Appendix F

Calculating Normal Vectors

This appendix describes how to calculate normal vectors for surfaces. You need to define normals to use OpenGL's lighting facility, which is described in **Chapter 6, "Lighting."** **"Normal Vectors"** introduces normals and the OpenGL command for specifying them. This appendix goes through the details of calculating them. It has the following major sections:

- **"Finding Normals for Analytic Surfaces"**
- **"Finding Normals from Polygonal Data"**

Since normals are perpendicular to a surface, you can find the normal at a particular point on a surface by first finding the flat plane that just touches the surface at that point. The normal is the vector that's perpendicular to that plane. On a perfect sphere, for example, the normal at a point on the surface is in the same direction as the vector from the center of the sphere to that point. For other types of surfaces, there are other, better means for determining the normals, depending on how the surface is specified.

Recall that smooth curved surfaces are approximated by a large number of small flat polygons. If the vectors perpendicular to these polygons are used as the surface normals in such an approximation, the surface appears faceted, since the normal direction is discontinuous across the polygonal boundaries. In many cases, however, an exact mathematical description exists for the surface, and true surface normals can be calculated at every point. Using the true normals improves the rendering considerably, as shown in **Figure F−1**. Even if you don't have a mathematical description, you can do better than the faceted look shown in the figure. The two major sections in this appendix describe how to calculate normal vectors for these two cases:

- **"Finding Normals for Analytic Surfaces"** explains what to do when you have a mathematical description of a surface.
- **"Finding Normals from Polygonal Data"** covers the case when you have only the polygonal data to describe a surface.

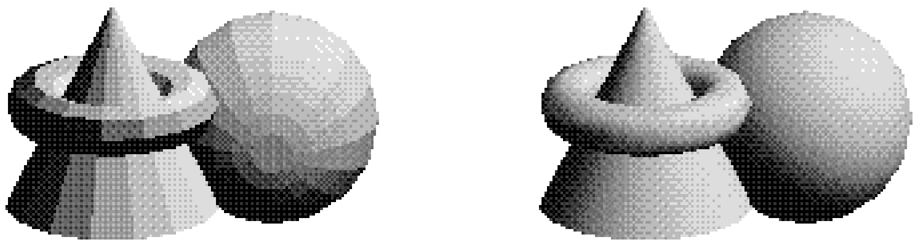

Figure F−1 Rendering with Polygonal Normals vs. True Normals

Finding Normals for Analytic Surfaces

Analytic surfaces are smooth, differentiable surfaces that are described by a mathematical equation (or set of equations). In many cases, the easiest surfaces to find normals for are analytic surfaces for which you have an explicit definition in the following form:

$V(s,t) = [\ X(s,t)\ \ Y(s,t)\ \ Z(s,t)\ ]$

where s and t are constrained to be in some domain, and X, Y, and Z are differentiable functions of two variables. To calculate the normal, find

$$\frac{\partial V}{\partial s} \text{ and } \frac{\partial V}{\partial t}$$

which are vectors tangent to the surface in the *s* and *t* directions. The cross product

$$\frac{\partial V}{\partial s} \times \frac{\partial V}{\partial t}$$

is perpendicular to both, and hence to the surface. The following shows how to calculate the cross product of two vectors. (Watch out for the degenerate cases where the cross product has zero length!)

$$\begin{bmatrix} v_x & v_y & v_z \end{bmatrix} \times \begin{bmatrix} w_x & w_y & w_z \end{bmatrix} = \begin{bmatrix} (v_y w_z - w_y v_z) & (w_x v_z - v_x w_z) & (v_x w_y - w_x v_y) \end{bmatrix}$$

You should probably normalize the resulting vector. To normalize a vector [*x*,*y*,*z*], calculate its length

$$\text{Length} = \sqrt{x^2 + y^2 + z^2}$$

and divide each component of the vector by the length.

As an example of these calculations, consider the analytic surface

$V(s,t) = [\ s2\ t3\ 3-st\ ]$

From this we have

$$\frac{\partial V}{\partial s} = \begin{bmatrix} 2s & 0 & -t \end{bmatrix},\ \frac{\partial V}{\partial t} = \begin{bmatrix} 0 & 3t^2 & -s \end{bmatrix},\text{ and } \frac{\partial V}{\partial s} \times \frac{\partial V}{\partial t} = \begin{bmatrix} -3t^3 & 2s^2 & 6st^2 \end{bmatrix}$$

So, for example, when $s=1$ and $t=2$, the corresponding point on the surface is (1, 8, 1), and the vector (−24, 2, 24) is perpendicular to the surface at that point. The length of this vector is 34, so the unit normal vector is (−24/34, 2/34, 24/34) = (−0.70588, 0.058823, 0.70588).

For analytic surfaces that are described implicitly, as $F(x, y, z) = 0$, the problem is harder. In some cases, you can solve for one of the variables, say $z = G(x, y)$, and put it in the explicit form given previously:

$$\mathbf{V}(s,t) = \begin{bmatrix} s & t & \mathbf{G}(s,t) \end{bmatrix}$$

Then continue as described earlier.

If you can’t get the surface equation in an explicit form, you might be able to make use of the fact that the normal vector is given by the gradient

$$\nabla F = \begin{bmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} & \frac{\partial F}{\partial z} \end{bmatrix}$$

evaluated at a particular point (x, y, z). Calculating the gradient might be easy, but finding a point that lies on the surface can be difficult. As an example of an implicitly defined analytic function, consider the equation of a sphere of radius 1 centered at the origin:

$x^2 + y^2 + z^2 - 1 = 0$

This means that

$F(x, y, z) = x^2 + y^2 + z^2 - 1$

which can be solved for z to yield

$$z = \pm\sqrt{1 - x^2 - y^2}$$

Thus, normals can be calculated from the explicit form

$$\mathbf{V}(s,t) = \begin{bmatrix} s & t & \sqrt{1 - s^2 - t^2} \end{bmatrix}$$

as described previously.

If you could not solve for z, you could have used the gradient

$$\nabla F = \begin{bmatrix} 2x & 2y & 2z \end{bmatrix}$$

as long as you could find a point on the surface. In this case, it's not so hard to find a point —for example, (2/3, 1/3, 2/3) lies on the surface. Using the gradient, the normal at this point is (4/3, 2/3, 4/3). The unit−length normal is (2/3, 1/3, 2/3), which is the same as the point on the surface, as expected.

Finding Normals from Polygonal Data

As mentioned previously, you often want to find normals for surfaces that are described with polygonal data such that the surfaces appear smooth rather than faceted. In most cases, the easiest way for you to do this (though it might not be the most efficient way) is to calculate the normal vectors for each of the polygonal facets and then to average the normals for neighboring facets. Use the averaged normal for the vertex that the neighboring facets have in common. **Figure F−2** shows a surface and its polygonal approximation. (Of course, if the polygons represent the exact surface and aren't merely an approximation —if you're drawing a cube or a cut diamond, for example —don't do the averaging. Calculate the normal for each facet as described in the following paragraphs, and use that same normal for each vertex of the facet.)

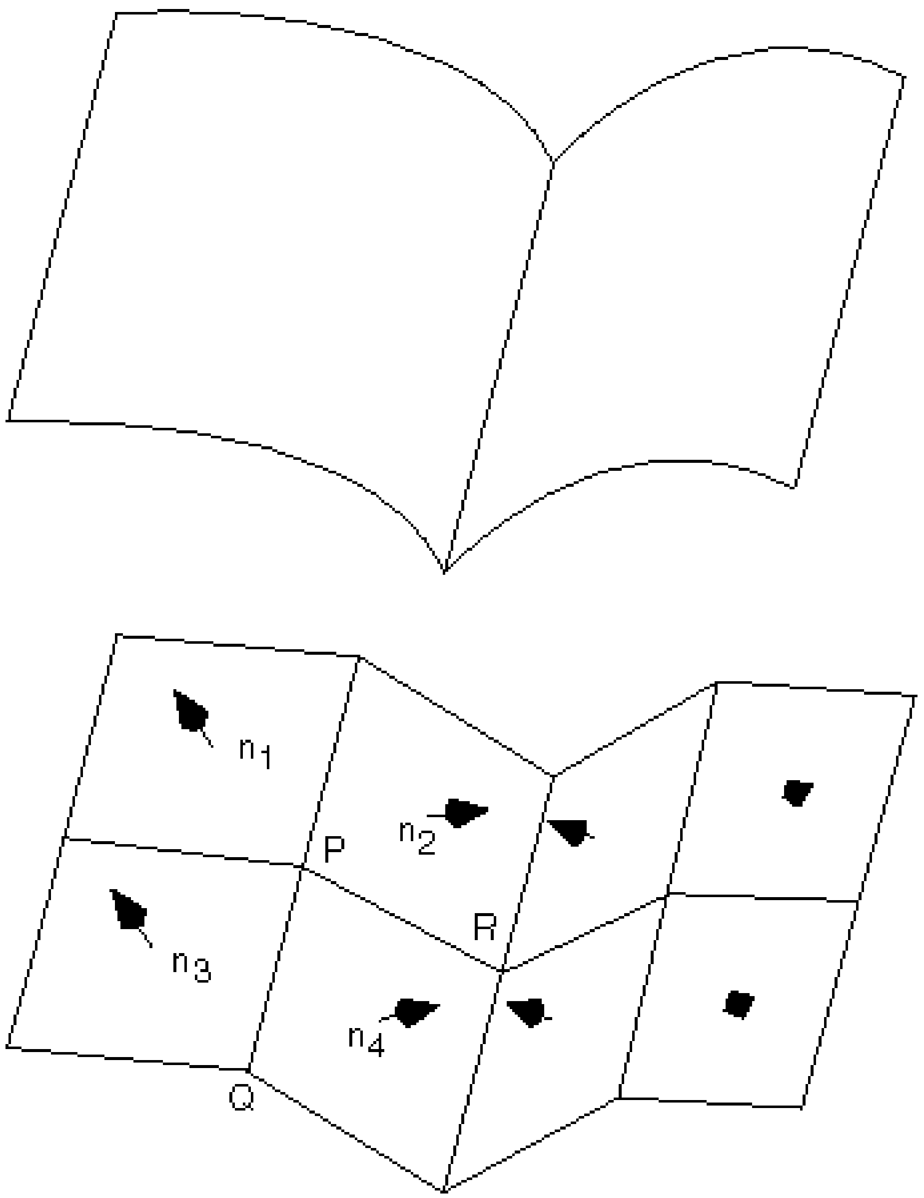

Figure F–2 Averaging Normal Vectors

To find the normal for a flat polygon, take any three vertices $v1$, $v2$, and $v3$ of the polygon that do not lie in a straight line. The cross product

$$[v_1 - v_2] \times [v_2 - v_3]$$

is perpendicular to the polygon. (Typically, you want to normalize the resulting vector.) Then you need to average the normals for adjoining facets, to avoid giving too much weight to one of them. For instance, in the example shown in **Figure F–2**, if n_1, n_2, n_3, and n_4 are the normals for the four polygons meeting at point P, calculate $n_1+n_2+n_3+n_4$ and then normalize it. The resulting vector can be used as the normal for point P.

Sometimes, you need to vary this method for particular situations. For instance, at the boundary of a surface (for example, point Q in **Figure F−2**), you might be able to choose a better normal based on your knowledge of what the surface should look like. Sometimes the best you can do is to average the polygon normals on the boundary as well. Similarly, some models have some smooth parts and some sharp corners (point R is on such an edge in **Figure F−2**). In this case, the normals on either side of the crease shouldn't be averaged. Instead, polygons on one side of the crease should be drawn with one normal, and polygons on the other side with another.