International Journal of Computer Vision 20(3), 243–261 (1996)

Exploiting the Generic Viewpoint Assumption

WILLIAM T. FREEMAN

Received July 20, 1993; Revised January 18, 1995

Abstract. The "generic viewpoint" assumption states that an observer is not in a special position relative to the scene. It is commonly used to disqualify scene interpretations that assume special viewpoints, following a binary decision that the viewpoint was either generic or accidental. In this paper, we apply Bayesian statistics to *quantify* the probability of a view, and so derive a useful tool to estimate scene parameters.

Generic variables can include viewpoint, object orientation, and lighting position. By considering the image as a (differentiable) function of these variables, we derive the probability that a set of scene parameters created a given image. This *scene probability equation* has three terms: the *fidelity* of the scene interpretation to the image data; the *prior probability* of the scene interpretation; and a new *genericity* term, which favors scenes likely to produce the observed image. The genericity term favors image interpretations for which the image is stable with respect to changes in the generic variables. It results from integration over the generic variables, using a low-noise approximation common in Bayesian statistics.

This approach may increase the scope and accuracy of scene estimates. It applies to a range of vision problems. We show shape from shading examples, where we rank shapes or reflectance functions in cases where these are otherwise unknown. The rankings agree with the perceived values.

1. Introduction

A major task of visual perception is to find the scene which best explains visual observations. Bayesian statistics are a powerful tool for this (Witkin, 1981; Geman and Geman, 1984; Szeliski, 1989; Bulthoff, 1991; Kersten, 1991; Jepson and Richards, 1992; Heeger and Simoncelli, 1992; Knill et al., 1996; Belhumeur, 1996). Assumptions are expressed in terms of *prior probabilities*. Using a model for how a scene relates to the observation, one forms the *posterior probability* for the scene, given the observed visual data. After choosing a criterion of optimality, one can calculate a best interpretation. Other computational techniques, such as regularization (Tikhonov and Arsenin, 1977; Poggio et al., 1985; Terzopoulos, 1986) and minimum description length analysis (Darrell et al., 1990; Pentland, 1990b), can be posed in a Bayesian framework (Szeliski, 1989; Leclerc, 1989). In this paper, we show how the commonly encountered conditions of "generic viewpoint" influence the posterior probabilities to give additional information about the scene.

The generic view assumption (Binford, 1981; Biederman, 1985; Lowe and Binford, 1985; Malik, 1987; Richards et al., 1987; Nakayama and Shimojo, 1992; Albert and Hoffman, 1995) postulates that the scene is not viewed from a special position. Figure 1 shows an example. The square in (a) could be an image of a wire-frame cube (b) viewed from a position where the line segments of the front face hid those behind them. However, that would require an unlikely viewpoint, and given the image in (a), one should infer a square, not a cube. The generic view assumption has been invoked to explain perceptions involving stereo and transparency (Nakayama and Shimojo, 1992), linear shape from shading (Pentland, 1990c), object parts and illusory contours (Albert and Hoffman, 1995), and feature or object identification (Koenderink and van Doorn, 1979; Lowe and Binford, 1985; Biederman, 1985; Richards et al., 1987; Malik, 1987; Jepson and Richards, 1992; Dickinson et al., 1992). Often, researchers assume a

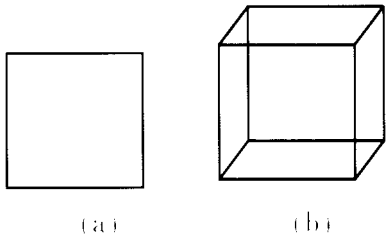

Figure 1. An example of use of the generic view assumption for binary decisions. The image (a) could be of a square, or it could be an "accidental view" of the cube in (b). Since a cube would require a special viewing position to be seen as the image in (a), we reject that possible interpretation for (a).

view is either generic, and therefore admissible, or accidental, and therefore rejected. Some have pointed out that it should be possible to quantify the degree of accidentalness or have done so in special cases (Lowe and Binford, 1985; Malik, 1987; Leclerc and Bobick, 1991; Nakayama and Shimojo, 1992; Jepson and Richards, 1992; Dickinson et al., 1992).

In this paper we quantify generic view probabilities in a general case; we find the probability of a given scene hypothesis under the assumption of generic viewpoint. We do not restrict ourselves to viewpoint; the generic variable can be, for example, object orientation or lighting position. Scene parameters can be reflectance function, shape, and velocity. We show that the generic view assumption can strongly influence the scene interpretation.

The key to quantifying the generic view probabilities is to find how the visual data would change were the generic variables (e.g., the viewpoint) to change. We will show that large image changes correspond to unlikely scenes. Our approach employs an established approximation in Bayesian statistics (approximating the log likelihood function as a gaussian (Laplace, 1812; Fisher, 1959; Jeffreys, 1961; Johnson, 1970; Lindley, 1972; Box and Tiao, 1973; Berger, 1985; Gull, 1989; Skilling, 1989; MacKay, 1992)), allowing convenient marginalization over the generic variables. Szeliski (1989) used related ideas to set regularization parameters by maximum likelihood estimation. See Weinshall et al. (1994) for a related non-Bayesian approach. Marginalization over the generic variables can also be interpreted using the loss functions of Bayesian decision theory (Berger, 1985), discussed in Section 4.2 and in Freeman and Brainard (1995), Freeman (1996), Yuille and Bulthoff (1996).

Our Bayesian framework also takes into account the fidelity of a rendered scene to the image data and the prior probability of the scene. The conditional probability we will derive gives a new objective function for a vision algorithm to optimize. Including the generic view probabilities may lead to more powerful vision algorithms, or better models of human perception.

We show applications to the shape from shading problem, yielding new results. Using a two-parameter family of reflectance functions, we show how to find the probability of a reflectance function from a single image. We find shape and lighting direction estimates under conditions where many estimates would account for the image data equally well. We show how a scene hypothesis which accounts less well for image data can be more likely. This method also applies to other vision problems, such as motion (Freeman, 1994) and stereo (Yuille and Bulthoff, 1996).

We motivate our approach in the remainder of the introduction. In Section 2 we derive the *scene probability equation*, the conditional probability for a scene interpretation given the observed visual data. Then we show the applications to shape from shading. Shorter reports of this work include (Freeman, 1993, 1994).

1.1. Example

Different shapes and reflectance functions can explain a given image. Figure 2 shows an example. The image (a) may look like a cylinder (c) painted with a Lambertian reflectance function (b) (shown on a hemisphere). However, it could also have been created by the flatter shape of (f), painted with a shiny reflectance function (e). If both interpretations account for the data, how can we choose between them? We should use whatever information we have about their prior probabilities, but we may not know those well (Nakayama and Shimojo, 1992).

We can distinguish between the two scene hypotheses if we imagine rotating them. The Lambertian shaded image would change little for small rotations, Fig. 2(d), while the shiny image would change considerably, (g). Thus, for the Lambertian solution, for a large range of object poses we would see the image of (a). For the shiny solution, we would see that image over a smaller range of poses.

This also holds if we reverse the roles of the shiney and Lambertian objects, as shown in Figs. 3. The image data, Fig. 3(a), may look like a shiny cylinder, but, again, it can be explained by either a Lambertian reflectance function, shape (c) painted with the reflectance function shown in (b), or a shiny one, the

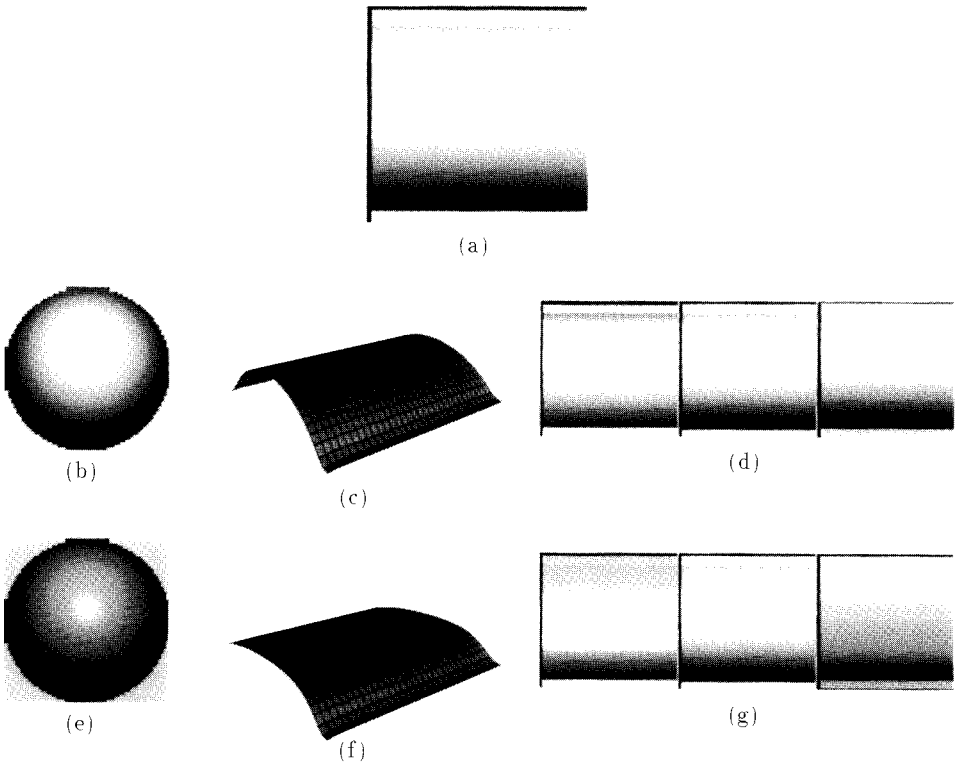

Figure 2. The image (a) appears to be a cylinder (c) painted with a Lambertian reflectance function (b) (shown on a hemisphere). However the flatter shape of (f) and a shiny reflectance function (e) also explain the data equally well. We can distinguish between the competing accounts for (a) by imagining rotating each shape. Images of each shape at three nearby orientations are shown in (d) and (g). We see that the image made assuming a Lambertian reflectance function (b) is more stable than that made assuming a shiny reflectance function (e). The reflectance function of (b) provides more angles over which the image looks nearly the same. If all viewpoints are equally likely, and the shapes and reflectances of (b)-(c) and (e)-(f) are equally likely to occur in the world, then (b)-(c) is a more probable interpretation than (e)-(f).

shape (f) painted with (e). Note that the shape for the Lambertian function is taller than that of the shiny reflectance function. When we rotate both shapes, in (d) and (g), it is the Lambertian image, (d), which changes more than the shiny one (g), because of the parallax induced as the tall shape moves back and forth.

Thus in each case the shape and reflectance functions which correspond to our perception of the image create a more stable image under imagined rotations of the rendered scene. A small image derivative with respect to object orientation means that the image will look nearly the same over a relatively large range of object poses. If all object orientations are equally probable, then the probability of an object is proportional to the range of angles over which it looks nearly the same as the image data. We will use a measurement noise model to specify what it means for two images to "look nearly the same". In our analysis, the image derivatives will arise from expanding the image in a Taylor series in the generic view variable.

2. The Scene Probability Equation

In this section we derive the probability densities for scene parameters given observed data. Let $\mathbf{y}$ be a vector of observations (boldface symbols will indicate vectors). This can be image intensities, or measures derived from them, such as spatial or temporal derivatives or normal velocities. For simplicity, we will call this "the image".

Let the vector β be the scene parameters we want to estimate. This vector can describe, for example, the object shape and reflectance function or the image velocities.

Let $\mathbf{x}$ be an M dimensional vector of the generic variables. These are the variables over which we will marginalize. For the example of Section 1.1 this was the object pose angle. Generic variables can be, for example, viewpoint position, object orientation, or lighting position. The probability density of $\mathbf{x}$, $P_{\mathbf{x}}(\mathbf{x})$, will typically be uniform, $P_{\mathbf{x}}(\mathbf{x}) = k$, but it need not be.

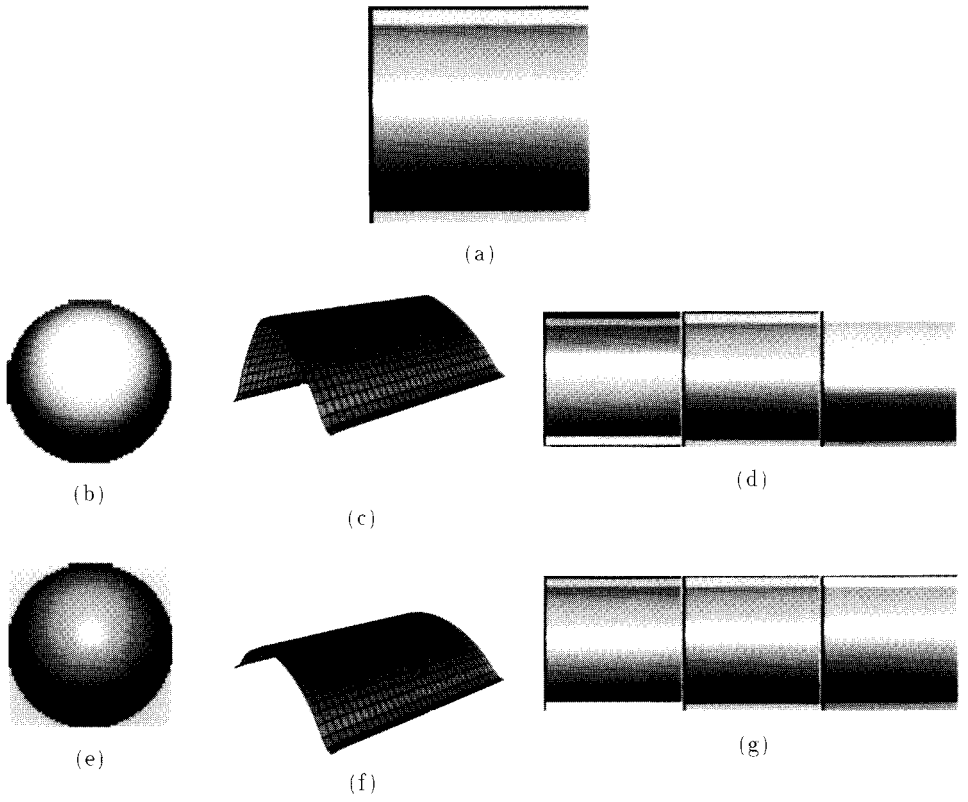

Figure 3. The image, (a), can be accounted for in two different ways. The shape (c) and the Lambertian reflectance function shown in (b) will create the image (a), as will the shape (f) and a shiny reflectance function (e). We can distinguish between the shiny and Lambertian explanations for (a) if we imagine rotating each shape. The greyscale images show each shape at three different orientations. The image made from the shiny reflectance function, (e), changes only a little, while the parallax caused by the rotation of the tall shape of the Lambertian solution causes a larger image change. The reflectance function of (e) provides more angles over which the image looks nearly the same. If all viewpoints are equally likely, and the shapes and reflectances of (b)-(c) and (e)-(f) are equally likely to occur in the world, then (e)-(f) is more probable than (b)-(c). In Section 2 we make this precise.

(The notation $P_a(a)$ denotes the probability density function on the variable a as a function of a. For brevity, we omit the subscript for conditional probability functions.)

The scene parameters β and generic variables $\mathbf{x}$ determine the ideal observation (image), $\tilde{\mathbf{y}}$, through the "rendering function", $\mathbf{f}$:

$$\tilde{\mathbf{y}} = \mathbf{f}(\mathbf{x}, \beta) \tag{1}$$

For the cylinders of Section 1.1 the rendering function was the computer graphics calculation which gave the image as a function of surface shape, β, and incident light angle, $\mathbf{x}$.

We postulate some measurement noise, although we will often examine the limit where its variance goes to zero. The observation, $\mathbf{y}$, is the rendered ideal image $\tilde{\mathbf{y}}$ plus the measurement noise, $\mathbf{n}$:

$$\mathbf{y} = \tilde{\mathbf{y}} + \mathbf{n}. \tag{2}$$

Let $P_{\mathbf{n}}(\mathbf{n})$ be the probability density function of the noise. We will assume that the measurement noise is a set of Gaussian random variables with mean zero and standard deviation σ. Here we assume the noise is identically distributed, but we extend the results to non-identical distributions in Section 4.1. Thus

$$P_{\mathbf{n}}(\mathbf{n}) = \frac{1}{(\sqrt{2\pi\sigma^2})^N} \exp \frac{-\|\mathbf{n}\|^2}{2\sigma^2}, \tag{3}$$

where N is the dimension of the observation and noise vectors and $\|\mathbf{n}\|^2 = \mathbf{n} \cdot \mathbf{n}$. For $\lim \sigma \to 0$, the noise term allows us to examine the local behavior of

the rendering function. For finite σ, it allows us to handle noisy or uncalibrated images.

The posterior distribution, $P(\beta, \mathbf{x} \mid \mathbf{y})$, gives the probability that scene parameter β (e.g., shape) and generic variable $\mathbf{x}$ (e.g., light direction) created the visual data $\mathbf{y}$ (the image). From $P(\beta, \mathbf{x} \mid \mathbf{y})$, we will find the marginal probability $P(\beta \mid \mathbf{y})$.

We use Bayes' theorem to evaluate $P(\beta, \mathbf{x} \mid \mathbf{y})$:

$$P(\beta, \mathbf{x} \mid \mathbf{y}) = \frac{P(\mathbf{y} \mid \beta, \mathbf{x}) P_\beta(\beta) P_\mathbf{x}(\mathbf{x})}{P_\mathbf{y}(\mathbf{y})}, \qquad (4)$$

where we have assumed that $\mathbf{x}$ and β are independent. The denominator is constant for all models β to be compared.

To find $P(\beta, \mathbf{x} \mid \mathbf{y})$, independent of the value of the generic variable $\mathbf{x}$, we integrate the joint probability of Eq. (4) over the possible $\mathbf{x}$ values:

$$P(\beta \mid \mathbf{y}) = \frac{P_\beta(\beta)}{P_\mathbf{y}(\mathbf{y})} \int P(\mathbf{y} \mid \beta, \mathbf{x})\, P_\mathbf{x}(\mathbf{x})\, d\mathbf{x}. \qquad (5)$$

$P(\mathbf{y} \mid \beta, \mathbf{x})$ is large where the scene β and the value $\mathbf{x}$ give an image similiar to the observation $\mathbf{y}$. Equation (5) integrates the area of $\mathbf{x}$ for which β roughly yields the observation $\mathbf{y}$.

For our noise model,

$$P(\mathbf{y} \mid \beta, \mathbf{x}) = \frac{1}{(\sqrt{2\pi\sigma^2})^N} e^{\frac{-\|\mathbf{y}-\mathbf{f}(\mathbf{x},\beta)\|^2}{2\sigma^2}}. \qquad (6)$$

For the low noise limit, we can find an analytic approximation to the integral of Eq. (6) in Eq. (5). We expand $\mathbf{f}(\mathbf{x}, \beta)$ in Eq. (6) in a second order Taylor series,

$$\mathbf{f}(\mathbf{x}, \beta) \approx \mathbf{f}(\mathbf{x}_0, \beta) + \sum_i \mathbf{f}'_i\, [\mathbf{x} - \mathbf{x}_0]_i + \frac{1}{2} \sum_{i,j} [\mathbf{x} - \mathbf{x}_0]_i\, \mathbf{f}''_{ij}\, [\mathbf{x} - \mathbf{x}_0]_j, \qquad (7)$$

where $[\cdot]_i$ indicates the ith component of the vector in brackets, and

$$\mathbf{f}'_i = \left. \frac{\partial \mathbf{f}(\mathbf{x}, \beta)}{\partial x_i} \right|_{\mathbf{x}=\mathbf{x}_0}, \qquad (8)$$

and

$$\mathbf{f}''_{ij} = \left. \frac{\partial^2 \mathbf{f}(\mathbf{x}, \beta)}{\partial x_i \partial x_j} \right|_{\mathbf{x}=\mathbf{x}_0}. \qquad (9)$$

We take $\mathbf{x}_0$ to be the value of $\mathbf{x}$ which can best account for the image data for a given β; i.e., the $\mathbf{x}$ for which $\|\mathbf{y} - \mathbf{f}(\mathbf{x}, \beta)\|^2$ is minimized. The derivatives $\mathbf{f}'_i$ and $\mathbf{f}''_{ij}$ must exist, so this approximation does not apply to non-differentiable image representations.

Using Eqs. (6)–(9) to second order in $\mathbf{x} - \mathbf{x}_0$ in the integral of Eq. (5) yields the posterior probability for the scene parameters β given the visual data $\mathbf{y}$ (see Appendix A):

$$\begin{aligned} & P(\beta \mid \mathbf{y}) \\ & = k \exp\left(\frac{-\|\mathbf{y} - \mathbf{f}(\mathbf{x}_0, \beta)\|^2}{2\sigma^2}\right) \quad [P_\beta(\beta) P_\mathbf{x}(\mathbf{x_0})] \quad \frac{1}{\sqrt{\det(\boldsymbol{A})}} \\ & = k \quad \text{(fidelity)} \quad \text{(prior probability)} \quad \text{(genericity)}, \end{aligned} \qquad (10)$$

where the i and jth elements of the matrix $\boldsymbol{A}$ are

$$A_{ij} = \mathbf{f}'_i \cdot \mathbf{f}'_j - (\mathbf{y} - \mathbf{f}(\mathbf{x}_0, \beta)) \cdot \mathbf{f}''_{ij}. \qquad (11)$$

We call Eq. (10) the *scene probability equation*. It has two familiar terms and a new term. The term $\exp(\frac{-\|\mathbf{y}-\mathbf{f}(\mathbf{x}_0,\beta)\|^2}{2\sigma^2})$ penalizes scene hypotheses which do not account well for the original data (hypotheses β for which the squared difference of $\mathbf{f}(\mathbf{x}_0, \beta)$ from the image data $\mathbf{y}$ is large). We call this the *image fidelity* term. (This may also be called the "likelihood of $\mathbf{x}_0$ and β with respect to $\mathbf{y}$"). The *prior probability* term $P_\beta(\beta)$ came from Bayes' law and incorporates prior assumptions. These two terms (the prior and a squared error term) are familiar. $\frac{1}{\sqrt{\det(\boldsymbol{A})}}$, is the new term, arising from the generic view assumption. If the rendered image changes quickly with the generic view variables, the image derivatives of Eq. (11) will be large. Then the generic view term $\frac{1}{\sqrt{\det(\boldsymbol{A})}}$ will be small, causing the scene hypothesis β to be unlikely. This $\frac{1}{\sqrt{\det(\boldsymbol{A})}}$ term quantifies our intuitive notion of generic view, and we call it the *genericity* term. The scene probability equation gives the probability that a scene interpretation β generated the visual data, $\mathbf{y}$, based on fidelity to the data, prior probability, and the probability that the scene would have presented us with the observed visual data.

We have combined the constants which do not depend on β into the normalization constant k. We usually examine relative probabilities; then k doesn't matter. If the model accounts exactly for the image, then $\mathbf{y} - \mathbf{f}(\mathbf{x}_0, \beta) = 0$ and the second derivative term of Eq. (11) can be ignored. Even if $\mathbf{y} \neq \mathbf{f}(\mathbf{x}_0, \beta)$, $\mathbf{f}(\mathbf{x}_0, \beta)$ may differ from $\mathbf{y}$ through random noise in a way which is uncorrelated with the image. Then the dot product of

the second derivatives are likely to be zero (Press et al., 1992). In many cases it is straightforward to calculate the value $\mathbf{x}_0$ in $\mathbf{f}(\mathbf{x}_0, \beta)$.

The approach of Section 4.2 handles cases where the matrix $\boldsymbol{A}$ in the denominator of the genericity term is not of full rank. In general, there will be only one or few generic variables, so the dimensionality $(M \times M)$ of the matrix $\boldsymbol{A}$ is low. We derive the scene probability equation for generic object pose in 3-d in Appendix B.

The quantification of the genericity of a view in Eq. (10) follows established techniques in Bayesian statistics. The matrix $\boldsymbol{A}$ is called the conditional Fisher information matrix (Fisher, 1959; Berger, 1985). It is used to approximate the likelihood locally as a Gaussian (Fisher, 1959; Jeffreys, 1961; Box and Tiao, 1973) and can be used in integration over a loss function or in marginalization (Lindley, 1972; Berger, 1985). For example, Box and Tiao (1964) employ this approximation when they integrate out nuisance parameters from a joint posterior, as we have done here. Gull (1988) calls $\frac{1}{\sqrt{\det(\boldsymbol{A})}}$ the Occam factor and he, Skilling (1989), and MacKay (1992) use it as we have here and in other ways.

The case of only one generic variable and $\|\mathbf{y} - \mathbf{f}(x_0, \beta)\| = 0$ shows the role of the image derivatives more clearly. Then the scene probability equation becomes:

$$P(\beta \mid \mathbf{y}) = c \ \exp\left(\frac{-\|\mathbf{y} - \mathbf{f}(x_0, \beta)\|^2}{2\sigma^2}\right) P_\beta(\beta) \times \frac{1}{\sqrt{\sum_i \left(\frac{\partial f_i(x,\beta)}{\partial x}\Big|_{x=x_0}\right)^2}}, \tag{12}$$

The probability of a parameter vector β varies inversely with the sum of the squares of the image derivatives with respect to the generic variable.

The scene probability densities in Eqs. (10) and (12) are the crux of a Bayesian analysis. Once $P(\beta \mid \mathbf{y})$ is known, the best estimate for β can be found using a number of standard criteria of merit (Papoulis, 1984). The parameter vector which minimizes the expected squared error, β_{MMSE}, is the conditional mean of β:

$$\beta_{\mathrm{MMSE}} = \int P(\beta \mid \mathbf{y}) \ \beta \ d\beta. \tag{13}$$

The *maximum a posteriori* (MAP) estimate is the β which maximizes the conditional probability,

$$\beta_{\mathrm{MAP}} = \underset{\beta}{\operatorname{argmax}} \ P(\beta \mid \mathbf{y}). \tag{14}$$

Alternatively, one can pass a representation of the entire probability density function $P(\beta \mid \mathbf{y})$ on to a higher level of processing.

Including the generic view term provides a better statistical model of the world. Using it should increase the accuracy of scene estimates. Starting from this framework, the next research direction is to develop algorithms which find the best β. Since the generic view term models a regularity that exists in the world, including it may give more powerful and accurate vision algorithms.

3. Shape from Shading Examples

We apply the scene probability equation to some problems in shape from shading. Given a shaded image, lighting conditions and the reflectance function, there are many algorithms which can compute a shape to account for the shaded image; see (Horn, 1989; Horn and Brooks, 1989) for reviews.

Most shape from shading algorithms require specification of the lighting and object surface characteristics. There are a number of methods that can infer these given more than one view of the object (Horn et al., 1978; Woodham, 1980; Grimson, 1984; Pentland, 1990a). Finding the object shape from a single view without these parameters is not a solved problem. Methods have been proposed to estimate light source direction or overall albedo, assuming Lambertian surfaces (Pentland, 1984; Lee and Rosenfield, 1989; Zheng and Chellapa, 1991). Brooks and Horn (1989) proposed a more general scheme that iterated to find a shape and reflectance map that could account for the image data.

However, accounting for image data is not enough. For some classes of images, many shapes and reflectance functions can account equally well for an image (although some images which are impossible to explain by Lambertian shading have been found, (Horn et al., 1993; Brooks et al., 1992)). An infinite number of surface and light source combinations can explain regions of 1-dimensional intensity variations, since the solution just involves a 1-dimensional integration. The rendering conditions of "linear shading" (Pentland, 1990c) can be invoked to explain *any* image, as we discuss later. Thus, to explain a given image, one must choose between a variety of feasible surface shapes, reflectance functions and lighting conditions.

To make such choices, one could invoke preferences for shapes or reflectance functions. Some shape from

shading algorithms do this implicitly by using regularizing functionals. However, these preferences may not be known well. The scene probability equation enables one to use the additional information provided by the generic view assumption to choose between shapes and reflectance functions, lessening the reliance on the prior assumptions about shapes or reflectance functions.

We have not developed a shape from shading algorithm which uses the scene probability equation directly. Rather, we will use existing shape from shading algorithms (Bichsel and Pentland, 1992; Pentland, 1990c) to generate hypothesis shapes and use the scene probability equation to evaluate their probability. Future research can incorporate the scene probability equation, or an approximation to it, directly into a shape from shading algorithm.

3.1. Reflectance Function

We apply the scene probability equation, Eq. (10), to the 1-dimensional examples of Section 1.1, shown again in Figs. 5(a) and (c). This will allow us a principaled way to distinguish between reflectance functions that account equally well for the image data.

Our observation $\mathbf{y}$ is the image data. The parameter vector β we wish to estimate is the shape and reflectance function of the object. We use a two variable parameterization of reflectance functions, a subset of the Cook and Torrance model (1981). The parameters are surface roughness, which governs the width of the specular highlight, and specularity, which determines the ratio of the diffuse and specular reflections. Figure 4 gives a visual key.

We want to evaluate the probability $P(\beta \mid \mathbf{y})$ for each reflectance function in our parameterized space. A shape exists for each reflectance function which could have created the 1-d images of Figs. 5(a) and (c). For this example, we will assume a uniform prior for the reflectance functions and shapes, $P_\beta(\beta) = k$. We used a shape from shading algorithm (Bichsel and Pentland, 1992) to find the shape corresponding to each reflectance function. The boundary condition for the shape from shading algorithm was uniform height at the top edge of the image. For this image with one-dimensional intensity variations, the rendered shape accounts for the image data exactly, and the fidelity term of Eq. (10) for $P(\beta \mid \mathbf{y})$ is 1.

Now we consider the genericity term of the scene probability equation, the denominator of Eq. (10).

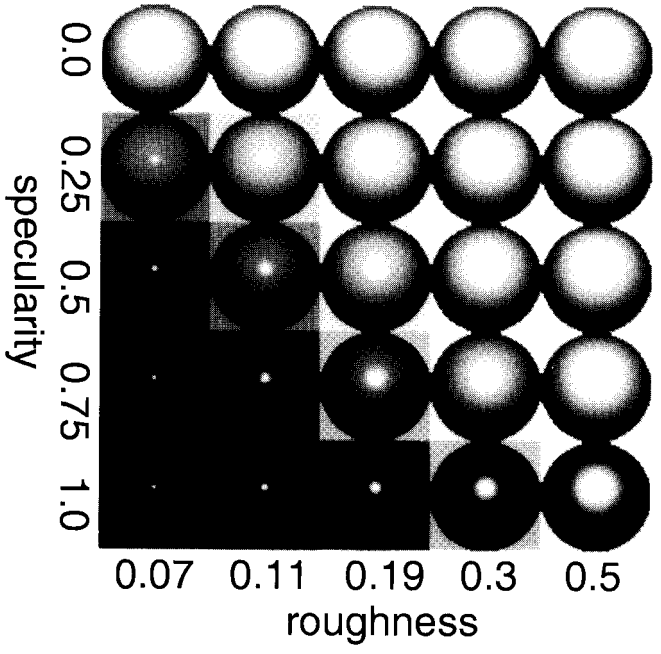

Figure 4. Key to reflectance function parameters of Fig. 5. Reflectance functions are displayed as they would appear on a hemisphere, lit in the same way as Fig. 5(a) and (c). The ratio of diffuse to specular reflectance increases in the vertical direction. The surface roughness (which only affects the specular component) increases horizontally. The sampling increments are linear for specularity and logarithmic for roughness.

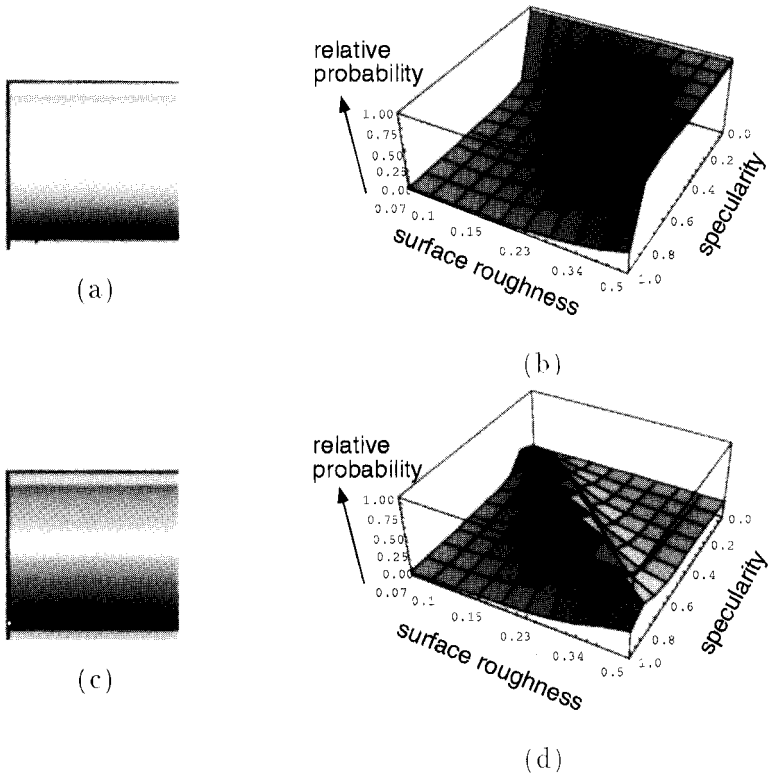

Figure 5. (a) Input image. (b) Probability that image (a) was created by each reflectance function and corresponding inferred shape. The probabilities are highest for the reflectance functions which look like the dull cylinder. See Fig. 4 for a visual guide to the reflectance function parameters of plots (b) and (d). (c) Input image. (d) Probability that (c) was created by each reflectance function and corresponding shape. The probabilities are highest for the reflectance functions which look like the shiny cylinder. All reflectance functions can account for the image data equally well and were assumed to be equally probable. The probability distinctions between reflectance functions came from the genericity term of the scene probability equation.

We will use both the vertical rotation of the object and the light position as the generic variables (the result for the case of generic vertical rotation alone is similar). We need the derivative of the image intensity $I(X, Y)$, at each position X, Y with respect to the rotation angle, ϕ and light position. We assume orthographic projection. The ϕ derivative is a special case of Eq. (27) of Appendix C,

$$\frac{dI}{d\phi} = \frac{\partial I}{\partial Y} Z + \frac{\partial m}{\partial q}(1 + q^2), \qquad (15)$$

where $q = \frac{\partial Z}{\partial Y}$, Z is the surface height, and m is the reflectance map. We have suppressed the X and Y dependence in Eq. (15); by $\frac{\partial m}{\partial q}$ we mean $\frac{\partial m(p,q)}{\partial q} \mid_{q=q(X,Y)}$.

We calculated numerically the image derivative with respect to light position. For the z value of the center of rotation we used the value which minimized the squared derivative of the image with respect to object rotation angle, see Appendix B.

Using the above in the scene probability equation, we plot in Figs. 5(b) and (d) the probability that each reflectance function generated the images of (a) and (c). Note that for each image, the high probabilities correspond to reflectance functions which look (see Fig. 4) more like the material of the image patches in Figs. 5(a) and (c). We have evaluated the relative probability that different reflectance functions created a given image. Note this was done from a single view and for a case where the reflectance function is otherwise completely unknown.

3.2. *Generic Light Direction*

The case of linear shading (Pentland, 1990c) is good for illustrating the benefits of this generic view approach.

Under linear shading, we assume that the image intensities I are linearly proportional to the surface slopes p and q:

$$I = k_1 p + k_2 q. \tag{16}$$

This equation approximates natural reflectance functions under conditions of shallow surface slopes and shallow illumination, or of a broad, linearly distributed light source. $\underline{\text{Arctan}}(k_1, k_2)$ tells the direction of the light, θ_l, and $\sqrt{k_1^2 + k_2^2}$ is proportional to the product of lighting strength and surface reflectance. The inferred surface slopes scale inversely with $\sqrt{k_1^2 + k_2^2}$. Without calibration information, k_1 and k_2 are unknown.

Pentland (1990c) has shown that a linear transformation relates the image I to a surface for which the slopes satisfy Eq. (16) above. Thus for any choice of k_1 and k_2, not both zero, we can find a surface which accounts for the observed image, I, by applying the appropriate linear transformation to it. Thus, assuming linear shading conditions, *any* assumed lighting direction and strength can explain an image by Eq. (16), each using a different inferred shape. How can we choose which shape and lighting parameters are best? The assumption of generic light direction provides a criterion.

Suppose the visual data is the image of Fig. 6(a). Perceptually, there are two possible interpretations: it could be a bump, lit from the left, or a dimple, lit from the right. Yet mathematically, using the linear shading equation, there are many interpretations to choose from. The image could arise from any of the shapes shown in (b), under the proper lighting conditions, which are indicated by the lighting direction arrow shown next to each shape. How should one choose between these competing explanations?

Without considering the generic variables, there are two criteria to evaluate an interpretation from the terms of Bayes rule for the posterior probability, Eq. (4): how well it accounts for the observed data, and the prior probability that the interpretation would exist in the world. If each shape accounts equally well for the image data, we are left with choosing based on prior probabilities. We could arbitrarily decide that we like bump shapes more than tube shapes but we may have no grounds for that. Such a decision could lead to an incorrect interpretation for some other image. What is missing?

For the three tube-like shapes shown, there is a suspicious alignment between the inferred surface structure and the assumed light direction. We would like to include this coincidence in our probability calculation. Figures 6(c) and (d) give an intuition for how the image derivatives of the scene probability equation, Eq. (10), measure the accidentalness of the surface and light direction alignments. If we imagine wiggling the assumed azimuthal light direction slightly, we see that for the shape of (c), the image changes quite a bit. For the shape of (d), we can observe the image of (a) over a much broader range of assumed light directions. There are more opportunities for the shape of (d) to have presented us with the image (a) than there are for the shape of (c).

For each assumed lighting direction (at constant lighting strength), we find the shape β which would create the observed image, $\mathbf{y}$, Fig. 6(a), using the linear shape from shading algorithm of Pentland (1990c) and the boundary conditions described therein.

To evaluate $P(\beta \mid \mathbf{y})$ in the scene probability equation, we need to find $\frac{\partial f_i(x,\beta)}{\partial x} \mid_{x=x_0} = \frac{\partial I}{\partial \theta_l}$. From Eq. (16) and the definition of θ_l, we have

$$\frac{\partial I}{\partial \theta_l} = -k_2 p + k_1 q. \tag{17}$$

Using the above Eq. (17) in the scene probability equation, Eq. (10) gives the probability for each candidate shape, plotted in Fig. 6(e). The bump and dimple shapes, which assume light coming from the left or right, are most likely, in agreement with the appearance of (a). We model variable contrast sensitivity effects for this example in Section 4.1.

Figure 7 shows the probabilities of shapes reconstructed assuming different light directions for an image of a nickel, assuming linear shading and the boundary conditions of Pentland (1990c). The most probable of those shapes assumes a light direction that is consistant with apparent light direction in the image.

3.3. Vertical Scale

We can use the assumption of generic object orientation to estimate the vertical scale in linearly shaded images where the scale is otherwise indeterminate. The intuition is as follows. If the object were very flat, it would require a very bright light at just the right angle to create the observed image. Any small change in the object pose would cause a large change in the image intensities, and that flat object would be unlikely, given the observed image. On the other hand, if the object

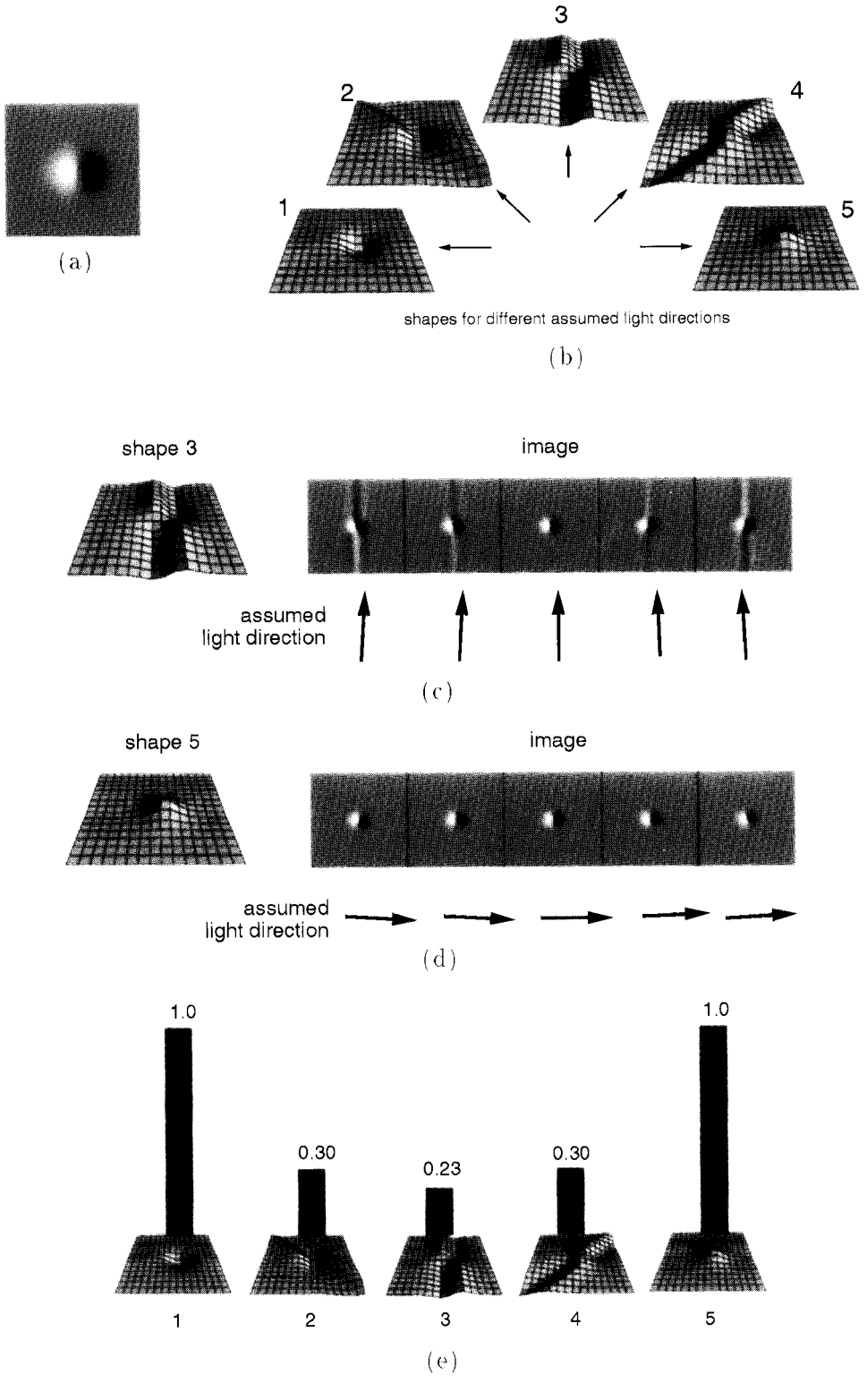

Figure 6. (a) Perceptually, this image has two possible interpretations. It could be a bump, lit from the left, or a dimple, lit from the right. (b) Mathematically, there are many possible interpretations. For a sufficiently shallow incident light angle, if we assume different light directions, we find different shapes, each of which could account for the observed image. Some of the shapes would require a coincidental alignment between the light direction and the inferred structure. (c) For the tube shape shown here, only a small range of light angles yields the observed image. (d) For the bump shape, a much larger range gives the observed image. (e) The scene probability equation allows us to quantify the degree of coincidence in the alignment of surface structure and light direction, by differentiating the observed image with respect to the assumed light direction. The resulting probabilities are shown here for the 5 different shapes. The results favor shapes which assume that the light comes from the left or right, in agreement with the perceptual appearance of (a). Reprinted with permission from *Nature* (Freeman, 1994).

were very tall and lit by a weak light, then, if the object were rotated, the image would change significantly because of parallax. In between those extremes, there should be a most probable lighting strength and corresponding shape.

The scene probability equation quantifies that intuition. We first need the derivatives of the observed image with respect to the generic variables. For the case of object pose in 3-dimensions, the generic variable is the rotation angle about all possible axes of rotation. We integrate over all possible rotation axes, as described in Appendix B. The resulting scene probability equation involves the image, the surface estimate and its spatial derivatives, and the reflectance map and its derivatives. The rotation origin is chosen to minimize the squared image derivative with respect to the rotation, see Appendix B.

Figure 7(c) shows the resulting probability tuning for vertical scale. In agreement with our intuition, very large and very small vertical scales are both unlikely. The distribution agrees well with the actual height (relative to picture widths) of the nickel. Note, however, that the tuning for vertical scale is very broad; the width at half maximum represents a factor of 64 in vertical scale. Nonetheless that is more information than we had before.

3.4. Why the Prior Probability Is Not Enough

Figure 8 shows an example where both the fidelity and prior probability terms favor a perceptually implausible explanation. The genericity term alone favors the perceptually plausible explanation and overwhelms the other two. Figure 8(a) shows an image, and (b1) and (c) are two possible explanations for it. (b1), lit at a grazing angle from the light direction shown above it, yields the image (d). (c), lit from a different direction, yields (e). (b2) shows the shape (b1) with the vertical scale exaggerated by a factor of 7. (We made this example by construction. Gaussian random noise at a 7 dB signal to noise ratio was added to (e) to make (a). (b1) was found from (a) using a shape from shading algorithm, assuming constant surface height at the left picture edge (Bichsel and Pentland, 1992). We evaluated the probabilities of (f) assuming both generic object orientation and generic azimuthal lighting direction. The actual noise variance was used for σ^2 in the fidelity term of Eq. (10), although a wide range of assumed variances would give the results we describe. The reflectance function was Lambertian.)

Perceptually, the shape Fig. 8(c) seems like a better explanation of (a) than the shape (b1), even though it doesn't account for all the noise. However, the fidelity

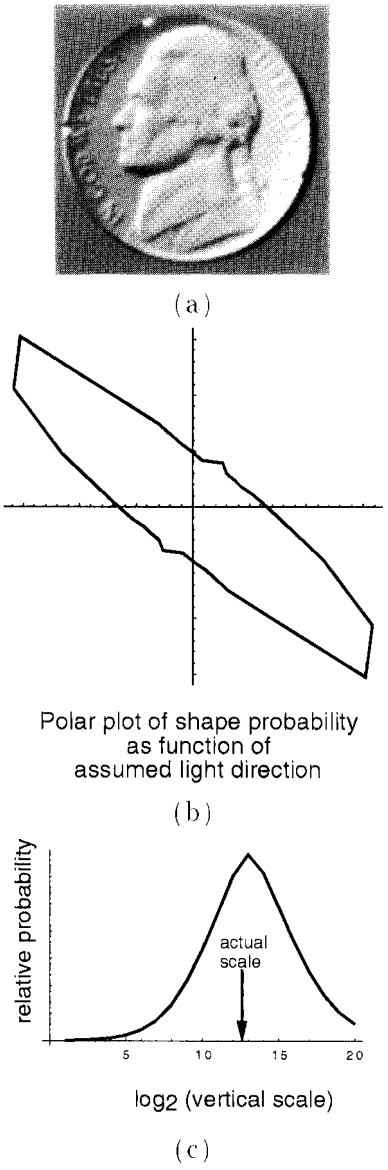

Figure 7. (a) Plaster mold from a nickel. We found a shape which yields the image (a) for various assumed azimuthal light directions. We assumed linear shading of constant lighting strength, with the boundary conditions of (Pentland, 1990). (b) Shows the probability for each shape and lighting combination from the genericity term of the scene probability equation, Eq. (10), under the assumption of generic azimuthal light direction. Each shape was assumed to be a priori equally probable. The probabilities are plotted as a function of the assumed light direction, showing that the shapes reconstructed assuming the correct light direction are more probable than those that were reconstructed assuming other light directions. (c) Under the linear shading approximation, many different vertical scalings can account for a given image, each assuming a different lighting strength. We inferred shapes which account for (a), using the same boundary conditions as before. (c) Shows the probability as a function of vertical scale for each of the shapes considered, obtained from the genericity term of the scene probability equation. While broadly tuned, this distribution agrees well with the actual height of the nickel (in terms of the picture width).

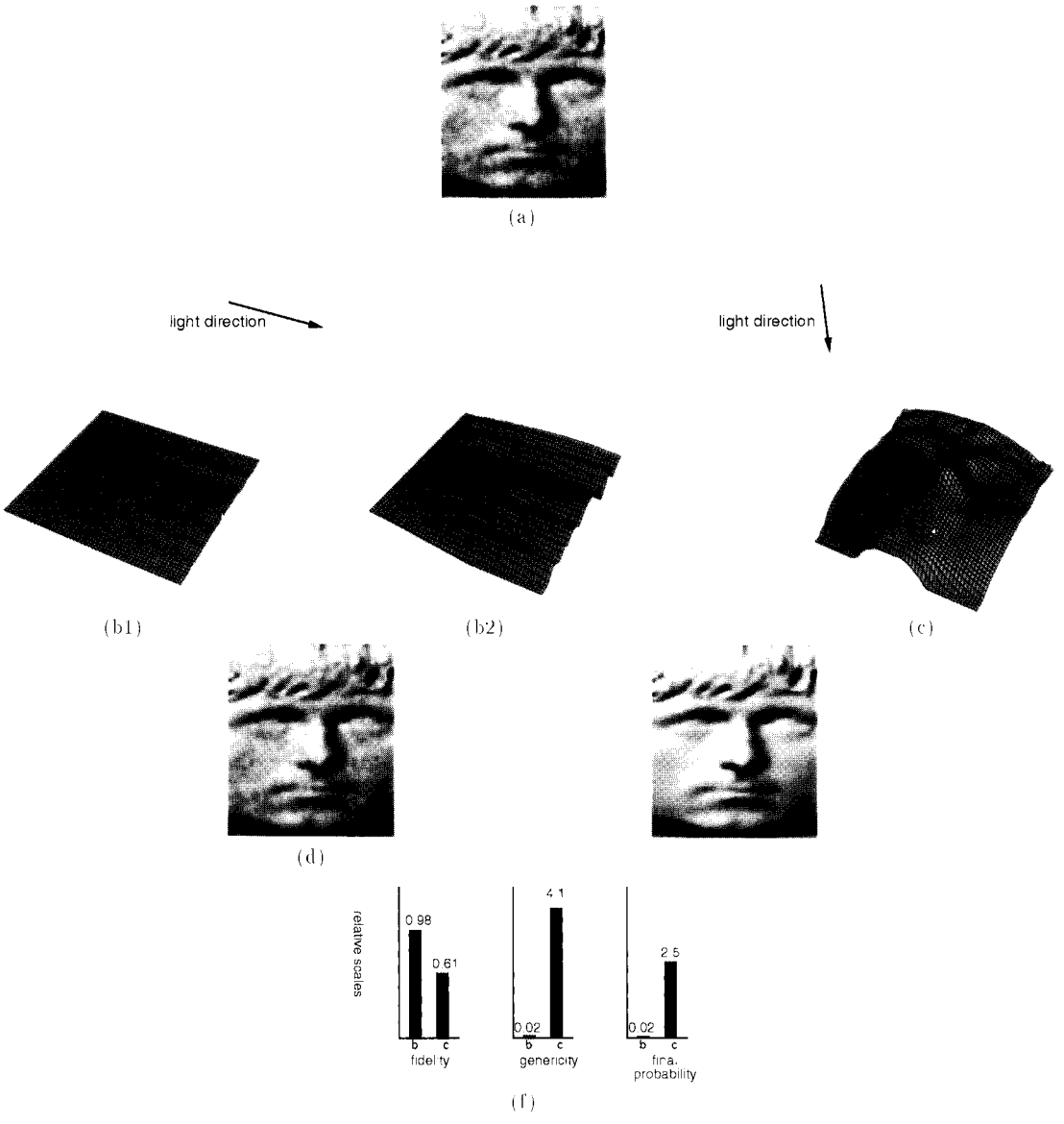

Figure 8. An example showing the need for the genericity term in Eq. (10). We compare the probability densities of two explanations for the image in (a). The surface (b2), lit from the left, yields the image (d). ((b2) Shows the same shape at 7 × vertical exaggeration.) Shape (c) is another possible interpretation. When lit from above, it yields (e), a less faithful version of the original image. The image fidelity term of Eq. (10) favors the shape (b1). The commonly used prior probability of surface smoothness (Poggio et al., 1985; Terzopoulos, 1986) also favors the shape (b1). However, the shape (b1) must be precisely positioned with respect to the light source to create the image (d). The genericity term of Eq. (10) penalizes this. Image (e) is stable with respect to lighting and object movements, giving a higher overall probability to shape (c), plotted in (f) without a surface smoothness prior.

term of Eq. (10) favors the flat interpretation, (b2), since it accounts better for the noisy details of (c). In a Bayesian or regularization approach, without considering the genericity, the only term left to evaluate the probability of an interpretation is the prior probability. A typical prior is to favor smooth surfaces, which would again favor the shape (b1), since it is much smoother than (e), as measured by the squared second derivatives of the surfaces.

We need some way to penalize the precise alignment between the light source and the object that is required to get the image (d) from the shape (b1). The genericity

term of the scene probability equation provides this. Because the image (e) is more stable with respect to object or lighting rotations, (c) has a higher overall probability than the shape (b1).

Whether or not the smooth shape (b1) would be more likely to exist in the world than the shape (c), it would be very unlikely to present the viewer with the image (d). Our approach takes both the prior probabilities and the probabilities of the viewing conditions into account to better model the conditional probability of each shape, given the image data.

4. Discussion

4.1. Other Image Error Metrics

The assumed observation noise model sets the penalty for differences between the rendered scene model and the observed data. The identically distributed Gaussian noise model of Eq. (3) corresponds to a squared error penalty for differences between images. While this may be adequate for many applications, it is not a good model of the human visual system's response (Schreiber, 1986). Our Bayesian framework can accomodate a different image error metric.

Viewers are more sensitive to intensity changes in regions of low image contrast. We will assume that the sensitivity for contrast detection is proportional to the local contrast, a model based on Weber's Law (Cornsweet, 1970). (Other local contrast response models (Albrecht and Geisler, 1991; Carandini and Heeger, 1994) could be used.) This fractionally scaled approach is consistent with the multiplicative impact on image intensities of changes in lighting intensity or small changes in surface slope.

It is convenient to model the contrast sensitivity differences as variations in the strength of the observation noise. We generalize our observation noise model to

$$P_{\mathbf{n}}(\mathbf{n}) = \frac{1}{(\sqrt{2\pi\sigma^2})^N \, |\mathbf{\Lambda}|^{\frac{1}{2}}} \times \exp \frac{-(\mathbf{y} - \mathbf{f}(\mathbf{x}, \beta))^T \, \mathbf{\Lambda}^{-1} \, (\mathbf{y} - \mathbf{f}(\mathbf{x}, \beta))}{2\sigma^2}, \tag{18}$$

where σ^2 is now a scale factor for a noise covariance matrix $\mathbf{\Lambda}$. We calculate the contrast as the square root of the local image variance $(\bar{I^2}) - (\bar{I})^2$, where overbar denotes a local spatial average and I are the image intensities. We then use $\Lambda = \text{diag}[(\bar{I^2}) - (\bar{I})^2]$, where diag places the elements of an N dimensional vector along the diagonal of an N by N matrix.

Following the steps of Eqs. (5)–(9) with the noise model of Eq. (18) yields a modified scene probability equation,

$$P(\beta \mid \mathbf{y}) = k \, \exp\left(\frac{-(\mathbf{y} - \mathbf{f}(\mathbf{x}_0, \beta))^T \mathbf{\Lambda}^{-1} (\mathbf{y} - \mathbf{f}(\mathbf{x}_0, \beta))}{2\sigma^2}\right) \times [P_\beta(\beta) P_{\mathbf{x}}(\mathbf{x_0})] \, \frac{1}{\sqrt{\det(\boldsymbol{A})}} \tag{19}$$

where the i and jth elements of the matrix $\boldsymbol{A}$ are

$$A_{ij} = \mathbf{f}_i' \, \mathbf{\Lambda}^{-1} \, \mathbf{f}_j' - (\mathbf{y} - \mathbf{f}(\mathbf{x}_0, \beta)) \, \mathbf{\Lambda}^{-1} \, \mathbf{f}_{ij}''. \tag{20}$$

In both the fidelity and genericity terms, squared image differences and derivatives are now scaled by the reciprocal of the local contrast variance.

We use the example of Fig. 6 to illustrate the usefulness of these modifications. On perturbing the light source direction, the tube-like shapes cause image changes where they are very detectable, in low-contrast regions of the image. Equations (19) and (20) will provide extra penalty for such image changes.

Figure 9(a) shows the input image. (b) is the local image variance. It is brightest near the center of the blob, as expected. The spatial averaging used was a 2.5 pixel standard deviation Gaussian blur (the image is 128×128 pixels). The dynamic range of the local noise variance image was restricted to be 100 to 1. (g) Shows the calculated probabilities for each shape, based on the contrast sensitivity model of Eq. (19). Note that the tube-like shapes are penalized much more with this varying contrast sensitivity model than they were in the calculation of Fig. 6, which assumed uniform contrast sensitivity.

4.2. Relationship to Loss Functions

The loss functions of Bayesian decision theory (Berger, 1985) provide an alternate interpretation of the genericity term in the scene probability equation. This analysis has been described by Freeman and Brainard (1995), Freeman (1996), Yuille and Bulthoff (1996).

We include the generic variables $\mathbf{x}$ as well as the scene parameters β into an augmented scene parameter variable, $\mathbf{z}$. A loss function $L(\mathbf{z}, \tilde{\mathbf{z}})$ specifies the penalty for estimating $\tilde{\mathbf{z}}$ when the true value is $\mathbf{z}$. Knowing the

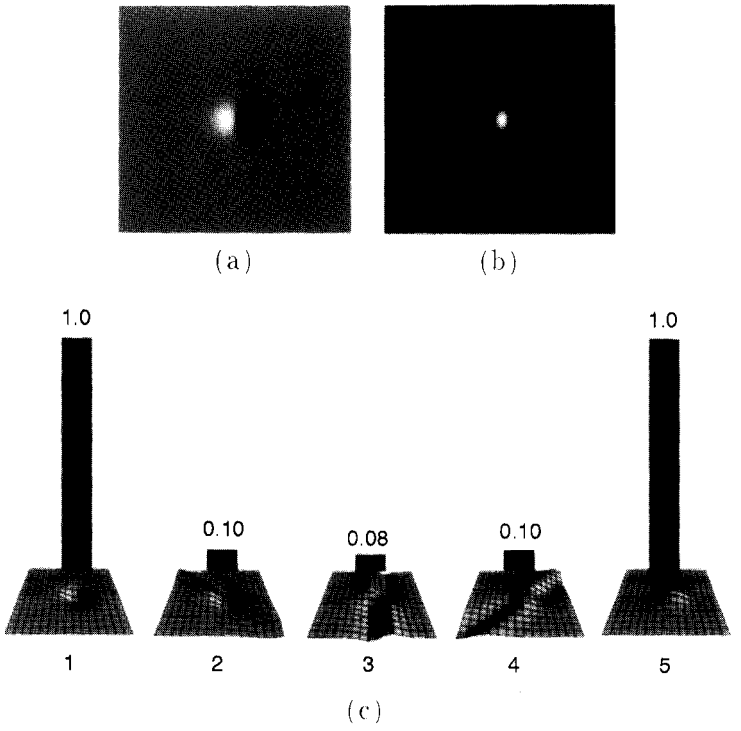

Figure 9. The effect of modeling contrast dependent noise sensitivity. (a) Input image. (b) Assumed observation noise variance, calculated from local image variance. This noise distribution will allow us to model contrast dependent sensitivity to image changes. (c) Resulting posterior probabilites. Note that the tube-like shapes are now penalized further than they were with the contrast independent noise sensitivity model of Fig. 6(e). Perturbing the light source position with shapes 2–4 causes the image to change in regions of low-contrast, which the sensitivity to changes is assumed to be high.

posterior probability, one can select the parameter values which minimize the expected loss for a particular loss function:

$$[\text{expected loss}] = \int [\text{posterior}]\,[\text{loss function}]\; d\,[\text{parameters}]$$

$$R(\tilde{\mathbf{z}} \mid \mathbf{y}) = -C \int \left[\exp\left[-\frac{\tau}{2\sigma^2} \|\mathbf{y} - \mathbf{f}(\mathbf{z})\|^2 \right] \times \mathbf{P}_{\mathbf{z}}(\mathbf{z}) \right] L(\mathbf{z}, \tilde{\mathbf{z}})\, d\mathbf{z}, \quad (21)$$

where we have substituted from Bayes' rule, Eq. (4), and the noise model, Eq. (3). The optimal estimate is the parameter $\tilde{\mathbf{z}}$ of minimum risk.

We have not specified what loss function to use with the posterior probability of the scene probability equation, Eq. (10). For this comparison, we will assume MAP estimation, Eq. (14), where we choose the scene parameters β which maximize the posterior probability. The comparison for other estimators is analogous.

The integral to be minimized for the expected loss in Eq. (21) can be made equivalent to the integral to be maximized for the marginal posterior in Eq. (5). We must choose the proper loss function: $L(\mathbf{z}, \tilde{\mathbf{z}}) = -\delta(\beta - \tilde{\beta})$. This means we don't care at all about the generic variables $\mathbf{x}$, but we care about the scene parameter components, β, to infinite precision. This loss function is plotted in Fig. 10(b). Figure 10(a) explains the loss function plot format. MAP estimation using the marginal posterior after integrating out the generic variables is equivalent to finding the parameter of minimum risk using the loss function of Fig. 10(b).

Figure 10(c) shows another possible form for the loss function, allowing different parameters to be estimated with different requirements for precision. Generic variables could be estimated with coarse precision, and scene parameters with high precision. See (Brainard and Freeman, 1994; Freeman and Brainard, 1995; Freeman, 1996; Yuille and Bulthoff, 1996) for examples of this approach. An advantage is that it avoids dividing the world parameters into two groups, generic variables and scene parameters. A disadvantage is that

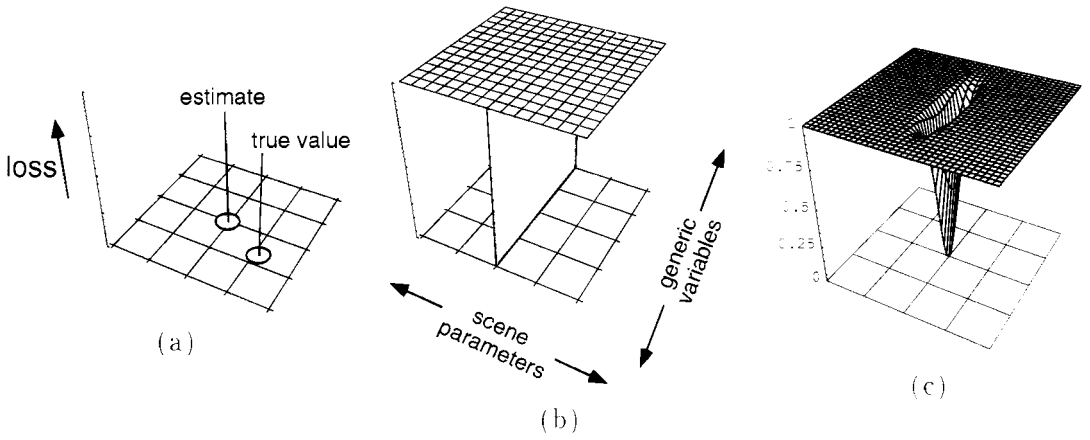

Figure 10. Loss function interpretation of generic viewpoint assumption. (a) Shows the general form for a shift invariant loss function. The function $L(\mathbf{z}, \tilde{\mathbf{z}})$ describes the penalty for guessing the parameter $\tilde{\mathbf{z}}$ when the acutal value was $\mathbf{z}$. The marginalization over generic variables of Eq. (5) followed by MAP estimation is equivalent to using the loss function of (b). (c) Shows another possible form for the loss function, discussed in (Brainard and Freeman, 1994; Freeman and Brainard, 1995; Freeman, 1996; Yuille and Bluthoff, 1996).

integration over scene parameters, as prescribed by the loss function of (c), might be difficult for scene parameters of high dimensionality.

5. Summary

The generic view assumption is commonly used to label scene interpretations as either "generic" or "accidental" in a world of geometrical objects. Here, we extend this to a complementary, continuous domain by assigning relative probabilities to different scene interpretations.

The visual input can be greyscale images or other visual data. We divide the parameters into two groups: scene parameters, and generic variables. Scene parameters are the parameters such as shape or velocity that we want to estimate. We marginalize over the generic variables, which can include lighting direction, object orientation, or viewpoint.

We apply this in a Bayesian framework. The prior probabilities for generic variables are typically well-known and simple. We integrate the joint posterior distribution over the generic variables to gain extra information about the scene parameters. We use a commonly employed low-noise approximation to obtain an analytic result. The resulting *scene probability equation* gives the probability of a set of scene parameters, given an observed image. It has three terms:

a *fidelity term*, which requires that the scene parameters explain the observed visual data;

the *prior probability*, which accounts for prior expectations of the scene parameters;

the *genericity term*, which quantifies how accidental our view of a particular scene is. It reflects the probability that a given scene would have presented us with the observed image. This term occurs in Bayesian analysis applied to other domains. Including its effects may lessen the reliance on the prior probabilities, for example, in choosing between explanations which account for the image data equally well.

We show various applications to shape from shading. The scene probability equation gives the probability of different shape and reflectance function combinations to explain a given image. The scene probability equation, Eq. (10), gives a principled way to select shape and light direction or reflectance function calibration in cases where these are otherwise ambiguous. The genericity term in the scene probability is important; one can have a shape from shading solution which is faithful to the data, but unlikely, and one which is less faithful but more likely. We draw connections between the scene probability equation and the loss functions of Bayesian decision theory.

This approach may have many applications in vision. The scene probability equation derived in this paper could be incorporated into algorithms of, for example, shape from shading, motion analysis, and stereo. This may result in vision algorithms of greater power and accuracy.

Appendix

A. *Asymptotic Expansion of Marginalization Integral*

We want to examine the asymptotic behavior of the integral of Eq. (6) in Eq. (5) when the observation noise covariance becomes small, or $\frac{1}{\sigma}$ becomes large. For an integral of the form

$$B(\tau) = \int \exp\left[-\tau\phi(\mathbf{x})\right] g(\mathbf{x})\, d\mathbf{x}, \tag{22}$$

one can show (Bleistein and Handelsman, 1986) that the leading order term in an asymptotic expansion for large τ is:

$$B(\tau) \approx \frac{e^{-\tau\phi(\mathbf{x}_0)}}{\sqrt{|\det(\phi_{x_i x_j}(\mathbf{x}_0))|}} \left(\frac{2\pi}{\tau}\right)^{\frac{n}{2}} g(\mathbf{x}_0), \tag{23}$$

where $\mathbf{x}_0$ minimizes $\phi(\mathbf{x})$ and n is the dimensionality of $\mathbf{x}$.

We can put our integral into the form of Eq. (22) if we identify $g(\mathbf{x}) = \mathbf{P}_\mathbf{x}(\mathbf{x})$, $\tau = \frac{1}{\sigma}$, and

$$\phi(\mathbf{x}) = \frac{1}{2\sigma^2}\|(\mathbf{y} - \mathbf{f}(\mathbf{x}, \beta))\|^2. \tag{24}$$

Substituting these into Eq. (23) gives Eq. (10). Twice differentiating $\phi(\mathbf{x})$ in Eq. (24) gives Eq. (11).

B. *Scene Probability Equation under General Object Pose*

We want to find the scene probability $P(\beta \mid \mathbf{y})$ for a shaded image under the condition of general object pose. Here the reflectance map $m(p, q)$ and the shape $Z(X, Y)$ make up the scene parameter β. (We use capital X, Y, and Z for the Cartesian coordinates of the object surface). The reflectance map tells the image brightness as a function of object slopes $p = \frac{\partial Z(X,Y)}{\partial X}$ and $q = \frac{\partial Z(X,Y)}{\partial Y}$.

The generic variable is the rotation angle ϕ about the unit vector rotation axis $\hat{\omega}$. We assume the prior probability density for rotation angle ϕ is uniform and we integrate over all possible axes, $\hat{\omega}$. Equations (5) and (6) give

$$P(\beta \mid \mathbf{y}) = k_1 P_\beta(\beta) \int_{\text{all unit } \hat{\omega}} \times \int_{\text{all } \phi} e^{\frac{-\|\mathbf{y}-\mathbf{f}(\hat{\omega},\phi,\beta)\|^2}{2\sigma^2}} d\hat{\omega} d\phi. \tag{25}$$

The expansion of Appendix A gives

$$P(\beta \mid \mathbf{y}) = k_2 P_\beta(\beta) \exp\frac{-\|\mathbf{y} - \mathbf{f}(\beta)\|^2}{2\sigma^2} \int_{\text{all unit } \hat{\omega}} \times \int_{\text{all } \phi} \frac{1}{\sqrt{\mathbf{f}' \cdot \mathbf{f}' - (\mathbf{y} - \mathbf{f}(\beta)) \cdot \mathbf{f}''}} d\hat{\omega} d\phi. \tag{26}$$

where we have written $\mathbf{f}(\phi_0 = 0, \beta) = \mathbf{f}(\beta)$. For this treatment, we set $(\mathbf{y} - \mathbf{f}(\beta)) \cdot \mathbf{f}'' = 0$, for the reasons cited below Eq. (11).

We seek $\mathbf{f}' = \frac{d\mathbf{I}}{d\phi}$. Given the surface height, Z and slopes p, q at each pixel, we want to find the derivative of the image intensity I with respect to rotation in ϕ about the unit vector $\hat{\omega}$. By straightforward manipulations we show in Appendix C that

$$\frac{dI}{d\phi} = Q\omega_X + R\omega_Y + S\omega_Z, \tag{27}$$

where ω_a is the a component of the unit vector $\hat{\omega}$ and

$$\begin{aligned} Q &= \frac{\partial I}{\partial Y} Z + pq\frac{\partial m}{\partial p} + (1 + q^2)\frac{\partial m}{\partial q} \\ R &= -\frac{\partial I}{\partial X} Z - pq\frac{\partial m}{\partial q} - (1 + p^2)\frac{\partial m}{\partial p} \\ S &= Y\frac{\partial I}{\partial X} - X\frac{\partial I}{\partial Y} + p\frac{\partial m}{\partial q} - q\frac{\partial m}{\partial p}. \end{aligned} \tag{28}$$

For brevity, we have suppressed the X and Y dependence of the symbols on both sides of Eq. (28). By $\frac{\partial m}{\partial p}$ we mean $\frac{\partial m(p,q)}{\partial p}\,|_{p=p(X,Y)}$.

One can parameterize the direction of the unit vector $\hat{\omega}$ by angle θ in the X-Y plane, and angle γ with the Z axis. The integral over all $\hat{\omega}$ of Eq. (26) is straightforward to evaluate numerically in terms of dot products of the images **Q**, **R**, and **S** which appear in the square root:

$$P(\beta \mid \mathbf{y}) = k_2 P_\beta(\beta) \exp\frac{-\|\mathbf{y}-\mathbf{f}(\beta)\|^2}{2\sigma^2} \int_0^{\pi} d\theta \int_0^{2\pi} \times d\gamma \frac{\sin(\gamma)}{\sqrt{2\pi\sigma^2\|\mathbf{Q}\cos\theta\sin\gamma + \mathbf{R}\sin\theta\sin\gamma + \mathbf{S}\cos\gamma\|^2}}. \tag{29}$$

If we add another generic variable, that of the light direction azimuthal angle ψ, we can follow an analogous derivation of the scene probability equation. The

result is

$$P(\beta \mid \mathbf{y}) = k_2 P_\beta(\beta) \exp \frac{-\|\mathbf{y} - \mathbf{f}(\beta)\|^2}{2\sigma^2} \int_0^\pi d\theta \int_0^{2\pi} \times d\gamma \frac{\sin(\gamma)}{\sqrt{2\pi\sigma^2 \det \begin{vmatrix} \frac{d\mathbf{I}}{d\phi} \cdot \frac{d\mathbf{I}}{d\phi} & \frac{d\mathbf{I}}{d\phi} \cdot \frac{d\mathbf{I}}{d\psi} \\ \frac{d\mathbf{I}}{d\phi} \cdot \frac{d\mathbf{I}}{d\psi} & \frac{d\mathbf{I}}{d\psi} \cdot \frac{d\mathbf{I}}{d\psi} \end{vmatrix}}}. \tag{30}$$

Only $\frac{d\mathbf{I}}{d\phi}$ is a function of θ or γ and numerical integration over θ and γ is straightforward.

Finally, we need to specify the origin, X_0, Y_0, Z_0 of the object's rotation. We set $X_0 = Y_0 = 0$, the center of the image. For the Z origin, we want a value which doesn't introduce spurious image change because of the origin placement. We take Z_0 to be that value which minimizes the average squared derivative over all orientations for $\omega_Z = 0$. That is the Z_0 which minimizes

$$\sum_{\text{pixels}} (Q^2 + R^2) = \sum_{\text{pixels}} \left(\frac{\partial I}{\partial Y}(Z - Z_0) + pq\frac{\partial m}{\partial p} + (1+q^2)\frac{\partial m}{\partial q} \right)^2 + \left(\frac{\partial I}{\partial X}(Z - Z_0) + pq\frac{\partial m}{\partial q} + (1+p^2)\frac{\partial m}{\partial p} \right)^2. \tag{31}$$

The dependence on the variables X and Y has been suppressed. Minimizing this quadratic equation with respect to Z_0 gives

$$Z_0 = \frac{1}{\sum_{\text{pixels}} \left(\left(\frac{\partial I}{\partial X}\right)^2 + \left(\frac{\partial I}{\partial Y}\right)^2 \right)} \times \left(\sum_{\text{pixels}} \frac{\partial I}{\partial Y} \left(\frac{\partial I}{\partial Y} Z + pq\frac{\partial m}{\partial p} + (1+q^2)\frac{\partial m}{\partial q} \right) + \sum_{\text{pixels}} \frac{\partial I}{\partial X} \left(\frac{\partial I}{\partial X} Z + pq\frac{\partial m}{\partial q} + (1+p^2)\frac{\partial m}{\partial p} \right) \right). \tag{32}$$

C. *Image Derivatives for General Object Pose*

Given the surface height, Z and slopes p, q at each pixel, we want to find $\frac{dI}{d\phi}$, the change in the image intensity with respect to rotation in the angle ϕ about an axis $\hat{\omega}$ under orthographic projection. We use this result in Appendix B and in Section 3.1. The change in image intensity comes from two effects:

1. The change in image intensity because a new surface element comes into view at the position X, Y.
2. The change in image intensity due to the change in slopes p, q caused by the rotation.

The total derivative of the image intensity is the sum of those two changes,

$$\frac{dI}{d\phi} = \left[\frac{\partial I}{\partial X}\frac{\partial X}{\partial \phi} + \frac{\partial I}{\partial Y}\frac{\partial Y}{\partial \phi} \right] + \left[\frac{\partial I}{\partial p}\frac{\partial p}{\partial \phi} + \frac{\partial I}{\partial q}\frac{\partial q}{\partial \phi} \right]. \tag{33}$$

Consider the first term of Eq. (33). The desired image intensity change is the dot product of the spatial gradient of the image with the projected velocity due to the rotation. The rotation velocity is $\hat{\omega} \times \mathbf{r}(X, Y)$, where $\mathbf{r}(X, Y)$ is the position vector of the point seen at X, Y. Its velocity relative to the stationary observed image is $-\hat{\omega} \times \mathbf{r}(X, Y)$. Thus

$$\frac{\partial I}{\partial X}\frac{\partial X}{\partial \phi} + \frac{\partial I}{\partial Y}\frac{\partial Y}{\partial \phi} = \frac{\partial I}{\partial X}(\omega_Z Y - \omega_Y Z) + \frac{\partial I}{\partial Y}(\omega_X Z - \omega_Z X). \tag{34}$$

Consider the second term of Eq. (33). To determine $\frac{\partial p}{\partial \phi}$ and $\frac{\partial q}{\partial \phi}$ we look at the change in the local surface normal vector, $\hat{\mathbf{n}}$, under rotation and then relate that to the change in p and q. From the definitions of p, q, and $\hat{n}$, we have

$$p = -\frac{n_X}{n_Z}$$
$$q = -\frac{n_Y}{n_Z}, \tag{35}$$

where $\hat{\mathbf{n}} = n_X \hat{i} + n_Y \hat{j} + n_Z \hat{k}$. For a rotation in angle ϕ about the unit vector $\hat{\omega}$, we have

$$\frac{d\hat{\mathbf{n}}}{d\phi} = \hat{\omega} \times \hat{\mathbf{n}}. \tag{36}$$

If we differentiate Eq. (35) for p and q with respect to ϕ and use Eq. (36) for the components of $\frac{d\hat{n}}{d\phi}$, we find

$$\frac{\partial p}{\partial \phi} = -\frac{n_Z(\omega_Y n_Z - n_Y \omega_Z) - n_X(\omega_X n_Y - n_X \omega_Y)}{n_Z^2}, \tag{37}$$

and

$$\frac{\partial q}{\partial \phi} = -\frac{n_Z(\omega_Z n_X - n_Z \omega_X) - n_Y(\omega_X n_Y - n_X \omega_Y)}{n_Z^2}. \tag{38}$$

Using Eq. (35) in Eqs. (37) and (38) above we have

$$\frac{\partial p}{\partial \phi} = pq\,\omega_X - (1+p^2)\,\omega_Y - q\,\omega_Z, \tag{39}$$

and

$$\frac{\partial q}{\partial \phi} = p\,\omega_Z + (1+q^2)\,\omega_X - qp\,\omega_Y. \tag{40}$$

Combining Eq. (34) for the first term of Eq. (33) with Eqs. (39) and (40) for the second we have,

$$\begin{aligned}\frac{dI}{d\phi} = {} & \frac{\partial I}{\partial X}(\omega_Z Y - \omega_Y Z) + \frac{\partial I}{\partial Y}(\omega_X Z - \omega_Z X) \\ & + \frac{\partial m}{\partial p}(pq\omega_X - \omega_Y(1+p^2) - q\omega_Z) \\ & + \frac{\partial m}{\partial q}(p\omega_Z + \omega_X(1+q^2) - qp\omega_Y)\end{aligned} \tag{41}$$

where we have substituted $\frac{\partial m}{\partial p} = \frac{\partial m(p,q)}{\partial p}\,|_{p=p(X,Y)}$ for $\frac{\partial I}{\partial p}$ (and similarly for q) in Eq. (33). Grouping these terms by components of $\hat{\omega}$ gives Eq. (28), as desired.

Acknowledgments

For helpful discussions and suggestions, thanks to: E. Adelson, D. Brainard, D. Knill, D. Mumford, K. Nakayama, A. Pentland, B. Ripley, E. Simoncelli, R. Szeliski and A. Yuille. Some of this research was performed at the MIT Media Laboratory and was supported by a contract with David Sarnoff Research Laboratories (subcontract to the National Information Display Laboratory) to E. Adelson.

References

Albert, M.K. and Hoffman, D.D. 1995. Genericity in spatial vision. In D. Luce (Ed.), *Geometric Representations of Perceptual Phenomena: Papers in Honor of Tarow Indow's 70th Birthday*. L. Erlbaum (in press).

Albrecht, D.G. and Geisler, W.S. 1991. Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience*, 7:531–546.

Belhumeur, P.N. 1996. A computational theory for binocular stereopsis. In D. Knill and W. Richards (Eds.), *Perception as Bayesian Inference*. Cambridge University Press.

Berger, J.O. 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer.

Bichsel, M. and Pentland, A.P. 1992. A simple algorithm for shape from shading. In *Proc. IEEE CVPR*, Champaign, IL, pp. 459–465.

Biederman, I. 1985. Human image understanding: Recent research and a theory. *Comp. Vis., Graphics, Image Proc.*, 32:29–73.

Binford, T.O. 1981. Inferring surfaces from images. *Artificial Intelligence*, 17:205–244.

Bleistein, N. and Handelsman, R.A. 1986. *Asymptotic Expansions of Integrals*. Dover.

Box, G.E.P. and Tiao, G.C. 1964. A Bayesian approach to the importance of assumptions applied to the comparison of variances. *Biometrika*, 51(1, 2):153–167.

Box, G.E.P. and Tiao, G.C. 1973. *Bayesian Inference in Statistical Analysis*. John Wiley and Sons, Inc.

Brainard, D.H. and Freeman, W.T. 1994. Bayesian method for recovering surface and illuminant properties from photosensor responses. In *Proceedings of SPIE*, vol. 2179, San Jose, CA.

Brooks, M.J. and Horn, B.K.P. 1989. Shape and source from shading. In B.K.P. Horn and M.J. Brooks (Eds.), *Shape from Shading*. MIT Press: Cambridge, MA, Chap. 3.

Brooks, M.J., Chojnacki, W., and Kozera, R. 1992. Impossible and ambiguous shading patterns. *Int. J. Comp. Vis.*, 7(2):119–126.

Bulthoff, H.H. 1991. Bayesian models for seeing shapes and depth. *Journal of Theoretical Biology*, 2(4).

Carandini, M. and Heeger, D.J. 1994. Summation and division by neurons in primate visual cortex. *Science*, 264:1333–1336.

Cook, R.L. and Torrance, K.E. 1981. A reflectance model for computer graphics. In *SIGGRAPH-81*.

Cornsweet, T.N. 1970. *Visual Perception*. Academic Press.

Darrell, T., Sclaroff, S., and A. Pentland. 1990. Segmentation by minimal description. In *Proc. 3rd Intl. Conf. Computer Vision*, Osaka, Japan, IEEE, pp. 112–116.

Dickinson, S.J., Pentland, A.P., and Rosenfeld, A. 1992. 3-d shape recovery distributed aspect matching. *IEEE Pat. Anal. Mach. Intell.*, 14(2):174–198.

Fisher, R.A. 1959. *Statistical Methods and Scientific Inference*. Hafner.

Freeman, W.T. 1993. Exploiting the generic view assumption to estimate scene parameters. In *Proc. 4th Intl. Conf. Comp. Vis.*, Berlin, IEEE, pp. 347–356.

Freeman, W.T. 1994. The generic viewpoint assumption in a framework for visual perception. *Nature*, 368(6471):542–545.

Freeman, W.T. and Brainard, D.H. 1995. Bayesian decision theory, the maximum local mass estimate, and color constancy. In *Proc. 5th Intl. Conf. Comp. Vis.*, Boston, IEEE, pp. 210–217.

Freeman, W.T. 1996. The generic viewpoint assumption in a Bayesian framework. In D. Knill and W. Richards (Eds.) *Perception as Bayesian Inference*. Cambridge University Press.

Geman, S. and Geman, D. 1984. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Pat. Anal. Mach. Intell.*, 6:721–741.

Grimson, E. 1984. Binocular shading and visual surface reconstruction. *Comp. Vis., Graphics, Image Proc.*, 28:19–43.

Gull, S.F. 1988. Bayesian inductive inference and maximum entropy. In G.J. Erickson and C.R. Smith (Eds.), *Maximum Entropy*

and Bayesian Methods in Science and Engineering, Kluwer, vol. 1.

Gull, S.F. 1989. Developments in maximum entropy data analysis. In J. Skilling (Ed.), *Maximum Entropy and Bayesian Methods*, Cambridge. Kluwer, pp. 53–71.

Heeger, D.J. and Simoncelli, E.P. 1992. Model of visual motion sensing. In L. Harris and M. Jenkin (Eds.), *Spatial Vision in Humans and Robots*. Cambridge University Press.

Horn, B.K.P. 1989. Height and gradient from shading. Technical Report 1105, MIT Artificial Intelligence Lab. MIT, Cambridge, MA 02139.

Horn, B.K.P. and Brooks, M.J. 1989. *Shape from Shading*. MIT Press: Cambridge, MA.

Horn, B.K.P., Szeliski R., and Yuille, A. 1993. Impossible shaded images. *IEEE Pat. Anal. Mach. Intell.*, 15(2):166–170.

Horn, B.K.P., Woodham, R.J., and Silver, W.M. 1978. Determining shape and reflectance using multiple images. Technical Report 490, Artificial Intelligence Lab. Memo. Massachusetts Institute of Technology, Cambridge, MA 02139.

Human Vision, Visual Processing and Digital Display V.

Jeffreys, H. 1961. *Theory of Probability*. Clarendon Press: Oxford.

Jepson, A.D. and Richards, W. 1992. What makes a good feature? In L. Harris and M. Jenkin (Eds.), *Spatial Vision in Humans and Robots*. Cambridge Univ. Press. See also MIT AI Memo. 1356 (1992).

Johnson, R.A. 1970. Asymptotic expansions associated with posterior distributions. *The Annals of Mathematical Statistics*, 41(3):851–864.

Kersten, D. 1991. Transparency and the cooperative computation of scene attributes. In M.S. Landy and J.A. Movshon (Eds.), *Computational Models of Visual Processing*. MIT Press: Cambridge, MA, Chapter 15.

Knill, D.C., Kersten, D., and Yuille, A. 1996. A Bayesian formulation of visual perception. In D. Knill and W. Richards (Eds.), *Perception as Bayesian Inference*. Cambridge University Press.

Koenderink, J.J. and van Doorn, A.J. 1979. The internal representation of solid shape with respect to vision. *Biol. Cybern.*, 32:211–216.

Laplace, P.S. 1812. *Theorie Analytique des Probabilites*. Courcier.

Leclerc, Y.G. 1989. Constructing simple stable descriptions for image partitioning. *Intl. J. Comp. Vis.*, 3:73–102.

Leclerc, Y.G. and Bobick, A.F. 1991. The direct computation of height from shading. In *Proc. IEEE CVPR*, Maui, Hawii, pp. 552–558.

Lee, C.-H. and Rosenfeld, A. 1989. Improved methods of estimating shape from shading using the light source coordinate system. In B.K.P. Horn and M.J. Brooks (Eds.), *Shape from Shading*. MIT Press: Cambridge, MA, Chapter 11.

Lindley, D.V. 1972. *Bayesian Statistics, A Review*. Society for Industrial and Applied Mathematics (SIAM).

Lowe, D.G. and Binford, T.O. 1985. The recovery of three-dimensional structure from image curves. *IEEE Pat. Anal. Mach. Intell.*, 7(3):320–326.

MacKay, D.J.C. 1992. Bayesian interpolation. *Neural Computation*, 4(3):415–447.

Malik, J. 1987. Interpreting line drawings of curved objects. *Intl. J. Comp. Vis.*, 1:73–103.

Nakayama, K. and Shimojo, S. 1992. Experiencing and perceiving visual surfaces. *Science*, 257:1357–1363.

Papoulis, A. 1984. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill: New York.

Pentland, A.P. 1984. Local shading analysis. *IEEE Pat. Anal. Mach. Intell.*, 6(2):170–187.

Pentland, A.P. 1990a. Photometric motion. In *Proceedings of 3rd International Conference on Computer Vision*.

Pentland, A.P. 1990b. Automatic extraction of deformable part models. *Intl. J. Comp. Vis.*, 4:107–126.

Pentland, T., Torre, V., and Koch, C. 1985. Computational vision and regularization theory. *Nature*, 317(26):114–139.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. 1992. *Numerical Recipes in C*. Cambridge University Press.

Richards, W.A., Koenderink J.J., and Hoffman, D.D. 1987. Inferring three-dimensional shapes from two-dimensional silhouettes. *J. Opt. Soc. Am. A*, 4(7):1168–1175.

Schreiber, W.F. 1986. *Fundamentals of Electronic Imaging Systems*. Springer Verlag.

Skilling, J. 1989. Classic maximum entropy. In J. Skilling (Ed.), *Maximum Entropy and Bayesian Methods*. Cambridge, Kluwer, pp. 45–52.

Szeliski, R. 1989. *Bayesian Modeling of Uncertainty in Low-Level Vision*, Kluwer Academic Publishers: Boston.

Terzopoulos, D. 1986. Regularization of inverse problems involving discontinuities. *IEEE Pat. Anal. Mach. Intell.*, 8(4):413–424.

Tikhonov, A.N. and Arsenin, V.Y. 1977. *Solutions of ll-Posed Problems*, Winston: Washington, DC.

Weinshall, D., Werman, M., and Tishby, N. 1994. Stability and likelihood of views of three dimensional objects. In *Proceedings of the 3rd European Conference on Computer Vision*, Stockholm, Sweden.

Witkin, A.P. 1981. Recovering surface shape and orientation from texture. *Artificial Intelligence*, 17:17–45.

Woodham, R.J. 1980. Photometric method for determing surface orientation from multiple images. *Optical Engineering*, 19(1):139–144.

Yuille, A.L. and Bulthoff, H.H. 1996. Bayesian decision theory and psychophysics. In D. Knill and W. Richards (Eds.), *Perception as Bayesian Inference*. Cambridge University Press.

Zheng, Q. and Chellapa, R. 1991. Estimation of illuminant direction, albedo, and shape from shading. *IEEE Pat. Anal. Mach. Intell.*, 13(7):680–702.