Precision Normals
(Beyond Phong)

T. Marc Olano
Terry S. Yoo

Department of Computer Science
University of North Carolina
Chapel Hill, NC 27599-3175, USA

Abstract

Almost all graphics architectures today support Gouraud shading, linear color interpolation between vertices; system designers aim toward a Phong shading model, linear interpolants of surface normals with a lighting model that supports both diffuse and specular components, as a superior means of rendering accurate images. However, the Phong model still retains serious artifacts. In this paper, we point out the shortcomings of linear interpolation of normals, and present a surface interrogation method for parametrically defined surfaces.

CR Categories: I.3.7 [Computer Graphics] Color, shading, shadowing, and texture; I.3.5 [Computer Graphics] Computational Geometry and Object Modeling - Curve, Surface and Geometric Algorithms; I.3.3 [Computer Graphics] Picture/Image Generation - Display algorithms.

Additional Keywords and Phrases: differential geometry, parametric surfaces, displacement mapping, bump mapping, linear normal interpolation, Phong lighting model, parametric normal interpolation.

Introduction

It has been said that good geometry can save bad rendering, but no amount of rendering can save bad geometry. But good rendering makes a big difference over bad rendering. Polygon based graphics systems use enhanced shading models to compensate for the polygonal nature of their images. We apply the principles of differential geometry and knowledge of parametric surfaces to the problems of shading and illumination, augmenting the already vast body of knowledge in texture mapping with a few observations on lighting models and their relationship to surface representation.

What are the effects of the surface normal on visual appearance? Following computer vision research, we address at this question with differential geometry . Applications of differential geometry to three dimensional curves and surfaces are discussed in Koenderink [Koenderink 90].

Much of the surface behavior in which we are interested can be understood by analyzing the principle curvatures (κ_1 and κ_2) at points along the surface. In particular, surfaces can be divided into regions based on the sign of their Gaussian curvature ($\mathbf{K} = \kappa_1\kappa_2$). Where **K** is positive the surface is concave or convex (*elliptic* or *synclastic*). Where **K** is negative the surface is saddle shaped (*hyperbolic* or *anticlastic*). Separating these two types of regions is the *parabolic curve*, the locus of points where one of the principle curvatures crosses zero as it changes sign. On smooth closed surfaces, the parabolic curve forms a continuous closed curve.

To visualize the behavior of these properties, it is helpful to consider the normal spherical map or Gauss map of the surface (Figure 1). This is the mapping of the surface onto the unit sphere where each point maps to the coordinates of its normal. For a particular light direction and viewing direction, there is one position on the Gauss map that corresponds to the normal direction where the specular highlight will appear. Parabolic curves correspond to folds in the Gauss map. The surface inflects and the rotation of the normal turns around as the principle curvature changes sign. On one side of that fold, there are two points on the surface with the same normal that will show the specular highlight. As the specular direction crosses the fold, these highlights move together, merge into a single highlight, then disappear. A similar effect occurs with the diffuse reflection but is much less noticeable.

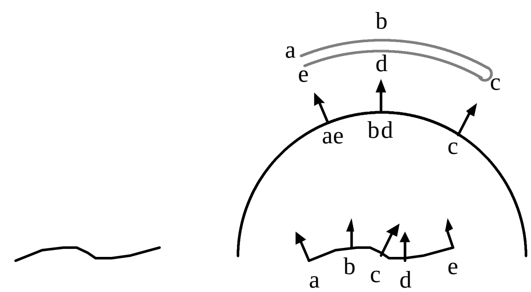

Figure 1. 2d example of Gauss map. Left: curve Right: mapping of normals on the curve to points along the unit circle

Consider a polygonalization of a surface where one polygon spans the parabolic curve. The normal direction should fold in the middle of the polygon, but with linear interpolation, this can only happen at the polygon boundaries (Figure 2). So creation and annihilation of specular points can only happen on polygon boundaries. Other normal-based computations (environment maps, anisotropic lighting models, ray traced reflections) also suffer and can generate artifacts much worse than incorrect specular reflections. A traditional Phong shading model with linear interpolation of normals is simply not sufficient.

What are the alternatives? There is a rich history in accurate rendering of parametric surfaces. Blinn performed much of the initial work [Blinn 78], and laid the foundations for what was to follow as the capabilities of graphics systems grew. Beyond the subdivision techniques that followed, there has been significant interest in the direct

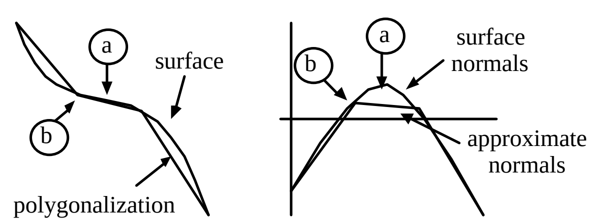

Figure 2. Left: :"Surface" and its "polygonalization." Right: vertical component of the normal as it folds and the Phong approximation to it. Note the points labeled a where the reflection events should occur and the points labeled b where they will occur.

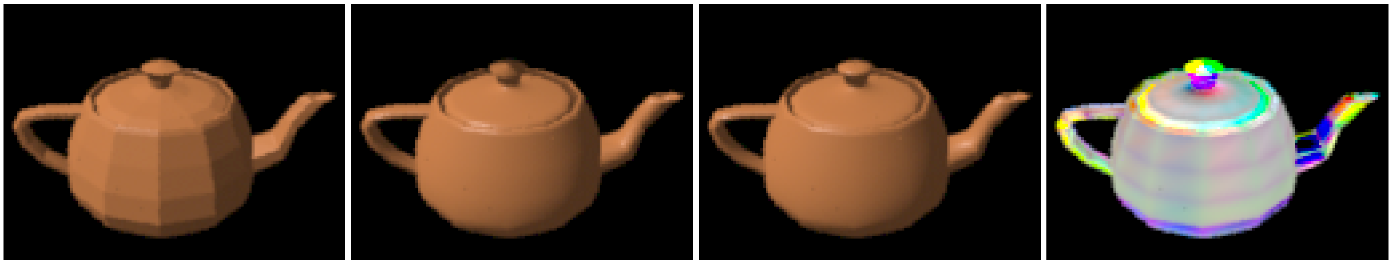

Figure 3. Comparison of linear normal interpolation with parametric normal interpolation. Left: 504 polygonal facets. Middle left: Shading with linear normal interpolation. Middle right: Parametric normal interpolation. Right: Difference of normals for middle images.

evaluation of the parametric forms. Kajiya discusses ray tracing parametric patches [Kajiya 82], solving for exact positions of ray intersection from the parametric description.

Shantz and Lien/Chang [Shantz 87][Chang 89] employ adaptive forward differencing to shade bicubic patches. Surfaces are incrementally split into several bicubic curves that leave no pixel-sized gaps. The remaining curves are then rendered, calculating the surface values from the parametric forms of the curve equations.

The behavior of geometric properties of surfaces has always been a concern of those interested in the generation of smooth surfaces. Moreton and Sequin [Moreton 1992] discuss the behavior of surface curvature at patch boundaries and describe a system for guaranteeing surface fairness.

Nelson Max [Max 89] discusses ensuring C^1 smoothness of normals across triangulated surfaces and generates smooth silhouettes and normals. We employ many of these ideas within the context of shading. Specifically we explore the problems of less than C^1 continuous normals, and apply the direct evaluation of parametric surfaces to the rendering of polygonal meshes.

Interpolants

We need higher order continuity, therefore we seek higher order interpolants for surface normals. Phong shading is linear, providing C^0 continuity of normals. With quadratic interpolation, the Gauss map could fold within the triangle, but matching the normal values and derivatives is overconstrained. Nelson Max presented a method using quadratic Bézier triangles to achieve C^1 normal continuity where he divided each triangle into six to provide the additional degrees of freedom. Max also used a quadratic function to interpolate depth and texture coordinates [Max 89]. Blinn suggests cubic Hermite interpolation of depth and texture coordinates to match their value and derivatives [Blinn 78]. He also suggests the texture coordinates of a spline patch could be used to directly evaluate the normal, though he preferred exact rendering of the surface. Schweitzer and Cobb implemented cubic normal interpolation for scanline rendering limited to bicubic surfaces [Schweitzer 82].

	Degree	
Surface type	surface	normals
Bézier triangles	n	2n - 2
Rational Bézier triangle	n	2n - 1
Bézier patch	n x m	(3n-2) x (3m-2)
Rational Bézier patch	n x m	(3n-1) x (3m-1)

Table 1: Degree of normal functions.

While it might not have been practical in 1978, cubic or higher interpolation of normals could easily be achieved in hardware today. Several architectures are already evaluating quadratic or cubic functions for various other purposes [Fuchs 89, Kirk 91, Akeley 92]. And for renormalization, HP has a hardware divide for z [Norrod 92] and square roots are similar in complexity to multiplies.

Direct Evaluation of Surface Normals

How close are the higher order approximations to the true normals of the underlying surface? For rational and non-rational spline patches, the non-normalized normal directions are given by non-rational spline (Table 1). For a bicubic patch, the normals are biquintic. For a forward differencing evaluation method, this would come to only two more additions per component over cubic interpolation. By using the exact normals, we avoid some of the more ellusive artifacts. For example, there are points along the parabolic curve with fourth order flatness called ruffles or nodal points where a pleat in the Gauss map ends at a cusp. At these locations, three specular points should come together. For the images in this paper, we used the normals of the underlying surface, evaluated from the linearly interpolated texture coordinates across the polygon. For improved images, we should use higher order interpolants for texture coordinates as well [Max 89]. Since the true correctness of our normals depends on the correctness of our parametric texture coordinates, we have called this parametric interpolation.

A simple example is shown in figure 3. The right most image shows the difference between the normals for the Phong and the parametric normals. Ignore the artifacts at the top of the knob on the lid (caused by incorrect normals in the model at a patch degeneracy). Note, however, the large errors on the curve of the spout, on the handle, and around the base of the knob, and also at areas of high curvature. Figure 4 is a close-up of the spout. The Phong version has a small specular point on the polygon boundary while the parametric normal versions

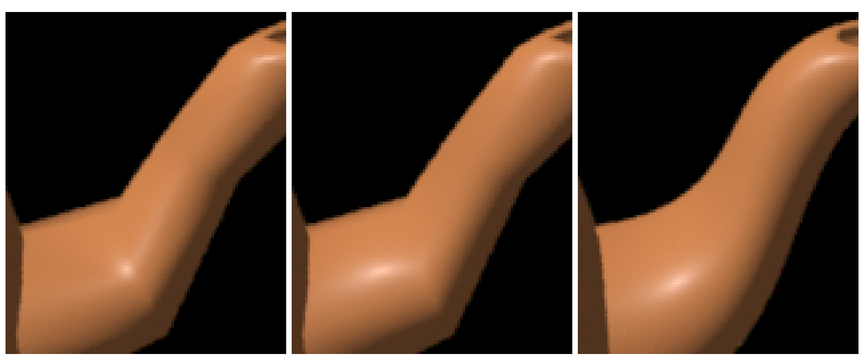

Figure 4. Close-up of teapot spout. Teapot rendered with 504 polygons. Left: Shading with linear normal interpolation. Middle: parametric surface normals. Right: True surface rendering. Note the spout "elbow": In the Phong model, specular highlights concentrate at polygon boundaries.

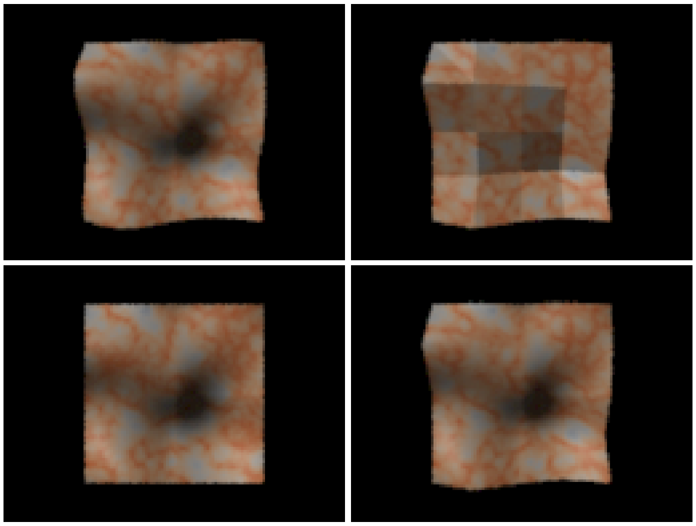

Figure 5. Band limited noise displacement map. Upper left: actual geometry of the surface. Upper right: surface positions of a single square polygon displaced by parametric map. Lower left: view of a single square polygon with normals perturbed by a parametric bump map. Lower right: polygon with a combined parametric displacement map and a parametric bump map.

has an the correct extended highlight within that polygon centered around the parabolic curve.

Shading is not limited to parametric patches, and there is no reason we should be limited either. Figure 5 shows a simple displacement map on a flat square. We created a bump map from the exact normals of the displacement map. The displacement map is applied to facet vertices and the bump map fills in the detail.

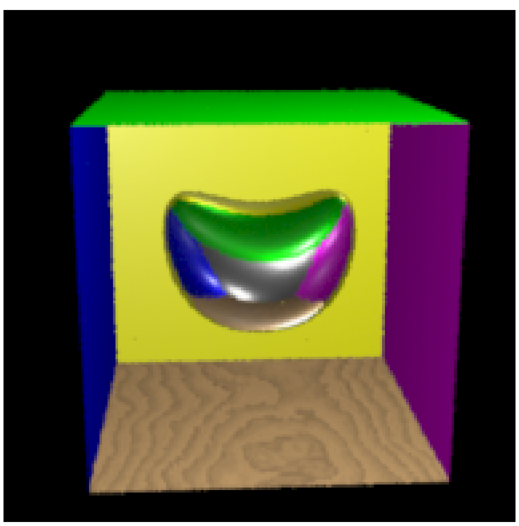

Figure 6. Representation of the test scenario—test figure rendered as a highly reflective surface with an environment map within a colored box. Facing surface is gray (removed in this image for visibility).

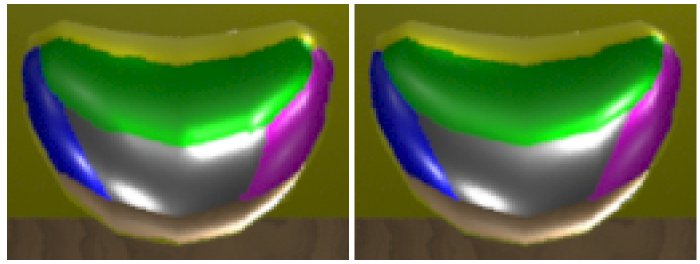

Figure 7. Comparison of the two interpolation methods using an environment map. Test object represented with 128 polygons. Left: linear normal interpolation. Right: parametric normal interpolation.

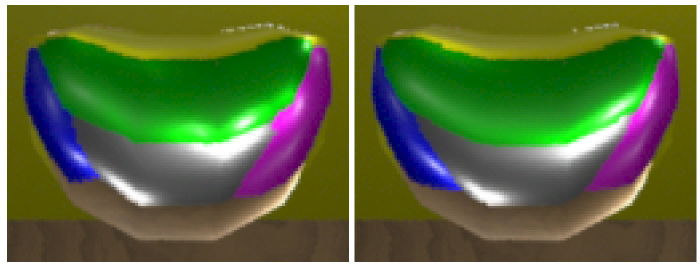

Figure 8. Comparison of the two interpolation methods using an environment map. Test object represented with 98 polygons. Left: linear normal interpolation. Right: parametric normal interpolation.

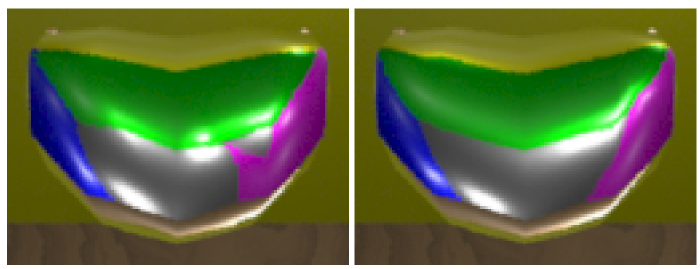

Figure 9. Comparison of the two interpolation methods using an environment map. Test object represented with 72 polygons. Left: linear normal interpolation. Right: parametric normal interpolation.

Results

While the behavior of surface normals is barely discernible on purely elliptical or purely hyperbolic regions, the nature of surface normals is very evident near inflection points between these regions. We present a highly reflective surface to demonstrate some of the most compelling reasons for higher order interpolants. In a highly reflective surface, reflection direction of all points on the surface are visible simultaneously.

Figure 6 shows a colored box with the test object (a sphere perturbed using a parametric displacement map). The surface of the test object is reflective with very little ambient light present, and no diffuse component.

Figures 7 through 9 show a progression of images of polygonalized forms of the test object through successively lower levels of geometric detail. The image pairs represent the test object using 128 triangles in Figure 7, 98 triangles in Figure 8, and 72 triangles in Figure 9 Within each image pair, the separate views were rendered using a different surface normal interpolation method; the left image was generated using a lighting model with linear Phong normal. The right image was prepared using parametric evaluation of surface normals from interpolated surface coordinates.

In each case, the model with the linear interpolant shows significant discontinuities of surface reflection and specular highlights at polygon boundaries. By contrast, parametric interpolation provides a more consistent representation as underlying geometric detail is removed. In each of the images, note the curved highlight along the border of the reflected green ceiling. The highlight distorts and eventually splits (See Figure 9) when using the linear interpolant.

As expected, the greatest degree of error is noticeable along the parabolic curve (the locus of zero Gaussian curvature) that divides the saddle shaped region from the remaining elliptical portions of the surface. The linear interpolant suffers from normals that attempt to follow polygon boundaries. The reflection artifact of the magenta side wall in Figure 9 is an example of normals erroneously following polygonal boundaries. Specular highlights are divided and regenerate on either side of a polygon. By contrast, the parametric interpolant does not exhibit these behaviors,; despite the low geometric complexity, is rendered relatively faithfully with respect to the test image.

Future Work

While these methods ensure C^1 continuity of surface normals, some artifacts persist due to the underlying interpolation of the parametric coordinates. Higher order continuity of normal values can be guaranteed by improving the interpolation of parametric coordinates. Values for δs/δsurface, and δt/δsurface can be stored or computed at each vertex, and a higher order interpolant implemented to provide even smoother transitions across polygonal boundaries.

Finally, the work presented here is constructed solely on generative and parametric surfaces: Hermite interpolants (bicubic patches in this case) or polynomial surface descriptions with procedural displacement maps. However, the techniques are not limited to procedural maps. Image-based surface maps controlling color and opacity are common; interpolation mechanisms to correct perspective distortions are well understood. An image based texture could be used to provide improved normals for a surface in a manner similar to bump mapping.

Conclusions

Achieving a Phong lighting model (with linear interpolation of surface normals) is considered the current goal of many designers of computer graphics architectures. The images presented in this paper indicate that when this has been achieved, there are still progressive improvements to be made. The non-fidelities in a linear interpolation scheme for surface normals cannot be ignored. Fortunately, processing power at rendering time is becoming more plentiful, and powerful texturing tools are no longer uncommon. These developments indicate that parametric normal interpolation may be the next hurdle beyond the upcoming generation of graphics systems.

Clearly, there is a step beyond the classic lighting model. For global illumination, there are the issues covered by more complex Bidirectional Reflection Functions (BDRFs). We submit that there are additional issues in interpolating surface normals, and that parametric evaluation of surface normals yields a more acceptable representation, with smoother transitions among normal values.

Acknowledgments

This work has been supported by the National Science Foundation (NSF), NSF Grant No. MIP-9000894, the Defense Advanced Research Projects Agency (DARPA), ISTO Order No. 7510, and through the NSF/DARPA Science and Technology Center for Computer Graphics and Scientific Visualization (NSF Cooperative Agreement No. ASC-8920219).

Special thanks to Pixar for access to a RenderMan™ license without which this research would have progressed much slower.

References

[Akeley 92] Akeley, Kurt. *Personal communication.*

[Blinn 78] Blinn, James F..*Computer Display of Curved Surfaces*, PhD Dissertation, December 1978, University of Utah.

[Chang 1989] Chang, Sheue-Ling, Shantz, Michael and Robert Rocchetti "Rendering Cubic Curves and Surfaces with INterger Adaptive Forward Differencing" *Computer Graphics*, Volume 23, Number 3 (Proceedings SIGGRAPH 1989), pp. 157-166, July 1989.

[Cook 1987] Cook, Robert L., Loren Carpenter, Edwin Catmul, "The Reyes Rendering Architecture," *Computer Graphics*, Volume 21, Number 4 (Proceedings SIGGRAPH 1987), pp. 95-102, July 1987.

[Farin 90] Farin, Gerald. *Curves and Surfaces for Computer Aided Geometric Design*. 2nd Edition. San Diego, California: Academic Press, 1990.

[Fuchs 89] Fuchs, Henry, et. al. "Pixel-Planes 5: A Heterogeneous Multiprocessor Graphics System Using Processor-Enhanced Memories," *Computer Graphics*, Volume 23, Number 3 (Proceedings SIGGRAPH 29). pp. 79-88, July 1989

[Kajiya 82] Kajiya, James "Raytracing Parametric Patches," *Computer Graphics*, Volume. 16, Number 3 (Proceedings SIGGRAPH 1982), pp. 245-254, July 1982.

[Kirk 90] Kirk, David and Douglas Voorhies, "The Rendering Architecture of the DN10000VS," *Computer Graphcis*, Volume 24, Number 4 (Proceedings SIGGRAPH 1990), pp. 299-272, August 1990

[Koenderink 90] Koenderink, Jan J. *Solid Shape.* Cambridge, Massachusetts: the MIT Press, 1990.

[Max 89] Max, Nelson, "Smooth Appearance for Polygonal Suraces," *The Visual Computer*, Volume ?, Number 5, Springer-Verlag, pp. 160-173, 1989.

[Moreton 92] Moreton, Henry P. and Carlo H. Sequin, "Functional Optimization for Fair Surface Design" *Computer Graphcis*, Volume 26, Number 2 (Proceedings SIGGRAPH 1992), pp. 167-176, July 1992

[Norrod 91] Norrod, Forrest. *Personal Communication.*

[Shantz 1987] Shantz, Michael and Sheue-Ling Lien., "Shading Bicubic Patches," *Computer Graphics*, Volume 21, Number 4 (Proceedings SIGGRAPH 1987), pp. 189-196, July 1987.

[Schweitzer 82] Schweitzer , Dino and Elizabeth Cobb. "Scanline Rendering of Parametric Surfaces," *Computer Graphics*, Volume. 16, Number 3 (Proceedings SIGGRAPH 1982), pp. 265-273, July 1982.

[Upstill 90] Upstill, Steve. *The Renderman Companion.* Reading, Massachusetts: Addison-Wesley Publishing Company, 1990.