MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

# The Generic Viewpoint Assumption in a Framework for Visual Perception

William T. Freeman

TR93-19a December 1993

**Abstract**

A visual system makes assumptions in order to assign interpretations of shape, lighting, or motion to visual data. The assumption of g̈eneric viewK̈oenderink79,Binford81,Biederman85,Nakayama92 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 special viewpoints Lowe85a,Malik87,Richards87,Pentland90,Leclerc91,Jepson92. Here we show how to use the generic view assumption to quantify the likelihood of a view, adding a new term to the probability of a given image interpretation. The resulting framework better models what the eye sees and reduces the reliance on other prior assumptions. It may lead to computer vision algorithms of greater power and accuracy or better models of human vision. We show applications to the problem of inferring shape from a shaded image.

*Nature, vol. 368, pp. 542–545, 1994*

1. First printing, MN93-19, September, 1993
2. Revised, as MN93-19a, January, 1994

A visual system makes assumptions in order to interpret visual data. The assumption of "generic view" [1, 2, 3, 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, 6, 7, 8, 9, 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 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 (b) could explain the image. How should a visual system choose?

We note that the ridges in shapes 2 – 4 of (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 (a) only for a small range of assumed light directions. The bump in (c) (shape 5) presents images like (a) over a broader range of light directions. If all azimuthal light directions are equally likely, shape 5 has more chances to create image (a) than does shape 3.

To quantify such probabilities, we use a Bayesian framework (e.g., [11]). 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 between probability densities, $P$). Generic variables can include viewpoint,

lighting direction, or object pose. These are variables which 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 probabilites.

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 $P(\vec{\beta}, \vec{x} \mid \vec{y})$, independent of the value of the generic variable $\vec{x}$, we integrate the joint probability of Eq. (1) over the possible $\vec{x}$ values:

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

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 similiar to the observation $\vec{y}$. The integral of Eq. (2) integrates the area of $\vec{x}$ for which $\vec{\beta}$ yields the observation. In our example, it effectively counts the frames in Figure 1 (c) or (d) where the rendered image is similar to the input data.

We assume zero mean Gaussian observation noise of variance $\sigma^2$, 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,

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

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 Eq. 2. We expand $\vec{f}(\vec{x}, \vec{\beta})$ in Eq. (3) in a second order Taylor series,

$$\vec{f}(\vec{x}, \vec{\beta}) \approx \vec{f}(\vec{x}_0, \vec{\beta}) + \boldsymbol{A}(\vec{x} - \vec{x}_0) + \frac{1}{2}(\vec{x} - \vec{x}_0)^T \boldsymbol{B}(\vec{x} - \vec{x}_0), \tag{4}$$

where the $i$ and $j$th elements of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are:

$$A_{ij} = \frac{\partial f_j(\vec{x}, \vec{\beta})}{\partial x_i} \mid_{\vec{x}=\vec{x}_0}, \tag{5}$$

and

$$B_{ij} = \frac{\partial^2 \vec{f}(\vec{x}, \vec{\beta})}{\partial x_i \partial x_j} \mid_{\vec{x}=\vec{x}_0}. \tag{6}$$

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

Using Eqs. (3)–(6) to second order in $\vec{x} - \vec{x}_0$ in the integral of Eq. (2), we find the posterior probability for the scene parameters $\vec{\beta}$ given the visual data $\vec{y}$:

$$\begin{aligned} P(\vec{\beta} \mid \vec{y}) &= k \ \exp\left(\frac{-\|\vec{y} - \vec{f}(\vec{x}_0, \vec{\beta})\|^2}{2\sigma^2}\right) \quad P_{\vec{\beta}}(\vec{\beta}) \quad \frac{1}{\sqrt{\det(\boldsymbol{C})}} \\ &= \qquad \text{(fidelity)} \quad \text{(prior probability)} \quad \text{(generic view)}, \end{aligned} \tag{7}$$

where the $i$ and $j$th elements of the matrix $\boldsymbol{C}$ are

$$C_{ij} = (\boldsymbol{A}^T \boldsymbol{A})_{ij} - (\vec{y} - \vec{f}(\vec{x}_0, \vec{\beta})) \cdot B_{ij}. \tag{8}$$

We call Eq. (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, favors scene hypotheses which 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 favors interpretations which can generate the observed image over a relatively large range of generic variables, by penalizing high 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 Eq. (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 color perception (e.g. [16]). 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 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 favors hypotheses which could have generated the observed data many ways. See also [18], a related non-Bayesian approach.

Using the scene probability equation, Eq. (7), we plot in Fig. 1 (e) the relative probability of shapes $1-5$ of (b). Note the agreement with the bump/dimple shapes perceived to be the true explanation of (a). (Presumably, these are perceptually favored because they are more probable). Without the generic view term, one would have to state an arbitrary preference for bumps or dimples to choose between the candidate shapes.

In Figure 2 we use the scene probability equation to choose between surface reflectance

Figure 3 shows an example where both the fidelity and the prior probability terms favor a perceptually implausible explanation. Only when the generic view term of Eq. (7) is included does the perceptually favored explanation rank higher.

In Figure 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 which ranks scene interpretations, such as the scene probability equation, Eq. (7), one can develop vision algorithms which 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.

## Acknowledgements

For helpful discussions and suggestions, thanks to: E. Adelson, D. Knill, K. Nakayama, E. Simoncelli, and R. Szeliski. 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.

# Figure Captions

## Figure 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 [8]. Shapes 2 – 4 require coincidental alignment with the assumed light direction. For shape 3, (c), the rendered image changes quickly with assumed light angle; only a small range of light angles yields an image like (a). The generic view term of the scene probability equation, Eq. (7), penalizes an interpretation which has high image derivatives with respect to the generic variable, in this case light direction. For shape 5, (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 Eq. (7), are shown in (e). The results favor shapes 1 and 5, in agreement with the perceptual appearance of (a).

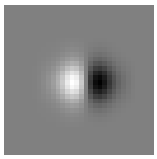
(a) and (b) show two images with intensity variations along only one dimension. Such images can be explained by many different combinations of surface reflectance function and shape. We use a two-parameter family of reflectance functions (a subset of the model of [21]) and a fixed light position to generate a family of possible shape and reflectance function explanations for each of (a) and (b). (c) provides a visual key to the parameters by showing the appearance of the surface reflectance functions, rendered on the surface of a sphere. For every specularity and roughness, shapes exist which 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 light angle and object orientation. The scene probability equation calculates their relative probabilities [22]. The 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 which look like the material of the corresponding original image (compare with (c)).

**Figure 3**

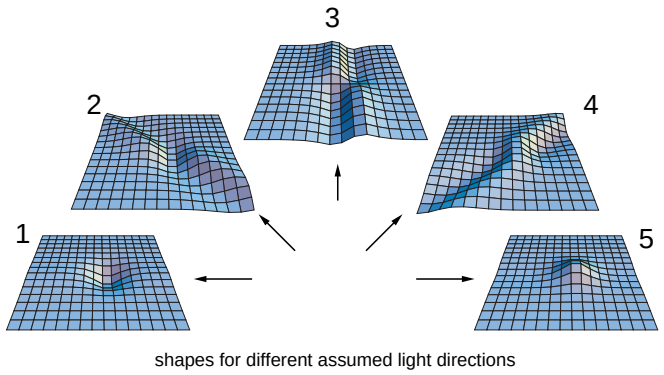
Showing the need for the generic view term of Eq. (7). We compare the probability densities of two explanations for the image (a). The surface (b) (shown at 7x vertical exaggeration), lit at a grazing angle, yields the image (d). The surface (c) gives the image (e), which accounts less well for the image (a). Thus, based on an image fidelity criterion, (b) is a better explanation. The common prior assumption of a smooth surface [14] would also favor (b) (the surface is very smooth at the true vertical scale). However, the object and light source must be precisely positioned for the shape (b) to give the image (d); the generic view term of the scene probability equation, Eq. (7), penalizes this. Including the generic view term makes the overall probability densities, shown in (f), favor 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. 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 $\sigma^2$ in the fidelity term of Eq. (7), although a wide range of assumed variances would give the results shown here).

**Figure 4**

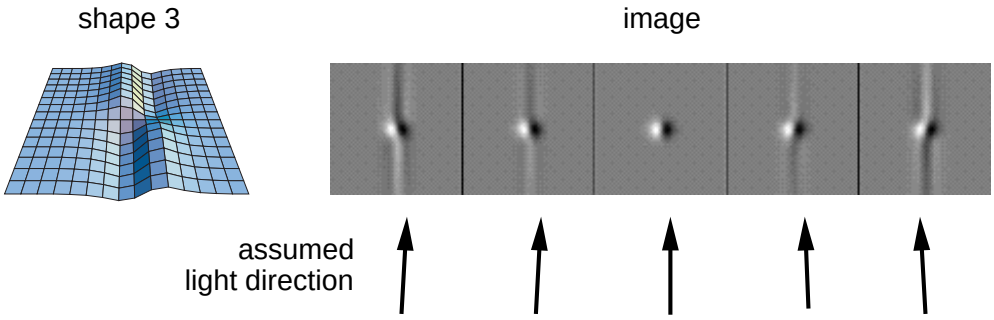
Application of the scene probability equation to velocity estimation. (a) Within a local aperture, the object velocity direction is ambiguous [19]. $V_\perp$, the component of velocity normal to the local contrast, is constrained by the measurement, while $V_\parallel$ is unconstrained. (b) Line in velocity 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 $\theta$ between the object velocity and the orientation of local contrast. The scene probability equation, Eq. (7), penalizes high derivatives of the normal velocity with respect to contrast orientation. (c) shows the resulting posterior probability for $V_\parallel$, showing a bias in favor of the normal velocity ($V_\parallel = 0$). This bias is consistent with psychophysical observations [20].



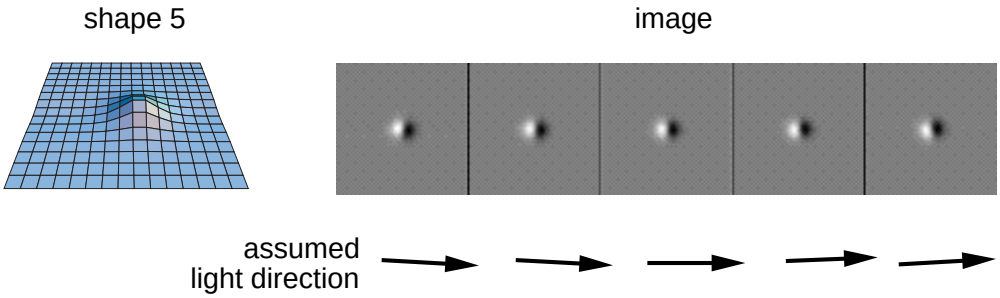
(a)



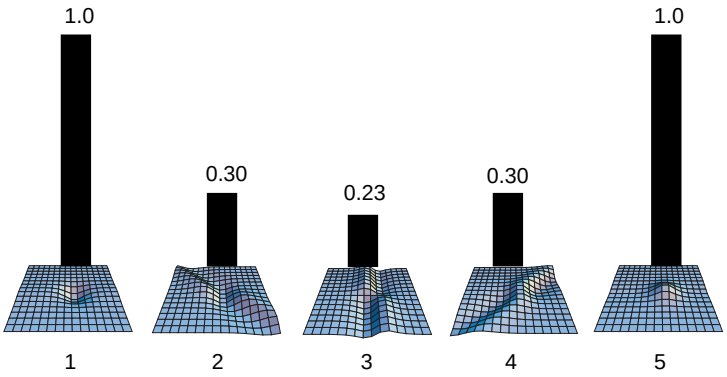

(b)




(c)




(d)

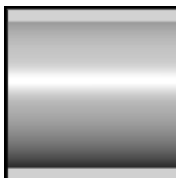



(e)

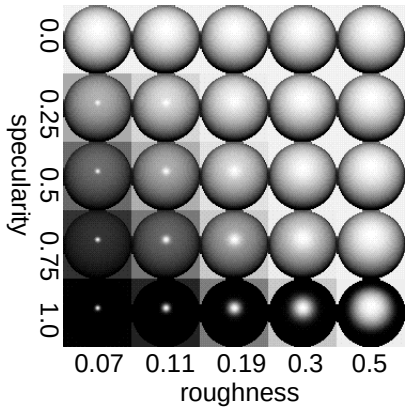


**Figure 1:**

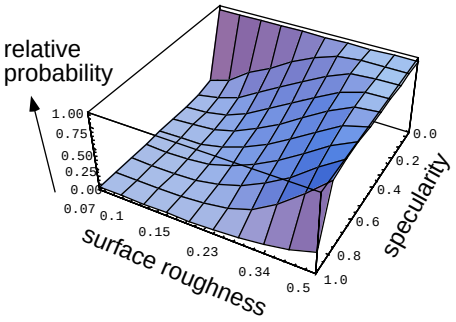
(a)



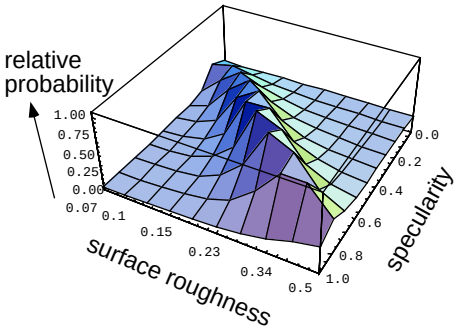
(b)




(c)



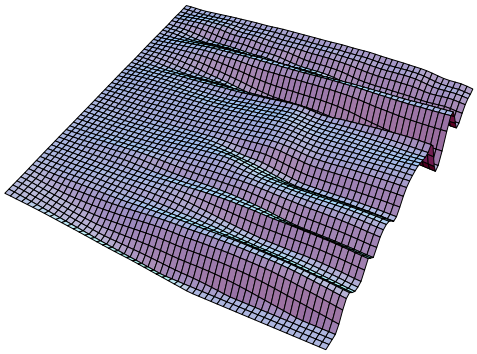

(d)



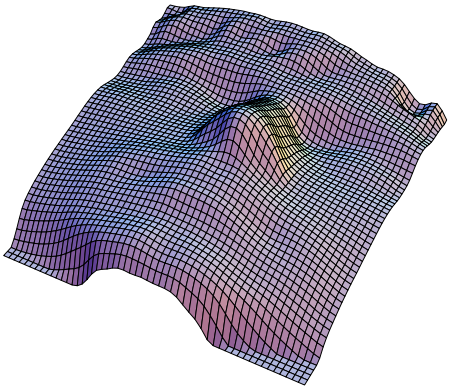

(e)

**Figure 2:**

(a)



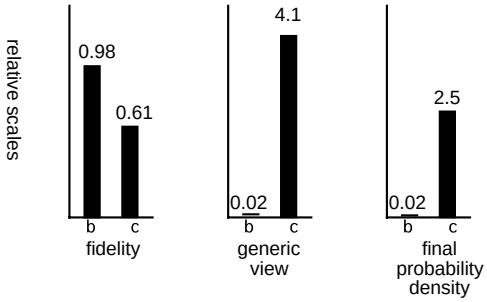
(b)



(c)

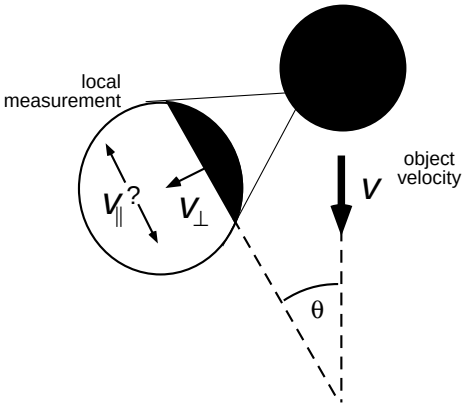
(d)

(e)

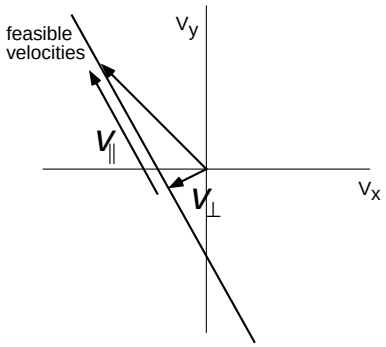



(f)

**Figure 3:**

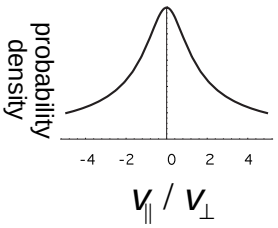



(a)




velocity space

(b)




(c)

**Figure 4:**

# References


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

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

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

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

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

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

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

[8] A. P. Pentland. Linear shape from shading. *Intl. J. Comp. Vis.*, 1(4):153–162, 1990.

[9] Y. G. Leclerc and A. F. Bobick. The direct computation of height from shading. In *Proc. IEEE CVPR*, pages 552–558, Maui, Hawaii, 1991.

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

[11] J. O. Berger. *Statistical decision theory and Bayesian analysis*. Springer-Verlag, 1985.

[12] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-posed Problems*. Winston, Washington, DC, 1977.

[13] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317(26):314–139, 1985.

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

[15] R. Szeliski. *Bayesian Modeling of Uncertainty in Low-level Vision*. Kluwer Academic Publishers, Boston, 1989.

[16] D. C. Marr. *Vision*. W H Freeman and Company, 1982.

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

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

[19] B. K. P. Horn and B. G. Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[20] K. Nakayama and G. H. Silverman. The aperture problem–I. perception of nonrigidity and motion direction in translating sinusoidal lines. *Vision Research*, pages 739–746, 1999.

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

[22] W. T. Freeman. Exploiting the generic view assumption to estimate scene parameters. In *Proc. 4th Intl. Conf. Computer Vision*, pages 347 – 356, Berlin, Germany, 1993. IEEE.

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