MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo 820 January, 1985

# Shape and Source from Shading

**Michael J. Brooks**
**Berthold K.P. Horn**

**Abstract:** Well-known methods for solving the shape-from-shading problem require knowledge of the reflectance map. Here we show how the shape-from-shading problem can be solved when the reflectance map is not available, but is known to have a given form with some unknown parameters. This happens, for example, when the surface is known to be Lambertian, but the direction to the light source is not known. We give an iterative algorithm that alternately estimates the surface shape and the light source direction. Use of the unit normal in parameterizing the reflectance map, rather than the gradient or stereographic coordinates, simplifies the analysis. Our approach also leads to an iterative scheme for computing shape from shading that adjusts the current estimates of the local normals toward or away from the direction of the light source. The amount of adjustment is proportional to the current difference between the predicted and the observed brightness. We also develop generalizations to less constrained forms of reflectance maps.



This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the System Development Foundation. This work was done while Michael Brooks was a visiting scientist at MIT on leave from Flinders University.

## 1. Introduction

Given an image, $E$, and a reflectance map, $R$, the shape-from-shading problem may be regarded as that of recovering a smooth surface, $z(x, y)$, satisfying the image irradiance equation

$$E(x,y) = R\left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right)$$

over the domain $\Omega$ of $E$. Any given boundary conditions on $z(x, y)$ should also be satisfied. The problem takes the form of a first-order partial differential equation (Horn, 1970 & 1975).

Implicit in this formulation are a number of assumptions, the principal one being that the brightness of a surface patch does not depend on its position in space. Another is that an image depicts a smooth surface of homogeneous reflectance. Several algorithms have been devised to tackle the problem, notably those of Horn (1975), Strat (1979), and Ikeuchi & Horn (1981).

One of the many difficulties these schemes face in practice is that the reflectance map is typically not known. The reflectance map specifies how the brightness of a surface patch depends on its orientation under given circumstances. It therefore encodes information about the reflecting properties of the surface, and information about the distribution and intensity of the light sources. In fact, the reflectance map can be computed from the bidirectional reflectance-distribution function and the light source arrangement, as shown by Horn & Sjoberg (1979).

When encountering a new scene, the information required to determine the reflectance map is usually not available. Yet without this information, the shape-from-shading problem cannot be formulated, much less solved. The dilemma may be resolved if a calibration object of known shape appears in the scene, since the reflectance map can be computed from its image. Here we wish to consider the situation where we are not that fortunate.

It is interesting to evaluate how some basic assumptions can resolve this impasse. Pentland (1984) has looked at the problem of recovering shape from shading under the assumption that the image depicts a Lambertian surface illuminated by a point source whose direction is unknown. Under the additional assumption that the surface is locally umbilical, surface normals are shown to be recoverable by a local operation. This method does not depend on the iterative propagation of information across the image.

There are some serious drawbacks to the local approach, however. One problem is that the umbilical assumption is very restrictive. In fact, spheres are the only surfaces whose points are all umbilical. So this method naturally computes incorrect normals for other shapes, although the errors for approximately spherical surfaces, such as ellipsoids of low eccentricity, may be acceptable. Further, the constraining effect of known occluding boundary normals cannot be incorporated into the local method. This is unfortunate because these normals provide powerful boundary conditions on the shape-from-shading problem, as shown by Bruss (1983). Finally, because a local method does not take into account neighbors when calculating the normal at a point, nearby normals may differ a great deal, particularly in the presence of noise.

We now present an alternative approach that does not suffer from these disadvantages.

## 2. An iterative scheme for shape and source direction

The task is to recover the shape of a smooth surface depicted in an image, $E$, that is defined over a region $\Omega$ in the $xy$-plane. Let the shape of the surface be characterized by the function, $\mathbf{n}$, that associates a unit normal with each point in $\Omega$. The problem is therefore to find $\mathbf{n}(x, y)$ over $\Omega$. Assume for now that the object has a Lambertian surface, and that it is illuminated by a single point source. If the vector $\mathbf{s}$ points to the source, and $\mathbf{n}(x, y)$ is the unit normal of a surface patch, then the apparent brightness of the patch is given by the reflectance map

$$R_{\mathbf{s}}(\mathbf{n}(x, y)) = \mathbf{n}(x, y) \cdot \mathbf{s}.$$

We do not, by the way, force $\mathbf{s}$ to be a unit vector; this allows for the possibility that the intensity of the source may be unknown. This way we can also deal with unknown sensor sensitivity and unknown surface albedo, provided it is uniform.

Our problem then becomes one of finding a smooth shape, $\mathbf{n}$, and source direction, $\mathbf{s}$, satisfying the image irradiance equation

$$E(x, y) = \mathbf{n}(x, y) \cdot \mathbf{s} \qquad \forall (x, y) \in \Omega.$$

In practice, brightness cannot be determined with perfect accuracy, and so it appears reasonable to transform this into a minimization problem (Ikeuchi & Horn, 1981; Horn & Brooks, 1985). There is another reason to consider this as a minimization problem: if we simply try to solve the image irradiance equation as it stands, we obtain a set of differential equations equivalent to the characteristic strip equations. Here, however, we seek an iterative scheme lending itself to a parallel implementation on a grid, as originally suggested by Horn (1970). Further, a shape-from-shading problem that has noisy image data may well not have a theoretical solution. A minimization approach will, however, enable the recovery of a shape that fits the given data best, in a sense determined by the functional chosen.

We seek a smooth shape, $\mathbf{n}$, and a source direction, $\mathbf{s}$, that minimize

$$\iint_{\Omega} \left(E(x, y) - \mathbf{n}(x, y) \cdot \mathbf{s}\right)^2 \, dx \, dy.$$

If a solution exists, and there are no errors in brightness measurements, then the image irradiance equation will have been satisfied (although there is no guarantee that the resulting $\mathbf{n}$ will be integrable; see Horn & Brooks, 1985).

We also adopt a regularizing component (Poggio & Torre, 1984) by incorporating the expression

$$\iint_{\Omega} \left(\mathbf{n}_x^2(x, y) + \mathbf{n}_y^2(x, y)\right) \, dx \, dy,$$

which is intended to select a particularly smooth solution from a possibly infinite set of candidates. Note that a subscript here denotes partial differentiation, and that squaring a vector is equivalent to taking the dot-product with itself. Finally, we wish to insist that normals have unit length. This is accomplished with the constraint

$$\mathbf{n}^2(x, y) = 1 \qquad \forall (x, y) \in \Omega.$$

Combining the three terms gives the composite functional

$$I(\mathbf{n},\mathbf{s}) = \iint_\Omega \left[ \left(E - \mathbf{n}\cdot\mathbf{s}\right)^2 + \lambda(\mathbf{n}_x^2 + \mathbf{n}_y^2) + \mu(x,y)(\mathbf{n}^2 - 1) \right] \, dx\, dy$$

which is to be minimized with respect to $\mathbf{n}$ and $\mathbf{s}$. Here, $\lambda$ is a scalar that weights the relative importance of the regularization term, while $\mu(x,y)$ is a Lagrangian multiplier function used to impose the constraint that $\mathbf{n}(x,y)$ be a unit vector (see Horn & Brooks, 1985).

Minimizing $I$ is a problem in the calculus of variations. First, assume that $\mathbf{s}$ is known and that $I$ is to be minimized by a suitable choice of $\mathbf{n}$. Extrema of functionals are the solutions of the associated Euler equations (see Courant and Hilbert, 1953). The functional

$$\iint_\Omega F(x,y,\mathbf{n},\mathbf{n}_x,\mathbf{n}_y)\, dx\, dy$$

has the Euler equation

$$F_{\mathbf{n}} - \frac{\partial}{\partial x} F_{\mathbf{n}_x} - \frac{\partial}{\partial y} F_{\mathbf{n}_y} = \mathbf{0}.$$

So, by substitution, it follows that $I$ has the Euler equation

$$(E - \mathbf{n}\cdot\mathbf{s})\mathbf{s} + \lambda \nabla^2 \mathbf{n} - \mu \mathbf{n} = \mathbf{0},$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

Now, a discrete approximation to the Laplacian operator is given by

$$\left\{\nabla^2 \mathbf{n}\right\}_{ij} \approx \frac{4}{\epsilon^2}(\bar{\mathbf{n}}_{ij} - \mathbf{n}_{ij}),$$

in which $\epsilon$ is the distance between adjacent picture cells in the image, and $\bar{\mathbf{n}}_{ij}$ is the local average of normals

$$\bar{\mathbf{n}}_{ij} = \tfrac{1}{4}(\mathbf{n}_{i,j+1} + \mathbf{n}_{i,j-1} + \mathbf{n}_{i+1,j} + \mathbf{n}_{i-1,j}).$$

Hence we may translate the Euler equation into the discrete form

$$(E_{ij} - \mathbf{n}_{ij}\cdot\mathbf{s})\mathbf{s} + \frac{4\lambda}{\epsilon^2}(\bar{\mathbf{n}}_{ij} - \mathbf{n}_{ij}) - \mu_{ij}\,\mathbf{n}_{ij} = \mathbf{0}.$$

Rearranging this in order to isolate $\mathbf{n}_{ij}$ on one side yields the iterative scheme

$$\mathbf{n}_{ij}^{k+1} = \frac{1}{1 + \mu_{ij}(\epsilon^2/4\lambda)} \left( \bar{\mathbf{n}}_{ij}^k + \frac{\epsilon^2}{4\lambda}(E_{ij} - \mathbf{n}_{ij}^k \cdot \mathbf{s})\mathbf{s} \right),$$

that computes shape, given the light source direction. Other approximations for the Laplacian may lead to improved results, at the cost of increased computation. For example, if we use the more accurate 9-point approximation for the Laplacian, in which

$$\bar{\mathbf{n}}_{ij} = \frac{1}{20}\big[4(\mathbf{n}_{i,j+1} + \mathbf{n}_{i,j-1} + \mathbf{n}_{i+1,j} + \mathbf{n}_{i-1,j}) \\ + (\mathbf{n}_{i-1,j-1} + \mathbf{n}_{i-1,j+1} + \mathbf{n}_{i+1,j+1} + \mathbf{n}_{i+1,j-1})\big],$$

then twice as many array accesses are needed (and the constant multiplier $\epsilon^2/4\lambda$ becomes $3\epsilon^2/10\lambda$). The simple 5-point approximation was adequate for our purposes.

Note that we have yet to solve for $\mu$, the Lagrangian multiplier. This can be avoided, however, by observing that the division of the right hand side by $\big(1 + \mu(\epsilon^2/4\lambda)\big)$ does not change the direction of the vector being computed. Since $\mu$ is intended to ensure that the result is normalized, we can simply do this explicitly, as in

$$\begin{cases} \mathbf{m}_{ij}^{k+1} = \bar{\mathbf{n}}_{ij}^{k} + \dfrac{\epsilon^2}{4\lambda}(E_{ij} - \mathbf{n}_{ij}^{k} \cdot \mathbf{s})\mathbf{s} \\ \mathbf{n}_{ij}^{k+1} = \mathbf{m}_{ij}^{k+1}/|\mathbf{m}_{ij}^{k+1}|. \end{cases}$$

Now consider the problem of minimizing $I$ with respect to $\mathbf{s}$, given that $\mathbf{n}$ is known. This is a problem in conventional calculus. Computing the partial derivative of $I$ with respect to $\mathbf{s}$, we have

$$\frac{\partial I}{\partial \mathbf{s}} = -\iint_{\Omega} 2(E - \mathbf{n} \cdot \mathbf{s})\mathbf{n}\; dx\, dy = \mathbf{0},$$

and so

$$-\iint_{\Omega} E\,\mathbf{n}\; dx\, dy + \iint_{\Omega} (\mathbf{n} \cdot \mathbf{s})\mathbf{n}\; dx\, dy = \mathbf{0}.$$

Noting that

$$(\mathbf{n} \cdot \mathbf{s})\mathbf{n} = (\mathbf{n}^T\mathbf{s})\mathbf{n} = \mathbf{n}(\mathbf{n}^T\mathbf{s}) = (\mathbf{n}\mathbf{n}^T)\mathbf{s},$$

then by substitution we have

$$\iint_{\Omega} E\,\mathbf{n}\; dx\, dy = \left[\iint_{\Omega} \mathbf{n}\mathbf{n}^T\; dx\, dy\right]\mathbf{s},$$

where $(\mathbf{n}\mathbf{n}^T)$, and also the integral of $(\mathbf{n}\mathbf{n}^T)$, are $3 \times 3$ matrices. From this we finally obtain the desired equation

$$\mathbf{s} = \left[\iint_{\Omega} \mathbf{n}\mathbf{n}^T\; dx\, dy\right]^{-1} \iint_{\Omega} E\,\mathbf{n}\; dx\, dy.$$

Here $^{-1}$ denotes the inverse of a matrix. A discrete version of this formula, in which the integrals are replaced by sums, is easily obtained. An iterative scheme in both $\mathbf{n}$ and $\mathbf{s}$

is then within grasp:

$$\begin{cases} \mathbf{m}_{ij}^{k+1} = \bar{\mathbf{n}}_{ij}^{k} + \dfrac{\epsilon^2}{4\lambda}(E_{ij} - \mathbf{n}_{ij}^{k} \cdot \mathbf{s}^{k})\mathbf{s}^{k} \\ \mathbf{n}_{ij}^{k+1} = \mathbf{m}_{ij}^{k+1} / |\mathbf{m}_{ij}^{k+1}| \\ \mathbf{s}^{k+1} = \left[ \displaystyle\sum_{i,j \in \Omega} \mathbf{n}_{ij}^{k+1} \mathbf{n}_{ij}^{k+1\,T} \right]^{-1} \displaystyle\sum_{i,j \in \Omega} E_{ij}\, \mathbf{n}_{ij}^{k+1}. \end{cases}$$

This takes advantage of the fact that $\mathbf{s}$ need not be a unit vector. If the brightness of the source is known, we can normalize $\mathbf{s}$ so that it is a unit vector. Then the determination of $\mathbf{s}$ becomes slightly more complex, since it involves a constrained minimization.

## 3. Properties and performance of the scheme

The iterative scheme has two components: one concerned with the recovery of shape, the other concerned with the determination of the source direction. The shape-recovery component has an intuitively satisfying form. In essence, a new normal is computed by taking a local average, and adjusting this either toward or away from the source. The magnitude and sign of the adjustment is determined by the brightness error of the current estimate.

For a given shape, a new source direction is computed by a single pass through the image; unlike shape-recovery, no iteration is necessary. The $3 \times 3$ matrix $(\mathbf{n}\mathbf{n}^T)$ is summed across the image, as is the vector $(E\,\mathbf{n})$. The source direction can then be computed using Gaussian elimination or even Cramer's method (see Korn and Korn, 1968). The source-recovery component has been tested on a number of images and shapes. When the data are free of noise, the estimate of source direction is extremely accurate. Furthermore, estimates remain very good in the face of significant noise. For example, a synthetic image was generated of a sphere illuminated by a point source in the direction $(-4, 3, 8)^T$. The image was quantized to 255 irradiance levels, and the correct surface normal was given for each of the 1250 image points. Gaussian noise was added to the image giving an average perturbation in irradiance values of 34. Despite this, the source-finder computed an estimate of source direction that was only 2.7° in error. Further trials gave similar results.

The source-direction estimates are robust because the whole image is used. Theoretically, the problem is highly over-determined, as source direction is recoverable from brightness values of only three different surface orientations. Using the whole image ensures, however, that noise effects are significantly reduced.

The shape-and-source-from-shading problem for a point light source and Lambertian surface has a natural two-way ambiguity. If the image irradiance equation is satisfied over $\Omega$ by the shape $\mathbf{n}_1$ and the source direction $\mathbf{s}_1$, it will also be satisfied by the dual shape $\mathbf{n}_2$ and source direction $\mathbf{s}_2$ where

$$\mathbf{n}_2 = 2\hat{\mathbf{z}}(\mathbf{n}_1 \cdot \hat{\mathbf{z}}) - \mathbf{n}_1 \qquad \text{and} \qquad \mathbf{s}_2 = 2\hat{\mathbf{z}}(\mathbf{s}_1 \cdot \hat{\mathbf{z}}) - \mathbf{s}_1,$$

and $\hat{\mathbf{z}}$ is the viewing direction. Here, both source direction and surface normals are reflected about the viewing direction. This is easily verified by observing that

$$\begin{aligned}\mathbf{n_2} \cdot \mathbf{s_2} &= \left(2\hat{\mathbf{z}}(\mathbf{n_1} \cdot \hat{\mathbf{z}}) - \mathbf{n_1}\right) \cdot \left(2\hat{\mathbf{z}}(\mathbf{s_1} \cdot \hat{\mathbf{z}}) - \mathbf{s_1}\right) \\ &= 4(\hat{\mathbf{z}} \cdot \hat{\mathbf{z}})(\mathbf{n_1} \cdot \hat{\mathbf{z}})(\mathbf{s_1} \cdot \hat{\mathbf{z}}) - 2(\mathbf{n_1} \cdot \hat{\mathbf{z}})(\mathbf{s_1} \cdot \hat{\mathbf{z}}) - 2(\mathbf{s_1} \cdot \hat{\mathbf{z}})(\mathbf{n_1} \cdot \hat{\mathbf{z}}) + \mathbf{n_1} \cdot \mathbf{s_1} \\ &= \mathbf{n_1} \cdot \mathbf{s_1}.\end{aligned}$$

Given some initial values for the normals, the shape-and-source scheme will head for one or the other of these solutions. The dual shape and source direction can then be determined immediately using the equations given above.

We now present two examples of the program at work. Each (synthetic) image considered depicted a Lambertian surface illuminated by a point source in the direction $(3, 2, 9)^T$. The images each contained more than 1000 points at which normals were to be determined. Normals were assigned an initial value of $(0, 0, 1)^T$, as was the source direction. Occluding boundary normals were given. The equations for $\mathbf{n}_{ij}$ could be solved sequentially using the Gauss-Seidel algorithm. Since we are ultimately interested in parallel implementation on a grid, we used the Jacobi method instead (despite the fact that the Gauss-Seidel method has slightly better convergence properties).

The first image portrayed a hemisphere viewed from directly overhead. After 100 iterations with $\lambda = 0.005$, the average angular difference between estimated and correct normals was less than 3°. The maximum such deviation was less than 2.5 times the average value. The estimate of the light source direction, at this time, had errors in azimuth and zenith angles of 1.4° and 1.6° respectively.

The second image depicted a cylinder with rounded, hemispherical ends, viewed from a direction perpendicular to its axis. After 60 iterations, this time with $\lambda = 0.003$, the average angular error in surface normal was less than 5°. A further 30 iterations brought this value down to 4°. The maximum error remained somewhat larger, however, due to the scheme's tendency to smooth the intersection between the cylinder and the hemispheres. The errors in azimuth and zenith angles for the source were 7.3° and 1.1° respectively, achieved after 90 iterations. These, too, improved slowly with further processing.

The scheme was sometimes slow in converging. After rapid initial improvements, the rate of progress would decrease appreciably. However, one might expect the scheme to be slower than some of the current iterative methods, given the disadvantage of not knowing the light source direction. Convergence could be accelerated by employing multigrid techniques that propagate information across the image more quickly (see Terzopoulos, 1984). Interestingly, in the examples considered, a reasonable estimate for the source direction was obtained after only a few iterations. Subsequent processing just improved the estimate.

## 4. Iterative schemes for other reflectance maps

We now present two more new iterative schemes: the first extends the shape-and-source finder to cover situations in which a simple model of the sky is also included; the second uses methods developed above to find shape from shading, given a general reflectance map, but does not recover source direction.

## 4.1. Incorporating a sky component

The reflectance map

$$R_{sky}(\mathbf{n}) = \tfrac{1}{2}(1 + \mathbf{n} \cdot \hat{\mathbf{z}})$$

captures the situation in which a Lambertian surface is illuminated by a hemispherical source of uniform radiance (Brooks, 1978; Horn & Sjoberg, 1979). A point source may be added to the map to give

$$R_{ss}(\mathbf{n}) = \alpha(\mathbf{n} \cdot \mathbf{s}) + \tfrac{1-\alpha}{2}(1 + \mathbf{n} \cdot \hat{\mathbf{z}})$$

Here, $\alpha$ controls the relative intensity of the sun and the sky. We can now generate a method of shape-and-source recovery, under the assumption that the image was formed in accordance with the reflectance map $R_{ss}$.

We seek to minimize

$$\iint_\Omega \left[ \left( E - \alpha(\mathbf{n} \cdot \mathbf{s}) - \tfrac{1-\alpha}{2}(1 + \mathbf{n} \cdot \hat{\mathbf{z}}) \right)^2 + \lambda(\mathbf{n}_x^2 + \mathbf{n}_y^2) + \mu(x, y)(\mathbf{n}^2 - 1) \right] \, dx \, dy,$$

with respect to both $\mathbf{n}$ and $\mathbf{s}$. Fixing $\mathbf{s}$ for the time being, we are required to minimize the above functional with respect to $\mathbf{n}$ alone. The Euler equation for this problem is

$$\left( E - \alpha(\mathbf{n} \cdot \mathbf{s}) - \tfrac{1-\alpha}{2}(1 + \mathbf{n} \cdot \hat{\mathbf{z}}) \right) \left( \alpha \, \mathbf{s} + \tfrac{1-\alpha}{2} \hat{\mathbf{z}} \right) + \lambda \nabla^2 n - \mu \mathbf{n} = 0.$$

Treating $\mu$ as before, the following scheme is obtained:

$$\begin{cases} \mathbf{m}_{ij}^{k+1} = \bar{\mathbf{n}}_{ij}^k + \dfrac{\epsilon^2}{4\lambda} \left( E_{ij} - R_{ss}(\mathbf{n}_{ij}^k) \right) \left( \alpha \, \mathbf{s}^k + \tfrac{1-\alpha}{2} \hat{\mathbf{z}} \right) \\ \mathbf{n}_{ij}^{k+1} = \mathbf{m}_{ij}^{k+1} / |\mathbf{m}_{ij}^{k+1}|. \end{cases}$$

Here, the reflectance map, $R_{ss}$, has been substituted back into the equation to improve the presentation.

We now assume $\mathbf{n}$ to be fixed and minimize the functional with respect to $\mathbf{s}$. This we do, as before, by differentiating with respect to $\mathbf{s}$ and equating the result to zero. Thus we have

$$-2 \iint_\Omega \left( E - \alpha(\mathbf{n} \cdot \mathbf{s}) - \tfrac{1-\alpha}{2}(1 + \mathbf{n} \cdot \hat{\mathbf{z}}) \right) \alpha \, \mathbf{n} \; dx \, dy = 0.$$

Expanding,

$$\iint_\Omega \left( E - \tfrac{1-\alpha}{2}(1 + \mathbf{n} \cdot \hat{\mathbf{z}}) \right) \mathbf{n} \; dx \, dy = \iint_\Omega \alpha(\mathbf{n} \cdot \mathbf{s}) \mathbf{n} \; dx \, dy.$$

Noting as before that $(\mathbf{n} \cdot \mathbf{s})\mathbf{n} = (\mathbf{n}\mathbf{n}^T)\mathbf{s}$, the equation becomes

$$\iint_\Omega \left( E - \tfrac{1-\alpha}{2}(1 + \mathbf{n} \cdot \hat{\mathbf{z}}) \right) \mathbf{n} \; dx \, dy = \alpha \left[ \iint_\Omega \mathbf{n}\mathbf{n}^T \; dx \, dy \right] \mathbf{s}.$$

Thus we obtain the equation in $\mathbf{s}$ given by

$$\mathbf{s} = \frac{1}{\alpha} \left[ \iint_\Omega \mathbf{n}\mathbf{n}^T \; dx \, dy \right]^{-1} \iint_\Omega \left( E - \tfrac{1-\alpha}{2}(1 + \mathbf{n} \cdot \hat{\mathbf{z}}) \right) \mathbf{n} \; dx \, dy.$$

This we may write in discrete form and combine with the iterative scheme for $\mathbf{n}$ derived previously to give

$$\begin{cases} \mathbf{m}_{ij}^{k+1} = \bar{\mathbf{n}}_{ij}^{k} + \dfrac{\epsilon^2}{4\lambda}\big(E_{ij} - R_{ss}(\mathbf{n}_{ij}^{k})\big)\big(\alpha\,\mathbf{s}^{k} + \tfrac{1-\alpha}{2}\hat{\mathbf{z}}\big) \\ \mathbf{n}_{ij}^{k+1} = \mathbf{m}_{ij}^{k+1}/|\mathbf{m}_{ij}^{k+1}| \\ \mathbf{s}^{k+1} = \dfrac{1}{\alpha}\Big[\displaystyle\sum_{i,j\in\Omega} \mathbf{n}_{ij}^{k+1}\mathbf{n}_{ij}^{k+1\,T}\Big]^{-1} \displaystyle\sum_{i,j\in\Omega}\Big(E_{ij} - \tfrac{1-\alpha}{2}(1 + \mathbf{n}_{ij}^{k+1}\cdot\hat{\mathbf{z}})\Big)\mathbf{n}_{ij}^{k+1}. \end{cases}$$

Note that $\alpha$ is assumed to be known. Interestingly, the computation of $\mathbf{s}$ proceeds as before, except that the contribution of the sky is subtracted from $E$. This does not render the scheme trivial, however, as the calculation of shape does not follow suit.

## 4.2. Recovery of shape for the general reflectance map

Recall that the iterative scheme for shape and source direction is composed of two parts. Either component of the scheme may stand alone in the event that shape is required from source, or source is required from shape. Indeed, if used in this way, the shape-recovery component may be generalized to incorporate any reflectance map, $R(\mathbf{n})$. In minimizing the functional

$$\iint_\Omega \Big[\Big(E(x,y) - R\big(\mathbf{n}(x,y)\big)\Big)^2 + \lambda(\mathbf{n}_x^2 + \mathbf{n}_y^2) + \mu(x,y)(\mathbf{n}^2 - 1)\Big]\; dx\,dy,$$

we obtain the Euler equation

$$(E - R)R_{\mathbf{n}} + \lambda\nabla^2\mathbf{n} - \mu\mathbf{n} = \mathbf{0},$$

from which we derive the scheme

$$\begin{cases} \mathbf{m}_{ij}^{k+1} = \bar{\mathbf{n}}_{ij}^{k} + \dfrac{\epsilon^2}{4\lambda}\big(E_{ij} - R(\mathbf{n}_{ij})\big)R_{\mathbf{n}}(\mathbf{n}_{ij}) \\ \mathbf{n}_{ij}^{k+1} = \mathbf{m}_{ij}^{k+1}/|\mathbf{m}_{ij}^{k+1}|. \end{cases}$$

This is perhaps the most appealing of the current shape-from-shading schemes that deal with a general reflectance map. It is simply derived, and is expressed elegantly in terms of unit normals, rather than a two-parameter system such as the stereographic $fg$ space of Ikeuchi and Horn.

## 5. An alternative use of the unit normal constraint

Recall that in deriving the shape-and-source finder, we avoided solving for the Lagrangian multiplier $\mu(x,y)$. Instead, we normalized the estimate for $\mathbf{n}$ after each iteration. We now derive a scheme in which the multiplier is dealt with explicitly.

In seeking to minimize the functional

$$\iint_\Omega \Big[\Big(E(x,y) - R\big(\mathbf{n}(x,y)\big)\Big)^2 + \lambda(\mathbf{n}_x^2 + \mathbf{n}_y^2) + \mu(x,y)(\mathbf{n}^2 - 1)\Big]\; dx\,dy,$$

we obtained the Euler equation

$$(E - R)R_{\mathbf{n}} + \lambda \nabla^2 \mathbf{n} - \mu \mathbf{n} = \mathbf{0}.$$

Taking the dot product of this with $\mathbf{n}$ we find that

$$\mu = (E - R)(R_{\mathbf{n}} \cdot \mathbf{n}) + \lambda(\nabla^2 \mathbf{n} \cdot \mathbf{n}),$$

and so substituting for $\mu$ in the Euler equation, we get

$$(E - R)\big[R_{\mathbf{n}} - (R_{\mathbf{n}} \cdot \mathbf{n})\mathbf{n}\big] + \lambda\big[\nabla^2 \mathbf{n} - (\nabla^2 \mathbf{n} \cdot \mathbf{n})\mathbf{n}\big] = \mathbf{0}.$$

Noting that

$$(\mathbf{x} \cdot \mathbf{n})\mathbf{n} = (\mathbf{n}^T \mathbf{x})\mathbf{n} = \mathbf{n}(\mathbf{n}^T \mathbf{x}) = (\mathbf{n}\mathbf{n}^T)\mathbf{x},$$

for any vector $\mathbf{x}$ and letting $\mathbf{M}$ be the $3 \times 3$ matrix

$$\mathbf{M} = \mathbf{I} - \mathbf{n}\mathbf{n}^T,$$

the Euler equation reduces to

$$\mathbf{M}\left[(E - R)R_{\mathbf{n}} + \lambda \nabla^2 \mathbf{n}\right] = \mathbf{0}.$$

This is an equation in components orthogonal to $\mathbf{n}$, since

$$\mathbf{M}\,\mathbf{n} = (\mathbf{I} - \mathbf{n}\mathbf{n}^T)\mathbf{n} = \mathbf{n} - (\mathbf{n}\mathbf{n}^T)\mathbf{n} = \mathbf{n} - \mathbf{n}(\mathbf{n}^T\mathbf{n}) = \mathbf{n} - \mathbf{n}(\mathbf{n} \cdot \mathbf{n}) = \mathbf{0}.$$

The equation thus provides only two constraints on the solution vector $\mathbf{n}$. The remaining constraint is $\mathbf{n}^2 = 1$. Note that, because $\mathbf{M}$ is singular, we cannot simply eliminate $\mathbf{M}$ from the equation above by multiplying through by its inverse.

Using a standard finite difference approximation such as

$$\nabla^2 \mathbf{n} \approx \frac{4}{\epsilon^2}(\bar{\mathbf{n}} - \mathbf{n}),$$

where $\bar{\mathbf{n}}$ is a local average of $\mathbf{n}$ given earlier, we can write the Euler equation in the discrete form

$$\mathbf{M}\left[(E - R)R_{\mathbf{n}} + \frac{4\lambda}{\epsilon^2}(\bar{\mathbf{n}} - \mathbf{n})\right] = \mathbf{0}.$$

(For the time being, we omit subscripts.) We can then develop an iterative scheme in which the new value, $\mathbf{m}$, say, is used for the center term in the above approximation, while all other terms are computed using the old value of $\mathbf{n}$. This way we obtain

$$\mathbf{M}\,\mathbf{m} = \mathbf{M}\left[\bar{\mathbf{n}} + \frac{\epsilon^2}{4\lambda}(E - R)R_{\mathbf{n}}\right].$$

Now let $\mathbf{m} = \mathbf{p} + \nu\mathbf{n}$, where $\mathbf{p} \perp \mathbf{n}$. Then $\mathbf{m}^2 = \mathbf{p}^2 + \nu^2$ and $\mathbf{M}\,\mathbf{m} = \mathbf{M}\,\mathbf{p} = \mathbf{p}$, since $\mathbf{M}\,\mathbf{n} = \mathbf{0}$, and so we get

$$\mathbf{p} = \mathbf{M}\left[\bar{\mathbf{n}} + \frac{\epsilon^2}{4\lambda}(E - R)R_{\mathbf{n}}\right].$$

The component $\nu\,\mathbf{n}$, parallel to the old normal vector $\mathbf{n}$, is computed using

$$\nu = \sqrt{1 - \mathbf{p}^2}.$$

Note that there are theoretically two solutions for $\nu$, one positive and one negative. The positive value leads to a new estimate close to the previous one, while the negative value gives rise to one almost opposite to the old one. It is clear that one should use the positive root.

Thus we finally have the scheme

$$\begin{cases} \mathbf{p}_{ij}^{k+1} = \left(\mathbf{I} - \mathbf{n}_{ij}^k \mathbf{n}_{ij}^{k\,T}\right) \left(\bar{\mathbf{n}}_{ij}^k + \dfrac{\epsilon^2}{4\lambda}\left(E_{ij} - R(\mathbf{n}_{ij}^k)\right) R_{\mathbf{n}}(\mathbf{n}_{ij}^k)\right) \\ \mathbf{n}_{ij}^{k+1} = \mathbf{p}_{ij}^{k+1} + \mathbf{n}_{ij}^k \sqrt{1 - (\mathbf{p}_{ij}^{k+1})^2}. \end{cases}$$

This may be coupled with the source-recovery component given earlier, when $R(\mathbf{n}) = \mathbf{n}\cdot\mathbf{s}$.

As we approach a solution with this scheme, $\mathbf{p}$ will be small, since $\bar{\mathbf{n}} \approx \mathbf{n}$ and $E \approx R$. However, in the early stages of iteration, it may be necessary to place an artificial limit on the amount of adjustment made away from the old normal. That is, one may have to limit the magnitude of $\mathbf{p}$ so that problems do not arise in computing $\nu$ using the equation above.

(The above method for solving the underdetermined equation $\mathbf{M}\mathbf{m} = \mathbf{0}$, under the constraint $\mathbf{m}^2 = 1$, can be arrived at most easily using the pseudoinverse of the matrix $\mathbf{M}$. We avoided this approach in the exposition here since the solution can also be found directly).

## 6. Summary

Most current methods for obtaining shape from shading assume complete knowledge of the reflectance map. Here, we considered the situation in which a Lambertian surface is illuminated by a point light source from an unknown direction. Thus we dealt with a parameterized rather than a fixed, reflectance map. The local approach to recovering shape from shading, which is also intended to deal with unknown source direction, was found to have several drawbacks, notably its restrictive assumption that surfaces are locally umbilical.

The adoption of unit normal vectors for describing surface orientation was important for the development of our method. This led to simple derivations and elegant presentations. The problem was cast as one of minimizing a positive-definite functional containing a brightness error term, a regularizing term, and a Lagrangian multiplier to enforce the condition that the normal be of unit length. The Euler equation for this calculs of variation problem was shown to be a second-order partial differential equation in the unknown surface-normal function. A convergent iterative scheme solved it in the discrete domain.

The direction of the light source can be determined in closed form if the surface shape is known. During any iteration a source-direction estimate can be obtained using the current estimate of the surface shape. The iterations for obtaining increasingly accurate estimates of the surface shape can be interlaced with estimation of the light-source direction.

We implemented and tested this method for recovering shape and source direction. We also discussed a two-way ambiguity that can appear in the solution. Further, we showed how to extend the shape-from-shading component of the iterative scheme to more general reflectance maps.

**Acknowledgements**

We thank Alan Yuille for many useful discussions. Comments on a draft by Steve Bagley, Eric Grimson, Tomás Lozano-Pérez and, especially, Demetri Terzopoulos were much appreciated. Mike Brooks wishes to express gratitude to the MIT AI laboratory for the opportunity to spend some months there, and to Flinders University for granting leave.

**References**

Brooks, M.J., "Investigating the Effects of Planar Light Sources," CSM 22, Dept. Computer Science, Essex University, 1978.

Bruss, A.R., "Is What You See What You Get?", *Proc. International Joint Conference on Artificial Intelligence*, Karlsruhe, August, 1983, pp. 1053–1056.

Courant, R. & Hilbert, D., *Methods of Mathematical Physics, Vol. 1*, Interscience Publishers, Inc., New York, 1953.

Horn, B.K.P., "Shape-from-Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from one View," MAC-TR-79 and also AI-TR-232, M.I.T., November 1970.

Horn, B.K.P., "Obtaining Shape from Shading Information," in *The Psychology of Computer Vision*, P.H.Winston (Ed.), McGraw-Hill, 1975.

Horn, B.K.P. & Sjoberg, R.W., "Calculating the Reflectance Map," *Applied Optics*, Vol. 18, No. 11, June 1979, pp. 1770–1779.

Horn, B.K.P. & Brooks, M.J., "A Variational Approach to Shape from Shading," A.I.M. 813, Artificial Intelligence Lab., M.I.T., January, 1985.

Ikeuchi, K. & Horn, B.K.P., "Numerical Shape from Shading and Occluding Boundaries," *Artificial Intelligence*, Vol. 17, 1981, pp. 141–185.

Korn, G.A. & Korn, T.M., *Mathematical Handbook for Scientists and Engineers*, 2nd edition, McGraw-Hill, 1968.

Pentland, A.P., "Local Shape Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, March, 1984, pp. 170–187.

Poggio, T. & Torre, V., "Ill-posed Problems and Regularization Analysis in Early Vision," A.I.M. 773, Artificial Intelligence Lab., M.I.T., 1984.

Strat, T.M., "A Numerical Method for Shape from Shading from a Single Image," M.S. thesis, Department of E.E.C.S., M.I.T., 1979.

Terzopoulos, D., "Efficient Multiresolution Algorithms for Computing Lightness, Shape from Shading, and Optical Flow," *Proc. International Joint Conference on Artificial Intelligence*, Austin, Texas, August, 1984, pp. 314–317.