Computational Vision
U. Minn. Psy 5036
Daniel Kersten
Lecture 15: Geometry & surfaces: depth & shape

Initialize

■ Spell check off

In[1]:=
```
Off[General::spell1];
```

In[2]:=
```
Off[General::spell1];
```

In[3]:=
```
<< VectorFieldPlots`;
```

In[4]:=
```
SetOptions[ArrayPlot, ColorFunction → "GrayTones", DataReversed → True,
  Frame → False, AspectRatio → Automatic, Mesh → False,
  PixelConstrained → {1, 1}, ImageSize → Small];
```

Outline

Last time

- **Extrastriate cortex--overview**
- **Scenes from images, scene-based modeling of images**

Today

- **Geometry, shape and depth: Mainly representation & generative models**
- **Lambertian model**

Surfaces, geometry & depth

Introduction

Recall two major tasks of vision: "Knowing what is where just by looking"--Marr's definition of vision.

Today: How can vision extract geometrical information about the world? Important for what and where.

General issues: Between-object, viewer-object, within-object geometry.

Coarse vs. dense estimations of geometrical relations.

Two basic classes of geometrical information for vision

- **Scene geometry--Spatial layout, large-scale surface structure**

  Where are objects relative to the viewer?

  Where are they relative to each other? Relative to a frame? (e..g. ground plane)

Object geometry--Surfaces & shape, small scale surface structure

How can we describe objects themselves in terms of their geometry--shape?

What is the relationship of parts of objects to each other?

Extrinsic vs. intrinsic geometrical descriptions

Where are objects? Spatial layout

■ Absolute

Distance of objects or scene feature points from the observer.

"Physiological cues": Binocular convergence--information about the distance between the eyes and the angle converged by the eyes. Crude, but constraining. Errors might be expected to be proportional to reciprocal distance. Closely related to accommodative requirements.

"Pictorial cue"--familiar size

Pattern of errors can depend on how human absolute depth is assess (e.g. verbal estimates vs. walking) (Loomis et al., 1992)

Important for reaching. (Marrotta and Goodale, 2001).

■ Relative

Distance between objects or object feature points. Important for scene layout, planning actions, navigation.

Processes include: Stereopsis (binocular parallax) and motion parallax.

Also information having to do with the "pictorial" cues:

occlusion, transparency, perspective, proximity luminance, focus blur, also familiar size & "assumed common physical size", "height in picture plane", cast shadows, texture & texture gradients for large-scale depth & depth gradients

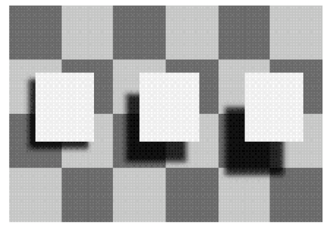
■ Examples of pictorial information for depth

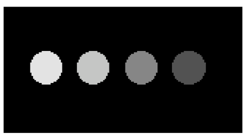

height in field of view


interposition

linear perspective


familiar size


cast shadows


proximity luminance

For some depth from shadow illusions, see: http://gandalf.psych.umn.edu/~kersten/kersten-lab/demos/shadows.html

■ More later

...over a dozen cues to depth. Later, we'll study theories of integration (e.g. stereo + cast shadows). Also theories of cooperativity (e.g. motion parallax <=> transparency).

Vision for spatial layout of objects, navigation, heading and for reach

What is geometrical shape?

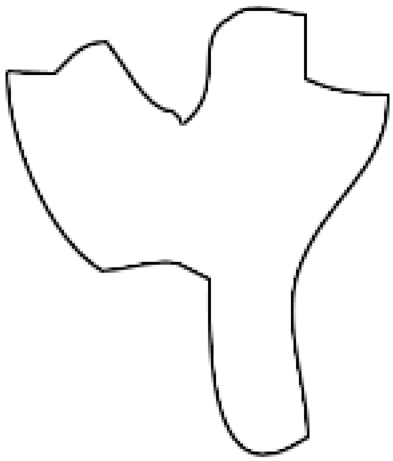
Important for determining what an object is--shape-based recognition, rather than where it is.

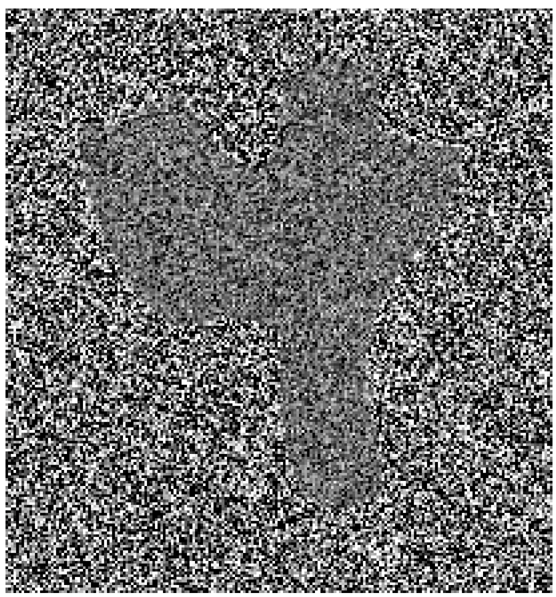
(Bear in mind other information for object identity, e.g. material, context)

■ Contours & region information

1D "lines" vs. 2D "fields", object contours vs. region information

2D and 3D shape

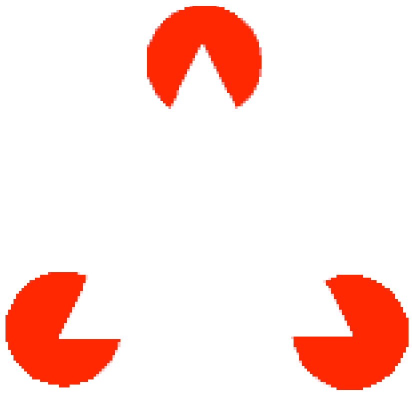




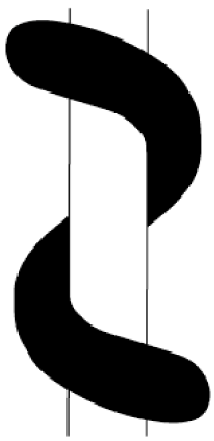
■ Cues to shape: Contour based

Contour based--contours at object boundaries (depth discontinuity) and at sharp object bends (orientation discontinuities, smooth self-occluding contours).

The Kanizsa triangle below illustrates contours at apparent depth discontinuities. The illusion illustrates the effectiveness of contours. Cartoons illustrate the effectiveness of line-drawings.



The contours below are interpreted as self-occluding contours, where the surface smoothly wraps out of view. Is the "worm" flat or round? Is the vertiical rod flat or round?

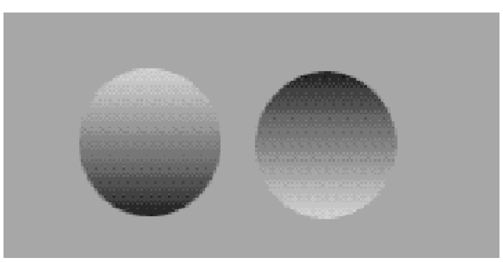


http://journalofvision.org/3/4/4/article.aspx

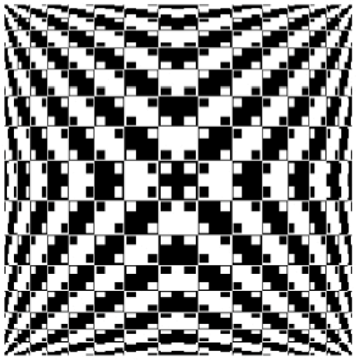
■ Cues to shape: Region based

Region based: shading, texture gradients, disparity gradients & fields, motion parallax fields

■ Example of shading & shape



■ Example of texture and shape



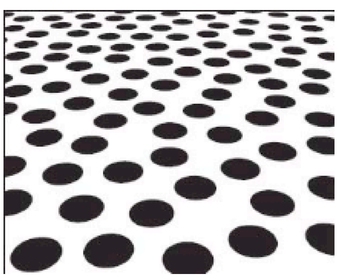
http://www.ritsumei.ac.jp/~akitaoka/index-e.html

Later, we'll look at the interactions between contour and region information in the perception of shape.

■ Local vs. global shape.

Global measures: aspect ratio, compactness

Dense vs. coarse fields. Interpolation.



Local shape fields & dense estimation

■ From dense depth to shape, normals & curvature

Below we will start with a mathematically simple representation of depth from the viewer: z = f(x,y), and show how to derive a simple dense shape measure in terms of the rate of change of depth from the viewer. This a viewer-centered depth representation. Later, we will discuss intrinsic object shape measures such as curvature.

An issue that is perceptually important and of theoretical interest is: How to go from a set of sparse measurements in a region to a dense field. This is a problem of surface interpolation. We'll see more of that later.

Local dense representations of surface regions

Shape-based generative modeling of images

In terms of understanding image information, it is simplest to start with a discussion of dense local representations of shape and understand how shape changes influence image intensity changes.

There are two aspects to image formation. The first has to do with geometrical manipulations including projection (which we will study later via matrix transformations on homogeneous coordinates). The second aspect is the determination of intensity at each image point from a scene description. In order to understand shape from shading, we need to understand the constraints implicit in the forward optics problem of "shading from shape"--i.e. the generative model. Here we can benefit from research in attempts in computer graphics to obtain photorealism.

Image formation constraints can be obtained by understanding how material properties, shape, and illumination interact to form an image. This is part of the field of computer graphics. The quest for physical realism in reasonable computing time is still a challenge in computer graphics (Greenberg, 1999; for eary history, see: Blinn, J. F., 1977; Cook, R., & Torrance, K., 1982).

One of the earliest models is the Lambertian shading equation which describes how intensity is distributed for curved matte surfaces with constant reflectance (e.g. arbitrarily defined as 1):

$$L(x,y) = \hat{E} \cdot \hat{N}(x,y)$$

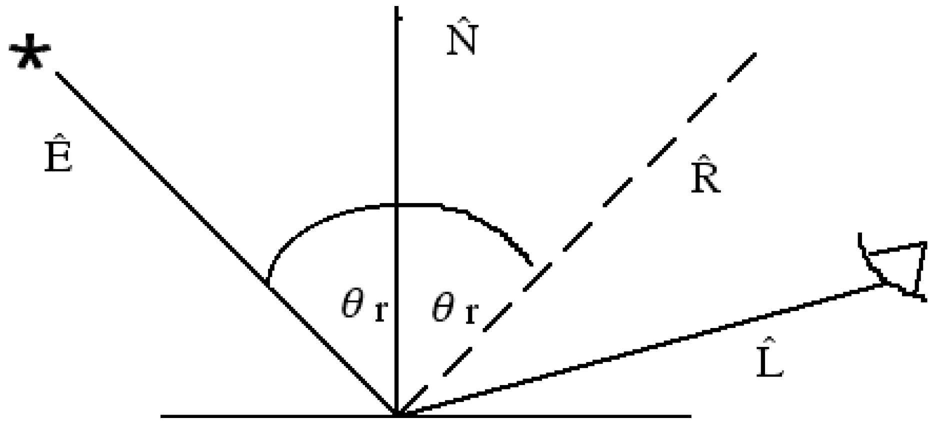
E is a vector representing light source direction. If at infinity, it has just two degrees of freedom and is constant over the surface. If light source intensity is normalized, then the E vector is a unit vector. Alternatively its vector length can be used to indicate the strength of the illumination.

■ An example of a simple computer graphics lighting model--"matte" & "plastic" world

An elaboration of the lambertian model for shading, based in part on physics, and in part on heuristics and "beauty pageant" observation, describes image luminance in terms of of a lambertian, specular, and ambient light components:

$$L = r_a E_a + \frac{E_p}{R+K} \left[r(\hat{E} \cdot \hat{N}) + r_s (\hat{R} \cdot \hat{L})^n \right]$$

where the lower case r's are the reflectivities (0 means no reflectance of the illumination, and 1 means complete reflectance), and the unit vectors E, N, and L, point in the directions of the light source, surface normal, and viewpoint respectively. Ep and Ea, are the strengths of the point source, and ambient illumination, respectively. R points in the direction that a ray would go if the surface was a mirror, i.e. purely specular. This form for the specular term is due to Phong, who pointed out that n could be used to control the degree of specularity. High values of n correspond to a perfect mirror surface, values between 1 and 200 are typically used to add some gloss to the rendering. Surfaces then look more plastic, or metallic.



The R+K term is a "fudge" term, reflecting common experience (not physics) in which surfaces (of the same reflectance, and orientation) that are farther away from the viewer are dimmer. K is a constant, and R is the distance from the object to the viewpoint. This is a weak, but nevertheless useful constraint for vision. A closely related idea is "proximity luminance" cues in vision (see Dosher, B. A., Sperling, G., & Wurst, S. A., 1986). Our viewpoint is likely to be close to the light source direction than opposite it.

For more on rendering, including physically correct phong-like model, see: Larson, G. W., & Shakespeare, R. (1998).

Note that as is, there are severe limitations to this image formation constraint. One of the major problems is the absence of cast shadows. Physically based models become much more complicated when one has to take into account multiple reflections. (See Foley, J. D., van Dam, A., Feiner, S. K., & Hughes, J. F., 1990; Fundamentals of Computer Graphics by Peter Shirley et al.).

The above equation also does not take into account how light bounces between objects (mutual illumination or indirect lighting)--the ambient "fudge" term is the crude approximation to model the overall effect of mutual illumination.

Material properties can be much more complicated than Lambertian plus a Phong specular term. Illumination patterns can also be much more complicated.

More on the complexities of the image formation model latter.

■ Preview of shape-from-shading

Later we will study the "shape-from-shading" problem. If one represents shape in terms of a dense distribution of surface normals, then a simplified version of the formal problem is to estimate N(x,y) given data L(x,y) such that the following simplification of the above equations holds:

$$L(x,y) = \hat{E} \cdot \hat{N}(x,y)$$

As it stands, this set of equations (one for each location x,y) is underconstrained or "ill-posed"--for every image intensity L, there are two numbers to estimate for N. Assuming the light source is a point at infinity simplifies things (same two numbers E for all surface points). Assuming surface smoothness and integrability also constrains the solution. But more on this later.

Representing shape

A central issue in object perception is how the shape of an object is represented by the visual system. Shape may be represented in a variety of ways that depend on the visual task and the stages of processing in a given task. Questions about shape representation can be classified along several dimensions. Two central questions have to do with whether the representation is local or global, and whether the representation depends on viewpoint.

Global and local representations of shape. A global representation of solid shape consists of a set of parameters or templates that describe a class of surfaces or ``parts". Several theories of object recognition assume that objects can be decomposed into elementary parts. These parts are drawn from a limited set of elementary shapes, such as generalized cylinders (Marr and Nishihara, 1978), geons (Biederman, 1987) or superquadrics (Pentland, 1990).

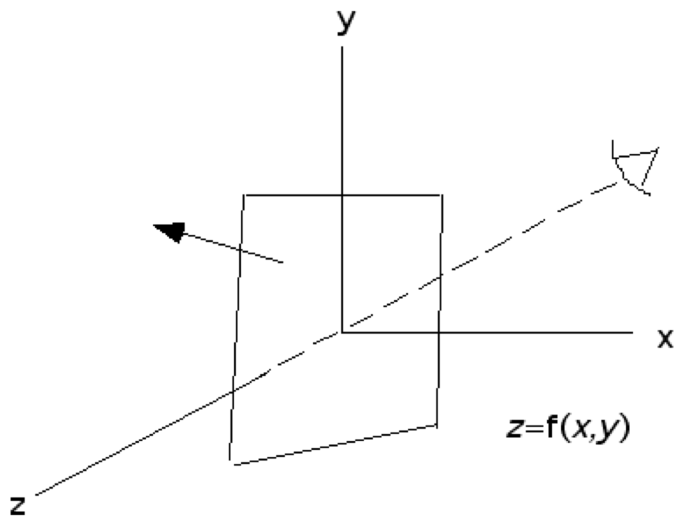
A representation in terms of parts is global in the sense that the change of each parameter which describes a part will affect the *whole* shape. In comparison, a local representation is a dense characterization of shape, such that a change of a parameter at one spatial location will not affect the shape at another location. A surface normal vector at each surface location is an example of a dense representation.

Viewpoint dependency in shape representations. A second major question is whether the shape representation depends on viewpoint. This debate has arisen in the context of models of object recognition (Tarr & Bülthoff, 1995). Local viewpoint dependent descriptors such as slant and tilt have been studied by a number of investigators (Todd & Mingolla, 1983; Mingolla & Todd, 1986; see below). But view-dependent descriptors have disadvantages for some tasks. One problem with slant and tilt is that the local slant of an oriented plane varies with viewpoint. So we have a discrepancy between apparent global flatness and the local variation in slant (Mamassian, 1995). For part-based object recognition, it would seem best to do part extraction using local view-independent descriptors. It is also reasonable that view-independent shape descriptors could support other types of visual processes. For instance, the manual prehension of an object requires one to locate stable grasp points on the surface, a task which only makes sense in an object-centered frame of reference. Nevertheless, the visual information is of course firstly described in a viewer-centered frame of reference, and the fundamental issue then becomes the tranformation of viewpoint dependent into viewpoint independent representation (Andersen, 1987). One local, intrinsic representation of solid shape describes the second order depth variation of the surface (Besl and Jain, 1986), or equivalently, the first order orientation variation (Rogers and Cagenello, 1989). For more on shape representations, see Koenderink (1990).

Gradient space & surface normals

Imagine a small planar surface patch.

One way of representing shape locally and with respect to the viewpoint of the observer, is to use **gradient space:**



Let φ(*x*,*y*,*z*)=f(*x*,*y*)-*z* Then,

$$\nabla\phi = \frac{\partial\phi}{\partial x}\,\hat{\mathrm{i}} + \frac{\partial\phi}{\partial \mathrm{y}}\,\hat{\mathrm{j}} + \frac{\partial\phi}{\partial \mathrm{z}}\,\hat{\mathrm{k}}$$

is a normal vector to the surface at *(x,y,f(x,y))* :

$$\hat{\mathrm{n}} = \frac{\nabla\phi}{||\nabla\phi||}$$

$$\nabla\phi = \frac{\partial\,\mathrm{f}}{\partial x}\,\hat{\mathrm{i}} + \frac{\partial\,\mathrm{f}}{\partial \mathrm{y}}\,\hat{\mathrm{j}} - \hat{\mathrm{k}} \;=\; p\hat{\mathrm{i}} + q\hat{\mathrm{j}} - \hat{\mathrm{k}} \longrightarrow (p,\, q,\, -1)$$

Gradient space is defined by the mapping of (*p*,*q*,*-1*)-->(*p*,*q*), ie the orthographic projection.

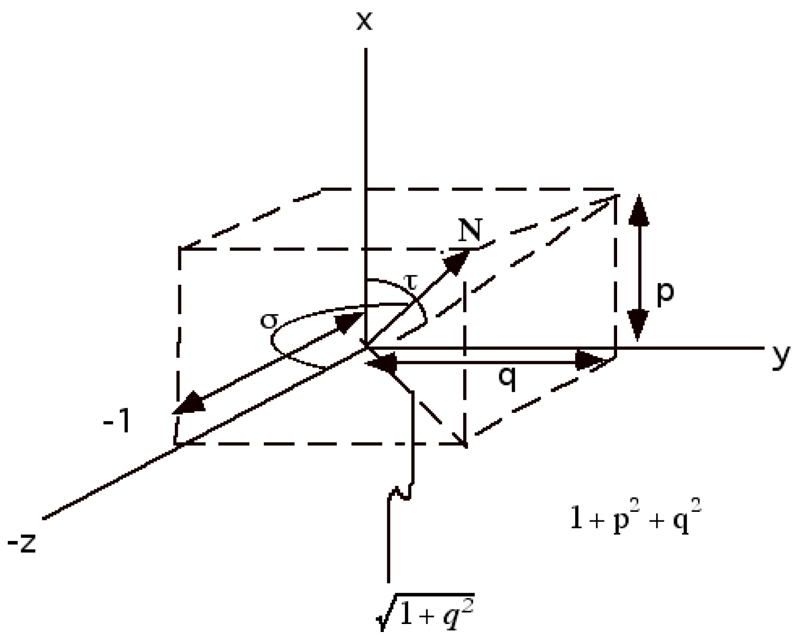
Slant and tilt

Again, imagine a small planar surface patch.

Shape can also be described by **slant** and **tilt.**

Used for both local dense estimation, and global near-planar surface attributes, e.g. for large-scale layout.

Let **N** be the surface normal expressed in terms of p and q:



The direction of steepest descent is the *tilt:*

$$\tau = \tan^{-1}\left(\frac{q}{p}\right)$$

Note: Sqrt[q^2+p^2] is the rate of change of z in the direction of maximum change (steepest descent).

The slant is:

$$\sigma = \tan^{-1}\left(\sqrt{q^2 + p^2}\right)$$

How to choose a representation?

How do we know what is the best representation to use? Slant and tilt seem to be important from a perceptual point of view.

One reason gradient space is useful is that it is related to relative depth in a straight forward way:

We can easily get back to distance by integrating:

$$\mathrm{d}z = p\ \mathrm{d}x + q\ \mathrm{d}y$$

$$z_1 - z_2 = \int_{\text{path } p_o}^{p_1} \nabla f \cdot \mathrm{d}\mathbf{r}$$

gives the relative distance, and a constant is lost in the process.

Other local representations (Koenderink, 1990; Mamassian & Kersten 1996, Mamassian, Knill & Kersten, 1996).

More on the Lambertian model: Using *Mathematica* to go from depth to normals to image intensities

This section gives you some practice using the Lambertian scene-based generative model.

Range data defines surface list

- **Range data to define rface list -- BIG 64x64 file.**

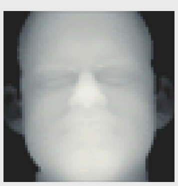
  Just initialize this cell--don't bother opening cell (from: http://sampl.eng.ohio-state.edu/)

  If you want to read in a different file, use the Import[ ] function in the Appendix

Intensity in a DensityPlot is proportional to the z (depth) from the camera

In[6]:=
```
size = Dimensions[rface][[1]]; hsize = size / 2;
ArrayPlot[rface]
```

Out[7]=



Fit continuous 3D function to range surface list

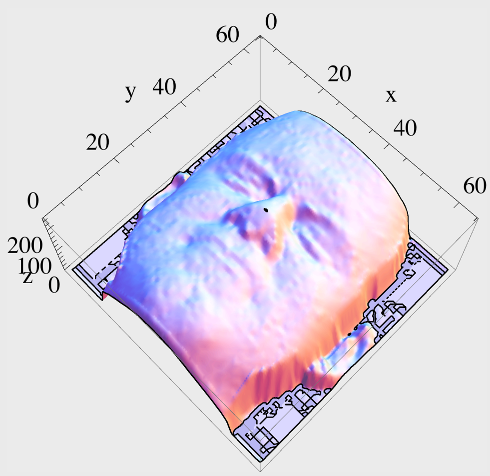
- **Spline-fit the first surface face to define a continuous function**

In[8]:=
```
crface = ListInterpolation[Transpose[rface]];
```

Plot 3D surface

In[9]:=
```
g1 = Plot3D[crface[x, y], {x, 1, size}, {y, 1, size}, PlotPoints → 64,
  PlotRange → {0, 255}, Mesh → False, AxesLabel → {"x", "y", "z"},
  ViewPoint → {1, -1, 3}, AspectRatio → 1,
  PlotRange → {{1, size}, {1, size}, {0, 255}}, ImageSize → Small]
```

Out[9]=




```
crface[32.2, 32]
```

249.144

Calculate surface normals of surface

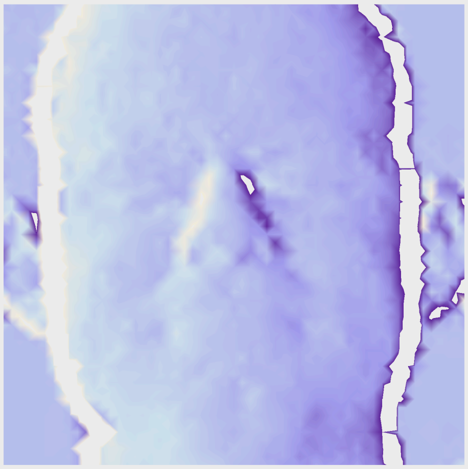
In[10]:=
```
Clear[x, y]
```

In[11]:=
```
nx[x_, y_] := Evaluate[D[crface[x, y], x]];
ny[x_, y_] := Evaluate[D[crface[x, y], y]];
```

nx[x,y] is $\frac{\partial z}{\partial x}$, and similarly for ny. The rate of change of depth range is greatest as the face slopes away from the viewpoint:

In[16]:=
```
DensityPlot[nx[x, y], {x, 1, size}, {y, 1, size}, Mesh → False,
 Frame → False, ImageSize → Small]
```

Out[16]=



As we saw for intensity, we can also make a plot of the magnitude of the gradient:

```
DensityPlot[Sqrt[nx[x, y]^2 + ny[x, y]^2], {x, 1, size}, {y, 1, size},
  Mesh → False, Frame → False, ImageSize → Small];
```

Out[14]=

Lambertian rendering: specification for normals, light, reflectance

■ Unit surface normals

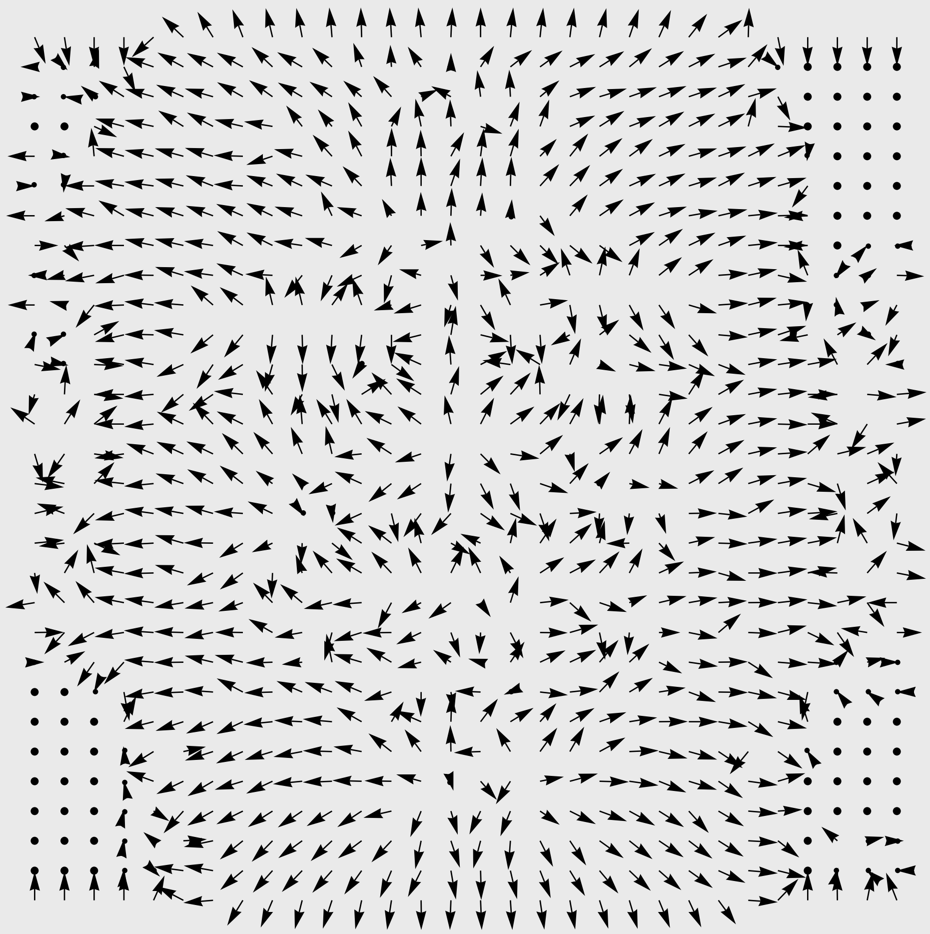
Normalize the surface normal vectors to unit length:

In[51]:=
```
normface[x_, y_] := -{nx[x, y], ny[x, y], 1} /
   Sqrt[nx[x, y]^2 + ny[x, y]^2 + 1];
```

■ **Here is a 2D sparse plot of the x and y components of the surface normals**

In[18]:=
```
VectorFieldPlot[{normface[x, y][[1]], normface[x, y][[2]]},
 {x, 1, size}, {y, 1, size}, PlotPoints → 30]
```

Out[18]=



■ **We can also plot the unit surface normals on the surface itself in 3D**

In[32]:=
```
facearray =
  Flatten[Table[N[{{x, y, crface[x, y]}, 5 * normface[x, y]}],
    {x, 1, size, 2}, {y, 1, size, 2}], 1];
```

In[39]:=
```
ListVectorFieldPlot3D[facearray, VectorHeads → True,
  ViewPoint → {1, -1, 3}, AspectRatio → 1,
  PlotRange → {{1, size}, {1, size}, {0, 255 * 2}}, BoxRatios → {1, 1, 3}]
```

Out[39]=

■ Point light source direction

Let's combine plots to show the direction of the illumination vector

s is a vector specifying the direction of the light source. The length can be used as to represent intensity. We normalize it.

In[52]:=
```
s = {100, 100, 10}; s = N[s / Sqrt[s.s]]
```

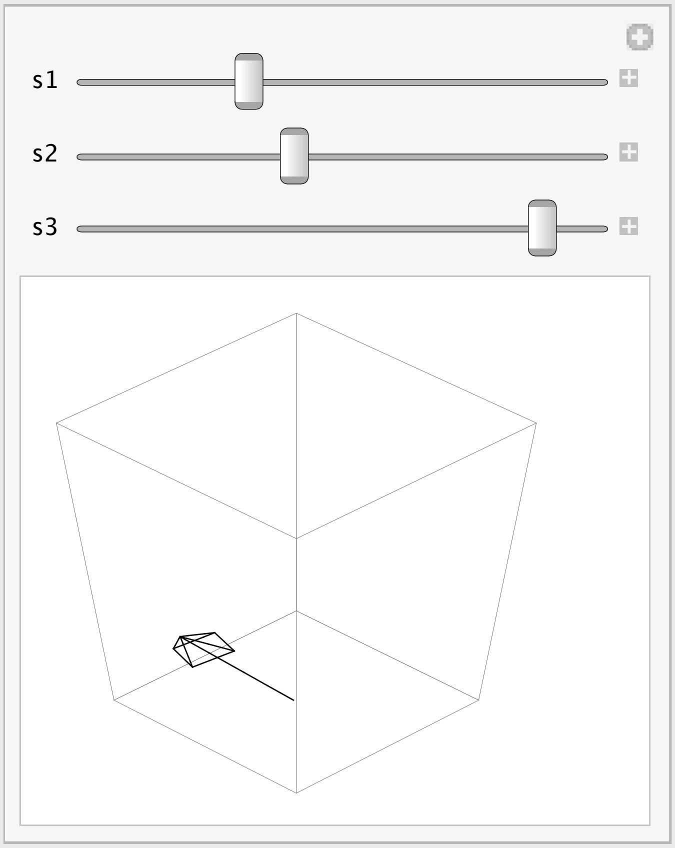
Out[52]= {0.705346, 0.705346, 0.0705346}

We can try to plot the light source vector together with the surface

In[61]:=

```
Manipulate[
 s = {s1, s2, s3};
 s = N[s / Sqrt[s.s]];
 g3 = ListVectorFieldPlot3D[{{{size / 2, size / 2, 0}, 60 * s}},
   ViewPoint → {1, -1, 3}, AspectRatio → 1,
   PlotRange → {{1, size}, {1, size}, {0, 255}}, VectorHeads → True,
   ImageSize → Small], {s1, -10, 10}, {s2, -10, 10}, {s3, -10, 10}]
```

Out[61]=

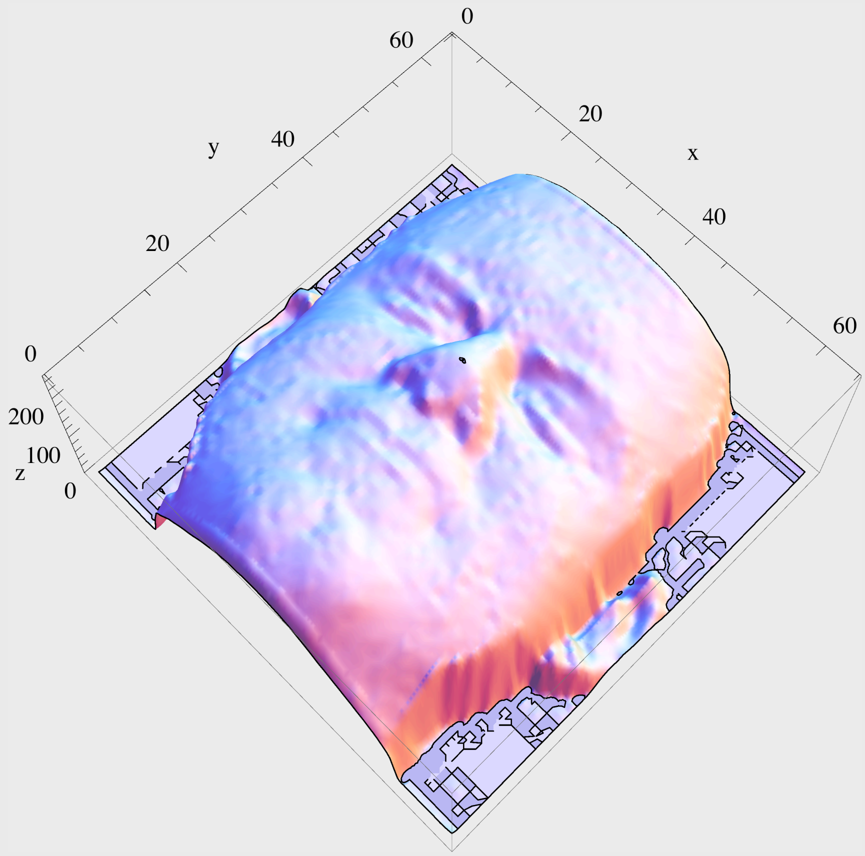


In[62]:=

```
s = {100, 100, 10}; s = N[s / Sqrt[s.s]];
g1 = Plot3D[crface[x, y], {x, 1, size}, {y, 1, size}, PlotPoints → 64,
   PlotRange → {0, 255}, Mesh → False, AxesLabel → {"x", "y", "z"},
   ViewPoint → {1, -1, 3}, AspectRatio → 1,
   PlotRange → {{1, size}, {1, size}, {0, 255}}];
```

In[64]:=
```
g12 = Show[{g1, g3}]
```

Out[64]=



■ Reflectance

The surface is modeled as having constant reflectivity of 1--i.e. "white".

In[65]:=
```
a[x_, y_] := 1;
```

■ Lambertian rendering model for first surface

In[74]:=
```
ss = {100, -100, 10}; ss = N[ss / Sqrt[ss.ss]];
imageface[x_, y_] := a[x, y] * normface[x, y].ss;
```

Render face surface

In[76]:=
```
DensityPlot[imageface[x, y], {x, 1, size}, {y, 1, size}, Mesh → False,
 PlotPoints → 32, Frame → False, ImageSize → Small]
```

Out[76]=

Next time

- **Shape from X, shape from shading**

References

Blake, A., & Bülthoff, H. H. (1990). Does the brain know the physics of specular reflection? Nature, 343, 165-169.

Blake, A., & Bülthoff, H. H. (1991). Shape from specularities: Computation and Psychophysics. Philosophical Transactions of the Royal Society (London) Series B, 331, 237-252.

Blinn, J. F. (1977). Models of light reflection for computer synthesized pictures. Computer Graphics, 11(2), 192-198.

Brady, M., & Yuille, A. (1984). An extremum principle for shape from contour. IEEE Transactions on pattern analysis and machine intelligence, PAMI-6(3), 288-301.

Cook, R., & Torrance, K. (1982). A reflectance model for computer graphics. ACM TOB, 1(1), 7-24.)

Dosher, B. A., Sperling, G., & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. Vision Research, 26(6), 973-990.

Dupuis, P., & Oliensis, J. (1994). An optimal control formulation and related numerical methods for a problem in shape reconstruction. The Annals of Applied Probability, 4(No, 2), 287-346.

Foley, J. D., van Dam, A., Feiner, S. K., & Hughes, J. F. (1990). Computer Graphics: Principles and Practice (2nd ed.). Reading, Massachusetts. Addison-Wesley.)

Freeman, W. T. (1994). The generic viewpoint assumption in a framework for visual perception. Nature, 368(7 April 1994), 542-545.

W. T. Freeman, Exploiting the generic viewpoint assumption, International Journal Computer Vision, 20 (3), 243-261, 1996. TR93-15.

W. T. Freeman, The generic viewpoint assumption in a Bayesian framework, in Perception as Bayesian Inference, D. Knill and W. Richards, eds., Cambridge University Press, 365 - 390, 1996.

Greenberg, D. P. (1989). Light reflection models for computer graphics. Science, 244, 166-173.

Greenberg, D. P. (1999). A framework for realistic image synthesis. J Commun. ACM, 42(8), 44-53

Ikeuchi, K., & Horn, B. K. P. (1981). Numerical shape from shading and occluding boundaries. Artificial Intelligence, 17,, 141-184.

Kersten, D., O'Toole, A. J., Sereno, M. E., Knill, D. C., & Anderson, J. A. (1987). Associative learning of scene parameters from images. Applied Optics, 26, 4999-5006.)

Kersten, D. (1999). High-level vision as statistical inference. In M. Gazzaniga (Ed.), The Cognitive Neurosciences Cambridge, MA: MIT Press.

Knill, D. C., & Kersten, D. (1990). Learning a near-optimal estimator for surface shape from shading. Computer Vision, Graphics and Image Processing., 50(1), 75-100.

Koenderink, J. J. (1990). Solid Shape . Cambridge, Massachusetts: M.I.T. Press.

Larson, G. W., & Shakespeare, R. (1998). *Rendering With Radiance: The Art and Science of Lighting Visualization*.

Langer, M. S., & Zucker, S. W. (1994). Shape from Shading on a Cloudy Day. Journal of the Optical Society of America A, 11(2), 467-478.

Lehky, S. R., & Sejnowski, T. J. (1988). Network model of shape from shading: neural function arises from both receptive and projective fields. Nature, 333, 452-454.

Loomis, J. M., Da Silva, J. A., Fujita, N., & Fukusima, S. S. (1992). Visual space perception and visually directed action. *J Exp Psychol Hum Percept Perform, 18*(4), 906-921.

Mamassian, P., & Kersten, D. (1996). Illumination, shading and the perception of local orientation. *Vision Res, 36*(15), 2351-2367.

Mamassian, P., Kersten, D., & Knill, D. C. (1996). Categorical local-shape perception. *Perception, 25*(1), 95-107.

Marotta, J. J., & Goodale, M. A. (2001). Role of familiar size in the control of grasping. *J Cogn Neurosci, 13*(1), 8-17.

Mingolla, E., & Todd, J. T. (1986). Perception of Solid Shape from Shading. 53, 137-151.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. Science, 257, 1357-1363.

Oliensis, J. (1991). Uniqueness in Shape From Shading. The International Journal of Computer Vision, 6(2), 75-104.

Ramachandran, V. S. (1988). Perceiving Shape from Shading. Scientific American, , 76-83.

Shirley, Peter, Michael Ashikhmin, Michael Gleicher, Stephen R. Marschner, Erik Reinhard, Kelvin Sung, William B. Thompson, Peter Willemsen (2005) . Fundamentals of Computer Graphics. 2nd edition.

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon-structural-descriptions or by multiple-views? Journal of Experimental Psychology: Human Perception and Performance, 21(6), 1494-1505.

Todd, J. T., & Mingolla, E. (1983). Perception of Surface Curvature and Direction of Illumination from Patterns of Shading. Journal of Experimental Psychology: Human Perception & Performance, 9(4), 583-595.

Todd, J. T., Norman, J. F., Koenderink, J. J., & Kappers, A. M. L. (1997). Effects of texture, illumination, and surface reflectance on stereoscopic shape perception. Perception, 26(7), 807-822.

Todd, J. T., & Reichel, F. D. (1989). Ordinal structure in the visual perception and cognition of smoothly curved surfaces. Psychological Review, 96(4), 643-657.

Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In K. D.C., & R. W. (Ed.), Perception as Bayesian Inference Cambridge, U.K.: Cambridge University Press.