# Revisiting Pentland's estimator of light source direction

**Wojciech Chojnacki and Michael J. Brooks**

*Department of Computer Science, University of Adelaide, SA 5005, Australia*

**Danny Gibbins**

*Discipline of Computer Science, Flinders University of South Australia, GPO Box 2100, Adelaide, SA 5001, Australia*

Received February 3, 1993; revised manuscript received July 22, 1993; accepted July 22, 1993

We examine the pioneering method of Pentland [J. Opt. Soc. Am. **72**, 448 (1982)] for automatically estimating the direction of the "sun" (light source) from a single image. It is shown that, under the assumptions used in the derivation of the method, the estimate of source direction is erroneous. Specifically, it is shown that an image-based expression used in calculating source direction diverges to infinity as the density of image points is increased and that the formula involving this expression is therefore incorrect. When the method is implemented, the flaw manifests itself in the undesirable dependence of the estimator on image resolution. Supporting experimental evidence is given for this. An alternative source-direction estimator that is free of these drawbacks is proposed.

## 1. INTRODUCTION

Much work has been done to develop automated techniques for estimating the direction of the light source, or "sun," from a single image of a scene (see Refs. 1–7). That work has been motivated by the desire to permit various existing shape-from-shading algorithms (see Ref. 8) to operate with less prerequisite information. Currently, most shape-from-shading techniques need to be given prior information relating to light source configuration and surface reflectance properties. A good preliminary estimator of source direction would reduce this dependence.

Estimating light source direction given only a single image is a highly ill-posed problem. As a result, techniques for solving this problem need to make strong assumptions about both the shape and the reflectance properties of the object depicted. For example, many estimators assume that the image depicts an object that has spherical shape and Lambertian reflectance. Inevitably, the performance of an estimator will depend in good measure on the extent to which the underlying assumptions are satisfied by the object in view. Not always will the assumptions well approximate the given circumstances. An implication here is that there can be no truly versatile or universal estimator of source direction.

The first and most influential attempt to develop a technique for automatic recovery of light source direction was made by Pentland.[3] In Pentland's estimator, light source direction is couched in terms of the directional derivatives of image irradiance. In this paper we show that Pentland's method contains a flaw. Specifically, we show that an image-based expression used in calculating source direction diverges to infinity as the density of image points is increased and that the formula involving this expression is therefore incorrect. When the method is implemented, the flaw manifests itself in the dependence of the corresponding source-direction estimator on the density of image points, or image resolution. The method gives inaccurate results even in the presence of data that are perfectly consistent with the assumptions under which the method is derived.

In attempting to remedy Pentland's method we have developed an estimator that we term the disk method. This method retains the image-based expressions of Pentland but combines them in such a way that the divergence problem is avoided and the quality of the estimate is improved. Indeed, perfect results are obtained for ideal data that satisfy the underlying assumptions. Experimental results are presented to support this claim.

To our knowledge, no earlier work has commented on the correctness of the mathematical derivation of Pentland. When Pentland's derivation has been described previously, it has been reproduced uncritically; when criticism of Pentland's method has previously arisen, it has concerned the nature of the assumptions adopted or deficiencies in the performance of the method (see Refs. 1, 2, 4, and 7). In particular, to our knowledge, criticism has never before extended to an observance of the dependency of the method on image resolution.

## 2. PENTLAND'S METHOD

### A. Technical Background

We start by introducing some notation. For any function $f$ on a subset $\Omega$ of the $x$–$y$ plane, let $\mathbb{E}^{\Omega}\{f\}$ be the expected value of $f$ given by

$$\mathbb{E}^{\Omega}\{f\} = \frac{1}{|\Omega|}\int_{\Omega} f(x, y)\mathrm{d}x\mathrm{d}y,$$

where $|\Omega|$ denotes the area of $\Omega$, and let $\mathrm{Var}^{\Omega}\{f\}$ be the variance of $f$ defined by

$$\mathrm{Var}^{\Omega}\{f\} = [\mathbb{E}^{\Omega}\{f^2\} - (\mathbb{E}^{\Omega}\{f\})^2]^{1/2}.$$

Given a vector $\mathbf{s} = (s_1, s_2)$ and $(x, y)$ in $\Omega$, let $f_s(x, y)$ denote

0740-3232/94/010118-07$06.00 

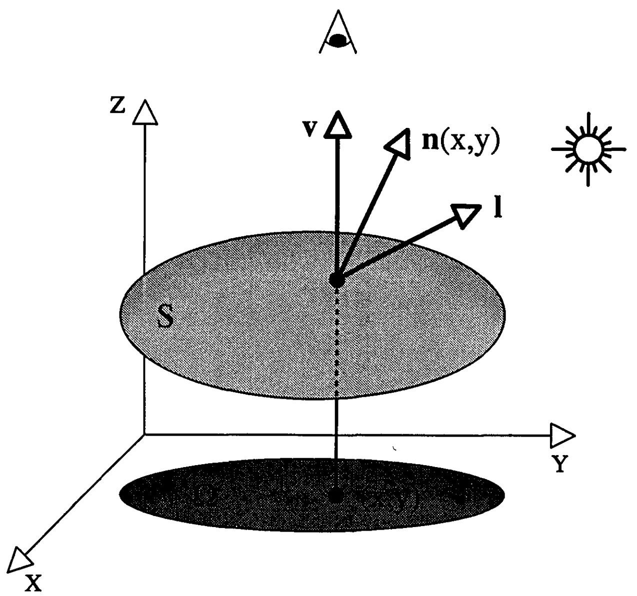


Fig. 1. Image formation.

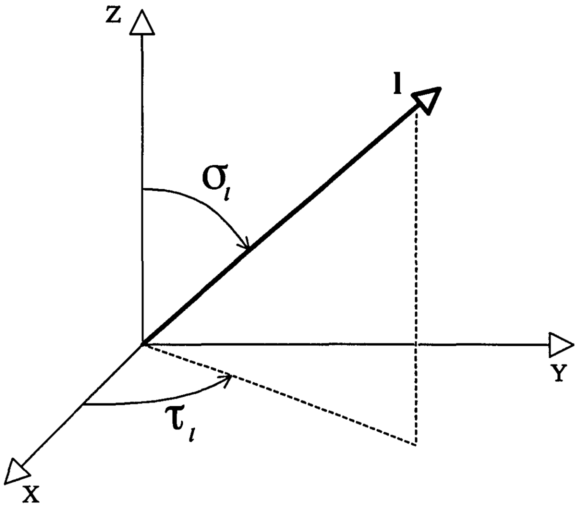


Fig. 2. Slant $\sigma_l$ and tilt $\tau_l$.

the partial derivative of $f$ at $(x, y)$ in direction $\mathbf{s}$ defined by

$$f_s(x, y) = s_1 f_x(x, y) + s_2 f_y(x, y), \tag{1}$$

where $f_x(x, y)$ and $f_y(x, y)$ denote the partial derivatives of $f$ at $(x, y)$ in directions $(1, 0)$ and $(0, 1)$, respectively.

We now consider the geometry and photometry of image formation and formulate the problem of estimating source direction. Suppose that a Lambertian surface $S$ with constant albedo $\mu$ is illuminated by an infinitely distant sun from the direction determined by the unit vector $\mathbf{l} = (l_1, l_2, l_3)$, as depicted in Fig. 1. Define the slant $\sigma_l$ and tilt $\tau_l$ of the source vector $\mathbf{l}$ by the formula

$$\mathbf{l} = (\sin \sigma_l \cos \tau_l, \sin \sigma_l \sin \tau_l, \cos \sigma_l), \tag{2}$$

with $0 \leq \sigma_l \leq \pi$ and $0 \leq \tau_l < 2\pi$ (see Fig. 2). Assume that $S$ is viewed from the direction determined by the vector $\mathbf{v} = (0, 0, 1)$. Let $\mathbf{n}(x, y)$ be the unit normal to $S$ such that the scalar product of $\mathbf{n}(x, y)$ and the viewing vector $\mathbf{v}$ is nonnegative at the point whose (orthographic) projection along the $z$ axis onto the $x$–$y$ plane coincides with $(x, y)$. Then the irradiance $E(x, y)$ of the image of $S$ on a plane parallel to the $x$–$y$ plane is given by

$$E(x, y) = \mu \mathbf{n}(x, y) \cdot \mathbf{l}. \tag{3}$$

This image irradiance equation holds over the image domain $\Omega$ defined as that portion of the image pattern where $E(x, y) > 0$; that is, $\Omega$ is the part of the image exhibiting nonzero brightness values, any unilluminated area being outside the image domain. The light source estimation problem may now be expressed simply as the need to determine $\mathbf{l}$, or equivalently $\sigma_l$ and $\tau_l$, given $E$.

Under the assumptions that $S$ is a hemisphere, $\mathbf{l}$ is such that $l_3 \geq 0$, and $\mathbf{s}$ is a unit vector (satisfying $s_1^2 + s_2^2 = 1$), Pentland proposed formulas for slant and tilt of the source direction that, in an integral form (as opposed to the original differential one), read as follows:

$$\sigma_l = \arccos\left[1 - \frac{(\mathbb{E}^{\Omega}\{E_x\})^2 + (\mathbb{E}^{\Omega}\{E_y\})^2}{(\mathrm{Var}^{\Omega}\{E_s\})^2}\right]^{1/2}, \tag{4}$$

$$\tau_l = \arctan \frac{\mathbb{E}^{\Omega}\{E_y\}}{\mathbb{E}^{\Omega}\{E_x\}}. \tag{5}$$

Here the slant formula is tacitly assumed to be independent of any particular choice of $\mathbf{s}$. Note that the tilt formula is arbitrarily aligned to the directions implied by the $x$ and $y$ axes.

In the original equations of Pentland, the integral mean values given above are replaced by least-squares estimates of finite-sum averages (over an unspecified number of elements that may depend on image resolution) of finite-difference approximations to directional derivatives of image irradiance. One obtains these estimates by sampling over various directions $\mathbf{s}$ and taking Eq. (1) for the regression relation. The resulting source-direction estimator provides a practical recipe for implementations and as such has been used in our implementation of Pentland's method. It should be noted, however, that recourse to the method of least squares as a preferred technique for statistical inference is somewhat arbitrary; for example, one might equally well consider employing Bayes's method. In contrast, Eqs. (4) and (5) describe the relationships between the quantities involved in a statistics-free form. As a result, these equations facilitate model idealization and have the advantage of stopping short of any implementation commitments.

If both Eq. (4) and Eq. (5) were correct, then one might reasonably expect exact estimates of source direction given images of spheres. This begs the question of the expected accuracy of the estimates in the face of images of nonspherical surfaces. In this situation one might expect reasonably accurate responses provided that the distributions of irradiance derivatives of images of nonspherical surfaces are sufficiently close to the distributions of irradiance derivatives of images of spheres. In fact, these expectations are not met. As we show below, even when applied to an ideal image of a sphere, the formula for slant in Eq. (4) is incorrect. In contrast, tilt formula (5) is mathematically sound, as the subsequent discussion verifies.

### B. Invalidity of the Slant Formula

We proceed to establish the invalidity of Eq. (4). Let $S$ be the hemisphere given by the graph of the function

$$z(x, y) = (R^2 - x^2 - y^2)^{1/2},$$

where $R$ is a positive number and $(x, y)$ runs over the closed disk D:

$$D_R = \{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 \leq R^2\}$$

in the $x$–$y$ plane centered at the origin with radius $R$. Assume that $S$ is viewed from the direction determined by the vector $\mathbf{v} = (0, 0, 1)$. Then the corresponding unit normal $\mathbf{n}$ at a point $(x, y, z)$ in $S$ takes the form $[x/R, y/R, (R^2 - x^2 - y^2)^{1/2}/R]$. Suppose that the source vector $\mathbf{l}$ is such that $l_3 \neq 0$. Now, in accordance with Eq. (3), the image of the hemisphere is given by

$$E(x, y) = \frac{\mu}{R}[l_1 x + l_2 y + l_3(R^2 - x^2 - y^2)^{1/2}] \tag{6}$$

over the domain

$$\Omega = \{(x, y) \in \mathbb{R}^2 \colon l_1 x + l_2 y + l_3(R^2 - x^2 - y^2)^{1/2} > 0,\ x^2 + y^2 \leq R^2\}.$$

We shall establish the invalidity of Eq. (4) by proving that, for each unit vector $\mathbf{s}$,

$$\mathrm{Var}^{\Omega}\{E_s\} = +\infty. \tag{7}$$

As the expected values $\mathbb{E}^{\Omega}\{E_x\}$ and $\mathbb{E}^{\Omega}\{E_y\}$ are finite, the last equality implies that if Eq. (4) were to hold, then the only admissible value of $\sigma_1$ would be 0.

Clearly, the finiteness of $\mathbb{E}^{\Omega}\{E_x\}$ and $\mathbb{E}^{\Omega}\{E_y\}$ implies that $\mathbb{E}^{\Omega}\{E_s\}$ is finite for each unit vector $\mathbf{s}$. Thus Eq. (7) will follow once we show that

$$\mathbb{E}^{\Omega}\{E_s^2\} = +\infty \tag{8}$$

for any unit vector $\mathbf{s}$. To prove this identity, note that, by rotating simultaneously the base of the hemisphere and the light source, we may assume (without loss of generality) that $\mathbf{s} = (1, 0)$. In view of Eq. (6), for each $(x, y)$ in the interior of $\Omega$,

$$E_x(x, y) = \frac{\mu}{R}\left[l_1 - \frac{l_3 x}{(R^2 - x^2 - y^2)^{1/2}}\right], \tag{9}$$

and hence

$$E_x^2(x, y) = \left(\frac{\mu}{R}\right)^2\left[l_1^2 - \frac{2l_1 l_3 x}{(R^2 - x^2 - y^2)^{1/2}} + \frac{l_3^2 x^2}{R^2 - x^2 - y^2}\right].$$

Taking into account that

$$\int_{D_R} \frac{|x|}{(R^2 - x^2 - y^2)^{1/2}} \mathrm{d}x\mathrm{d}y \leq \int_{D_R} \frac{R}{(R^2 - x^2 - y^2)^{1/2}} \mathrm{d}x\mathrm{d}y = 2\pi R \int_0^R \frac{r}{(R^2 - r^2)^{1/2}} \mathrm{d}r = 2\pi R^2$$

and the fact that $l_3 \neq 0$, we see that Eq. (8) will follow after we show that

$$\int_{\Omega} \frac{x^2}{R^2 - x^2 - y^2} \mathrm{d}x\mathrm{d}y = +\infty. \tag{10}$$

Let $U$ be an open semidisk of the form

$$\{(r \cos \theta, r \sin \theta) \colon 0 < r < R, \theta_0 < \theta < \theta_0 + \pi\},$$

for some $\theta_0$, contained in $\Omega$. (If the closure of $\Omega$ is a proper subset of $D_R$, then $U$ is uniquely determined by $\Omega$; otherwise, we are free to choose any open semidisk in $D_R$ with radius $R$.) Then, passing to polar coordinates, we have

$$\int_{\Omega} \frac{x^2}{R^2 - x^2 - y^2} \mathrm{d}x\mathrm{d}y \geq \int_U \frac{x^2}{R^2 - x^2 - y^2} \mathrm{d}x\mathrm{d}y = \int_{\theta_0}^{\theta_0 + \pi} \int_0^R \frac{r^3 \cos^2 \theta}{R^2 - r^2} \mathrm{d}r\mathrm{d}\theta = \frac{\pi}{2} \int_0^R \frac{r^3}{R^2 - r^2} \mathrm{d}r,$$

which together with the identities

$$\int_0^R \frac{r^3}{R^2 - r^2} \mathrm{d}r = \frac{1}{2} \int_0^{R^2} \frac{u}{R^2 - u} \mathrm{d}u = \frac{1}{2} \int_0^{R^2} \frac{R^2 - v}{v} \mathrm{d}v = \frac{R^2}{2}\left(\int_0^{R^2} \frac{1}{v} \mathrm{d}v - 1\right) = +\infty,$$

obtained by making the consecutive substitutions $r = \sqrt{u}$ and $u = R^2 - v$, implies Eq. (10). Thus Eq. (7) is established, and Pentland's formula for slant is shown to be invalid.

### C. Derivation of the Tilt Formula

Suppose again that $S$ is the Lambertian hemisphere with constant albedo, given by the graph of the function

$$z(x, y) = (R^2 - x^2 - y^2)^{1/2},$$

where $R$ is a positive number and $(x, y)$ runs over the closed disk in the $x$–$y$ plane centered at the origin with radius $R$. If the source vector $\mathbf{l}$ is such that $l_1^2 + l_2^2 > 0$, then one can define a map $\varphi = (\varphi_1, \varphi_2)$ by setting

$$\varphi_1(x, y) = \frac{l_1^2 - l_2^2}{l_1^2 + l_2^2} x + \frac{2l_1 l_2}{l_1^2 + l_2^2} y,$$

$$\varphi_2(x, y) = \frac{2l_1 l_2}{l_1^2 + l_2^2} x + \frac{l_2^2 - l_1^2}{l_1^2 + l_2^2} y.$$

Geometrically, $\varphi$ is the axial symmetry of the $x$–$y$ plane with respect to the line passing through the origin and the point $(l_1, l_2)$. Clearly,

$$[\varphi_1(x, y)]^2 + [\varphi_2(x, y)]^2 = x^2 + y^2, \tag{11}$$

$$l_1 \varphi_1(x, y) + l_2 \varphi_2(x, y) = l_1 x + l_2 y, \tag{12}$$

so $\varphi$ maps the image domain $\Omega$ into itself. Since the composition $\varphi \circ \varphi$ is the identity, it follows that $\varphi$ is one to one and maps $\Omega$ onto itself. Evidently, the Jacobian $J_\varphi$ of $\varphi$ is equal to $-1$. Thus, in view of Eq. (11),

$$\begin{aligned}\int_\Omega \frac{\varphi_1(x,y)}{(R^2-x^2-y^2)^{1/2}}\mathrm{d}x\mathrm{d}y &= \int_\Omega \frac{\varphi_1(x,y)}{[R^2-\varphi_1^2(x,y)-\varphi_2^2(x,y)]^{1/2}}\mathrm{d}x\mathrm{d}y \\ &= \int_\Omega \frac{\varphi_1(x,y)}{[R^2-\varphi_1^2(x,y)-\varphi_2^2(x,y)]^{1/2}}|J_\varphi|\mathrm{d}x\mathrm{d}y \\ &= \int_\Omega \frac{x}{(R^2-x^2-y^2)^{1/2}}\mathrm{d}x\mathrm{d}y, \end{aligned} \tag{13}$$

and, similarly,

$$\int_\Omega \frac{\varphi_2(x,y)}{(R^2-x^2-y^2)^{1/2}}\mathrm{d}x\mathrm{d}y = \int_\Omega \frac{y}{(R^2-x^2-y^2)^{1/2}}\mathrm{d}x\mathrm{d}y. \tag{14}$$

Combining Eqs. (13) and (14) with

$$l_2[\varphi_1(x,y)+x] = l_1[\varphi_2(x,y)+y],$$

we see that

$$\begin{aligned} l_2\int_\Omega \frac{x}{(R^2-x^2-y^2)^{1/2}}\mathrm{d}x\mathrm{d}y &= \frac{l_2}{2}\int_\Omega \frac{\varphi_1(x,y)+x}{(R^2-x^2-y^2)^{1/2}}\mathrm{d}x\mathrm{d}y \\ &= \frac{l_1}{2}\int_\Omega \frac{\varphi_2(x,y)+y}{(R^2-x^2-y^2)^{1/2}}\mathrm{d}x\mathrm{d}y \\ &= l_1\int_\Omega \frac{y}{(R^2-x^2-y^2)^{1/2}}\mathrm{d}x\mathrm{d}y. \end{aligned}$$

Hence, by Eq. (9) and by

$$E_y(x,y) = \frac{\mu}{R}\left[l_2 - \frac{l_3 y}{(R^2-x^2-y^2)^{1/2}}\right], \tag{15}$$

we have

$$\begin{aligned} l_2\mathbb{E}^\Omega\{E_x\} &= \frac{\mu l_1 l_2}{R} - \frac{l_2 l_3}{|\Omega|}\int_\Omega \frac{x}{(R^2-x^2-y^2)^{1/2}}\mathrm{d}x\mathrm{d}y \\ &= \frac{\mu l_1 l_2}{R} - \frac{l_1 l_3}{|\Omega|}\int_\Omega \frac{y}{(R^2-x^2-y^2)^{1/2}}\mathrm{d}x\mathrm{d}y \\ &= l_1\mathbb{E}^\Omega\{E_y\}. \end{aligned} \tag{16}$$

This together with Eq. (2) establishes Eq. (5) provided that $l_1 \neq 0$ and $\mathbb{E}^\Omega\{E_x\} \neq 0$.

We now show that

- $l_1 = 0$ if and only if $\mathbb{E}^\Omega\{E_x\} = 0$;
- if $\mathbb{E}^\Omega\{E_x\} = 0$, then $\mathbb{E}^\Omega\{E_y\} \neq 0$;
- if $\mathbb{E}^\Omega\{E_x\} = 0$ and $\mathbb{E}^\Omega\{E_y\} > 0$, then $\tau_l = \pi/2$;
- if $\mathbb{E}^\Omega\{E_x\} = 0$ and $\mathbb{E}^\Omega\{E_y\} < 0$, then $\tau_l = 3\pi/2$.

We start by establishing the following inequality:

$$l_1\mathbb{E}^\Omega\{E_x\} + l_2\mathbb{E}^\Omega\{E_y\} > 0. \tag{17}$$

To prove it, note that the boundary $\partial\Omega$ of $\Omega$ is the union of the two sets $\Gamma_1$ and $\Gamma_2$, where $\Gamma_1$ is the semicircle on the circumference of the disk $D_R$ defined by

$$\Gamma_1 = \left\{\left\{\frac{R(l_1\sin\theta + l_2\cos\theta)}{(l_1^2+l_2^2)^{1/2}}, \frac{R(l_2\sin\theta - l_1\cos\theta)}{(l_1^2+l_2^2)^{1/2}}\right\}: 0<\theta<\pi\right\},$$

and $\Gamma_2$ is the set given by

$$\Gamma_2 = \{(x,y)\in\mathbb{R}^2: l_1x + l_2y + l_3(R^2-x^2-y^2)^{1/2} = 0,\ x^2+y^2 \le R^2\},$$

all of whose interior points lie in the open disk centered at the origin with radius $R$. Observe that $E(x,y) > 0$ for each $(x,y)$ in $\Gamma_1$ and that $E(x,y) = 0$ for each $(x,y)$ in $\Gamma_2$. By Green's formula,

$$\begin{aligned} |\Omega|(l_1\mathbb{E}^\Omega\{E_x\} + l_2\mathbb{E}^\Omega\{E_y\}) &= \int_\Omega [l_1E_x(x,y) + l_2E_y(x,y)]\mathrm{d}x\mathrm{d}y \\ &= l_1\int_{\partial\Omega} E(x,y)\mathrm{d}y - l_2\int_{\partial\Omega} E(x,y)\mathrm{d}x, \end{aligned} \tag{18}$$

where $\partial\Omega$ is oriented counterclockwise. For $0<\theta<\pi$, let

$$\tilde{E}(\theta) = E\left\{\frac{R(l_1\sin\theta + l_2\cos\theta)}{(l_1^2+l_2^2)^{1/2}}, \frac{R(l_2\sin\theta - l_1\cos\theta)}{(l_1^2+l_2^2)^{1/2}}\right\}.$$

Then

$$l_1\int_{\Gamma_1} E(x,y)\mathrm{d}y - l_2\int_{\Gamma_1} E(x,y)\mathrm{d}x = R(l_1^2+l_2^2)^{1/2}\int_0^\pi \tilde{E}(\theta)\sin\theta\mathrm{d}\theta,$$

and, since $\tilde{E}(\theta) > 0$ and $\sin\theta > 0$ for $0<\theta<\pi$, it follows that

$$l_1\int_{\Gamma_1} E(x,y)\mathrm{d}y - l_2\int_{\Gamma_1} E(x,y)\mathrm{d}x > 0. \tag{19}$$

Since $E = 0$ on $\Gamma_2$, we see that

$$l_1\int_{\Gamma_2} E(x,y)\mathrm{d}y - l_2\int_{\Gamma_2} E(x,y)\mathrm{d}x = 0. \tag{20}$$

Now, comparing relations (18), (19), and (20), we obtain relation (17).

Proceeding to establish the validity of the four statements above, note that if $\mathbb{E}^\Omega\{E_x\} = 0$, then, by relation (17), $\mathbb{E}^\Omega\{E_y\} \neq 0$ and, by Eq. (16), $l_1 = 0$. Conversely, if $l_1 = 0$, then $l_2 \neq 0$ and, by Eq. (16), $\mathbb{E}^\Omega\{E_x\} = 0$. Thus $l_1 = 0$ if and only if $\mathbb{E}^\Omega\{E_x\} = 0$. Now, when $\mathbb{E}^\Omega\{E_x\} = 0$, then, by relation (17), $l_2\mathbb{E}^\Omega\{E_y\} > 0$ and hence $\mathbb{E}^\Omega\{E_y\} \neq 0$; moreover, if $\mathbb{E}^\Omega\{E_y\} > 0$, then $l_2 > 0$, and if $\mathbb{E}^\Omega\{E_y\} < 0$, then $l_2 < 0$. Therefore, in the case that $\mathbb{E}^\Omega\{E_x\} = 0$, if $\mathbb{E}^\Omega\{E_y\} > 0$, then $\tau_l = \pi/2$, and if $\mathbb{E}^\Omega\{E_y\} < 0$, then $\tau_l = 3\pi/2$.

Note that if $l_1 = l_2 = 0$, then the above derivation is inapplicable. This, however, is not a concern, since it may then be seen that the light source is overhead, and, by definition, tilt is undefined.

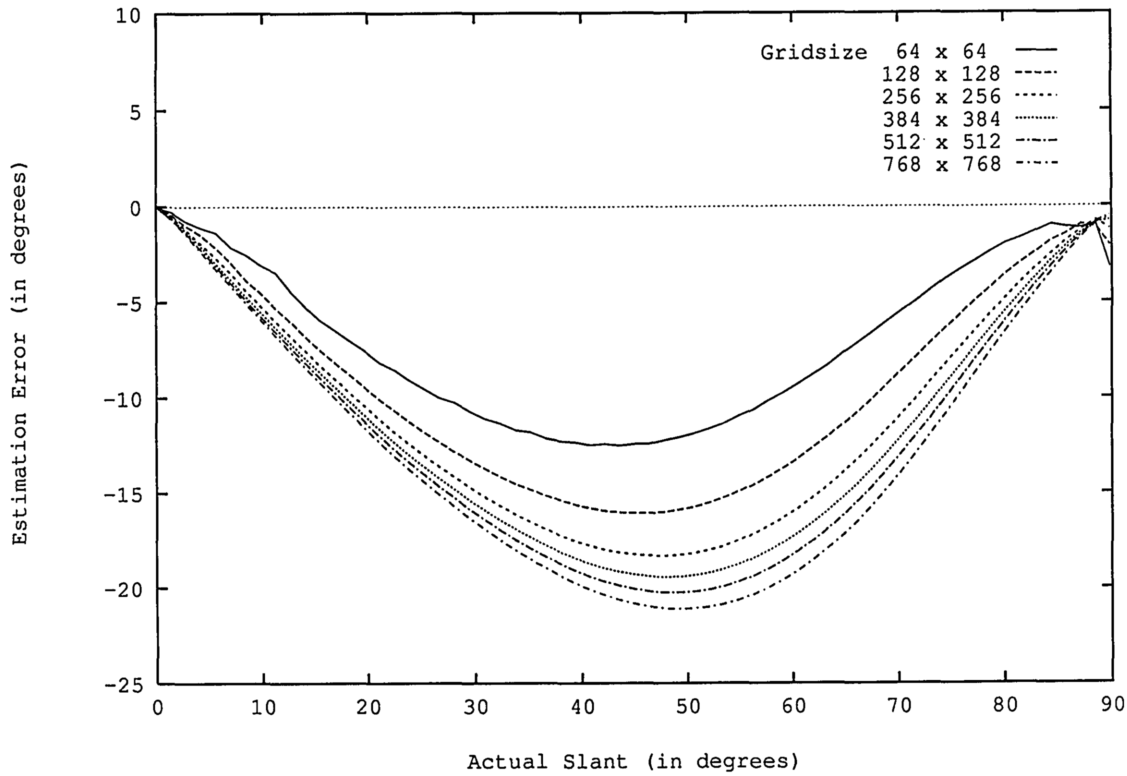


Fig. 3. Slant error against actual slant for Pentland's method.

### D. Experimental Results

In practice, if Pentland's technique is applied to an image of a sphere in such a way that image points are always confined within the boundary, then expected values for the square of image irradiance derivatives, obtained by means of finite differencing, will always be bounded. It might therefore seem that the divergence in the integral slant formula of Pentland has no practical impact. However, as the resolution of the image is increased (with a more-dense grid of image points), the expected value for the square of image irradiance derivatives, evaluated with use of finite differences, will increase without bound. This acute sensitivity of Pentland's slant formula to image resolution is clearly a highly undesirable property.

Figure 3 displays the change in performance of Pentland's slant estimator as a function of image resolution. For a range of image resolutions, we plot error in the estimated slant versus actual slant. The images used are those of a sphere illuminated from various directions. Each graph is derived from 65 slant estimates, where the actual slant ranges over 0 to 90 deg. Performance of the estimator is worse in the midrange, where it may be seen clearly that increased image resolution results in a poorer slant estimate. Of course, it is evident from Fig. 3 that the Pentland method is far from exact when it is confronted with ideal data.

## 3. DISK METHOD

### A. Presentation

To alleviate the shortcomings of Pentland's method, we propose an estimator, which we term the disk method, that adjusts the domain of integration in the formulas for slant and tilt so that points close to the edge of the image domain are no longer involved. This adjustment ensures that expected values of squared image derivatives remain finite.

Once again, assume that the image depicts the Lambertian hemisphere with constant albedo, with shape given by the graph of the function

$$z(x, y) = (R^2 - x^2 - y^2)^{1/2}$$

for some positive $R$. Assume, moreover, that the source vector $\mathbf{l}$ is such that $l_3 \geq 0$. Let $0 < \alpha < 1$ be such that the closed disk $D_{\alpha R}$, centered at the origin with radius $\alpha R$, is contained in the image domain. As above, let $\mathbf{s} = (s_1, s_2)$ be a unit vector. Then, as we prove below, the tilt and slant of source direction are given by

$$\tau_l = \arctan \frac{\mathbb{E}^{D_{\alpha R}}\{E_y\}}{\mathbb{E}^{D_{\alpha R}}\{E_x\}}, \tag{21}$$

$$\sigma_l = \arccos\left\{1 + \frac{\theta(\alpha)[(\mathbb{E}^{D_{\alpha R}}\{E_x\})^2 + (\mathbb{E}^{D_{\alpha R}}\{E_y\})^2]}{(\mathrm{Var}^{D_{\alpha R}}\{E_s\})^2}\right\}^{-1/2}, \tag{22}$$

where

$$\theta(\alpha) = -\frac{1}{2} - \frac{1}{2\alpha^2}\ln(1 - \alpha^2) = \frac{\alpha^2}{4} + \frac{\alpha^4}{6} + \frac{\alpha^6}{8} + \ldots. \tag{23}$$

As we can see, Pentland's formula for tilt also applies when the domain of integration is confined to a disk that is smaller than the entire image domain. A similarly confined domain is also used in the slant estimate; however, Pentland's formula for slant this time undergoes significant change.

### B. Derivation

We now derive the disk method's tilt and slant formulas (21) and (22). Retaining the notation from Subsection 3.A, observe that Eqs. (9) and (15), and the identities

$$\int_{D_{\alpha R}} \frac{x}{(R^2 - x^2 - y^2)^{1/2}} \mathrm{d}x\mathrm{d}y = \int_{D_{\alpha R}} \frac{y}{(R^2 - x^2 - y^2)^{1/2}} \mathrm{d}x\mathrm{d}y = 0$$

imply that

$$\mathbb{E}^{D_{\alpha R}}\{E_x\} = \frac{\mu}{R} l_1,$$

$$\mathbb{E}^{D_{\alpha R}}\{E_y\} = \frac{\mu}{R} l_2. \tag{24}$$

Hence

$$\frac{l_2}{l_1} = \frac{\mathbb{E}^{D_{\alpha R}}\{E_y\}}{\mathbb{E}^{D_{\alpha R}}\{E_x\}},$$

which, in view of Eq. (2), yields Eq. (21).

To establish the disk method's slant formula (22), note that, by Eqs. (24), we have

$$\mathbb{E}^{D_{\alpha R}}\{E_s\} = \frac{\mu}{R}(s_1 l_1 + s_2 l_2). \tag{25}$$

Since

$$E_s^2(x, y) = \frac{\mu^2}{R^2}\left[(s_1 l_1 + s_2 l_2)^2 - \frac{2 l_3 (s_1 l_1 + s_2 l_2)(s_1 x + s_2 y)}{(R^2 - x^2 - y^2)^{1/2}} + \frac{l_3^2 (s_1^2 x^2 + 2 s_1 s_2 x y + s_2^2 y^2)}{R^2 - x^2 - y^2}\right],$$

$$\int_{D_{\alpha R}} \frac{x}{(R^2 - x^2 - y^2)^{1/2}} \mathrm{d}x\mathrm{d}y = \int_{D_{\alpha R}} \frac{y}{(R^2 - x^2 - y^2)^{1/2}} \mathrm{d}x\mathrm{d}y = \int_{D_{\alpha R}} \frac{xy}{R^2 - x^2 - y^2} \mathrm{d}x\mathrm{d}y = 0,$$

$$\int_{D_{\alpha R}} \frac{x^2}{R^2 - x^2 - y^2} \mathrm{d}x\mathrm{d}y = \int_{D_{\alpha R}} \frac{y^2}{R^2 - x^2 - y^2} \mathrm{d}x\mathrm{d}y = \pi \int_0^{\alpha R} \frac{r^3}{R^2 - r^2} \mathrm{d}r,$$

it follows that

$$\mathbb{E}^{D_{\alpha R}}\{E_s^2\} = \frac{\mu^2}{R^2}\left[(s_1 l_1 + s_2 l_2)^2 + \frac{l_3^2}{\alpha^2 R^2} \int_0^{\alpha R} \frac{r^3}{R^2 - r^2} \mathrm{d}r\right].$$

By making the consecutive substitutions $r = Ru$, $u = \sqrt{v}$, and $v = 1 - w$, we see that

$$\frac{1}{\alpha^2 R^2} \int_0^{\alpha R} \frac{r^3}{R^2 - r^2} \mathrm{d}r = \frac{1}{\alpha^2} \int_0^{\alpha} \frac{u^3}{1 - u^2} \mathrm{d}u = \frac{1}{2\alpha^2} \int_0^{\alpha^2} \frac{v}{1 - v} \mathrm{d}v = \frac{1}{2\alpha^2} \int_{1-\alpha^2}^{1} \frac{1 - w}{w} \mathrm{d}w = \theta(\alpha),$$

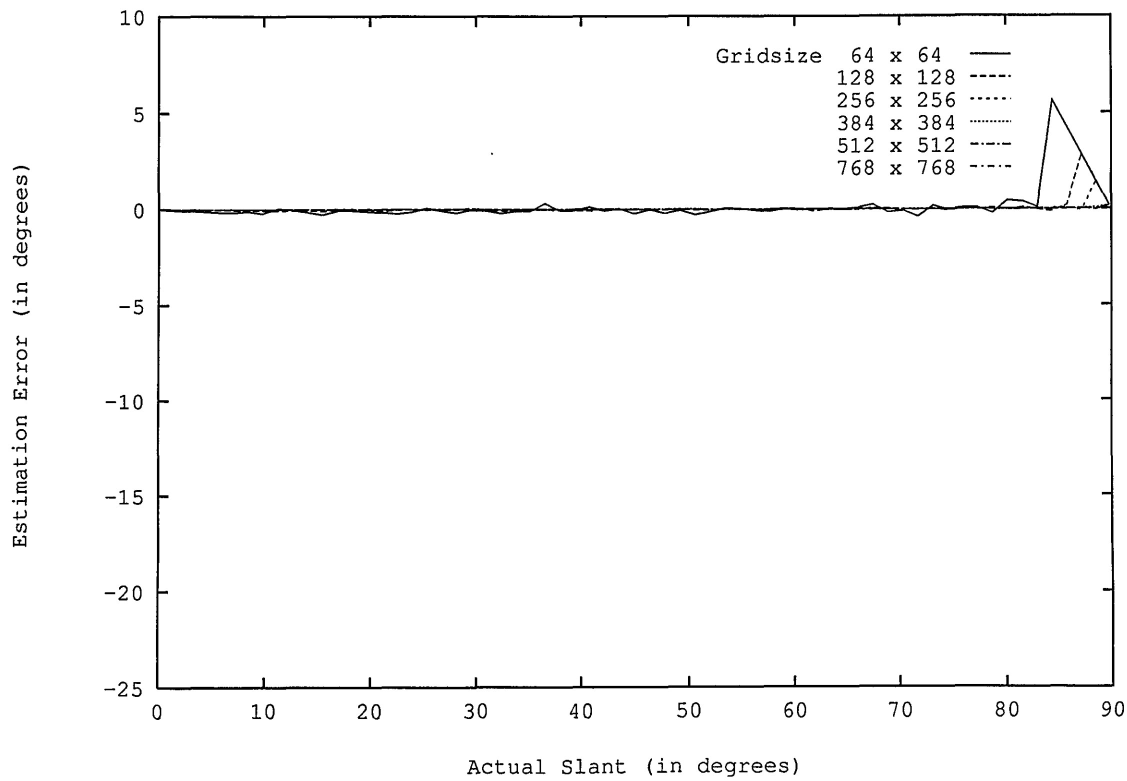


Fig. 4. Slant error against actual slant for the disk method.

where $\theta(\alpha)$ is given by Eq. (23). Hence

$$\mathbb{E}^{D_{\alpha R}}\{E_s^2\} = \frac{\mu^2}{R^2}\left[(s_1 l_1 + s_2 l_2)^2 + \theta(\alpha) l_3^2\right],$$

and further, by Eq. (25),

$$(\mathrm{Var}^{D_{\alpha R}}\{E_s\})^2 = \mathbb{E}^{D_{\alpha R}}\{E_s^2\} - (\mathbb{E}^{D_{\alpha R}}\{E_s\})^2 = \frac{\mu^2}{R^2}\theta(\alpha) l_3^2.$$

This jointly with Eqs. (24) yields

$$\begin{aligned}\frac{\mu^2}{R^2} &= \frac{\mu^2}{R^2}(l_1^2 + l_2^2 + l_3^2)\\ &= (\mathbb{E}^{D_{\alpha R}}\{E_x\})^2 + (\mathbb{E}^{D_{\alpha R}}\{E_y\})^2 + \frac{(\mathrm{Var}^{D_{\alpha R}}\{E_s\})^2}{\theta(\alpha)}\end{aligned}$$

whence, if we remember that $l_3 \geq 0$,

$$l_3 = \left\{1 + \frac{\theta(\alpha)[(\mathbb{E}^{D_{\alpha R}}\{E_x\})^2 + (\mathbb{E}^{D_{\alpha R}}\{E_y\})^2]}{(\mathrm{Var}^{D_{\alpha R}}\{E_s\})^2}\right\}^{-1/2}.$$

The latter equality together with Eq. (2) implies Eq. (22).

### C. Experimental Results

Figure 4 shows the disk method's slant estimator at work. To permit comparison, we have scaled the graphs of Figs. 3 and 4 similarly. We confirm experimentally that, for an image of a sphere, estimated slant errors are very small and are not significantly affected by increased resolution. It should be stressed that we developed the disk method simply to highlight and overcome the deficiencies of the Pentland method. We do not claim that the disk method is especially suitable for widespread utility. An analysis of the performance of the disk method may be found in Ref. 9.

## 4. CONCLUSION

We have revisited the method of Pentland for estimating the direction of the sun from a single image and have shown that the original derivation of the method was flawed. One consequence of this flaw is that any implementation of the method will lead to slant estimates that are dependent on image resolution: as the density of brightness points in an image increases, the slant estimate will worsen.

Various factors will affect performance of an estimator. For example, if an estimator is derived under the assumption that the depicted object has a particular shape, adverse consequences will result when the assumption is violated; the presence of any noise in the image will have a similarly deleterious effect. These factors will inevitably bias the outcome of any method. In contrast, however, the additional factor revealed in the current paper, whereby slant estimation is dependent on image resolution, is one that is avoidable. Whereas no estimator could be designed that is completely immune to image noise or is unaffected by the choice of the underlying surface model, an effective estimator should not be unduly influenced by change in image resolution.

We have tracked an error in Pentland's reasoning as a result of our endeavor to put Pentland's method within a sound mathematical framework. In the course of a formal rederivation of the method, we have reestablished the validity of Pentland's tilt estimator and have devised an amended method for slant estimation, in the spirit of the original approach. Unlike Pentland's slant estimator, this revised estimator gives perfect results in the presence of ideal data. An extensive performance analysis of the disk method can be found in Ref. 9.

## ACKNOWLEDGMENTS

W. Chojnacki and M. J. Brooks acknowledge the support of the Australian Research Council. Two anonymous referees gave valuable suggestions for improving the paper.

## REFERENCES

1. N. E. Hurt, "Mathematical methods in shape-from-shading: a review of recent results," Acta Appl. Math. **23,** 163–188 (1991).
2. C. H. Lee and A. Rosenfeld, "Improved methods of estimating shape from shading using the light source coordinate system," Artif. Intell. **26,** 125–143 (1985). See also revised version in Ref. 8 below, pp. 323–347.
3. A. P. Pentland, "Finding the illuminant direction," J. Opt. Soc. Am. **72,** 448–455 (1982).
4. A. P. Pentland, "Linear shape from shading," Int. J. Comput. Vis. **4,** 153–162 (1990).
5. D. Weinshall, "The shape of shading and the direction of illumination from shading on occluding contours," A. I. Memo 1264, Artificial Intelligence Laboratory, and C.B.I.P. Memo 60, Center For Biological Information Processing (Massachusetts Institute of Technology, Cambridge, Mass., 1990). A short version appears in *Proceedings of the 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society Press, Champaign, Ill., 1992), pp. 716–718.
6. Y. Yang and A. Yuille, "Sources from shading," in *Proceedings of the 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society Press, Maui, Hawaii, 1991), pp. 534–539.
7. Q. Zheng and R. Chellappa, "Estimation of illumination direction, albedo, and shape from shading," IEEE Trans. Pattern Anal. Mach. Intell. **13,** 680–702 (1991).
8. B. K. P. Horn and M. J. Brooks, eds., *Shape from Shading* (MIT Press, Cambridge, Mass., 1989).
9. D. Gibbins, M. J. Brooks, and W. Chojnacki, "Light source direction from a single image: a performance analysis," Aust. Comput. J. **23,** 165–174 (1991).