A Simple Method for Ray Tracing Diffraction

Emmanuel Agu[1] and Francis S. Hill Jr.[2]

[1] Department of Computer Science, Worcester Polytechnic Institute,
Worcester, MA 01609, USA
emmanuel@cs.wpi.edu

[2] Department of Electrical and Computer Engineering, University of Massachusetts,
Amherst, MA 01002, USA
hill@ecs.umass.edu

Abstract. Diffraction and interference are optical phenomena which split light into its component wavelengths, hence producing a full spectrum of iridescent colors. This paper develops computer graphics models for iridescent colors produced by diffractive media. Diffraction gratings, certain animal skins and the crystal structure of some precious stones are known to produce diffraction. Several techniques can be employed to derive solutions to the diffraction problem including: (1)Electromagnetic boundary value methods (2)Applying the Huygens-Fresnel principle (3)Applying the Kirchoff-Fresnel theorem (4)Fourier optics. Previous work in developing diffraction models for computer graphics has used boundary value methods and Fourier optics but no models using Huygens-Fresnel principle have been published. This paper derives a set of diffraction solutions based on the Huygens-Fresnel principle, which are then used to extend well-known illumination models. We then use our new models to render images in a ray tracer.

1 Introduction

Humans live in a world filled with beautiful colors. Iridescent colors refer to the different colors which some surfaces radiate at different light source or viewer angles. New colors which were neither visible in the incident light nor the object being observed appear to have been created. These surfaces are sometimes said to "shimmer" as they are rotated. Common sources of iridescent colors include rainbows, shiny CD-ROM surfaces, opals, hummingbird wings, some snake skins, oil slicks and soap bubbles. Four mechanisms are known to produce iridescent color, namely, dispersive refraction, scattering, interference and diffraction.

Diffraction occurs when light encounters some obstacle or aperture of a dimension comparable with its wavelength. For example, a diffraction grating has a series of finely ruled parallel grooves. Different wavelengths are diffracted at different angles; hence different colors are produced at different angles, leading to the phenomenon of iridescence. Common sources of iridescent colors include diffraction gratings, opals and some liquid crystals, hummingbird winds and some snakeskins. Diffraction, also referred to as a wavefront splitting phenomenon is distinguished from interference, an amplitude-splitting phenomenon [Gon94].

V. Kumar et al. (Eds.): ICCSA 2003, LNCS 2669, pp. 336–345, 2003.

A common problem in optics is that of determining the outgoing light intensity, wavelength and color, given that light is incident on the diffraction surface at a certain angle and intensity and is comprised of specified wavelengths. In particular, a closed-form expression relating incoming and outgoing light permits diffraction surfaces to be elegantly modelled in computer graphics. Several techniques have been employed in the optics literature to derive solutions to this problem [Str67] including (1) Solutions derived by applying electromagnetic boundary value methods (2) Solutions based on Huygens-Fresnel principle (3) Kirchoff-Fresnel based solutions (4) Solutions using Fourier optics.

Two main bodies of work have been identified which attempt to model diffraction in computer-generated imagery. Thorman [Tho96] first developed a simple computer graphics model for diffraction by using the grating equation derived by applying electromagnetic boundary value methods. Stroke [Str67] derives the geometrical conditions for iridescence. Thorman [Tho96] then uses Stroke's results and addresses the specific issue of modelling iridescent colors produced by diffraction in computer graphics. Thorman's work focuses on determining directions in which the grating equation would produce peak responses, and hence render images accordingly. Specifically, the grating equation used by Thorman is not a continuous function but only gives the directions for perfect constructive interference. No information is given on the behavior of the reflected light at angles which are away from those of perfect constructive interference. Also, Thorman erroneously assumes that all peak intensity values are equal. A continuous function needs to be derived which gives the behavior of light in all outgoing directions.

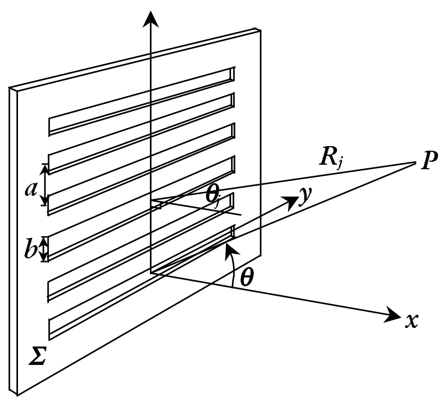
Stam [Sta99] has published work using Fourier analysis. Fourier optics solutions are believed to give the most accurate but also most complex solutions. However, detailed information about both the scene configuration and diffraction surface profile are vital before Fourier solutions can be evaluated and in fact, each solution is valid only for one configuration. In order to arrive at a Fourier optics solution, Stam makes assumptions about the scene configuration, as well as the profile and distribution of the grating surface. The fact that Fourier solutions are configuration-dependent increases their complexity.

In what follows, we will apply the Huygens-Fresnel principle to derive a continuous function that defines the behavior of light in all outgoing directions and then include it in a complete illumination model, showing clearly how to use the model to render pictures of surfaces with diffraction.

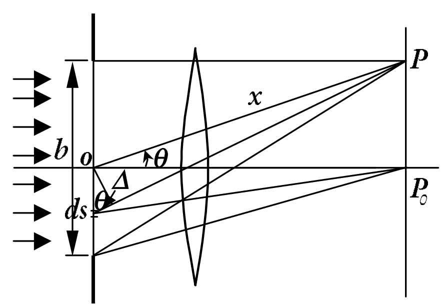
2 Our Optics Model

The Huygens-Fresnel principle states that every unobstructed point of a wavefront, at a given instant in time, serves as a source of spherical secondary wavelets (with the same frequency as that of the primary wave). The amplitude of the optical field at any point beyond is the superposition of all these wavelets (considering their amplitudes and relative phases). This principle arises because all vibrating particles exert a force on their neighbors and thus act as point sources. It explains why when a wave passes through an aperture or obstruction, it always spreads to some extents to regions that were not directly exposed to the oncoming waves and can be used to derive useful approximate solutions to the diffraction problem.

In optics, distinction is also made between *near-field* or *Fresnel* and *far-field* or *Fraunhofer* diffraction. In Fresnel diffraction, the point of observation is so close to the aperture that the image formed bears a close resemblance to the aperture, the emergent waves are spherical and intensities received at a given point vary as one travels along the aperture width. For Fraunhofer diffraction on the other hand, the point of observation is so far that the image formed bears almost no resemblance to the actual aperture, the emergent waveforms can be approximated as planar waves with uniform intensities from any point on the aperture width. We deal only with Fraunhofer diffraction in this paper since it encompasses almost all configurations of practical interest in computer graphics.



(a) Diffraction System with N Slits



(b) Diagram of a single slit

Fig. 1.

We shall now apply the Huygens-Fresnel principle to derive an expression for the irradiance from an N-slit diffraction grating where N is sufficiently large and the point of observation is sufficiently far from the grating surface. Note that the following derivations apply to gratings with slits, transmissive and reflection gratings. Applying the Huygens-Fresnel principle, each slit or grating edge of a width much less than λ, now acts as a secondary source. The number of slits, N (typically in the thousands per inch of diffraction grating) is tremendously large and their separation is small. Figure 1(a) is an example of such a grating. Only a few slits of the grating are shown in the figure for ease of illustration. We wish to derive a closed form expression for the net contributions of these N slots of a diffraction grating at an arbitrary location in space P. The final closed form derivation will be approached in a simple two-stage process. First we derive the effect of one of these N slits at the point P, following which we shall consider the net effect of vector addition of several of these slits at the point P.

Figure 1(b) shows a single slit of the diffraction grating. ds is the elemental width of the wave front in the plane of the slit, at a distance s from the center O, which we shall refer to as the origin. The wavelet emitted by the element ds, observed at the point P, will be proportional to the length of ds and inversely proportional to the distance x [Hec87], [Jen57], [Lon67]. The general equation for a spherical wave can be written as:

$$y = \frac{a}{r}\sin(wt - kr) \tag{1}$$

where a is the amplitude at a unit distance from the source and r is the distance of the observation point from the source. Hence, from Figure 1(b), the infinitesimal displacement produced at the point P by the infinitesimal element ds can be written as

$$dy_0 = \frac{ads}{x}\sin(wt - kx) \tag{2}$$

The displacement varies in phase as the position of ds changes by a factor, due to the different path lengths to P_0, which can be expressed as $D = s\sin q$. So, at a given point s below the origin, the contribution will be

$$dy_s = \frac{ads}{x}\sin[wt - kx - ks.\sin(\theta)] \tag{3}$$

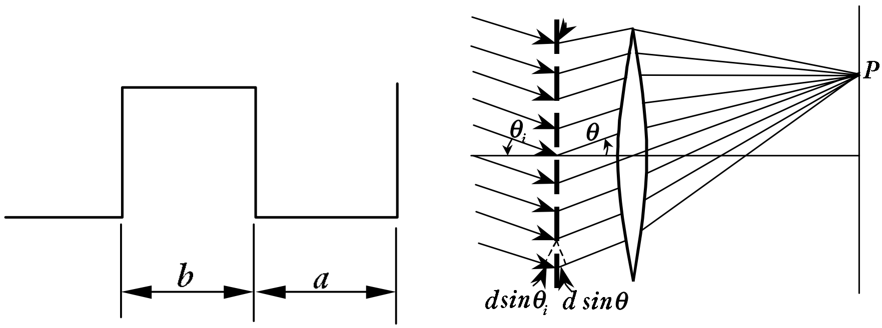
Integrating equation 3 from one edge of the slit to the other, we get

$$y = \frac{ab}{x}\frac{\sin\beta}{\beta}\sin(wt - kx) \tag{4}$$

Hence, the resultant vibration is a simple harmonic one, the amplitude of which varies with position P. Thus, the intensity at the screen due to one slit is then

$$I \approx A_0^2\left(\frac{sin\beta}{\beta}\right)^2 \tag{5}$$

Now, considering a diffraction grating system of N slits, with apertures of width a, and width b of the opaque portion separating apertures (See figure 2(a)). Next, we shall determine the vector sum of several of these single slits at the arbitrary observation point P. The incident light is still at a normal angle of incidence and the phase difference, d, between disturbances from corresponding strips of adjacent apertures is again $\delta = \left(\frac{2\pi}{\lambda}\right) d\sin(\theta)$ where $d = a + b$ specifies the distance between similar points in adjacent apertures as shown in figure 2(a). Expressed as a complex quantity, the phase difference between disturbances from corresponding strips of adjacent apertures is $e^{-i\delta}$. Hence, adding the net contributions of multiple slits of amplitude $\left(\frac{\sin\beta}{\beta}\right)^2$ as determined in the derivation



(a) Diagram showing a and b in adjacent slits

(b) Diagram showing oblique incidence

Fig. 2.

for a single slit and observed at a point P in space, the complex amplitude of the resultant is given by

$$z = \left(\frac{\sin\beta}{\beta}\right)\left[1 + e^{-i\delta} + e^{-2i\delta} .... + e^{-(N-1)i\delta}\right] \tag{6}$$

which gives

$$z = \left(\frac{\sin\beta}{\beta}\right)^2 \frac{\sin^2 N\alpha}{\sin^2\alpha} \text{ where } \alpha = \frac{\delta}{2} = \frac{\pi d \sin\theta}{\lambda} = \frac{\pi(a+b)\sin\theta}{\lambda} \tag{7}$$

Normalizing our result, we get

$$I = I_0 \frac{1}{N^2}\left(\frac{\sin\beta}{\beta}\right)^2\left(\frac{\sin^2 N\alpha}{\sin^2\alpha}\right) \tag{8}$$

here I_0 is the intensity of the incoming light ray in the $q = 0$ direction. Equation 8 is our final expression for irradiance from an arbitrary N-slit diffraction grating. Finally, we modify our results to take into account oblique incoming and outgoing angles. Consider the following figure 2(b) above. In the case of oblique angles, the general expression for irradiance stays the same. However, the phase difference between disturbances (contributions) from successive slots is given by

$$\delta = \left(\frac{2\pi}{\lambda}(a+b)(sin\theta - \sin\theta_i)\right) \tag{9}$$

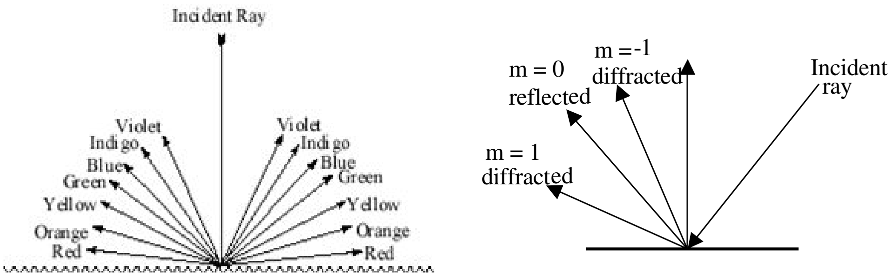
and

$$\beta = \pi b\frac{(\sin\theta - sin\theta_i)}{\lambda} = \qquad \text{and} \qquad \alpha = \pi(a+b)\frac{(\sin\theta - \sin\theta_i)}{\lambda} \tag{10}$$

Since $\alpha = \frac{\pi d}{\lambda}(\sin\theta - \sin\theta_i)$

$$d(\sin\theta - sin\theta_i) = m\lambda \tag{11}$$

where $m = 0, \pm1, \pm2...$ Equation 11 is widely known as the *grating equation* and gives the locations of maxima.



(a) White light incident perpendicular to the surfaces of a grating and the first order diffraction spectrum on each side

(b) View in the diffraction plane of a monochromatic incident ray and the diffracted rays of first order on each side

Fig. 3.

We can see from equation 11 that different wavelengths (and hence different colors) will peak at different angles with different $\sin q - \sin q_i$ as shown in figure 3(a). Likewise, we can see from figure 3(b) that according to equation 11, different modes peak at different angles.

3 Our Illumination Model

In this section, we shall outline our new illumination models which are based on the Huygens-Fresnel principle, include diffraction and can render iridescent surfaces. We can express our diffraction illumination model as

$$I = Ambient + Diffuse + Diffraction \tag{12}$$

We introduce a new diffraction component in equation 12 to account for both the directional specular and diffraction effects. The ambient and diffuse components in equation 12 are the same as those used in the Phong model [Pho75].The diffraction component in equation 12, $I(\theta)$ is expressed as (see equation 8)

$$I(\theta) = \sum_{\lambda} I_{0,\lambda} \frac{1}{N^2} \left(\frac{\sin\beta}{\beta}\right)^2 \left(\frac{\sin N^2\alpha}{\sin^2\alpha}\right) \tag{13}$$

where equation 13 is a summation of equation 11 over a discrete set of wavelengths and α and β equals the values expressed in equations 10 and $I_{o,\lambda}$ is the light intensity at a summation wavelength, λ. The summation above is over a chosen set of discrete λ, $[\lambda]$ in the visible range. We recall that the visible spectrum has wavelengths in the 380nm to 700nm range. In our illumination model of equation 12, we have replaced the specular term that was used previously in the Phong model by a more general diffraction component that also incorporates the Phong (or Cook-Torrance) specular component. In rendering the model, positions of peak wavelengths for each position are pre-calculated and care taken to include the peak intensity in the rendered image.

4 Rendering Our Models

In this section, we discuss how to use our illumination model to render pictures with diffractive surfaces. A ray tracer is used as our rendering system. However, first, we shall outline issues which need to be taken into consideration before these models can be used.

Thus far, our expressions for our illumination model have not reflected the fact that the grating or diffractive surface may be transformed into new arbitrary 3D positions of the user's choice. A convenient way of incorporating transformations and including the 3D case, is by expressing our illumination model using the half vector. It is interesting to note that in the case where the diffraction grating is coincident with the z-axis, the term $\sin q + \sin q_i$ in our diffraction expression is equal to the z-component of the normalized halfway vector, H. Hence, we can simplify equation 13 by replacing $\sin q - \sin q_i$ with $H.z$. We can thus re-write equations 9 and 10 as

$$\delta = \left(\frac{2\pi}{\lambda}\right) d \times H.z = \left(\frac{2\pi}{\lambda}\right) (a+b) \times H.z \tag{14}$$

and β and α become

$$\beta = \pi b \frac{H.z}{\lambda} = \frac{kb}{2} \times H.z \qquad \text{and} \qquad \alpha = \pi d \frac{H.z}{\lambda} = \frac{k(a+b)}{2} \times H.z \tag{15}$$

This single substitution is very powerful and greatly simplifies our expression and eases manipulation within a rendering system. It is also possible to create patterns using by introducing an arbitrary twist in the grating direction or by creating checkerboard patterns. Color for display in computer graphics is usually specified by relative amounts of a set of primary colors (e.g. Red, Green and Blue or their RGB values) which they contain. Since, the models for iridescence which were described in the preceding sections as well as the accompanying trigonometry, were specified on a wavelength basis, conversion from wavelength to RGB becomes necessary. Furthermore, following conversion, some colors which are readily specified in wavelengths may fall outside the gamut of RGB tuples which the CRT can display. In such a case, these colors need to be systematically converted or transformed to triples which can be displayed by the CRT. The underlying process of truncation or transformation is known as color clipping. In our ray tracer, we have investigated and implemented three alternative schemes for color clipping. These are clamping, intensity scaling and constant intensity scaling. Hall [3] has a complete discussion of these three scaling techniques.

In a ray tracer, the illumination models are evaluated per pixel and it is sufficient to describe what steps are taken to render a pixel. For each pixel that hits a diffraction surface, the following steps are taken:

1. Determine hit point: determine the first object the eye sees through this pixel while looking into the scene to the ray traced.
2. Compute ambient and diffuse components
3. Build light and eye vectors and compute normalized Half Vector, H
4. Transform grating direction vector, normal vector, twist grating vector
5. Replace $(\sin q + \sin q_i)$ with $H.z$ (half vector component in the grating direction) in equation 13
6. Search for all visible modes and corresponding wavelengths that peak at this angle. If no modes or wavelengths are visible, return diffuse + ambient color. If mode zero is visible, return specular component.
7. Evaluate equation 13 at peak wavelengths.
8. Convert peak wavelengths to RGB colors as described in section 4 and evaluate iridescent color.

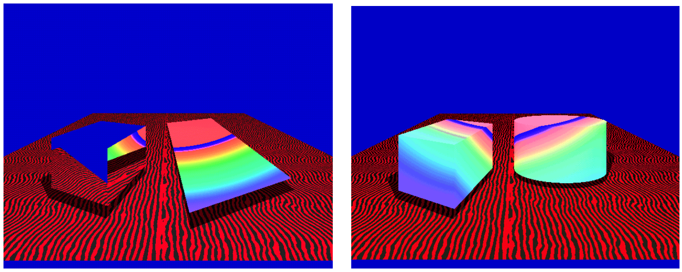
While steps 1-3 above are the same as previous ray tracers, steps 4-8 are related to our diffraction models.

5 Results

Figures 4(a) through 6 are all images incorporating our models and show some iridescent colors. Diffraction gratings with and without an arbitrary twist in the grating direction, have been demonstrated. We have also demonstrated variants of the checkerboard pattern alternating with a diffuse surface, as well as alternating the twist angles. The patterns created were extremely colorful and resembled real life iridescent diffraction patterns. Additionally, we have shown other geometries, such as the cube and cylinder that have been made from these diffractive gratings or patterns. Finally, we have also rendered a CD-ROM surface exhibiting iridescence. Simple animations were also produced to illustrate the color variance as viewer and surface orientations were altered.

6 Conclusion

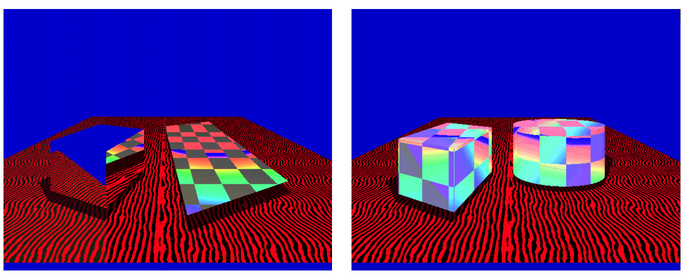
In this paper, we have defined the concept of angle-dependent iridescent coloration of certain materials, due to the optical phenomenon of diffraction. Two earlier attempts by Thorman and Stam, to develop diffraction illumination models were reviewed and their shortcomings clearly stated. We have developed



(a) Diffraction grating and reflective cube on wood textured surface

(b) Diffractive cube and cylinder on wood textured surface

Fig. 4.



(a) Checkerboard diffraction grating on wood textured surface

(b) Checkerboard diffractive cube and cylinder with alternated groove twist angle

Fig. 5.

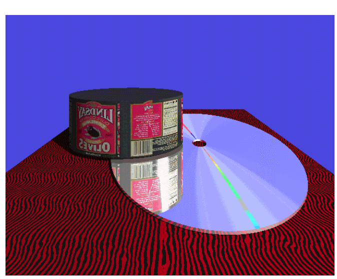


Fig. 6. CD-ROM showing iridescence, and image-mapped can on wood textured surface

diffraction shading models for computer graphics in two distinct phases; first, we developed an optics model, which describes the interaction of incident light with our diffractive surface. Next, we included our optics model in our complete illumination model.

In developing our optics model, we have applied the Huygens-Fresnel principle. In our derivations, we have made assumptions which make sense for computer graphics applications. These include assumptions that incident light is non-polarized, that the grating is several wavelengths away from the point of observation such that emergent waves can be approximated as plane waves (Fraunhofer diffraction).

We have incorporated our optics model into a complete illumination model for computer graphics by adding diffuse and ambient terms, similar to those used in the Phong and Cook-Torrance illumination models. We have rendered these models using a ray tracer and practical issues encountered discussed. Photorealistic scenes with diffractive surfaces, including diffraction gratings, checkerboard patterns and CD-ROMs, have been produced. Possible areas for future research include extending our work to include:

1. Efficient, radiosity-based solution which effectively tracks and renders surface inter-reflections for added realism.
2. Different light sources with different geometries and constituent wavelengths.
3. Three-dimensional gratings such as crystals and gemstones.
4. Diffraction in animals such as snakeskins and butterflies which exhibit iridescence.

References

[Agu02] Agu, E.,Diffraction Shading Models in Computer Graphics, (PhD Dissertation, U. of Massachusetts-Amherst, September 2001).

[Gon94] J. S. Gondek, G. W. Meyer and J. G. Newman, Wavelength Dependent Reflectance Functions, Proc. SIGGRAPH 1994, pp 213-220.

[Hal89] R. Hall, Illumination and Color in Computer Generated Imagery, (Springer-Verlag, 1989).

[Hec87] E. Hecht, Optics, Second Edition, (Addison-Wesley Publishing Co., 1987).

[Jen57] F. A. Jenkins and H. E. White, Fundamentals of Optics, Third Edition, (McGraw-Hill, 1957).

[Lon67] R. S. Longhurst, Geometrical and Physical Optics, Second Edition, John Wiley and Sons, 196.

[Pho75] B. Phong, Illumination for Computer-Generated Pictures, Comm. of ACM, 18(6), pp 311-17, 1975.

[Sta99] J. Stam, Diffraction Shaders Proc. ACM Computer Graphics (SIGGRAPH '99), pp 101-110.

[Str67] G. W. Stroke, Diffraction Gratings pp 426-754, Handbuck Der Physik (Encyclopaedia of Physics), (Springer-Verlag, 1967).

[Tho96] S. Thorman, Diffraction based models for iridescent colors in computer generated imagery (PhD Dissertation, U. of Massachusetts-Amherst, 1996).