|F|^2$ is the Fresnel reflectivity, G is a geometric factor, S is a shadowing/masking factor and D is the microfacet distribution function. Again, the cosine terms are included to account for the reduction in projected surface area as the angle increases.

$\rho_{bd,ud}$ is the uniform diffuse component and has the form:

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

Where $a(\lambda)$ is a free parameter with an upper bound determined by energy conservation,

The specular contribution was converted into an intensity basis by the multiplication of $I_i cos\theta_i d\omega_i$, thus becoming independent of the incident solid angle. The resulting form:

$$\text{Specular term} = \rho_s I_i \Delta$$

is well known and suitable for ray tracing. For rough surfaces, this term is reduced by the roughness factor, contained within ρ_s, and the shadowing/masking factor, S. For smooth surfaces, where the wavelength of the incident light is much greater than the projected surface roughness of the surface, this term is not attenuated. When the wavelength of the incident light is similar to the projected surface roughness, the first surface reflection process introduces interference and diffraction effects, and thus the directional diffuse term becomes important. The uniform diffuse term is small for metallic surfaces with shallow roughness slopes, but may be significant for rougher surfaces or non-metals.

4.2 Colour

Illumination models rely on colour models to convert their results to a displayable form. Most illumination models are applied at three wavelengths, usually in the red, green and blue regions, and the resultant values passed directly to the display medium as RGB values. The new physically based illumination model produces a spectrum as its end result. In order to determine the colour of a given spectrum it is necessary to utilise a colour model which is based on our knowledge of the human visual system. This section discusses the CIEXYZ colour model and explains why it was used in preference to the RGB model.

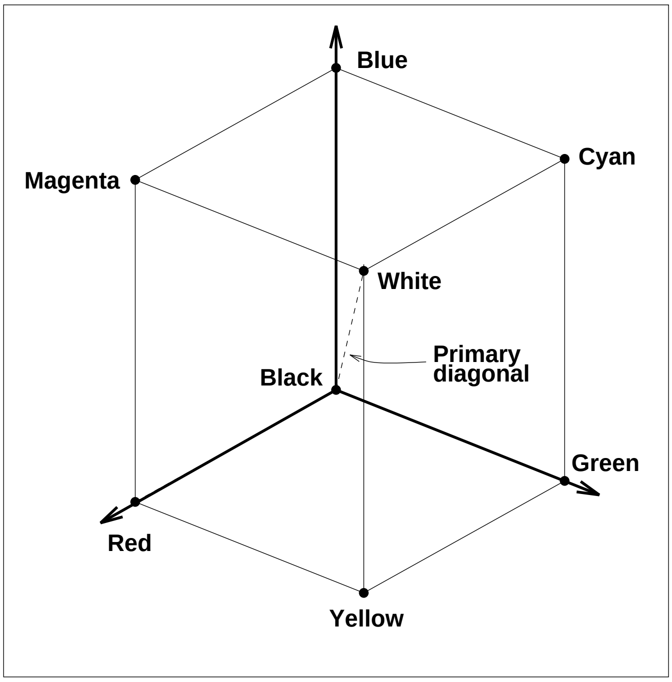

Figure 4.2: RGB Colour Space

4.2.1 RGB

Device gamut models [42, 27] are those used to describe the colour produced by a device in terms of those values input directly into it, the RGB model is an example. The RGB model specifies colour in terms of red, green and blue values passed to the monitor. Since the three values must be in the range zero to unity, representing no voltage and full voltage across the electron guns in the monitor, the RGB colour space can be represented as a cube , as in figure 4.2, where the primary diagonal represents a grey scale.

The RGB model is not suitable for the accurate representation of real colours because:

- The types of phosphors used in the production of monitors varies widely, even individual manufacturers may use different phosphors for each model.

Since the exact colour produced by an RGB triplet depends on the properties of the phosphors used in the monitor, a given RGB triplet will rarely produce the same colour on different monitors.

- It is hardware orientated and in no way does it consider the way in which colour is perceived by the human eye and brain. This means that it is very difficult to specify an exact colour or colour change using the RGB model.
- The model colour space does not represent the total range of colours detectable by the human eye.
- The model colour space is perceptually non-uniform, meaning that uniform intervals along a line in RGB space will not represent uniform colour increments.

In order to accurately represent colour, we must consider its perception by a human observer. Any colour model which is to be used for converting light spectra to a displayable form must be based on this information.

4.2.2 The Perception of Colour

Light energy is converted to nervous impulses by the *retina*, as can be seen in figure 4.3, which lines the inner surface of the eye. There are two types of light sensitive receptor cell in the retina, *rods* and *cones*. Rods are sensitive to very low light levels, reaching their maximum output at only moderate levels, thereafter giving constant output regardless of any increases in light intensity. Cones are less sensitive than rods and can therefore deal with higher light intensities, but become inactive at low intensities.

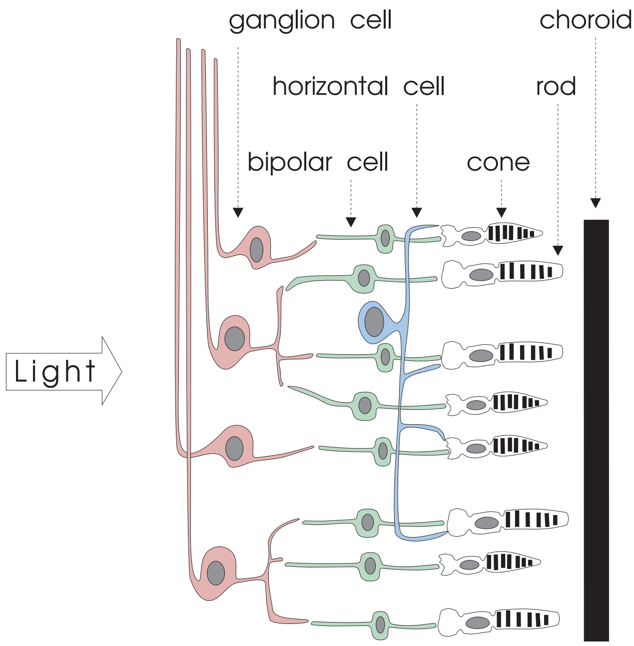

Figure 4.3: The Human Retina (courtesy of Chris Lilley)

The relative efficiency of the retina can be plotted as a function of wavelength, as can be seen in figure 4.4, this is known as the ISO standardized *luminous efficiency*. It represents the perceived brightness of a light as the wavelength is varied, while maintaining a constant intensity level. The portion of the graph between 450 and 500 nm corresponds roughly to blue, 500 to 550 nm corresponds to green and red is between 600 and 700 nm. At low light levels only the rods are active, this is called *scotopic* vision and is most sensitive in the green region of the spectrum. When the light intensity is higher, the rods become overloaded and the cones become active, this is called *photopic* vision [2] and is most sensitive in the yellow/green region of the spectrum.

In addition to being active at higher light intensities, cones are also responsible for the sensation of colour. There are three types of cone in the retina, called S, M and L cones. Each cone type is sensitive to a different range of wavelengths, as can be seen in figure 4.5. The output signals of these three types of cone

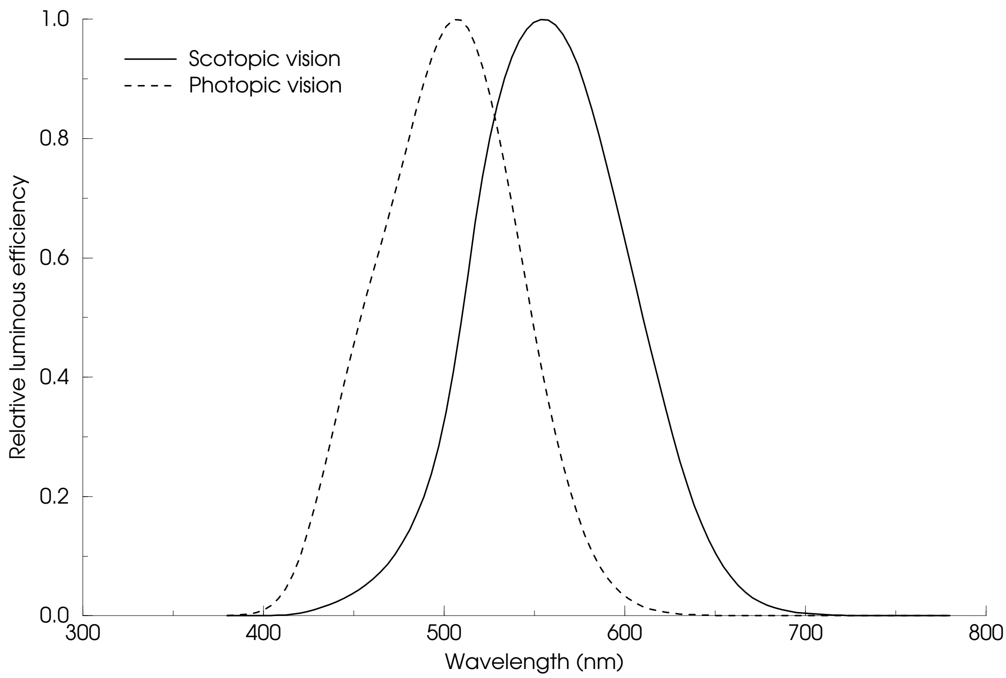

Figure 4.4: Luminous efficiency (courtesy of Chris Lilley)

determine the colour perceived by the human observer.

4.2.3 CIEXYZ

In order to standardize the measurement of colour a different colour model is necessary. Based on a series of experiments, in 1931 the International Lighting Committee (Committee Internationale de l'Eclairage, CIE) defined a *standard observer* [27], this is a set of data defining three primary colours for colour measurement. It states, for each wavelength interval, the amount of each primary required to match any spectrally pure colour for a statistically normal observer. A graph of the amount of each primary required to match any pure spectral colour is called a *matching function*, and is shown in figure 4.6. The three primaries, X, Y and Z are imaginary colours, in that they do not correspond to any colours visible to the human eye. The spectral matching function of the Y primary was

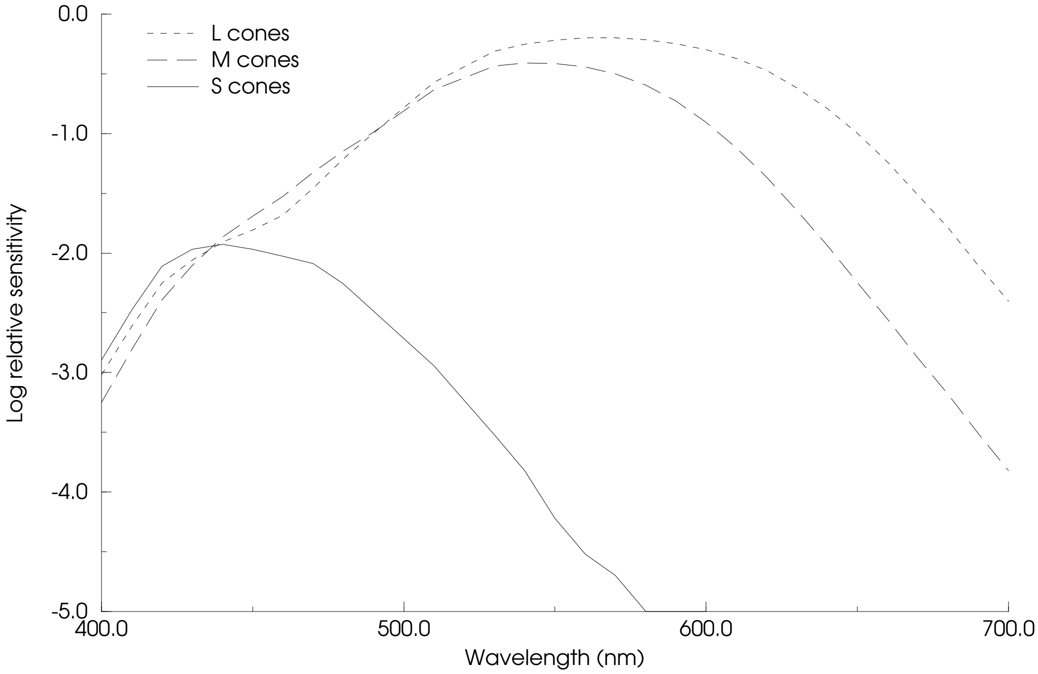

Figure 4.5: Cone sensitivities on a log scale (courtesy of Chris Lilley)

chosen to exactly match the CIE standard photopic luminious efficiency function shown in figure 4.4 and therefore carries all the luminance information about the colour.

The CIEXYZ colour model is far more suitable for the representation of real colours because it allows any colour to be specified in a device independent fashion. It also has the advantage that, all visible colours are within its colour space, meaning that any visible light spectrum can be reduced to a set of XYZ values. The CIE colour space is additive, so it is possible to perform simple operations on colour. This means that colour rendering equations developed for ray-tracing can be applied to colours in CIEXYZ space.

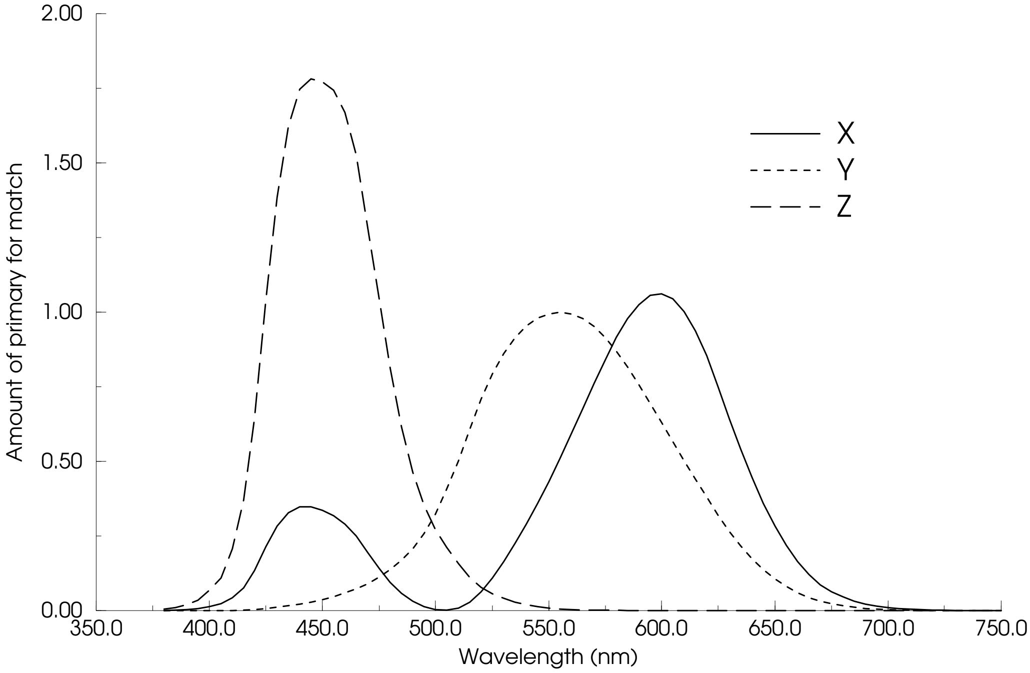

Figure 4.6: CIEXYZ matching functions (courtesy of Chris Lilley)

Chapter 5

Implementation of the new illumination model

In order to compare the two ilumination models, the new model was included in a ray tracer. Details of the implementation of both the ray tracer, and the terms in the model, are given in this chapter.

5.1 Ray tracing

5.1.1 Overview

One of the most important image synthesis methods is based on a technique used in physics. When designing optical systems, physicists trace the path of light rays, from a light source and through the system, this process is called *ray tracing*. In computer graphics, tracing rays from a light source to the observer

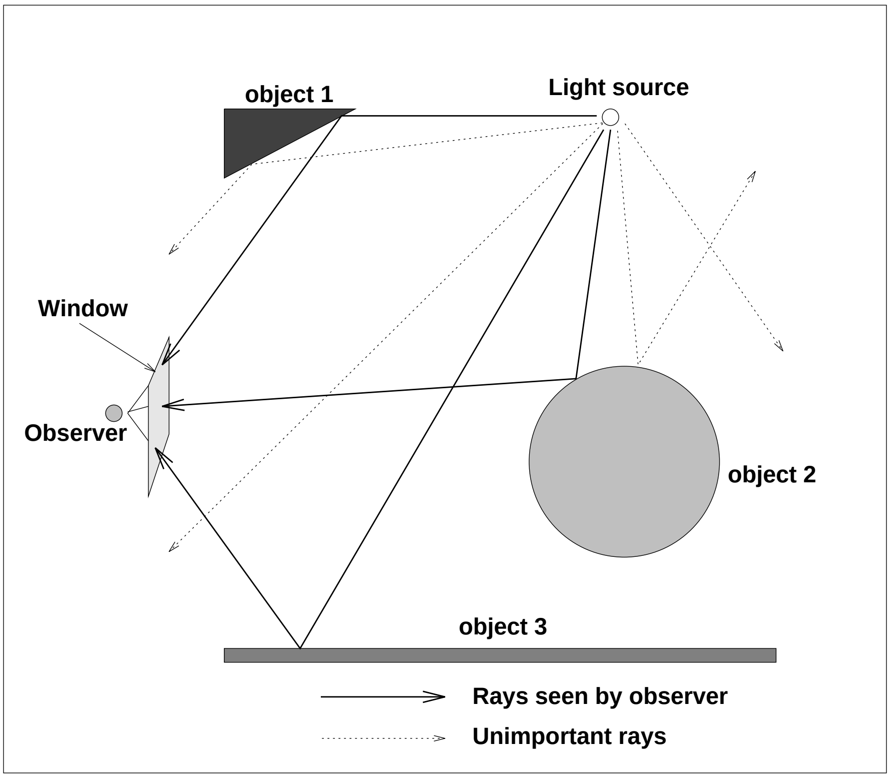

Figure 5.1: Forward ray tracing

is known as *forward ray tracing*. The main problem with forward ray tracing is that many of the rays do not reach the observer, meaning that the computer spends much of its time tracing unimportant rays, as can be seen in figure 5.1. In order to speed up the ray tracing process, we only want to consider those rays which eventually reach the observer. The best way to do this is to start from the observer and trace rays back into the scene, this is known as *backward ray tracing* and is far more efficient, as can be seen in figure 5.2.

In order to ray trace a scene, the position of the observer and a window, containing an arbitrary number of pixels, must be defined, as can be seen in figure 5.3. Pixels are elementary areas of the window and their size limits the resolution of the image. The position of the observer will be the origin of all rays cast into the scene, with at least one ray being cast through each pixel. Some scenes may

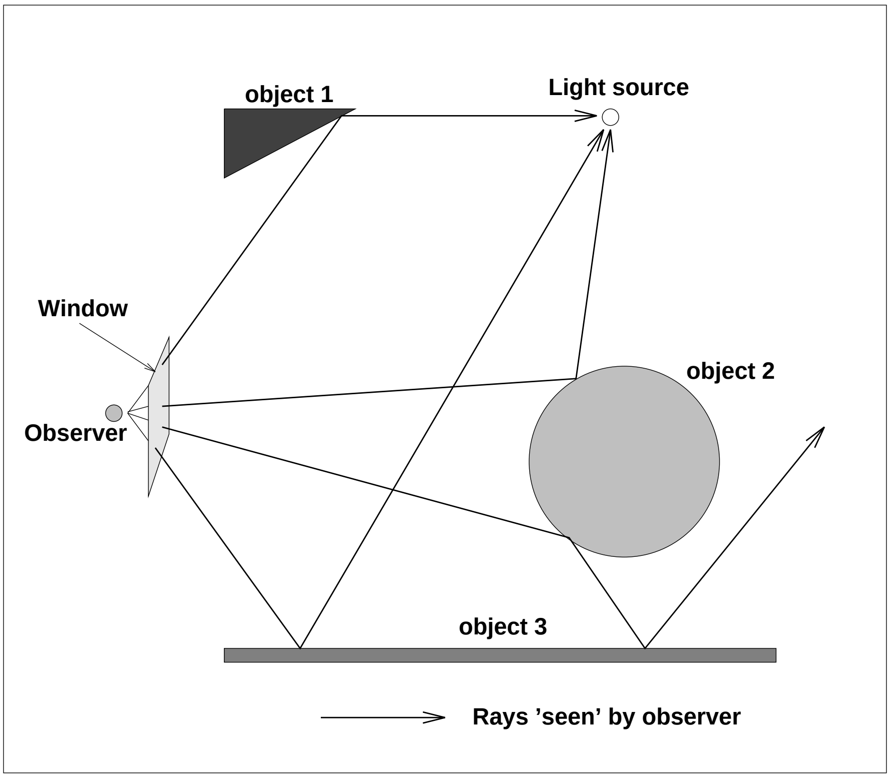

Figure 5.2: Backward ray tracing

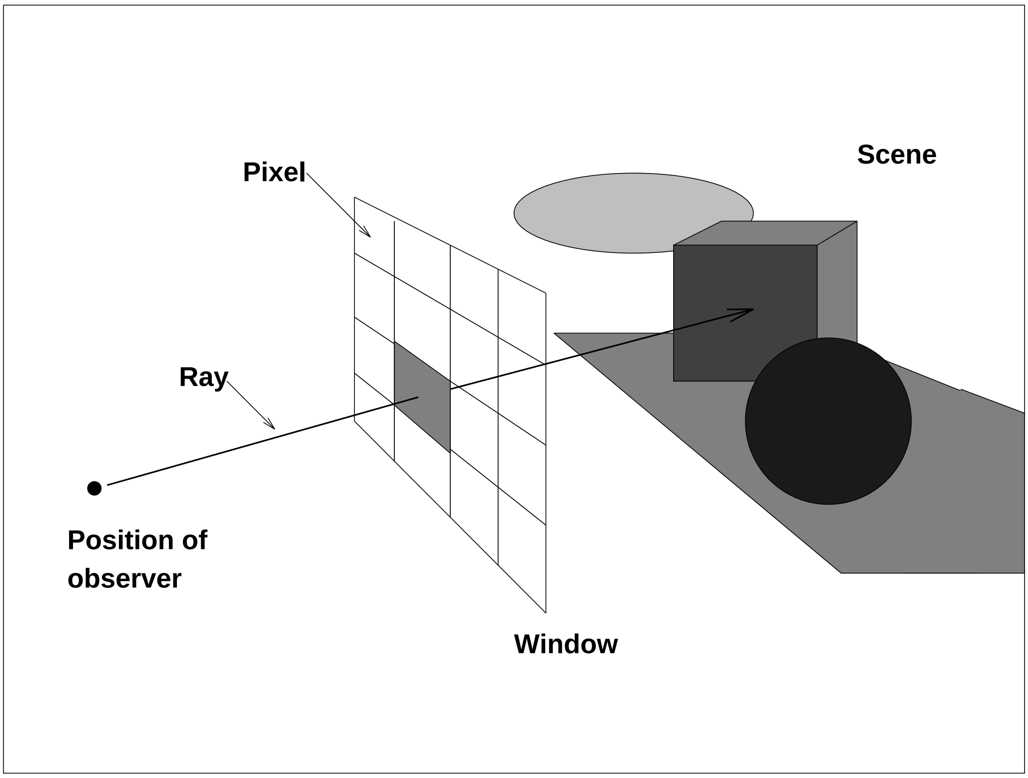

Figure 5.3: Ray tracing a scene

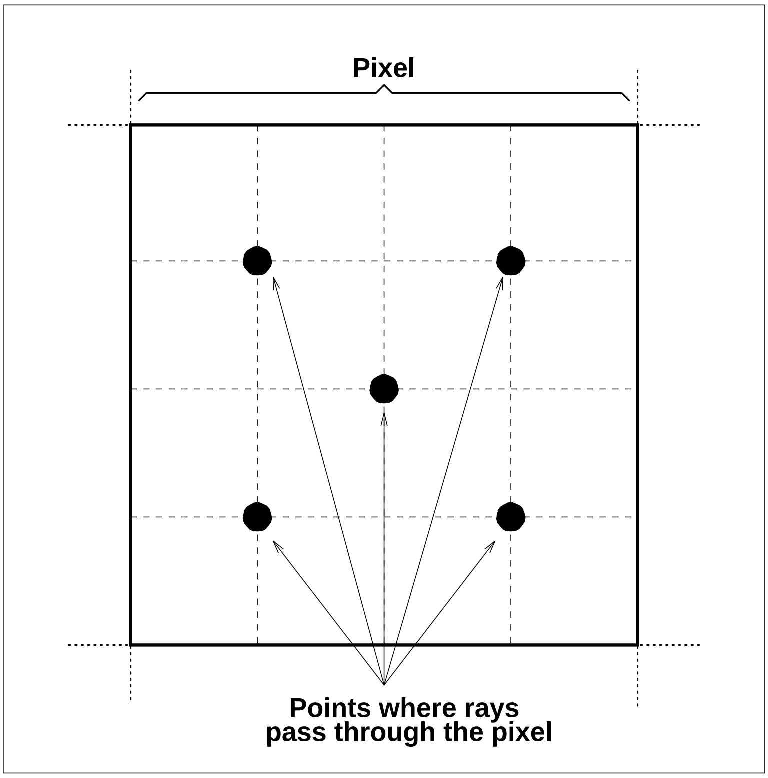

Figure 5.4: Pattern of rays cast through each pixel

be very complicated and multiple rays are often cast through each pixel in order to improve sampling quality. The simplicity of the scenes used to test the new illumination model mean that only five rays per pixel were required. The pattern of the five rays through each pixel was chosen because it provided a good coverage of the window, as can be seen in figure 5.4.

If the ray cast into the scene intersects an object, the illumination model is applied and its result stored. The final colour of each pixel is determined from the average of the results for each of the five rays cast through it. The pseudocode for a simple ray tracer is shown in figure 5.5.

A single reflected ray was used together with point and ambient light sources.

```
for each pixel in the image do
    begin
        determine ray from observer through pixel;
        for each object in the scene do
            if object is intersected and intersection is the closest so far
            then store intersection details;
        apply illumination model using the stored intersection details
        and determine the pixel colour from its result
    end;
```

Figure 5.5: Pseudocode for a simple ray tracer

The reflected intensity is given by:

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

Where:

n is the total number of light sources in the scene.

Subscript i denotes the ith light source.

ρ_{hd} is the hemispherical directional reflectivity of the surface.

5.1.2 Sphere intersection

At the heart of every ray tracer is a set of intersection routines, which determine whether a ray intersects a certain object. For the sake of simplicity, only a sphere intersection routine was included in the ray tracer. There are two methods to determine sphere intersection, algebraic and geometric. The algebraic solution is simpler, but the geometric solution is more efficient, because it determines

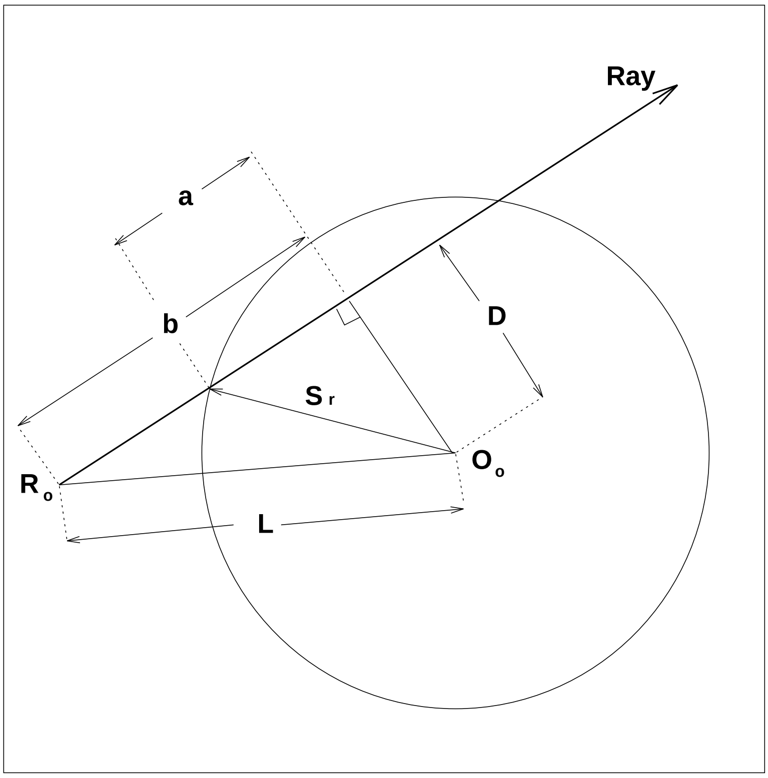

Figure 5.6: Geometry of sphere intersection

whether there is an intersection before calculating the intersection point. The stages in the geometric solution to sphere intersection are as follows and reference should be made to figure 5.6:

- Determine if the origin of the ray is outside the sphere. This is done by finding the distance between the ray origin R_o, and the centre of the sphere O_o, and comparing it to the radius of the sphere S_r. If the distance L, between them is less than the radius S_r, the ray originates from within the sphere and the calculation can be stopped.

- If the distance $a^2 = S_r^2 - L^2 + b^2$ is greater than zero, the ray must hit the sphere with an intersection distance of $t = b - \sqrt{a}$.

5.2 Scene description

In order to describe the lights and objects within a scene, the Manchester Scene Description Language (MSDL) [31] was used. However, due to the unusual input parameters required by the new illumination model, MSDL had to be slightly extended. MSDL data files are plain text files, the text file in appendix B shows the data necessary to define a copper sphere at and a D6500 point light source. The MSDL reader parses the file, filling its internal data structures, which were eventually passed to my application.

5.3 Spectral samples

The proposed model is wavelength dependent and so in order to implement it, the spectra of the light sources and the surface spectra of any objects in the scene must be sampled at a number of wavelengths [32]. To produce an accurate image the spectra should be sampled as many times as possible, as can be seen in figure 5.7. However, for each point to be shaded in the image, the illumination model must be applied once for each spectral sample, and so the number of samples should be kept to a minimum in order to reduce computation. Therefore, the number of samples must be a compromise between these two factors, Hall [18, 19] showed that nine spectral samples is a suitable number.

The wavelengths of the nine spectral samples [29] must be chosen carefully in order to extract the maximum amount of information from the spectra. It was decided that the wavelengths should be those which allow the most accurate

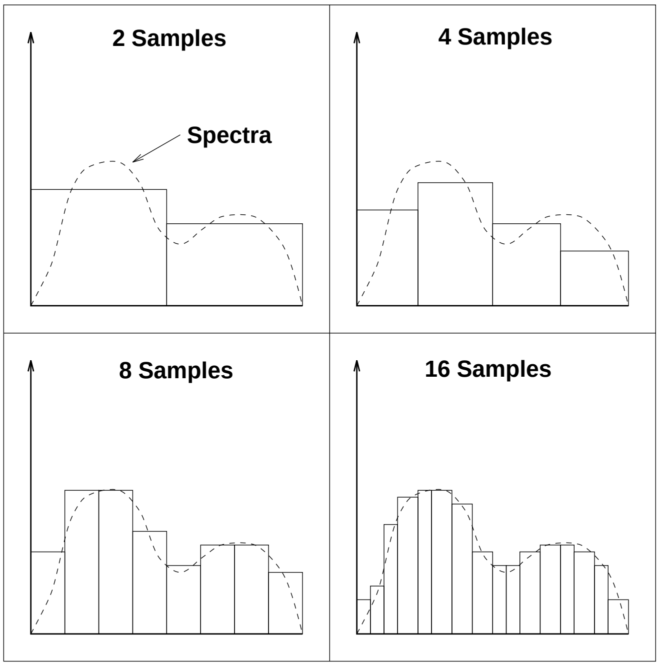

Figure 5.7: The Effects of Different Numbers of Spectral Samples

conversion of the spectra into CIE tristimulus values. The wavelengths therefore had to be spaced so as to evenly sample the XYZ matching functions, as can be seen in figure 5.8. The wavelengths selected are shown in table 5.1.

Wavelengths in nm
415.0
445.0
475.0
500.0
520.0
550.0
580.0
600.0
635.0

Table 5.1: Selected wavelengths

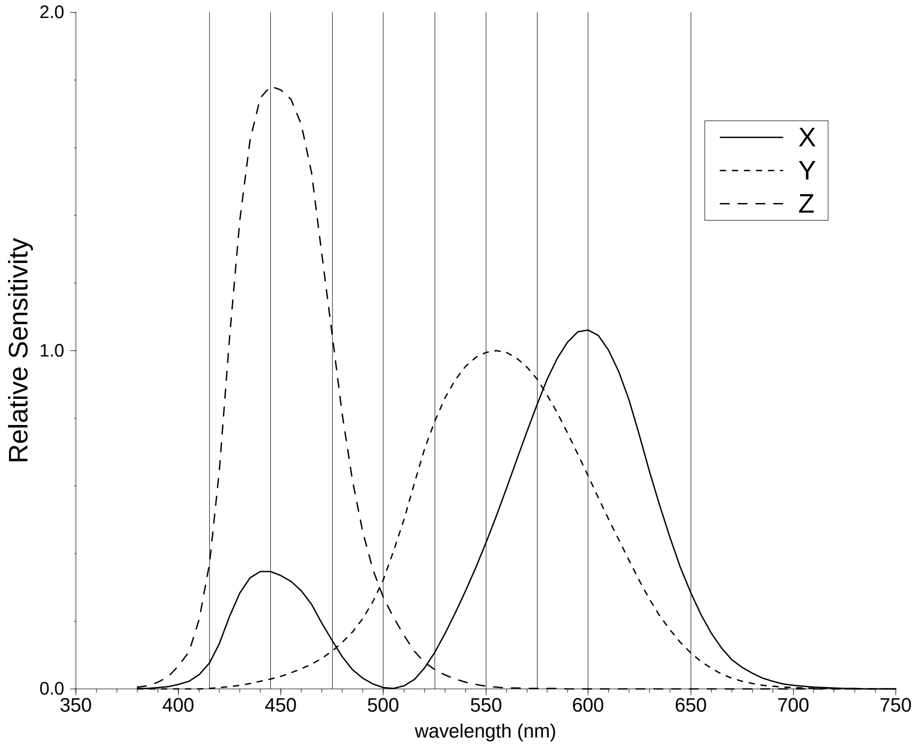

Figure 5.8: Wavelength selection (courtesy of Chris Lilley)

5.4 Conversion of spectral samples to an RGB triplet

The spectrum resulting from the application of the illumination model must be converted into an RGB triplet [13] for display on a monitor. In order to do this, the spectra must first be converted into CIE tristimulus values. This involves two stages, the first is the multiplication of the spectra by each of the matching functions, thus producing three intermediate spectra. These intermediate spectra must then be integrated, with each integral yielding one of the CIE tristimulus values, as can be seen in figure 5.9. The tristimulus values now represent the colour which would have been detected if the initial spectra was observed by the eye.

These values must now be converted to an RGB triplet which when displayed

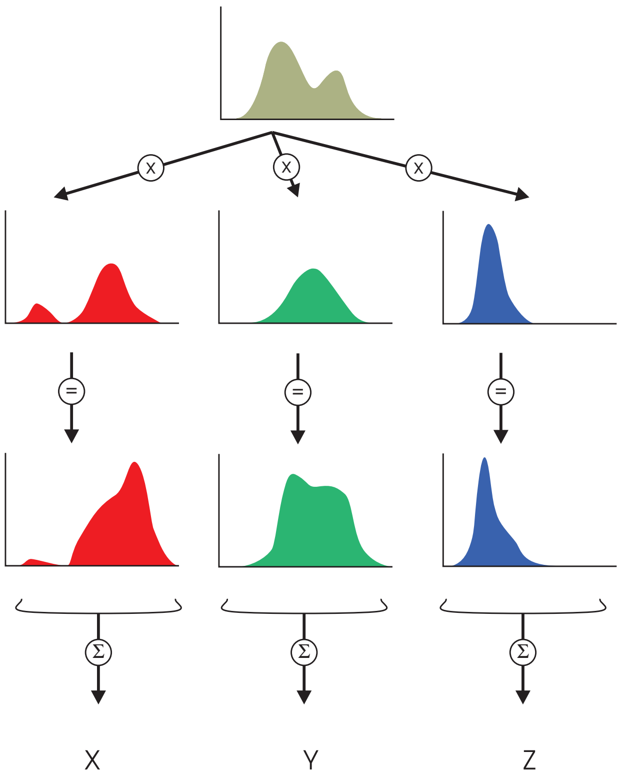

Figure 5.9: Computing CIE tristimulus values (courtesy of Chris Lilley)

x_R	y_R	x_G	y_G	x_B	y_B	x_W	y_W
0.625	0.340	0.285	0.605	0.150	0.065	0.283	0.298

Table 5.2: Details of Monitor Phosphors

on a monitor, will generate the colour represented by the initial spectra. To do this, the exact colour of the monitor phosphors must be known. Manufacturers supply information, in terms of CIE tristimulus values, about the exact colour of the three monitor phosphors and the white colour produced when the red, green and blue are all at maximum intensity. Table 5.2 shows the information for a Hewlett-Packard 730 with A1097C monitor (post November 1991 type).

Since the chromacity values are given in x,y pairs, the z components must be calculated using:

$$z = 1 - x - y$$

These values can then be inserted into a matrix, A:

$$A = \begin{vmatrix} x_r & x_g & x_b \\ y_r & y_g & y_b \\ z_r & z_g & z_b \end{vmatrix}$$

This matrix must then be inverted to give a new matrix, B:

$$B = A^{-1}$$

Since absolute luminance is not required, the monitor white point is converted to relative XYZ values by setting Y to unity:

$$X_w = \frac{x_w}{y_w}$$

$$Y_w = 1$$

$$Z_w = \frac{z_w}{y_w}$$

These values can then be used to determine the column vector, C:

$$C = B \cdot \begin{vmatrix} X_w \\ Y_w \\ Z_w \end{vmatrix}$$

The elements of C are then used to generate a new matrix, D:

$$D = \begin{vmatrix} C_1 & 0.0 & 0.0 \\ 0.0 & C_2 & 0.0 \\ 0.0 & 0.0 & C_3 \end{vmatrix}$$

The matrix D, must then be multiplied by the matrix A, in order to produce a new matrix E, the inverse of which F, can be used to convert from XYZ to RGB:

$$E = A \cdot D$$

$$F = E^{-1}$$

$$\begin{vmatrix} R \\ G \\ B \end{vmatrix} = F \cdot \begin{vmatrix} X \\ Y \\ Z \end{vmatrix}$$

The resulting RGB triplet can then be passed to the monitor for display.

5.5 Gamma correction

Figure 5.10 shows a graph of the perceived intensity, I_{RGB}, of a pixel on a monitor, plotted against V_{gun}, the voltage across the cathode ray tube being used to direct electrons onto the monitor phosphors. The value of V_{gun} is a device dependent number proportional to the RGB values pased to it. However, the perceived intensity is not proportional to V_{gun} and so neither is it proportional to the RGB values passed to it. The relationship between I_{RGB} and V_{gun} is:

$$I_{RGB} = kV_{gun}^{\gamma}$$

Where k and γ are both constants and γ usually lies between 2.2 and 2.5 .

In order to account for this non-linear proportionality, the images produced by the new illumination model were gamma corrected using a value of 2.2 [12]. Modern image display software allows the user to perform this correction as the image is displayed.

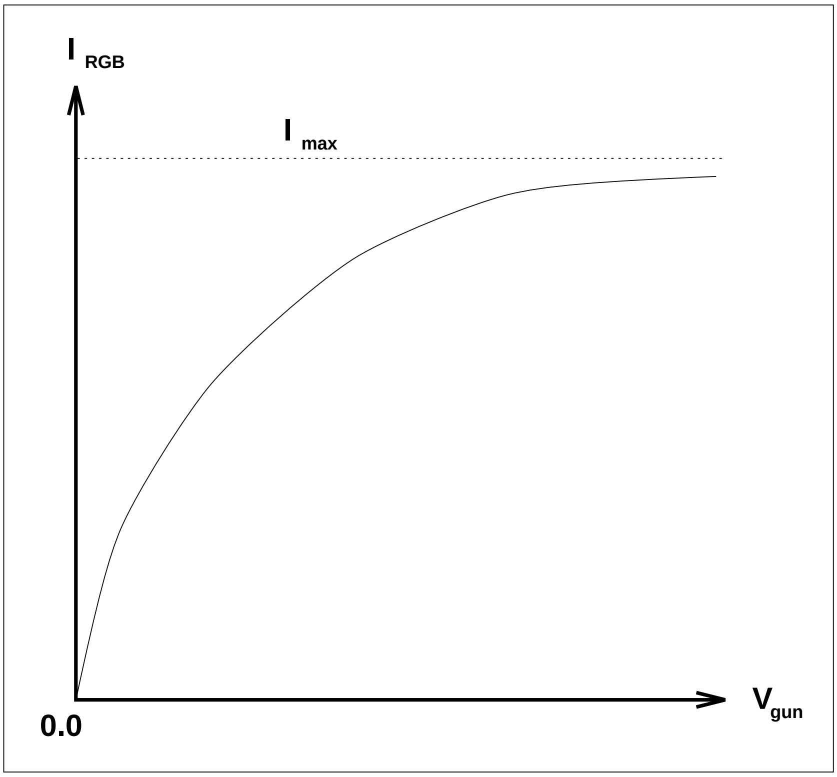

Figure 5.10: The Need for Gamma Correction

5.6 Fresnel term

The Fresnel term determines the fraction of light reflected and transmitted at a boundary, and is therefore of great importance in a physically based illumination model. The calculation is not a complicated one, but it requires information which is not available. It is therefore necessary to approximate this term using the available information.

5.6.1 Theory

The fraction of light reflected from a surface depends upon the refractive indices of the material and the angle of the incoming light to the surface normal, as can be seen in figure 5.12. It also depends on the polarization of the incident light, but this model assumes incident light to be randomly polarized and so this dependence is not significant.

For dielectrics the Fresnel relationship has the form:

$$F = \frac{1}{2}\left(r_{parallel}^2 + r_{perpendicular}^2\right)$$

Where $r_{parallel}$ and $r_{perpendicular}$ are the ratios of the amplitudes of the reflected wave to the incident wave, for light polarized parallel and perpendicular to the plane of the **L** and **N** vectors. For light passing from air to a material of refractive index n:

$$r_{parallel} = \frac{n(\mathbf{N}\cdot\mathbf{L}) + (\mathbf{N}\cdot\mathbf{T})}{n(\mathbf{N}\cdot\mathbf{L}) - (\mathbf{N}\cdot\mathbf{T})}$$

$$r_{perpendicular} = \frac{(\mathbf{N}\cdot\mathbf{L}) + n(\mathbf{N}\cdot\mathbf{T})}{(\mathbf{N}\cdot\mathbf{L}) - n(\mathbf{N}\cdot\mathbf{T})}$$

The Fresnel relationship for conductors is far more complicated, because conductors absorb light and are therefore opaque. A useful approximation to the Fresnel relationship for conductors is given below and is valid when $n^2 + k^2 \gg 1$, where k is the absorption coefficient of the conductor.

$$r_{parallel} = \frac{(n^2+k^2)(\mathbf{N}\cdot\mathbf{L})^2 - 2n(\mathbf{N}\cdot\mathbf{L}) + 1}{(n^2+k^2)(\mathbf{N}\cdot\mathbf{L})^2 + 2n(\mathbf{N}\cdot\mathbf{L}) + 1}$$

$$r_{perpendicular} = \frac{(n^2+k^2) - 2n(\mathbf{N}\cdot\mathbf{L}) + (\mathbf{N}\cdot\mathbf{L})^2}{(n^2+k^2) + 2n(\mathbf{N}\cdot\mathbf{L}) + (\mathbf{N}\cdot\mathbf{L})^2}$$

For more information on Fresnel reflectance refer to [18, 21, 15].

5.6.2 Implementation

The above equations require the values of refractive index,n, and absorption coefficient k, at every wavelength being sampled in the illumination model, since this information is rarely available a new approach must be found. The Fresnel reflectivity of a surface is far more readily available than the previously mentioned quantities. However, published spectral reflectance curves [47], such as those in figure 5.13, tend to be measured at only one angle of incidence, usually only a few degrees, so a method had to be found for determining the reflectance at other angles from this data.

A suitable approximation method was proposed by Cook [9, 10]. It requires a spectral curve, an average refractive index and, for metals, an average coefficient of absorption. An average reflectance $F_{ave}(\theta)$, was calculated at the current angle of incidence, using the Fresnel relationships and average values previously mentioned. This was then used as a basis for interpolation, the results of which can be seen in figure 5.11..

The Fresnel reflectance $F(j,\theta)$ for a particular spectral sample at a given angle θ, is interpolated from the spectral curve samples at normal incidence $F(j,0)$. The average reflectance at normal incidence $F_{ave}(0)$, and the average reflectance at the given angle $F_{ave}(\theta)$ are also used:

$$F(j,\theta) = F(j,0) + (1 - F(j,0))\left[\frac{F_{ave}(\theta) - F_{ave}(0)}{1 - F_{ave}(0)}\right]$$

This approximation works well for incident light in air. The spectral curves for

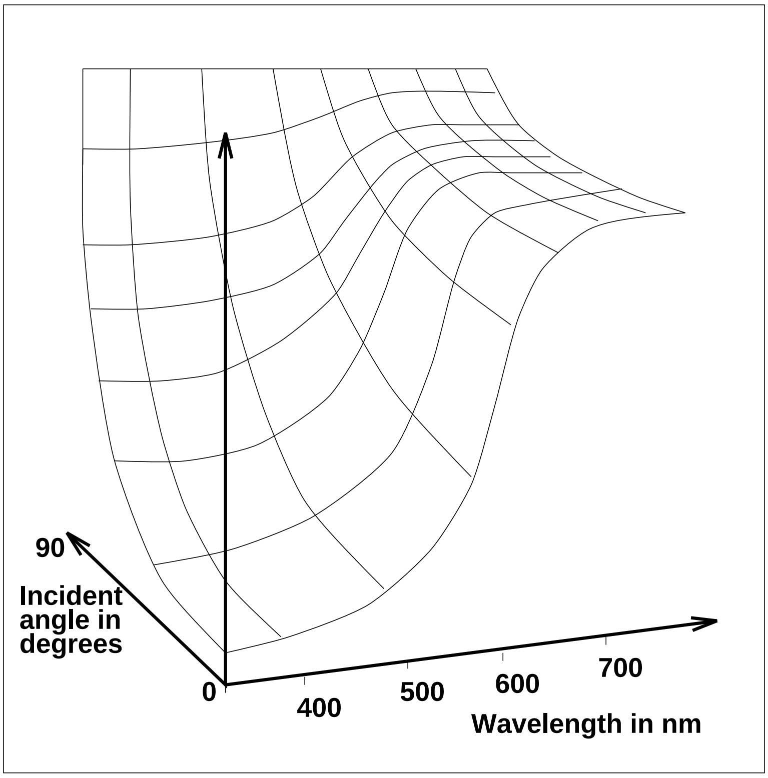

Figure 5.11: Approximated Fresnel Reflectance for Copper using the measured values of n = 0.617 and k = 2.630

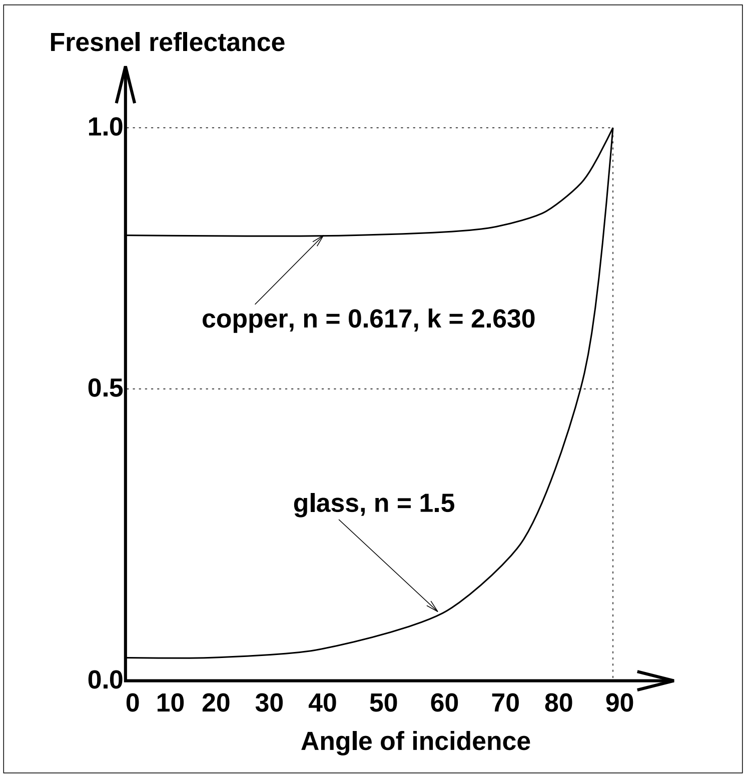

Figure 5.12: Variation of Reflectance with Angle

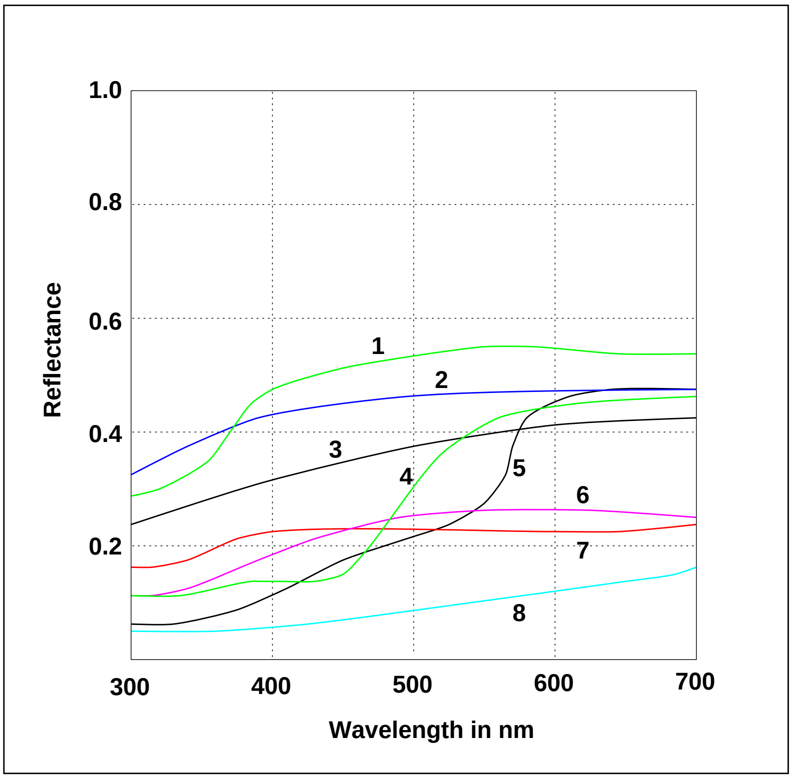

Figure 5.13: Spectral Reflectance Curves: (1) galvanized steel; (2) aluminium; (3) stainless steel; (4) brass; (5) copper; (6) zinc; (7)lead; (8) magnesium alloy

dielectrics are almost constant across the visible region of the spectrum. In this case the approximation should be very close to the actual behaviour of the material. Conductors have more complex spectral curves due to absorption, but the approximation is still fairly close to the actual behaviour. The approximation produces correct results at normal and grazing incidence, the error in approximated behaviour cannot be estimated until more comprehensive spectral data is available.

5.7 Facet distribution function

The facet distribution function is a measure of the number of facets facing in a given direction, and is therefore an indication of the smoothness of the surface.

The distribution function is an important term is the directional diffuse component of the new illumination model. It depends on surface roughness and the wavelength of the incident light, and has the following form:

$$D = \frac{\pi^2\tau^2}{4\lambda^2} \cdot \sum_{m=1}^{\infty} \frac{g^m e^{-g}}{m! \cdot m} \cdot exp\left(-\frac{v_{xy}^2 \tau^2}{4m}\right)$$

The infinite summation term makes this difficult to implement. In order to approximate its true value, a threshold had to be set, below which the summation was ignored. The threshold was set to one percent of the value produced with $m = 1$, as this produced reasonable results.

5.8 Effective surface roughness

The effective surface roughness is the roughness of the illuminated and visible parts of the surface. Its form is very complicated, and therefore its implementation is worthy of discussion.

It has the following form:

$$\sigma = \frac{\sigma_0}{\sqrt{1 + \frac{Z_0^2}{\sigma_0^2}}}$$

Where σ_0 is the actual roughness of the surface and Z_0 is the root of the following equation:

$$\sqrt{\frac{\pi}{2}} Z = \sigma_0 K \cdot exp\left(\frac{-Z^2}{2\sigma_0}\right)$$

Where:

$$K = K_i + K_r$$

$$K_i = \frac{1}{4} \tan\theta_i \cdot erfc\left(\frac{\tau}{2\sigma_0} \cos\theta i\right)$$

Parameter	value
p	0.327591100
a_1	0.254829592
a_2	-0.284496736
a_3	1.421413741
a_4	-1.453152027
a_5	1.061405429

Table 5.3: Parameters used in $erfc(\beta)$ approximation

$$K_r = \frac{1}{4}\tan\theta_r \cdot erfc\left(\frac{\tau}{2\sigma_0}\cos\theta r\right)$$
$$erfc(\beta) = \frac{2}{\sqrt{\pi}}\int_{\beta}^{\infty} e^{-t^2} dt$$

There are two things which make this very hard to implement. The first is the fact that the expression for $erfc(\beta)$ is insoluble and must therefore be approximated. The second is the fact that the solution to the above equation for Z is numerical and must be determined via an iterative process.

The approximation used for the $erfc(\beta)$ function has the form:

$$erfc(\beta) = \left(a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5\right) exp(-\beta^2)$$

Where $t = \frac{1}{1+px}$ and the values of all other parameters are shown in table 5.3.

The error in this approximation has been estimated [1] to be less than 10^{-7}.

By looking at the equation for Z, it is clear that the solution lies in the range zero to $\sigma_0 K\sqrt{\frac{2}{\pi}}$. The pseudocode for the iterative process is shown in figure 5.14.

```
Set max and min for soulution
set tolerance
set guess to average of max and min
substitute guess into equation for Z
WHILE difference between two sides is greater than the tolerance DO {
        IF right hand side ia greater than the left hand side {
                min = guess;
        } ELSE {
                max = guess;
        }
        guess = average of max and min;
        substitute guess into equation for Z
}
```

Figure 5.14: Pseudocode of iterative procedure

5.9 Specular term

The specular component from the first surface reflection process $\rho_{bd,sp}$ (reflection by mean surface) has the form:

$$\rho_{bd,sp} = \frac{\rho_s \Delta}{cos\theta_i d\omega_i}$$

As mentioned previously, Δ is a delta function which is unity inside the specular cone of reflection and zero otherwise. In order to determine whether the reflected ray is inside the specular cone, its solid angle must be considered. The reflected solid angle is the same as the incident solid angle. The view vector **V**, must be subtracted from the reflection vector **R**, about which the specular cone extends. Since both of these vectors are of unit magnitude, the difference vector is a measure of the proximity of the two vectors, as can be seen in figure 5.15. The magnitude of the difference vector must then be squared and multiplied by π in

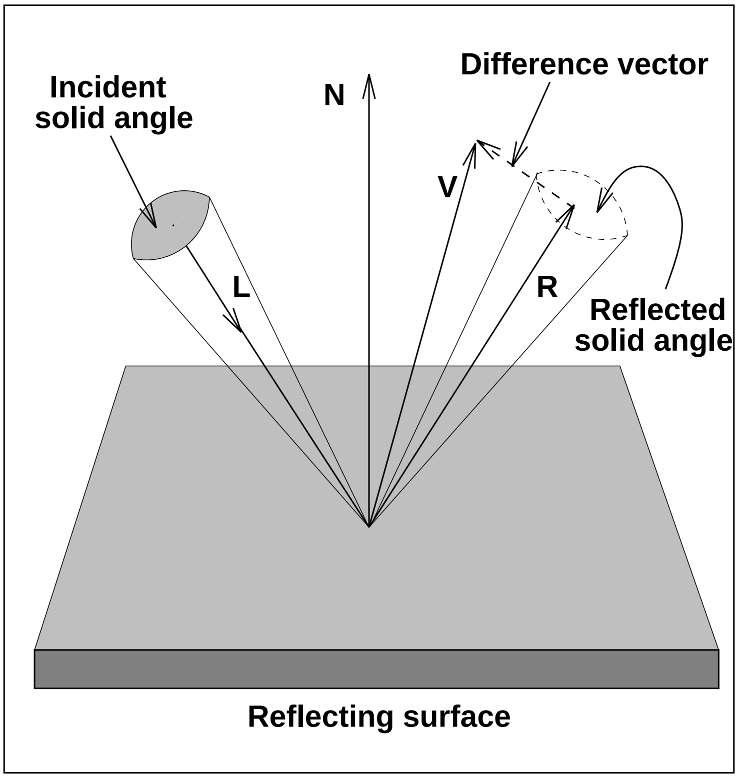

Figure 5.15: Specular Reflection from a Surface

order to determine the area of the circle it sweeps out. If the area of this circle is less than the solid angle of the specular cone, then the view vector lies within the cone and a highlight will be observed.

5.10 Ambient term

In order to calculate a physically correct ambient term for the illumination model, the bidirectional reflectance distribution function (BRDF) must be integrated over the incident hemisphere, refer to figure 5.16. However, the BRDF is very complicated, as can be seen in appendix A, and the calculation would be computationally expensive. The physically correct ambient term would be of the

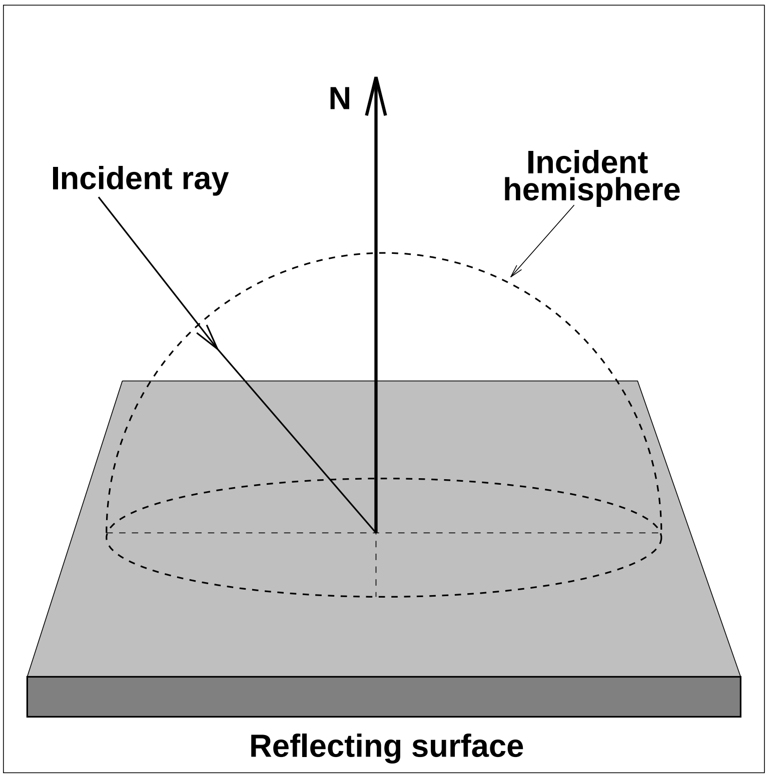

Figure 5.16: The incident hemisphere

form:

$$I_r(\lambda) = I_a(\lambda) \cdot \rho_{hd}(\lambda)$$

Where:

$I_a(\lambda)$ is the ambient light intensity in the scene.

$\rho_{hd}(\lambda)$ is the hemispherical directional reflectivity of the surface and is the result of integrating the BRDF over the incident hemisphere.

However, if the surface were to reflect all incident light equally, the ambient term would produce a white surface. The Fresnel reflectivity, $F(\lambda)$, must clearly be taken into account when approximating the ambient term in this model. Since ambient illumination, by definition, comes from every direction, the angular variation of the Fresnel reflectivity is of no importance. It is therefore possible to

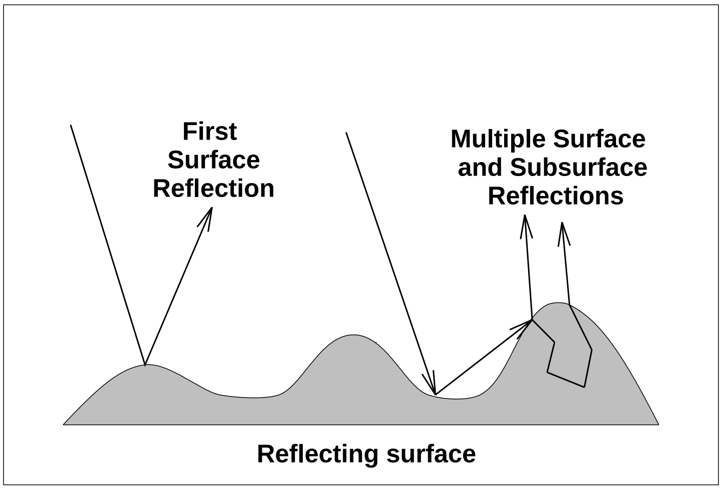

Figure 5.17: Reflection at a surface

approximate the ambient bidirectional reflectance, $\rho_{bd,a}$, to:

$$\rho_{bd,a} = \kappa \cdot F(\lambda)$$

Where κ is a constant in the range zero to one.

5.11 Uniform diffuse term

Uniform diffuse reflection is due to multiple surface and subsurface reflections, as can be seen in figure 5.17. Light reflected in this way is is more difficult to describe than light reflected by the first surface reflection process. Using geometrical optics, the multiple reflection process within surface V-grooves has been estimated [22, 38]. Also, estimates of the subsurface scattering process have been made [36]. The results suggest that the light reflected by these two processes is almost directionally uniform. Therefore, the contribution to the reflected light from these two processes is approximated as uniform diffuse reflection.

The uniform diffuse reflection coefficient, $a(\lambda)$, can be estimated using one of the following methods:

- *Theoretically*: if the V-groove geometry is applicable or if the subsurface scattering parameters are known.
- *Experimentally*: The bidirectional reflectance distribution function is integrated over the reflecting hemisphere. The resulting value is equal to the hemispherical directional reflectivity ρ_{dh}, which can be easily measured using an *integrating sphere reflectometer*.

However, in the absence of subsurface scattering parameters and experimental measurements, $a(\lambda)$ was assumed to be a constant.

Chapter 6

Results and conclusions

6.1 Results

The new illumination model was implemented as described in chapter 3 and the resulting images are shown here, along with similar images from PHIGS. The plates are organised into groups, each of which illustrates a well defined effect, such as ambient lighting. Comparisons will be made within each group.

The illumination models were applied to different scenes, although both scenes contained spherical surfaces. The differences in the scenes are unimportant, it is the realism of the surfaces which is of interest.

The scene used for the PHIGS images consisted of red, green and blue spheres in front of a white polygon, with multiple light sources. Spot light sources illuminated the corner of the polygon and the upper surface of the spheres, a point light source was located in the bottom right corner of the scene and there was

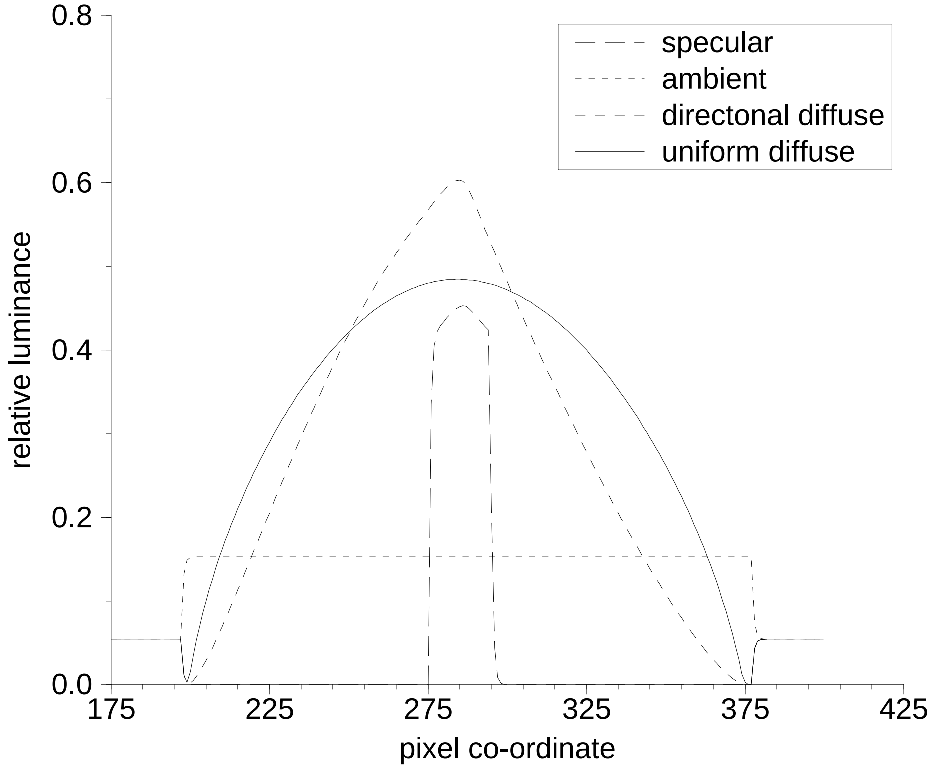

Figure 6.1: Specular, directional diffuse, uniform diffuse and ambient terms for Copper Sphere

also an ambient light source.

The scene used to illustrate the new illumination model consisted of copper and aluminium spheres with a blue background. A D6500 point light source was situated in front of the copper sphere, and an ambient light source was also included.

The contributions from each of the terms will now be considered separately. In order to clarify the discussion of these terms, graphs are included which show the relative luminance of the pixels on a line through each of the spheres. Figures 6.4 and 6.3 show the aluminium sphere, and figures6.2 and 6.1 show the copper sphere.

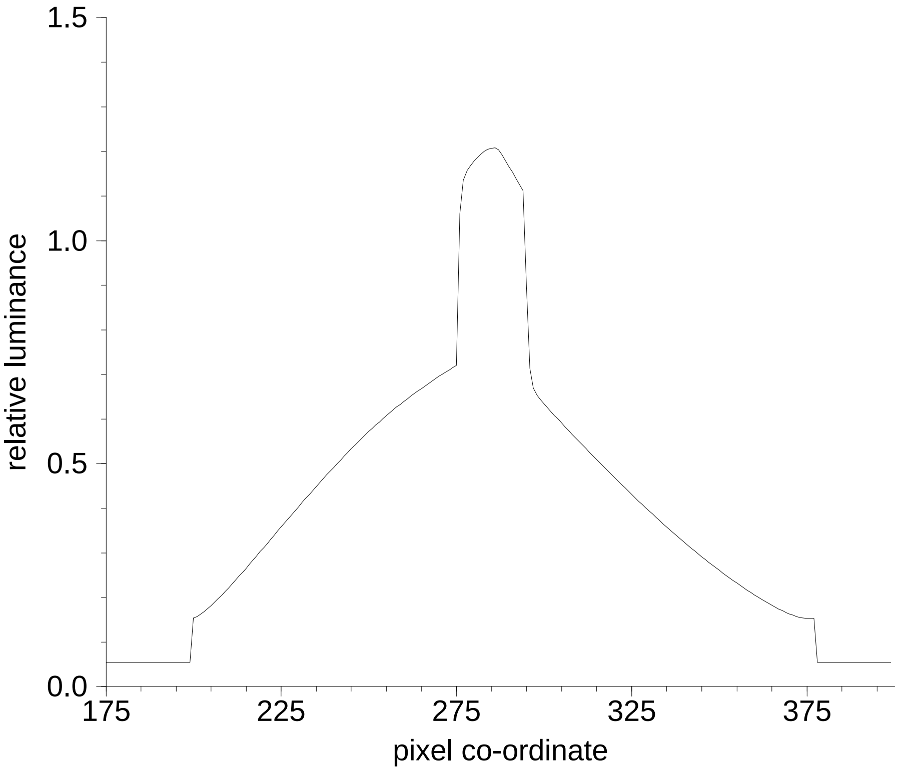

Figure 6.2: Sum of specular, directional diffuse, uniform diffuse and ambient terms for copper sphere

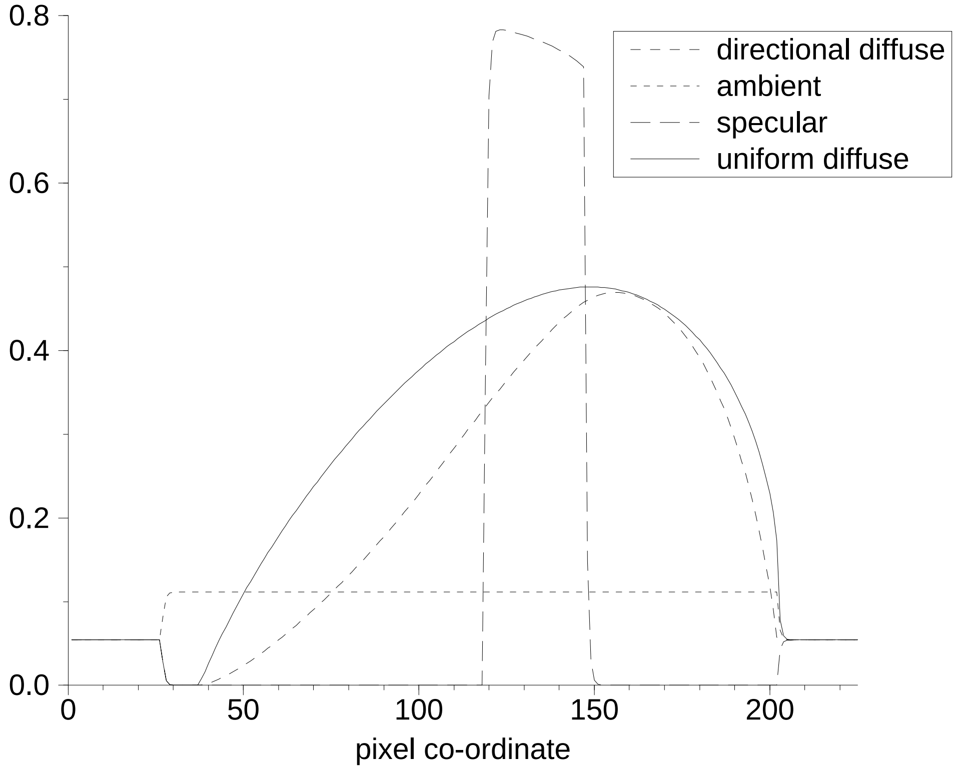

Figure 6.3: Specular, directional diffuse and ambient terms for aluminium sphere

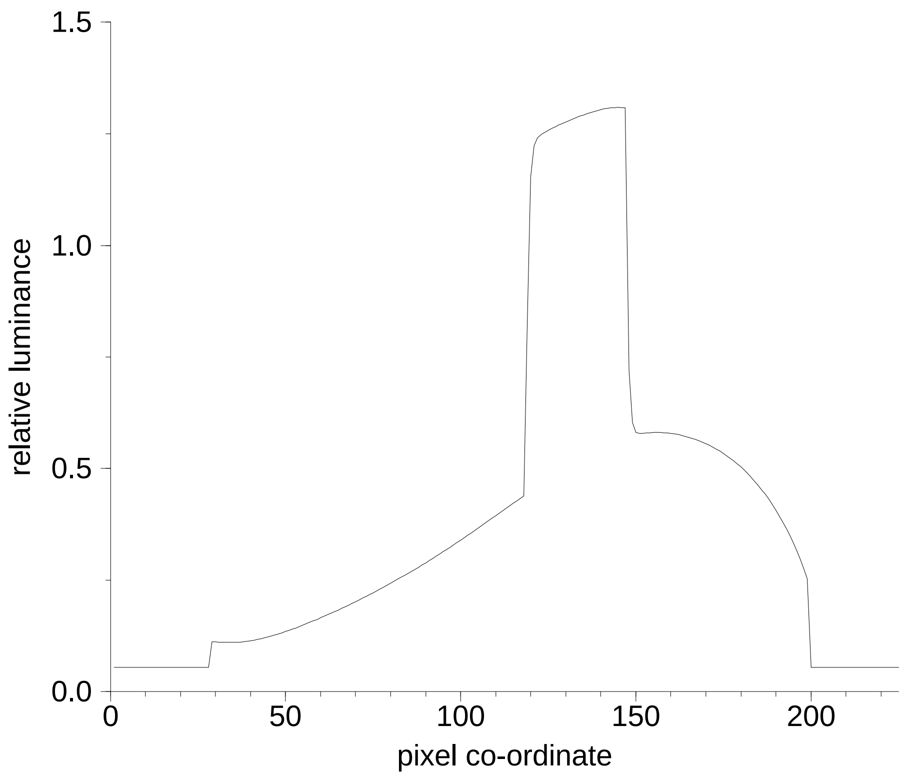

Figure 6.4: Sum of specular, directional diffuse and ambient terms for aluminium sphere

6.1.1 Ambient images

The ambient terms of the two models are very similar, as can be seen in plates 1 and 2. This is because both models use similar approximations to generate these terms. The only difference being that the physically based model uses the Fresnel reflectivity of the surface to produce the correct colour, and the PHIGS model uses red, green and blue ambient reflection coefficients.

6.1.2 Diffuse images

The uniform diffuse term of the new model produces monochrome images, as can be seen in plate 3. The lack of colour is not due to any short fall in the model itself, but rather a lack of the subsurface scattering parameters and experimental physical ambient

Plate 1: Physically based ambient term

PHIGS ambient

Plate 2: PHIGS ambient term

physical uniform diffuse

Plate 3: Physically based uniform diffuse term

PHIGS diffuse

Plate 4: PHIGS diffuse term

physical directional diffuse

Plate 5: Physically based directional diffuse term

measurements required by the model. This lead to the term being approximated to a constant, and the resulting wavelength independent reflection yielded light of the same colour as the light source. The diffuse term of the PHIGS model produces far better images, as can be seen in plate 4. Although both images show similar shading, the presence of colour in the PHIGS image makes it appear more realistic.

The directional diffuse term of the new model produces very good results, as can be seen in plate 5. This picture is similar to the PHIGS diffuse image, but differs in certain important aspects. The main difference is that the directional diffuse term shows a slight increase in intensity where the specular component would be expected. This can be seen more clearly in figure 6.1, which shows the relative luminosity of each pixel along a line through the copper sphere. This increase in

PHIGS specular

Plate 6: PHIGS specular term

intensity was expected from the directional diffuse term.

6.1.3 Specular images

The empirically based specular term use by PHIGS produces reasonable results, but on close examination it can be seen that the shape of the highlight is incorrect, as can be seen in plate 6. I suspect that this is due to the way in which a sphere is generated in PHIGS, a polygon mesh must be defined, as can be seen in plate 7, which is then interpolated to produce a smooth sphere. However, the specular component is calculated from the mesh and not the smooth surface, resulting in a polygonal highlight.

The specular term produced by the new illumination model is constant within

PHIGS mesh

Plate 7: PHIGS polygon mesh

physical specular

Plate 8: Physical based specular term

the specular cone, as can be seen in plate 8. Figures 6.1 and 6.3 show that there is a slight reduction in luminance towards the edge of the specular cone, but not enough to be apparent in the image.

A realistic specular component would tail off at its edges, so as to blend into the other components. Therefore, the delta term should not be needed to generate an accurate specular highlight, and its inclusion in the new illumination model means that there will always be a discontinuity in the image around the specular highlight.

In order to produce a realistic highlight within the specular cone, the terms in the specular equation would have to be very sensitive to angle, so as to vary considerably within the cone. The Fresnel term does have an angular dependence, but this only becomes significant as the angle approaches grazing. This will therefore produce very little variation within the specular cone. The shadowing/masking function decreases smoothly as the angle approaches grazing, and the surface roughness term increases gradually with angle. Although the specular term of the new model is very unrealistic, it is what would be expected from the specular equation. The specular term is a simple approximation of a complicated effect, and the approximation is inadequate.

6.1.4 Complete Images

The final image produced by PHIGS contains ambient, diffuse and specular components, as can be seen in plate 9. The overall result is a good one. However, it is very difficult to generate a surface of specific roughness, because the specular

PHIGS final

Plate 9: PHIGS final image

physical final

Plate 10: Physically based final image

physical final, without specular

Plate 11: Physical based final image without specular term

reflection coefficient and the Phong exponent have no physical foundation and therefore a trial and error method must be used. It is also difficult to specify exact colours for this model, although this is due to the use of RGB colour and is not a fault of the illumination model.

The final image produced by the new model can be seen in plate 10, and does not include the uniform diffuse term. The specular term gives rise to unrealistic looking surfaces, as can be seen in figures 6.2 and 6.4, which show that the specular term produces a discontinuity in the relative luminance of the image.

Plate 11 shows an image with only directional diffuse and ambient terms. The absence of the specular term improves the realism of the image, but the spheres appear matte and dull, as would be expected without a specular component.

6.1.5 Computation times

The rendering time for the new model was 4 min 46 sec, whereas PHIGS rendered its scene in 1 sec. The PHIGS model has hardware support whereas the new physically based model was implemented solely in software. In addition, the rendering time of the new model includes the time taken to ray trace the scene. Therefore, the relative computation times should not be taken as a direct measure of the amount of computation involved with each model.

6.2 Conclusion

Although the new illumination model produces reasonable results, the image quality does not justify the computation time. The main failings of the new model are the specular component and the lack of published data required by the uniform diffuse term, but I believe that with further work, the physically based model could produce good results. However, in their present state, PHIGS contains the better of the two illumination models.

6.3 Further work

There is still work to be done on the implementation of the new illumination model. The main areas for possible improvement are:

- The extension of the ray tracer to other object types. This involves the addition of new object intersection routines and surface normal determination procedures.

- The extension of the ray tracer to allow texture mapping of surfaces and environment mapping of the scene.
- The modifications required to allow global effects to be modelled. This would involve recursive calls of the ray tracer with depth limits and weighting.

There is also work to be done on the new illumination model. If any attempt is made to improve on the images it produces, the main area of concern *must* be the specular component. This can be improved in a number of ways:

- The physical basis of the term can be re-evaluated and should possibly contain the facet distribution function, as this would increase the angular sensitivity of the specular term, especially for smooth surfaces.
- The existing term can be multiplied by a masking function, to make it tail off smoothly. This would lessen the increase in relative luminance at the edges of the specular component.
- The entire term can be replaced with another term which is known to produce better results (the Phong model might be appropriate).

However, the last two solutions reduce the physical validity of the model and should therefore only be implemented as a last resort.

In order to accurately model uniform diffuse reflection, measurements of subsurface scattering parameters for common materials must be made. With this information, the uniform diffuse term of the new model would produce far better results.

There is still much work to be done in the area of physically based illumination models, much of which will be aimed at reducing the computational time and improving realism, two conflicting interests. A reduction in computational expense would allow far more complicated scenes to be rendered, with light rays breaking into their spectral components on reflection from a surface

Appendix A

Mathematics of the New Illumination Model

The BRDF was proposed to consist of three components, specular, directional diffuse and uniform diffuse:

$$\rho_{bd} = \rho_{bd,sp} + \rho_{bd,dd} + \rho_{bd,ud}$$

Where:

$\rho_{bd,sp}$ is the specular component from the first surface reflection process (reflection by mean surface) and has the form:

$$\rho_{bd,sp} = \frac{\rho_s \Delta}{cos\theta_i d\omega_i}$$

$\rho_{bd,dd}$ is the directional diffuse component (diffraction scattering from rough surface) and has the form:

$$\rho_{bd,dd} = \frac{|F|^2 GSD}{\pi cos\theta_i cos\theta_r}$$

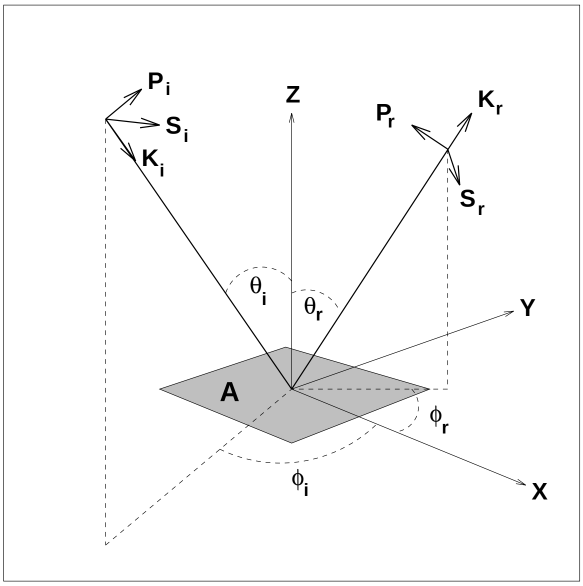

Figure A.1: Vectors used by Physical Model

$\rho_{bd,ud}$ is the uniform diffuse component and has the form:

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

Δ is a delta function which is unity inside the specular cone and zero outside.

$a(\lambda)$ is a free parameter with an upper bound determined by energy conservation, this term is used to represent uniform diffuse reflection.

$|F|^2$ is the Fresnel reflectivity.

S is a shadowing/masking factor.

G is a geometrical factor.

ρ_s, is the specular reflectivity of the surface and has the following form:

$$\rho_s = |F|^2 \cdot e^{-g} \cdot S$$

g is a function of the effective surface roughness, given by:

$$g = \left[\left(\frac{2\pi\sigma}{\lambda}\right)(\cos\theta_i + \cos\theta_r\Big)\right]^2$$

σ is the effective surface roughness and has the following form:

$$\sigma = \sigma_0 \cdot \left[1 + \left(\frac{z_0}{\sigma_0} \right)^2 \right]^{-\frac{1}{2}}$$

σ_0 is the actual surface roughness and is an input parameter required by the model.

$$\sqrt{\frac{\pi}{2}} z_0 = \frac{\sigma_0}{4} \left(K_i + Kr \right) \cdot exp \left(-\frac{z_0^2}{2\sigma_0^2} \right)$$

$$K_i = \tan \theta_i \cdot erfc \left(\frac{\tau \cos \theta_i}{2\sigma_0} \right)$$

$$K_r = \tan \theta_r \cdot erfc \left(\frac{\tau \cos \theta_r}{2\sigma_0} \right)$$

τ is the autocorrelation length of the surface, figure 3.2. This is another input parameter required by the model.

D is a distribution function which depends on surface roughness and the wavelength of the incident light, it has the following form:

$$D = \frac{\pi^2 \tau^2}{4\lambda^2} \cdot \sum_{m=1}^{\infty} \frac{g^m e^{-g}}{m! \cdot m} \cdot exp \left(-\frac{v_{xy}^2 \tau^2}{4m} \right)$$

v is the wave vector change given by:

$$v = k_r - k_i$$

where k_r and k_i are the reflected and incident wave vectors.

S is a shadowing and masking factor which depends on incident angle, reflected angle and surface roughness and has the form:

$$S = S_i(\theta_i) \cdot S_r(\theta_r)$$

$$S_i(\theta_i) = \frac{\left(1 - \frac{1}{2} erfc\left(\frac{\tau \cot \theta_i}{2\sigma_0}\right)\right)}{\Lambda(\cot \theta_i) + 1}$$

$$S_r(\theta_r) = \frac{\left(1 - \frac{1}{2} erfc\left(\frac{\tau \cot \theta_r}{2\sigma_0}\right)\right)}{\Lambda(\cot \theta_r) + 1}$$

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

and G is a geometrical factor of the form:

$$G = \left(\frac{v \cdot v}{v_z}\right)^2 \cdot \frac{1}{|k_r \times k_i|^4} \cdot \left[(s_r \cdot k_i)^2 + (p_r \cdot k_i)^2\right] \cdot \left[(s_i \cdot k_r)^2 + (p_i \cdot k_r)^2\right]$$

$$s_i = \frac{k_i \times \mathbf{N}}{|k_i \times \mathbf{N}|}$$

$$s_r = \frac{k_r \times \mathbf{N}}{|k_r \times \mathbf{N}|}$$

$$p_i = s_i \times k_i$$

$$p_r = s_r \times k_r$$

Appendix B

An MSDL data file

```
sphere                                          /* type of object */
    name copper                                 /* name of object */
    320.0 320.0 400.0 125.0                     /* position and radius of sphere */
              tau 3.0                           /* auto correlation length */
              sigma 1.0                         /* surface roughness */
              absorb 2.63                       /* absorption coefficient */
              ref_index 0.617                   /* average refractive index */
              spectra                           /* start of surface spectra information */
                        415 0.1750              /* spectral samples */
                        445 0.1875
                        475 0.2000
                        500 0.2150
                        520 0.2300
                        580 0.3625
```

```
                        600 0.4250
                        635 0.4750
    end
end

real_point    275.0 275.0 0.0     /* type of light and its position */
              415 0.9245          /* spectral samples */
              445 1.1095
              475 1.1540
              500 1.0940
              520 1.0480
              550 1.0400
              580 0.9580
              600 0.9000
              635 0.8350
```

References

[1] Abramowitz and Stegum. *Handbook of Mathematical Functions with Formulas, graphs and mathematical tables.* Dover publications, New York, 1970.

[2] Aguilar and Stiles. Saturation of the rod mechanism of the retina at high levels of stimulation. *Optical Acta*, 1(59), 1954.

[3] Amanatides. Realism in computer graphics: A survey. *IEEE Computer Graphics and Applications*, 7(1), January 1987.

[4] Beckmann. Shadowing of random rough surfaces. *IEEE Transactions on Antennas and Propagation*, May 1965.

[5] Beckmann and Spizzichino. *Scattering of ElectromagneticWaves from Rough Surfaces.* Macmillan, 1963.

[6] F Blinn, J. Models of light reflection for computer synthesized pictures. *Computer Graphics*, 11(2), 1977.

[7] J Bouknight, W. A procedure for generation of three dimensional half toned computer graphics presentations. *CACM*, 13(9), 1970.

[8] Brockleman and Hagfors. Note on the effect of shadowing on the backscattering of waves from random rough surfaces. *IEEE Transactions on Antennas*

and Propagation, 14(5), September 1966.

[9] Cook. A reflection model for realistic image synthesis. Master's thesis, Cornell University, Ithaca, 1981.

[10] Cook and Torrance. A reflectance model for computer graphics. *Computer Graphics*, 15(3), August 1981.

[11] Foley, VanDam, Feiner, and Hughes. *Computer Graphics, Principles and Practice.* Addison Wesley, 1990.

[12] N Gatenby. Radiosity in computer graphics: a proposed alternative to the heni-cube algorithm. Master's thesis, University of Manchester, January 1991.

[13] A Glassner. How to derive a spectrum from an RGB triplet. *IEEE Computer Graphics and Applications*, 9(4), 1989.

[14] A Glassner. *An Introduction to Ray Tracing.* Academic Press, 1989.

[15] Grant and Phillips. *Electromagnetism.* John Wiley and Sons, 1988.

[16] Greenberg. Light reflection models for computer graphics. *Science*, 244(4901), April 1989.

[17] Greenberg, Cohen, and Torrance. Radiosity: A method for computing global illumination. *The Visual Computer*, 2(5), 1986.

[18] R Hall. *Illumination and Colour in Computer Genetated Imagery.* Springer Verlag, 1989.

[19] R Hall. Manipulating colour. *SIGGRAPH*, August 1990.

[20] He, Torrance, Sillion, and Greenberg. A comprehensive physical model for light reflection. *Computer Graphics*, 25(4), July 1991.

[21] E Hecht. *Optics.* Addison Wesley, second edition, 1987.

[22] Hering and Smith. Apparent radiation properties of a rough surface. *Application to Thermal Design of Spacecraft*, 23, 1970.

[23] Hewlett Packard. *HP-PHIGS Graphics Techniques*, January 1991. Volume 2.

[24] Hopgood and Duce. *A Primer For PHIGS.* Wiley, 1991.

[25] Howard, Hewitt, Hubbold, and Wyrwas. *A Practical Introduction to PHIGS and PHIGS PLUS.* Addison Wesley, 1991.

[26] Information processing systems - computer graphics - Programmers Hierachical Interactive Graphics System (PHIGS), August 1988. ISO DP 9592.

[27] Colour in computer graphics : Student notes. ITTI Computer Graphics and Visualisation Module. Computer Graphics Unit, Manchester Computing Centre, University of Manchester.

[28] F Kip, A. *Fundamentals of Electricity and Magnetism.* McGraw Hill, Singapore, 1987.

[29] G Meyer. Wavelength selection for synthetic image generation. *Computer Vision, Graphics and Image Processing*, 41, 1988.

[30] Moravec. 3D graphics and the wave theory. *Computer Graphics*, 15(3), August 1981.

[31] Manchester Scene Descripton Language (MSDL). Information booklet. Published by the Computer Graphics Unit, Manchester Computing Centre, University of Manchester.

[32] M Peercy. Linear colour representations for full specral rendering. *Computer Graphics Proceedings, Anual Conference Series*, 1993.

[33] T Phong, B. Illumination for computer generated pictures. *ACM Comms*, 18(6), June 1975.

[34] J Popsel. Highlight shading: Lighting and shading in a PHIGS PLUS / PEX environment. *Computers and Graphics*, 14(1), 1990.

[35] P Shirley. Physically based lighting calculations for computer graphics. Master's thesis, University of Illinois, Urbana Champaign, November 1990.

[36] Siegel and Howell. *Thermal Radiation Heat Transfer*. Mcgraw Hill, second edition, 1981.

[37] Smith. Geometrical shadowing of a random rough surface. *IEEE Transactions on Antennas and Propagation*, 15(5), September 1967.

[38] Smith and Nichols. Effects of polarization on bidirectional reflectance of a one dimensional randomly rough surface. *Spacecraft Radiative Transfer and Temperature Control*, 83, 1981.

[39] A Stogryn. Electromagnetic scattering from finitely conducting surfaces. *Radio Science*, 18(6), 1967.

[40] Torrance and Sparrow. Off specular peaks in the diractional distribution of reflected thermal radiation. *Journal of Heat Transfer, Transactions of the*

ASME, May 1966.

[41] Torrance and Sparrow. Theory for off specular reflection from roughened surfaces. *Optical Society of America*, 57(9), 1967.

[42] D Travis. *Effective Colour Displays: Theory and Practice.* Academic Press, San Diego, 1991.

[43] T. S Trowbridge and K. P Reitz. Average irregularity representation of a roughened surface for ray reflection. *Journal of the Optical Society of America*, 65(5), 1967.

[44] A Watt. *Fundamentals of Three Dimensional Computer Graphics.* Addison Wesley, 1989.

[45] Whitted. An improved illumination model for shaded dispay. *ACM Comms*, 23(6), 1980.

[46] Wylie, Romney, Evans, and Erdahl. Half tone perspective drawings by computer. *Proc. AFIPS, Fall Joint Computer Conference*, 31, 1967.

[47] Wyzszecki and Stiles. *Colour Science: Concepts and Methods, Quantitative Data and Formulae.* John Wiley and Sons, Ney York, second edition, 1982.