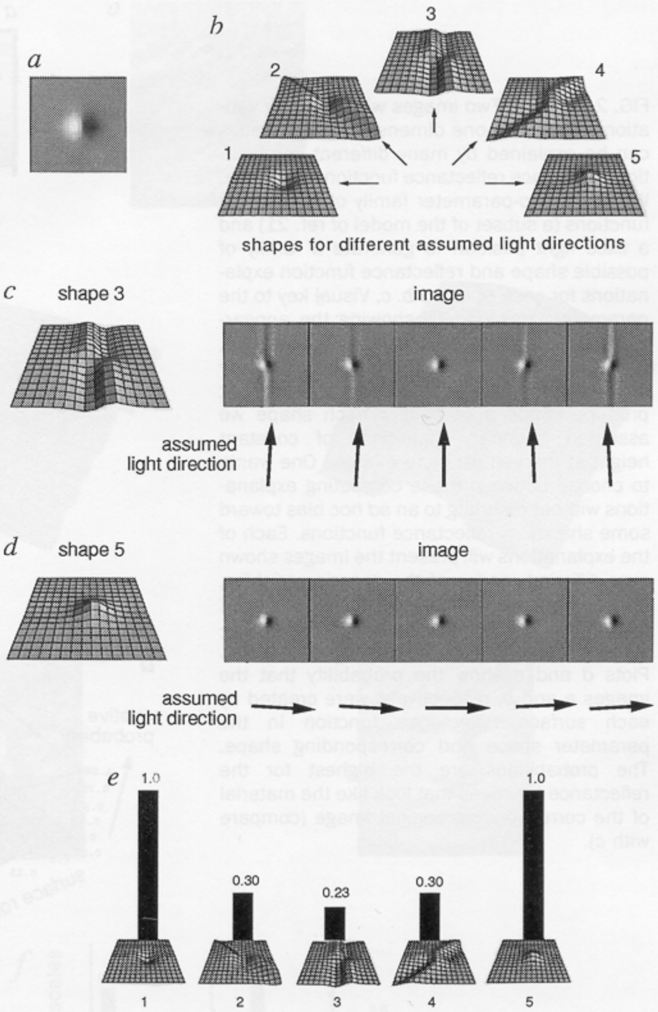


FIG. 1 *a*, Perceptually, this image has two possible interpretations. It could be a bump, lit from the left, or a dimple, lit from the right. *b*, Mathematically, there are many possibilities. The five shown here were found by a linear shape from shading algorithm, assuming shallow incident light from different azimuthal directions and the boundary conditions described in ref. 8. Shapes 2–4 require coincidental alignment with the assumed light direction. For shape 3 in *c*, the rendered image changes quickly with assumed light angle; only a small range of light angles yields an image like that shown in *a*. The generic view term of the scene probability equation, equation (7), penalizes an interpretation that has high image derivatives with respect to the generic variable, in this case light direction. For shape 5 in *d*, a much larger range of light angles gives the observed image. If all light directions are equally likely, shape 5 should be the preferred explanation. The probabilities of the candidate shapes, found using equation (7), are shown in *e*. The results favour shapes 1 and 5, in agreement with the perceptual appearance of *a*.

The generic viewpoint assumption in a framework for visual perception

William T. Freeman

Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, Massachusetts 02139, USA

A VISUAL system makes assumptions in order to interpret visual data. The assumption of 'generic view'[1–4] states that the observer is not in a special position relative to the scene. Researchers commonly use a binary decision of generic or accidental view to disqualify scene interpretations that assume accidental viewpoints[5–10]. Here we show how to use the generic view assumption, and others like it, to quantify the likelihood of a view, adding a new term to the probability of a given image interpretation. The resulting framework better models the visual world and reduces the reliance on other prior assumptions. It may lead to computer vision algorithms of greater power and accuracy, or to better models of human vision. We show applications to the problems of inferring shape, surface reflectance properties, and motion from images.

Consider the image of Fig. 1*a*. Perceptually, there are two possible interpretations: a bump, lit from the left, or a dimple, lit from the right. Yet many shapes and lighting directions (Fig. 1*b*) could explain the image. How should a visual system choose?

We note that the ridges in shapes 2–4 of Fig. 1*b* must line up with the assumed light direction. We can study the 'accidentalness' of this alignment by exploring how the image of the illuminated shape changes as we perturb the azimuthal light direction. Figure 1*c* shows that shape 3 presents images similar to that in Fig. 1*a* only for a small range of assumed light directions. The bump in Fig. 1*d* (shape 5) presents images like that in Fig. 1*a* over a broader range of light directions. If all azimuthal light directions are equally likely, shape 5 has more chances to create the image in Fig. 1*a* than does shape 3.

To quantify such probabilities, we use a bayesian framework (as in ref. 11, for example). This combines the data (Fig. 1*a*) with known or estimated prior probabilities to find the posterior probability of each candidate shape.

We treat the azimuthal light direction as a random variable, an example of what we call a generic variable, $\vec{x}$, with prior probability density $P_{\vec{x}}(\vec{x})$. (We use subscripts to distinguish

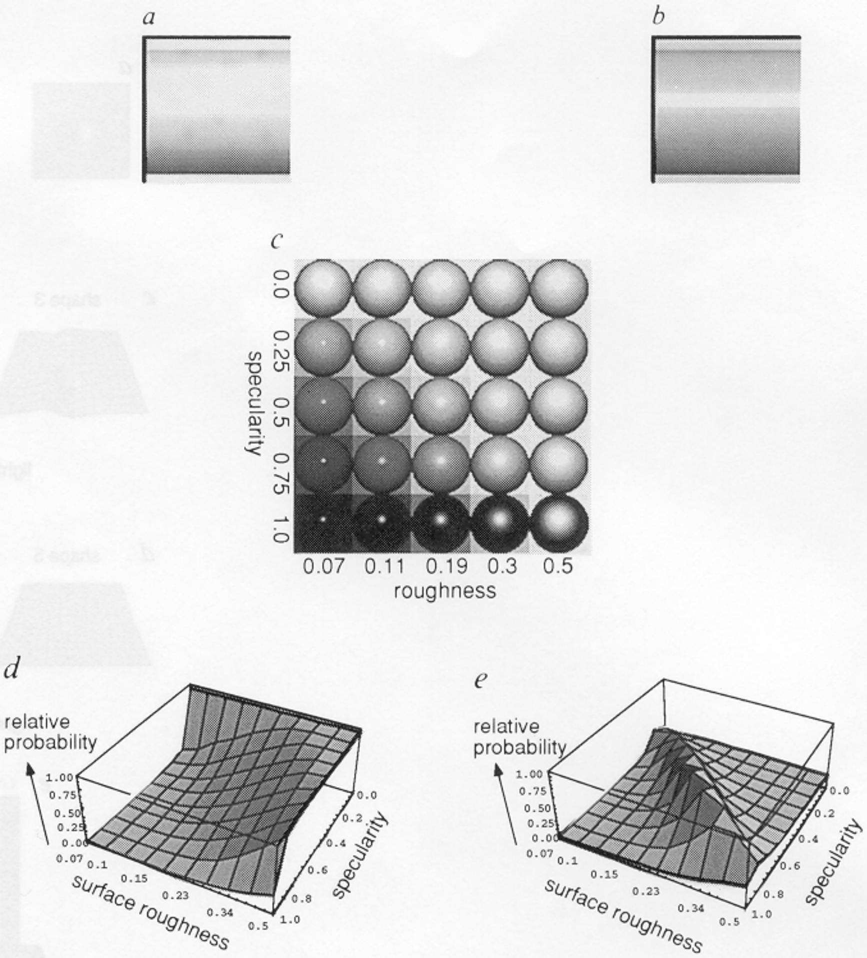


FIG. 2 *a* and *b*, Two images with intensity variations along only one dimension. Such images can be explained by many different combinations of surface reflectance function and shape. We use a two-parameter family of reflectance functions (a subset of the model of ref. 21) and a fixed light position to generate a family of possible shape and reflectance function explanations for each of *a* and *b*. *c*, Visual key to the parameters provided by showing the appearance of the surface reflectance functions, rendered on the surface of a sphere. For every specularity and roughness, shapes exist that produce image *a* or *b*. (For each shape we assumed boundary conditions of constant height at the vertical picture edge.) One wants to choose between these competing explanations without resorting to an *ad hoc* bias toward some shapes or reflectance functions. Each of the explanations will present the images shown over differing ranges of the generic variables, taken here to be vertical light angle and object orientation. The scene probability equation calculates their relative probability densities[22]. Plots *d* and *e* show the probability that the images *a* and *b*, respectively, were created by each surface reflectance function in the parameter space and corresponding shape. The probabilities are the highest for the reflectance functions that look like the material of the corresponding original image (compare with *c*).

between probability densities, P.) Generic variables can include viewpoint, lighting direction, or object pose. These are variables that we do not need to estimate precisely.

We assume a prior probability density, $P_{\vec{\beta}}(\vec{\beta})$, for the scene parameter $\vec{\beta}$ we want to estimate. For this example, shapes 1–5 are assigned equal probabilities. The posterior distribution, $P(\vec{\beta}, \vec{x} \mid \vec{y})$, gives the probability that scene parameter $\vec{\beta}$ (shape) and generic variable $\vec{x}$ (light direction) created the visual data $\vec{y}$ (Fig. 1*a*). From $P(\vec{\beta}, \vec{x} \mid \vec{y})$, we will find the posterior probability $P(\vec{\beta} \mid \vec{y})$.

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

$$P(\vec{\beta}, \vec{x} \mid \vec{y}) = \frac{P(\vec{y} \mid \vec{\beta}, \vec{x}) P_{\vec{\beta}}(\vec{\beta}) P_{\vec{x}}(\vec{x})}{P_{\vec{y}}(\vec{y})} \tag{1}$$

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

To find $P(\vec{\beta}, \vec{x} \mid \vec{y})$ independently of the value of the generic variable $\vec{x}$, we integrate the joint probability of equation (1) over the possible $\vec{x}$ values:

$$P(\vec{\beta} \mid \vec{y}) = \frac{P_{\vec{\beta}}(\vec{\beta})}{P_{\vec{y}}(\vec{y})} \int P(\vec{y} \mid \vec{\beta}, \vec{x}) P_{\vec{x}}(\vec{x})\, \mathrm{d}\vec{x} \tag{2}$$

We will assume that the prior probability $P_{\vec{x}}(\vec{x})$ of the generic variables is a constant. The generalization for other priors is straightforward. $P(\vec{y} \mid \vec{\beta}, \vec{x})$ is large where the scene $\vec{\beta}$ and the value $\vec{x}$ give an image similar to the observation $\vec{y}$. The integral of equation (2) integrates the area of $\vec{x}$ for which $\vec{\beta}$ yields the observation. In our example, it effectively counts the frames in Fig. 1*c* or *d*, where the rendered image is similar to the input data.

We assume zero mean gaussian observation noise of variance σ^2, which plays two roles. It measures the similarity between images as the probability that noise accounts for the differences. It can also model physical noise. For this noise model,

$$P(\vec{y} \mid \vec{\beta}, \vec{x}) = \frac{1}{(\sqrt{2\pi\sigma^2})^N} e^{-\|\vec{y} - \vec{f}(\vec{x}, \vec{\beta})\|^2 / 2\sigma^2} \tag{3}$$

where $\vec{f}(\vec{x}, \vec{\beta})$ is a known 'rendering function' which gives the image created by the generic and scene parameters $\vec{x}$ and $\vec{\beta}$, and N is the dimensionality of the visual data $\vec{y}$.

For the low noise limit, we can find an analytic approximation to the integral of equation (2). We expand $\vec{f}(\vec{x}, \vec{\beta})$ in equation (3) in a second-order Taylor series,

$$\vec{f}(\vec{x}, \vec{\beta}) \approx \vec{f}(\vec{x}_0, \vec{\beta}) + \sum_i \vec{f}'_i [\vec{x} - \vec{x}_0]_i + \tfrac{1}{2} \sum_{ij} [\vec{x} - \vec{x}_0]_i\, \vec{f}''_{ij} [\vec{x} - \vec{x}_0]_j \tag{4}$$

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

$$\vec{f}'_i = \left. \frac{\partial \vec{f}(\vec{x}, \vec{\beta})}{\partial x_i} \right|_{\vec{x} = \vec{x}_0} \tag{5}$$

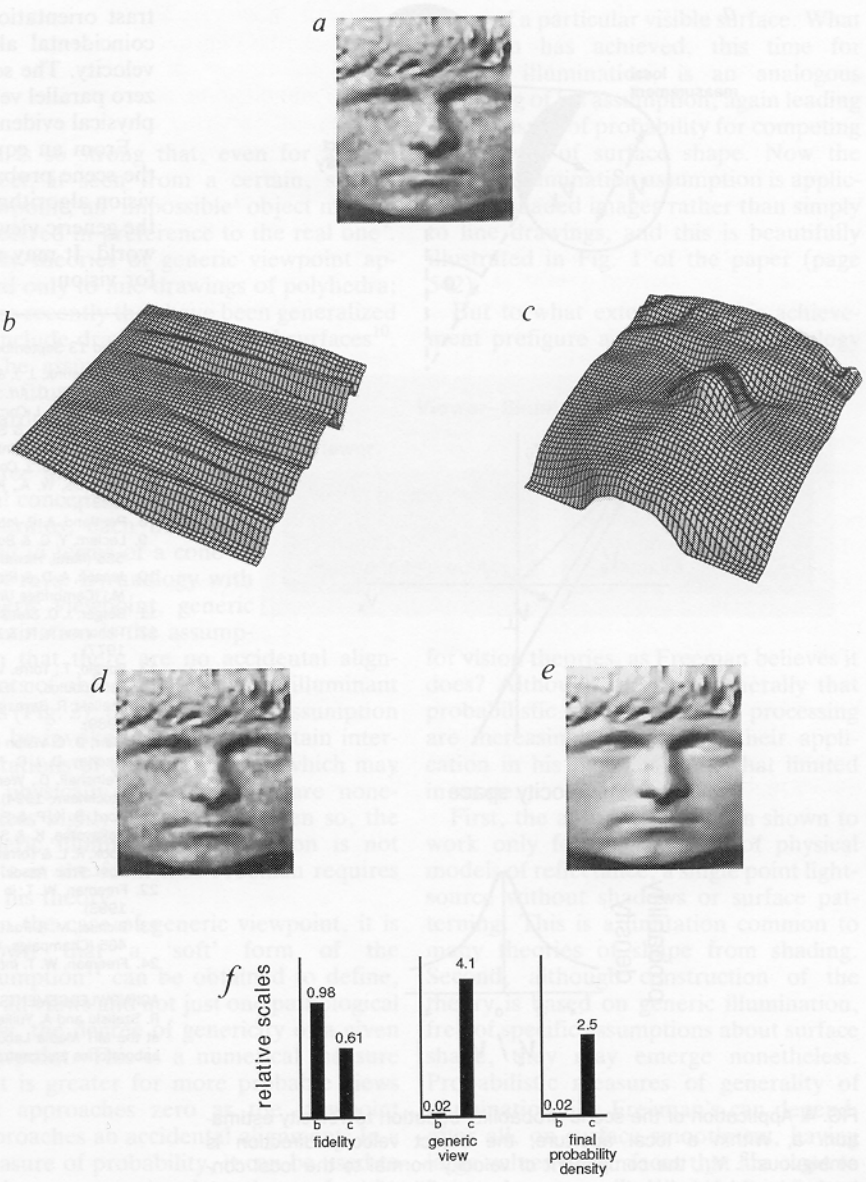
and

$$\vec{f}''_{ij} = \left. \frac{\partial^2 \vec{f}(\vec{x}, \vec{\beta})}{\partial x_i\, \partial x_j} \right|_{\vec{x} = \vec{x}_0} \tag{6}$$

We take $\vec{x}_0$ to be the value of $\vec{x}$ which can best account for the observed image data; that is, for which $\|\vec{y} - \vec{f}(\vec{x}, \vec{\beta})\|^2$ is minimized.

Using equations (3)–(6) to second order in $\vec{x} - \vec{x}_0$ in the integral of equation (2), we find the posterior probability for the

FIG. 3 Showing the need for the generic view term of equation (7). We compare the probability densities of two explanations for the image in *a*. The surface in *b* (shown at 7 × vertical exaggeration), lit at a grazing angle, yields the image in *d*. The surface in *c* gives the image in *e*, which accounts less well for the image in *a*. Thus, based on an image fidelity criterion, *b* is a better explanation. The common prior assumption of a smooth surface[14] would also favour *b* (the surface is very smooth at the true vertical scale). However, the object and light source must be precisely positioned for the shape in *b* to give the image in *d*; the generic view term of the scene probability equation (7) penalizes this. Including the generic view term makes the overall probability densities (shown in *f*), favour the perceptually reasonable explanation of shape *c* over shape *b*. (We made this example by construction. Gaussian random noise at a 7 dB signal-to-noise ratio was added to *e* to make *a*; *b* was found from *a* using a shape from shading algorithm, assuming constant surface height at the left picture edge[23]. We evaluated the likelihood of *b* and *c*, assuming both generic object pose and generic lighting direction[24]. The strength of a prior preference for smooth surfaces is arbitrary and none was included in the final densities. The actual noise variance was used for σ^2 in the fidelity term of equation (7), although a wide range of assumed variances would give the preferences shown here.)



scene parameters $\vec{\beta}$ given the visual data $\vec{y}$:

$$P(\vec{\beta} \mid \vec{y}) = k \exp\left(\frac{-\|\vec{y}-\vec{f}(\vec{x}_0, \vec{\beta})\|^2}{2\sigma^2}\right) P_{\vec{\beta}}(\vec{\beta}) \frac{1}{\sqrt{\det(\mathbf{C})}} \tag{7}$$

$$= k \text{ (fidelity) (prior probability) (generic view)}$$

where the i and jth elements of the matrix $\mathbf{C}$ are

$$C_{ij} = \vec{f}_i' \cdot \vec{f}_j' - (\vec{y} - \vec{f}(\vec{x}_0, \vec{\beta})) \cdot \vec{f}_{ij}'' \tag{8}$$

We call equation (7) the scene probability equation. The normalization constant k does not enter into comparisons between interpretations $\vec{\beta}$. The exponential term, which we call the image fidelity term, favours scene hypotheses that have a small mean-squared difference from the visual data. This and the prior probability term $P_{\vec{\beta}}(\vec{\beta})$ are familiar in computational vision. Regularization, from which many vision algorithms have been derived[12,13], finds the maximum probability density[14,15] using these two terms, when viewed in a bayesian context. The third, generic view term, accounts for the assumptions of generic viewpoint, pose or lighting position. The scene probability equation favours interpretations that can generate the observed image over a relatively large range of generic variables, by penalizing large image derivatives with respect to those variables. If the prior probability of the generic variable were not constant, then the factor $P_{\vec{x}}(\vec{x}_0)$ would be included in the prior term of equation (7).

The generic view term is especially useful when several explanations account equally well for visual data, as occurs commonly in problems of stereo, shape, motion and colour perception; ref. 16, for example. Then the image fidelity term is the same for the competing explanations. The prior probabilities may not be known well[4]. The generic view term allows a choice based on the relatively reliable assumptions of generic view, pose, or light source position.

Our approach relates to bayesian analyses of data interpolation, image restoration and other problems[11,15,17]. In that work, as in this, one favours hypotheses that could have generated the observed data in many ways (see also ref. 18, a related non-bayesian approach).

Using the scene probability equation (7), we plot in Fig. 1*e* the relative probabilities of shapes 1–5 of Fig. 1*b*. Note the agreement with the bump/dimple shapes perceived to be the true explanation of Fig. 1*a*. (Presumably, these are perceptually favoured because they are more probable.) Without the generic

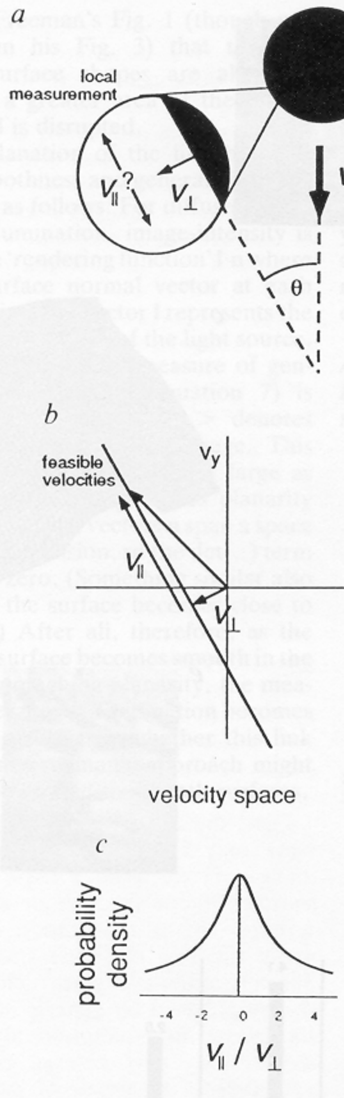


FIG. 4 Application of the scene probability equation to velocity estimation. *a*, Within a local aperture, the object velocity direction is ambiguous[19]. $V_\perp$, the component of velocity normal to the local contrast, is constrained by the measurement, while $V_\parallel$ is unconstrained. *b*, Line in velocity space of object velocities consistent with observed normal velocity. High values of $V_\parallel$ imply a concidental alignment of the local contrast orientation with the object velocity direction. In our framework, the measurement vector $\vec{y}$ is the normal velocity vector; the scene parameter $\vec{\beta}$ is $V_\parallel$; the generic variable $\vec{x}$ is the angle θ between the object velocity and the orientation of local contrast. The scene probability equation (7), penalizes high derivatives of the normal velocity with respect to contrast orientation. *c*, Resulting posterior probability for $V_\parallel$, showing a bias in favour of the normal velocity ($V_\parallel = 0$). This bias is consistent with psychophysical observations[20].

view term, one would have to state an arbitrary preference for bumps or dimples to choose between the candidate shapes.

In Fig. 2 we use the scene probability equation to choose between surface reflectance functions in a case where they would otherwise be indistinguishable. Figure 3 shows an example in which both the fidelity and the prior probability terms favour a perceptually implausible explanation. Only when the generic view term of equation (7) is included does the perceptually favoured explanation rank higher.

In Fig. 4, we apply the scene probability equation to the problem of estimating the local image velocity from local measurements of the velocity components normal to the contrast orientation[19]. All velocity components parallel to the local contrast orientation are possible, but high speeds would imply a coincidental alignment of the local contrast with the image velocity. The scene probability equation predicts a bias toward zero parallel velocity component, which is supported by psychophysical evidence[20].

From an equation that ranks scene interpretations, such as the scene probability equation (equation (7)), one can develop vision algorithms that find an optimum interpretation. Including the generic view term gives a better statistical model of the visual world. It may result in more powerful and accurate algorithms for vision. □

Received 13 September 1993; accepted 27 January 1994.

1. Koenderink, J. J. & van Doorn, A. J. *Biol. Cybern.* **32,** 211–216 (1979).
2. Binford, T. O. *Art. Intel.* **17,** 205–244 (1981).
3. Biederman, I. *Comp. Vis. Graph. Image Proc.* **32,** 29–73 (1985).
4. Nakayama, K. & Shimojo, S. *Science* **257,** 1357–1363 (1992).
5. Lowe, D. G. & Binford, T. O. *IEEE Pat. Anal. Mach. Intel.* **7,** 320–326 (1985).
6. Malik, J. *Intl. J. Comp. Vis.* **1,** 73–103 (1987).
7. Richards, W. A., Koenderink, J. J. & Hoffman, D. D. *J. Opt. Soc. Am. A.* **4,** 1168–1175 (1987).
8. Pentland, A. P. *Intl. J. Comp. Vis.* **1,** 153–162 (1990).
9. Leclerc, Y. G. & Bobick, A. F. in *Proc. IEEE Computer Vision and Pattern Recognition* 552–558 (Maui, Hawaii, 1991).
10. Jepson, A. D. & Richards, W. *Spatial Vision in Humans and Robots* (eds Harris, L. & Jenkin, M.) (Cambridge Univ. Press, UK, 1992).
11. Berger, J. O. *Statistical Decision Theory and Bayesian Analysis* (Springer, New York, 1985).
12. Tikhonov, A. N. & Arsenin, V. Y. *Solutions of Ill-posed Problems* (Winston, Washington DC, 1977).
13. Poggio, T., Torre, V. & Koch, C. *Nature* **317,** 314–319 (1985).
14. Terzopoulos, D. *IEEE Pat. Anal. Mach. Intell.* **8,** 413–424 (1986).
15. Szeliski, R. *Bayesian Modeling of Uncertainty in Low-level Vision* (Kluwer Academic, Boston, 1989).
16. Marr, D. C. *Vision* (Freeman, New York, 1982).
17. MacKay, D. J. C., *Neural Comput.* **4,** 415–447 (1992).
18. Weinshall, D., Werman, M. & Tishby, N. *Proc. 3rd Eur. Conf. Computer Vision* (Springer, Stockholm, 1994).
19. Horn, B. K. P. & Schunk, B. G. *Artif. Intel.* **17,** 185–203 (1981).
20. Nakayama, K. & Silverman, G. H. *Vision Res.* 739–746 (1988).
21. Cook, R. L. & Torrance, K. E. in *SIGGRAPH-81* (Special Interest Group on Computer Graphics Conf. Proc. Assoc. for Computing Machinery, New York, 1981).
22. Freeman, W. T. in *Proc. 4th Intl. Conf. Computer Vision* 347–356 (IEEE, Berlin, Germany, 1993).
23. Bichsel, M. & Pentland, A. P. in *Proc. IEEE Computer Vision and Pattern Recognition* 459–465 (Champaign, IL, 1992).
24. Freeman, W. T. *Intl. J. Comp. Vis.* (in the press).

ACKNOWLEDGEMENTS. I thank E. Adelson, D. Brainard, D. Knill, K. Nakayama, E. Simoncelli, R. Szeliski and A. Yuille for discussion and suggestions. Much of this research was performed at the MIT Media Laboratory and was supported by a contract from David Sarnoff Research Laboratories (subcontract to the National Information Display Laboratory) to E. Adelson.

VISION

Improbable views

Andrew Blake

THE problem of shape-from-shading analysis — estimating the relief of a visible surface from its pattern of light and shade in an image — has vexed researchers in vision for around 20 years[1]. From an experimental point of view, psychologists seem to disagree whether the human visual system can[2] or cannot[3] make effective use of shading information to estimate even the orientation of a surface, let alone its entire shape. Horn's[4] elegant demonstration of the reconstruction of a human face by growing height contours outwards from shading highlights showed that inferring shape from shading is theoretically possible, at least under certain conditions. The difficulty comes when, as is usually the case, some parameters of the physical environment are unknown, for example the directions and strengths of the sources of illumination. The parameters of a single light-source can be recovered if certain fairly strong assumptions (isotropy[5] or smoothness[6]) are made about the shape of the surface, but William Freeman (page 542 of this issue[7]) has now devised a theory founded on the far more general assumption of 'generic illumination'.

This assumption is akin to the concept of 'generic viewpoint' (Fig. 1), already familiar from computational and psychological theories of vision. A viewpoint is generic if it avoids accidental alignments, in the image, of features on a visible object. The assumption of generic viewpoint seems to underlie human visual analysis, explaining why certain line-drawings, for example the well-known Penrose triangle, are perceived as paradoxical[8] or 'impossible'. The assumption is so strong that, even for a solid object, if seen from a certain, special viewpoint, an 'impossible' object may be perceived in preference to the real one[9]. Early theories of generic viewpoint applied only to line-drawings of polyhedra; more recently they have been generalized to include drawings of curved surfaces[10].

The assumption of generic illumination can be regarded as the dual of generic viewpoint. The light source and the viewer are dual concepts, as the Table illustrates, each being defined in terms of a cone of light-rays. By analogy with generic viewpoint, generic illumination is the assumption that there are no accidental alignments of object features along illuminant rays (Fig. 2). In principle, this assumption can be invoked to rule out certain interpretations of a shaded image which may be physically consistent but are nonetheless highly improbable. Even so, the generic illumination assumption is not quite in the form that Freeman requires for his theory.

Viewer–illuminant duality

Viewer	Illuminant
Viewer position	Light-source position
Visible surface	Illuminated surface
Silhouette	Cast shadow
Blurred image outline	Penumbra

In the case of generic viewpoint, it is known that a 'soft' form of the assumption[11] can be obtained to define, for all views and not just one pathological view, the degree of genericity of a given viewpoint. This is a numerical measure that is greater for more probable views but approaches zero as the viewpoint approaches an accidental alignment. In a measure of probability, it can be used to evaluate competing hypotheses for the shape of a particular visible surface. What Freeman has achieved, this time for generic illumination, is an analogous softening of his assumption, again leading to a measure of probability for competing hypotheses of surface shape. Now the generic illumination assumption is applicable to shaded images rather than simply to line drawings, and this is beautifully illustrated in Fig. 1 of the paper (page 542).

But to what extent does this achievement prefigure a powerful methodology for vision theories, as Freeman believes it does? Although it is true generally that probabilistic models of visual processing are increasingly influential, their application in his paper is somewhat limited in scope.

First, the approach has been shown to work only for the simplest of physical models of reflectance, a single point light-source without shadows or surface patterning. This is a limitation common to many theories of shape from shading. Second, although construction of the theory is based on generic illumination, free of specific assumptions about surface shape, they may emerge nonetheless. Probabilistic measures of generality of illumination like Freeman's can depend, after all, on surface smoothness, having large values for surfaces that are close to being planar or cylindrical. Indeed it does

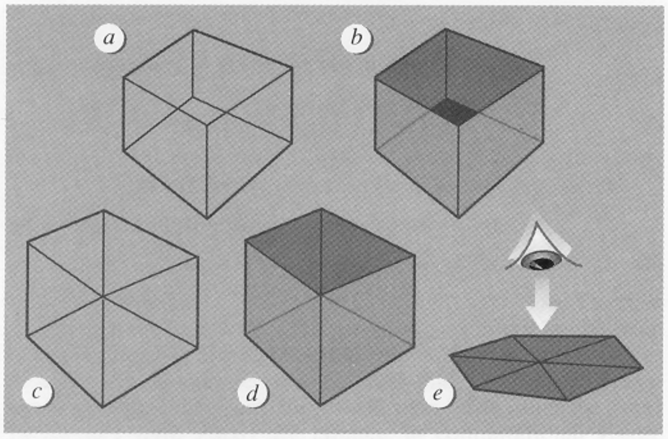


FIG. 1 The assumption of generic viewpoint. The drawing depicted in *a* may be interpreted as a line drawing of a three-dimensional object such as *b*. The drawing *c* could be a line-drawing of an object such as *d*. This interpretation implies, however, that the viewpoint is a special one in which two vertices have coincided in the image. A more plausible interpretation of *c* would therefore be one that involves no accidental alignment, for instance a planar shape such as that shown by *e*.

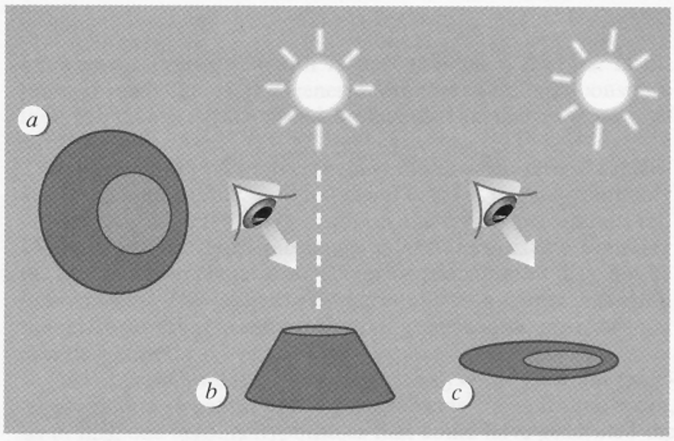


FIG. 2 Generic illumination. Just as the viewer can be in a special position with respect to an object, so can the source of illumination. The image *a* consists of two regions of constant-intensity shading. This could have arisen as the image of a cone *b* but only if the light-source is in a special position, on the axis of the cone. Any perturbation of the light-source's position away from the cone axis will lead to shaded regions whose intensity is no longer constant. A more likely explanation is that *a* is the image of a planar pattern such as *c*.

appear in Freeman's Fig. 1 (though not obviously in his Fig. 3) that the less plausible surface shapes are also less smooth — a greater area of the planar background is disrupted.

One explanation of the link between surface smoothness and generality of illumination is as follows. For diffuse ('Lambertian') illumination, image-intensity is given by the 'rendering function' $\mathbf{l}\cdot\mathbf{n}$ where $\mathbf{n}$ is the surface normal vector at each image position and vector $\mathbf{l}$ represents the strength and direction of the light source. In that case Freeman's measure of genericity of illumination (equation 7) is $1/\sqrt{\det(\langle \mathbf{n}\mathbf{n}^{T}\rangle)}$, where $\langle\_\rangle$ denotes spatial averaging over the image. This measure becomes unboundedly large as the visible surface approaches planarity because the normal vectors $\mathbf{n}$ span a space of only one dimension, so the det(..) term approaches zero. (Something similar also happens as the surface becomes close to cylindrical.) After all, therefore, as the underlying surface becomes smooth in the sense of approaching planarity, the measure of generality of illumination becomes large. The question is whether this link suggests that Freeman's approach might somehow be limited to smooth surfaces.

Overall, however, Freeman's specific achievement is impressive, both in seeing a way to express probabilistically the general illumination assumption and to some extent verifying experimentally its power to reject incorrect hypotheses about surface shape. The claim for greater generality is not altogether convincing as yet. It will be interesting to see whether computational experiments can establish more precisely the breadth of conditions over which the theory is viable. □

Andrew Blake is in the Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK.

1. Horn, B. K. P. & Brooks, M. J. (eds) *Shape from Shading* (MIT Press, Cambridge, Mass., 1989).
2. Bulthoff, H. H. & Mallot, H. A. *J. opt. Soc. Am.* **A5**, 1749–1758 (1988).
3. Mingolla, E. & Todd, J. T. *Biol. Cybern.* **53**, 137–151 (1986).
4. Horn, B. K. P. *Robot Vision* (McGraw-Hill, New York, 1986).
5. Pentland, A. P. *J. opt. Soc. Am.* **72**, 448–455 (1982).
6. Horn, B. K. P. & Brooks, M. J. *Comp. Vis. Graph. Im. Processing* **33**, 174–208 (1986).
7. Freeman, W. T. *Nature* **368**, 542–545 (1994).
8. Sugihara, K. *Machine Interpretation of Line Drawings* (MIT Press, Cambridge, Mass., 1986).
9. Gregory, R. L. *The Intelligent Eye* (World University, London, 1970).
10. Malik, J. *Intl J. Comp. Vis.* **1**, 73–103 (1987).
11. Witkin, A. P. *Artif. Intel.* **17**, 17–45 (1981).