



# Mixture models and the probabilistic structure of depth cues

David C. Knill

*Center for Visual Sciences, University of Rochester, 274 Meliora Hall, Rochester, NY 14627, USA*

Received 1 March 2002; received in revised form 19 September 2002

---

## Abstract

Monocular cues to depth derive their informativeness from a combination of perspective projection and prior constraints on the way scenes in the world are structured. For many cues, the appropriate priors are best described as mixture models, each of which characterizes a different category of objects, surfaces, or scenes. This paper provides a Bayesian analysis of the resulting model selection problem, showing how the mixed structure of priors creates the potential for non-linear, cooperative interactions between cues and how the information provided by a single cue can effectively determine the appropriate constraint to apply to a given image. The analysis also leads to a number of psychophysically testable predictions. We test these predictions by applying the framework to the problem of perceiving planar surface orientation from texture. A number of psychophysical experiments are described that show that the visual system is biased to interpret textures as isotropic, but that when sufficient image data is available, the system effectively turns off the isotropy constraint and interprets texture information using only a homogeneity assumption. Human performance is qualitatively similar to an optimal estimator that assumes a mixed prior on surface textures—some proportion being isotropic and homogeneous and some proportion being anisotropic and homogeneous.

© 2003 Elsevier Science Ltd. All rights reserved.

---

## 1. Introduction

Prior knowledge of statistical regularities in the environment allows the visual system to accurately estimate the three-dimensional layout of surfaces in a scene even in images with seemingly impoverished information. Specific models of this type of knowledge, in the form of “apriori” constraints, play a major role in computational theories of how the visual system estimates three-dimensional surface shape from a variety of cues. Examples include motion (rigidity (Ullman, 1979)), surface contours (isotropy (Brady & Yuille, 1984), symmetry (Kanade, 1981), lines of curvature (Stevens, 1981), geodesics (Knill, 1992)), shape from shading (lambertian reflectance, point light source (Ikeuchi & Horn, 1981)) and texture (homogeneity (Garding, 1992; Malik & Rosenholtz, 1995), isotropy (Blake & Marinos, 1989; Garding, 1995; Witkin, 1981)). Computational theories typically build on a single prior constraint; however, most cues admit multiple plausible prior models, each one of which accurately describes a limited class of objects, scenes or physical processes.

This gives rise to a problem of “model selection”—which prior constraint should be used to interpret a visual cue?

Yuille and Bulthoff introduced the problem of model selection in perception with their notion of competitive priors (Yuille & Bulthoff, 1996); however, since then it has remained an under-appreciated problem in understanding human three-dimensional perception. In some instances, specific categorical modes may lead to easily detected, diagnostic features in the image (Jepson, Richards, & Knill, 1996; Richards, Jepson, & Feldman, 1996) (e.g., parallelness); however, perspective distortion and noise in the image often renders independent detection of the features difficult. Reliable performance often requires that the visual system select prior models and estimate 3D surface layout cooperatively and simultaneously.

The current paper describes a Bayesian framework for cooperative model selection and estimation and analyzes the qualitative features of the problem that generate meaningful and psychophysically tractable questions about human perceptual performance. The first part of the paper develops the framework and analyzes its implications for perceptual performance. The analysis results in specific predictions relating the

---

*E-mail address:* [knill@cvs.rochester.edu](mailto:knill@cvs.rochester.edu) (D.C. Knill).

uncertainty in image information to human ability to selectively apply prior scene constraints. The second part of the paper demonstrates the usefulness of the framework by testing some of these predictions applied to a particular problem, perceiving planar surface orientation from texture.

## 2. Computational theory

In a Bayesian framework, the information provided by a set of image data about a scene is represented by a conditional probability density function,  $p(\vec{S}|\vec{I})$ , where  $\vec{S}$  represents the scene parameters being estimated and  $\vec{I}$  represents the available image data. The current analysis looks at how the modal (categorical) properties of the environment structure the posterior density function and how this impacts the problem of estimating scene properties from image data. The first subsection characterizes the estimation problem and briefly reviews different approaches to estimation in the presence of multi-modal structure. The third subsection describes the computational principles, based on a Bayesian form of Occam's razor, that allow the visual system to reliably solve the model selection problem when only one cue is available. The second subsection analyzes in more detail the implications of multi-modal cue structure for cue integration and shows how optimal estimation naturally leads to a number of non-linear cue integration strategies such as cue vetoing (Landy, Maloney, Johnston, & Young, 1995). The final subsection summarizes the implications of the analysis for psychophysical models of human perceptual performance.

### 2.1. Estimation with mixture models—basics

The posterior conditional density function,  $p(\vec{S}|\vec{I})$ , may be computed from a model of the image formation process and a model of the structure of the environment using Bayes' rule,

$$p(\vec{S}|\vec{I}) = \frac{p(\vec{I}|\vec{S})p(\vec{S})}{p(\vec{I})}, \quad (1)$$

where  $\vec{S}$  is a vector of parameters describing those aspects of a scene being estimated and  $\vec{I}$  is a set of image data. Assuming a flat prior on  $\vec{S}$ , the likelihood function fully characterizes the information content of the image data. For a number of reasons, the true likelihood function for a problem is often a mixture of model likelihood functions for the different generative processes that could have given rise to the image data,

$$p(\vec{I}|\vec{S}) = \sum_{i=1}^n \phi_i p_i(\vec{I}|\vec{S}), \quad (2)$$

where  $\phi_i$  is the prior probability of model  $i$  and  $p_i(\vec{I}|\vec{S})$  is the corresponding likelihood function. We will refer to likelihood functions of this type as *mixed* likelihood functions. An example of such a problem is shape-from-shading. Shading patterns depend on both the shapes of surfaces and their material properties. Since different materials reflect light in qualitatively different ways, a complete likelihood function for shape-from-shading should include component models for different classes of materials (metallic, plastic, matte, etc.).

Mixed likelihood functions have a characteristic, multi-modal structure, as illustrated in Fig. 1. Fig. 1(b) shows a likelihood function that presents a somewhat

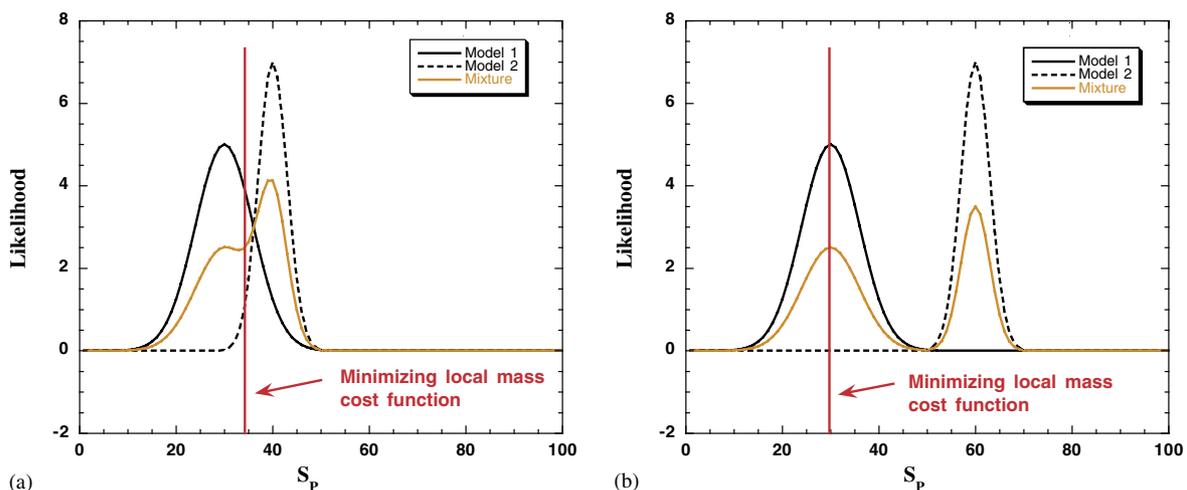


Fig. 1. Two examples of mixed likelihood functions. In (a) two models are consistent with overlapping interpretations of a scene, while in (b) they are mutually inconsistent. An estimator that minimizes a local mass cost function (Freeman, 1996) (a function that is quadratic for small errors, but remains constant for errors above a threshold value) selects the global mean when the model likelihood functions are close to one another, but selects the mean of the most likely model when they are far apart. Another approach would be to first select the most likely model and then apply an estimator of choice to that likelihood function (Mackay, 1992).

intractable problem to an estimator. Clearly, such a likelihood function does not support reliable scene estimates and would give rise to a significant number of gross errors, or bistable percepts. The rarity of such errors or of bistable perceptual modes in natural vision suggests that visual information typically suffices to disambiguate the model selection problem; that is, to squash one or another of the modes present in the likelihood function. The next two sections describe features of visual information that serve this function. The first involves the inherent selectivity within a single cue that results from an Occam's razor type effect. Mixed likelihood functions are inherently biased toward interpretations that are consistent with simpler models. The image data associated with a cue often leads to a mixed likelihood function with one mode substantially dominating others. When this is not true, other cues in the image data can enhance one mode while suppressing the other, effectively disambiguating the information from the first cue.

## 2.2. Hidden priors, nested models and Occam's razor

### 2.2.1. Basic principles

The categorical structure of the environment creates the problem that most image cues can be interpreted according to one of many different prior constraints. This is a principle source of the multi-modality discussed in the previous section. Yet, the visual system seems to reliably determine which prior constraints to use for cue interpretation even when no other cues are available in an image to disambiguate the choice. How is it possible for the image data associated with a single cue to disambiguate which prior constraints to use for interpreting the cue? A particular problem is posed by the fact that many plausible constraints are nested within one another; that is, one constraint is simply a more limiting form of another. Examples include matte surfaces being a more constrained subclass of specular surfaces (with the specular coefficient set to 0), rigid motion being a subclass of elastic motion and isotropic textures being a subclass of homogeneous textures. In these cases, the most general of a set of nested constraints will necessarily appear to fit the image data better, in the sense that fitting all free parameters of the general model will give a higher absolute likelihood than the free parameters of a more constrained model. How, then, does visual information allow one to select one model over another. Clearly, some form of Occam's razor needs to be applied—one would like to give preferential weighting to simpler models. The analysis in this section builds on prior work by Mackay (1992) showing how an Occam's razor type of effect falls naturally out of the probability calculus. In the current context, the effect derives from a process known as marginalization (see

also Freeman (1996) for an application of this technique to generalize the notion of generic views).

The first step in our chain of logic is to note that not all scene variables are equally important to an observer. When estimating surface shape from shading, for example, observers may be interested in the shape of a surface, but not in its reflectance or in the lighting conditions, though all of these contribute to the shading pattern in an image. Thus, one can decompose a scene (or surface) parameterization into two sub-vectors,  $\vec{S} = [\vec{S}_p, \vec{S}_s]$ , the first representing the *primary* variables that an observer estimates, and the second representing *secondary* variables, which the observer does not care about. The posterior density function that characterizes the estimation problem posed to an observer is the marginal posterior only on the primary variables,  $p(\vec{S}_p|\vec{I})$ . In order to derive this from the full posterior on all scene parameters, one must marginalize, or integrate, over the secondary variables (a process that statisticians use to deal with nuisance parameters (Ripley, 1996)). This is equivalent to marginalizing the likelihood function over the secondary variables, giving (assuming that the primary and secondary variables are independent)

$$p(\vec{S}_p|\vec{I}) = \frac{p(\vec{I}|\vec{S}_p)\mathcal{P}(\vec{S})}{p(\vec{I})} \quad (3)$$

$$= \frac{\left[ \int_{\vec{S}_s} p(\vec{I}|\vec{S}_p, \vec{S}_s)\mathcal{P}(\vec{S}_s) d\vec{S}_s \right] p(\vec{S}_p)}{p(\vec{I})}. \quad (4)$$

The prior constraints that make image cues to 3D surface geometry informative are often on secondary scene variables. Because they appear within the integral, we will refer to them as “hidden” priors. Table 1 lists some examples. The modal structure of cues derives from the existence of multiple, distinct priors on the secondary variables (what Yuille and Clark refer to as competitive priors (Yuille & Clark, 1993)). In particular, the likelihood function becomes

$$p(\vec{I}|\vec{S}_p) = \int_{\vec{S}_s} p(\vec{I}|\vec{S}_p, \vec{S}_s)\mathcal{P}(\vec{S}_s) d\vec{S}_s \quad (5)$$

$$= \sum_{i=1}^n \phi_i \int_{\vec{S}_s} p(\vec{I}|\vec{S}_p, \vec{S}_s)p_i(\vec{S}_s) d\vec{S}_s, \quad (6)$$

where  $p_i(\vec{S}_s)$  is the prior on the secondary variables for model  $i$ . The integrals inside the summation are the marginalized likelihood functions for each of the prior models on the secondary variables. Integrating over different priors gives model likelihood functions with different spreads and heights. For a given model, the integral is computed only over those model parameters that are free to vary within the model. Thus, the likelihood function for a model with a large number of free parameters is computed by integrating over a larger number of free variables than a more constrained model (e.g., specular vs. matte reflectance models). This gives

Table 1

Examples of 3D cues to surface shape and pose with associated secondary variables and a few of the constraints on those secondary variables that have been proposed in the computational literature for interpreting the cues

Image cue	Secondary variables	Hidden constraints
Motion	Object and observer motion	Rigid motion (Ullman, 1979) Affine motion (Koederink & van Doorn, 1991) Elastic motion (Aggarwal, Cai, Liao, & Sabata, 1998)
Shading	Reflectance and lighting	Lambertian reflectance (Ikeuchi & Horn, 1981) Linear reflectance (Pentland, 1990) General reflectance function (Bakshi & Yang, 1997) Point light source (Ikeuchi & Horn, 1981) Hemispheric light source (Langer & Zucker, 1994)
Contour	Shapes of surface curves	Symmetric curves (Kanade, 1981) Lines of curvature (Stevens, 1981) Geodesics (Knill, 1992; Stevens, 1981) Planar cuts (Horn & Brady, 1988) Isotropic curves (Brady & Yuille, 1984; Weiss, 1988)
Texture	Surface texture properties	Homogeneous textures (Garding, 1992; Malik & Rosenholtz, 1995) Isotropic textures (Blake & Marinos, 1989; Garding, 1995; Witkin, 1981) Follows lines of curvature (Li & Zaidi, 2000) (for texture flow)

rise to an implicit Occam's razor effect that favors simpler models over more complex ones (Mackay, 1992).

To illustrate the Occam's razor effect, we consider a simple example in which both the primary and secondary variables are scalar quantities. Returning to the earlier shape-from-shading example, we will assume the primary variable,  $S_p$ , is surface curvature and the secondary variable,  $S_s$ , is a parameter specifying the shininess of the surface, which we will let vary between  $-1$  and  $1$ , with  $0$  indicative of a matte surface. Matte surfaces are thus a special case of metallic surfaces. Suppose that the world consists of two types of surfaces, matte and metallic, with some non-zero proportion of surfaces belonging to each class. The likelihood function for curvature under the constrained, matte model is equal to the joint likelihood function on curvature and specularly with the specular parameter fixed at  $0$ . The likelihood function for curvature under the metallic model, on the other hand, is the integral of the joint likelihood function over the full range of possible values for the specular parameter.

Fig. 2 illustrates the difference in how the two likelihood functions are calculated. The integral used to compute the likelihood function for the metallic model computes the *average* of the joint likelihood function for each value of the specular parameter (assuming a flat prior on the parameter). Because the likelihood is low for most values of the specular parameter, this tends to shrink the marginal likelihood. The likelihood function for the matte model, on the other hand, is simply a slice through the joint likelihood function and is not penalized by this averaging process.

In Fig. 2, the joint likelihood function on  $S_p$  and  $S_s$  has a peak near  $S_s = 0$ , which matches the matte constraint. Accordingly, the matte model has a much higher

peak likelihood. Fig. 3 illustrates a case in which the peak of the joint likelihood function is at a very different value of  $\vec{S}_s$ . In this case, the  $S_s = 0$  line intersects the joint likelihood function in a low probability region of the parameter space, making the metallic model the more likely of the two models.

The implication of the analysis is that optimal estimators will exhibit natural biases toward more constrained prior models. Consider, for example, an estimator that selects the mean of the posterior distribution as its estimate of a set of scene properties. Such an estimator selects as its estimate a weighted average of the means of the likelihood functions associated with the component models,

$$\hat{\vec{S}}_p = \sum_{i=1}^n w_i \hat{\vec{S}}_{p_i}, \quad (7)$$

where  $\hat{\vec{S}}_{p_i}$  is the mean of the likelihood function for the  $i$ th model. The weights,  $w_i$  are given by

$$w_i = \frac{\pi_i \mathcal{I}_i}{\sum_{i=1}^n \pi_i \mathcal{I}_i}, \quad (8)$$

where  $\pi_i$  is the prior probability on model  $i$  and  $\mathcal{I}_i$  is the integral of the likelihood function for model  $i$  (the likelihood that model  $i$  is true given the data)

$$\mathcal{I}_i = \int_{\vec{S}_p} p_i(I|\vec{S}_p) p(\vec{S}_p) d\vec{S}_p. \quad (9)$$

A comparison of Figs. 2 and 3 reveals that the weights in the average are not constant, but vary as the peak of the likelihood function shifts away from an interpretation consistent with the constrained model. Since, on average, the peak of the likelihood function will be near the true value of the scene parameters, this implies that, on average, as a scene deviates more and

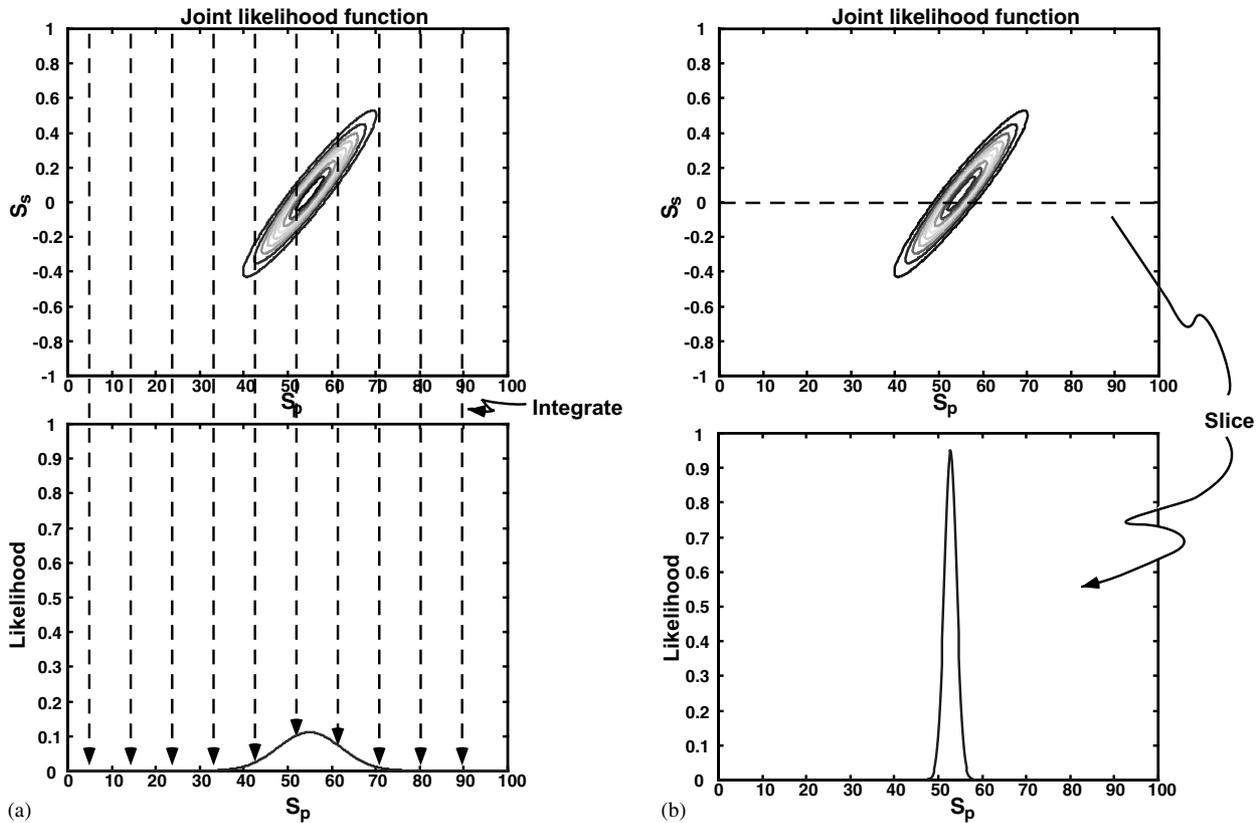


Fig. 2. (a)  $S_p$  and  $S_s$  are highly correlated. Marginalizing over  $S_s$  amounts to calculating for each value of  $S_p$  the average likelihood over all possible values of  $S_s$  (shown in bottom left panel). (b) The matte surface model has a built-in assumption that  $S_s = 0$ ; thus, the likelihood on  $S_p$  for this model is simply the slice through the joint likelihood function at  $S_s = 0$ . Since this slices the joint likelihood function near the peak, the model likelihood function has a much higher peak than the likelihood function for the unconstrained metallic surface model shown in (a). It also has significantly lower variance, a common property of more constrained models.

more from a constrained model, the weight given to the constrained model will shrink. In particular, for nested Gaussian models, the problematic case illustrated by Fig. 1(b), in which the likelihood functions for unconstrained and constrained models lead to a strongly bimodal distribution, often will not arise. When a scene deviates from a constrained model by enough to cause potentially fatal ambiguities, the image data effectively turns down the gain on the influence of the constrained model on an estimator's interpretation. An optimal estimator will be biased by the constrained model only for images of scenes that deviate from the constrained model by small amounts. As scenes deviate more from the model, the bias will shrink to nothing.

2.2.2. Effects of image uncertainty

The level of uncertainty in the image data should clearly effect the strength of biases toward constrained interpretation. Appendix A derives the relationship between uncertainty and bias for Gaussian likelihood functions. The result supports what would be our natural intuition. First, the proportional bias toward constrained interpretations (the weights in Eq. (7)) for

scenes that deviate by a fixed amount from a constrained model increases with the uncertainty in the image data. Second, the range of scenes for which the constrained model will dominate interpretations increases with image uncertainty. This result forms the basis of several of the psychophysical predictions outlined in Section 2.4.

2.3. Cue integration: cooperative model selection

While many scenes will support reliable model selection from individual cues, a range of conditions always exists in which the appropriate model to use is ambiguous. In these conditions, other cues to a scene's three-dimensional layout that are normally available in natural images can serve to disambiguate the appropriate model to use for interpreting the cue. Within a Bayesian context, the logic of this process is straightforward. The likelihood function for a pair of cues, when the cues are conditionally independent (e.g., the noise on the cues is independent), is simply the product of the likelihood functions for the individual cues; thus, the likelihood function for one cue can selectively amplify one mode of the likelihood function for another cue while depressing

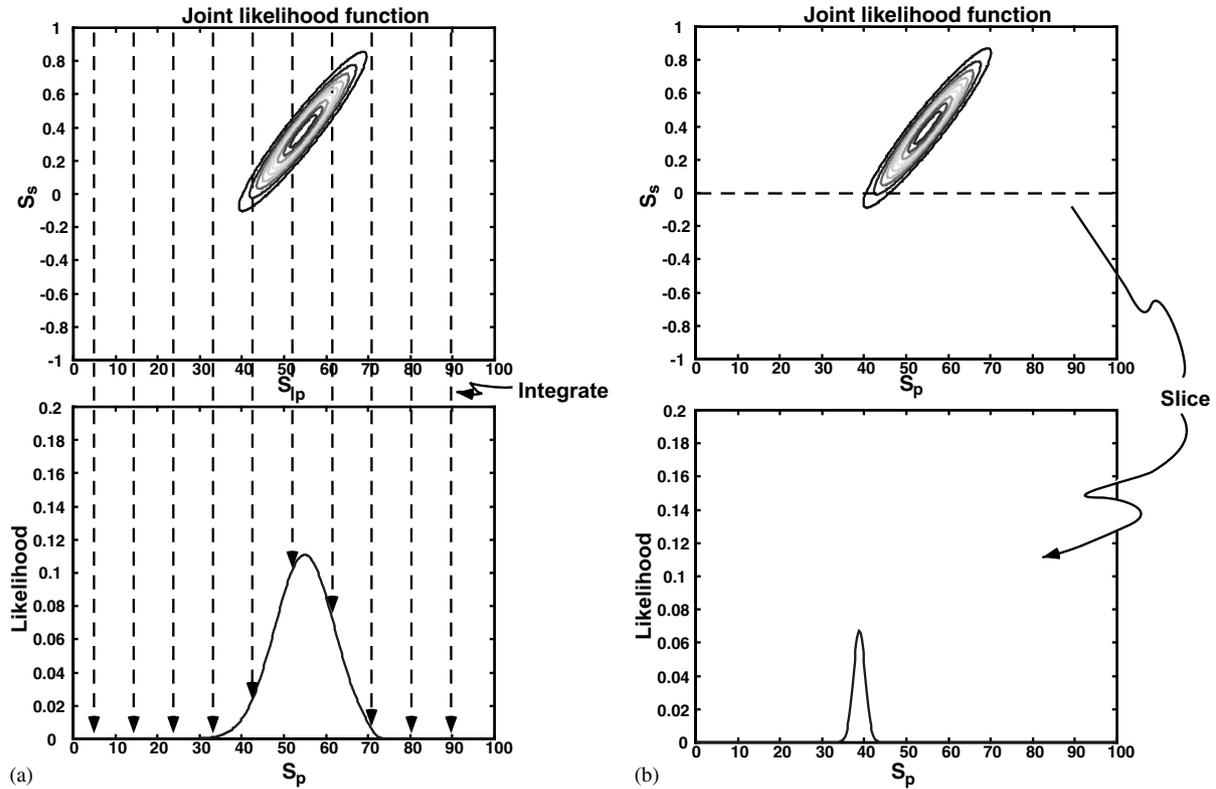


Fig. 3. When the image data is most consistent with a value of  $\bar{S}_s$  very different from 0, the unconstrained model has a higher peak likelihood. In this illustration, the likelihood function for the unconstrained model has not changed, whereas the likelihood function for the constrained model has shifted, reflecting the bias induced by the incorrect assumption that  $\bar{S}_s = 0$ , and scaled down, because the model's slice of the joint likelihood function is more distant from the peak. Note that the scale of the coordinate axis is different from that shown in Fig. 2, to match the lower peak likelihoods found in this condition.

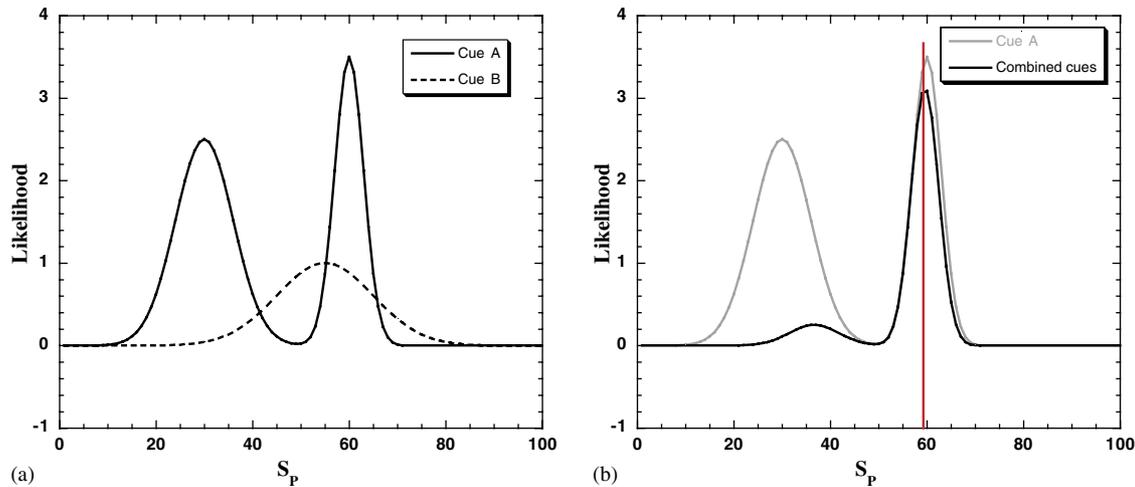


Fig. 4. (a) Likelihood functions for two cues, one of which (cue A) relies on two different generative models. (b) The joint likelihood for the two cues is simply the product of the individual likelihood functions. The red line indicates the estimate derived by minimizing the local mass cost function applied to the joint likelihood function.

the other. Fig. 4 illustrates the effect, here showing how a relatively weak cue can strongly impact a perceptual estimate by effectively selecting which of two models appropriately fits the data from another cue. This is a form of “cooperative model selection.”

A special case of cooperative model selection provides a rational account for the cue vetoing strategy proposed by Landy and Malony as a form of robust cue integration (Landy et al., 1995). Cue vetoing occurs when one of the models corresponding to a cue is

degenerate; that is, has an extremely broad likelihood function. Such models exist for most monocular cues (e.g., non-rigid motion for structure-from-motion, random curves on a surface for shape-from-contour, etc.). Fig. 5 shows an illustrative example. The likelihood function for cue A is a mixture of likelihood functions derived from an uninformative model (the flat model likelihood function) and from an informative model (the Gaussian likelihood function). When cue B suggests an interpretation that is close to the interpretation suggested by the informative model for cue A, that model is effectively selected for the interpretation of cue A, and the integrated estimate is well approximated by a weighted linear combination of the two interpretations. When the interpretation suggested by cue B moves far away from the constrained interpretation suggested by the informative model for cue A, the integrated estimate

is equivalent to the estimate derived from cue B alone; that is, cue A is effectively turned off.

Fig. 6 shows the predicted pattern of estimates of  $S_p$  as the interpretation suggested by the constrained version of cue A moves away from that suggested by cue B. This is exactly the cue vetoing pattern predicted by Landy and Maloney. The Bayesian interpretation of the behavior, however, is very different. Rather than viewing the behavior as “turning off” a cue, the mixture model formulation treats it as selecting an uninformative hidden prior over an informative one. Li et al., for example, described a case in which large conflicts between stereo and motion lead to a down-weighting of motion information (Li, Maloney, & Landy, 1997). The Bayesian interpretation of their result is not that the motion cue was turned off, but rather that in the face of the conflict, subjects reinterpreted the motion to be non-rigid in 3D.

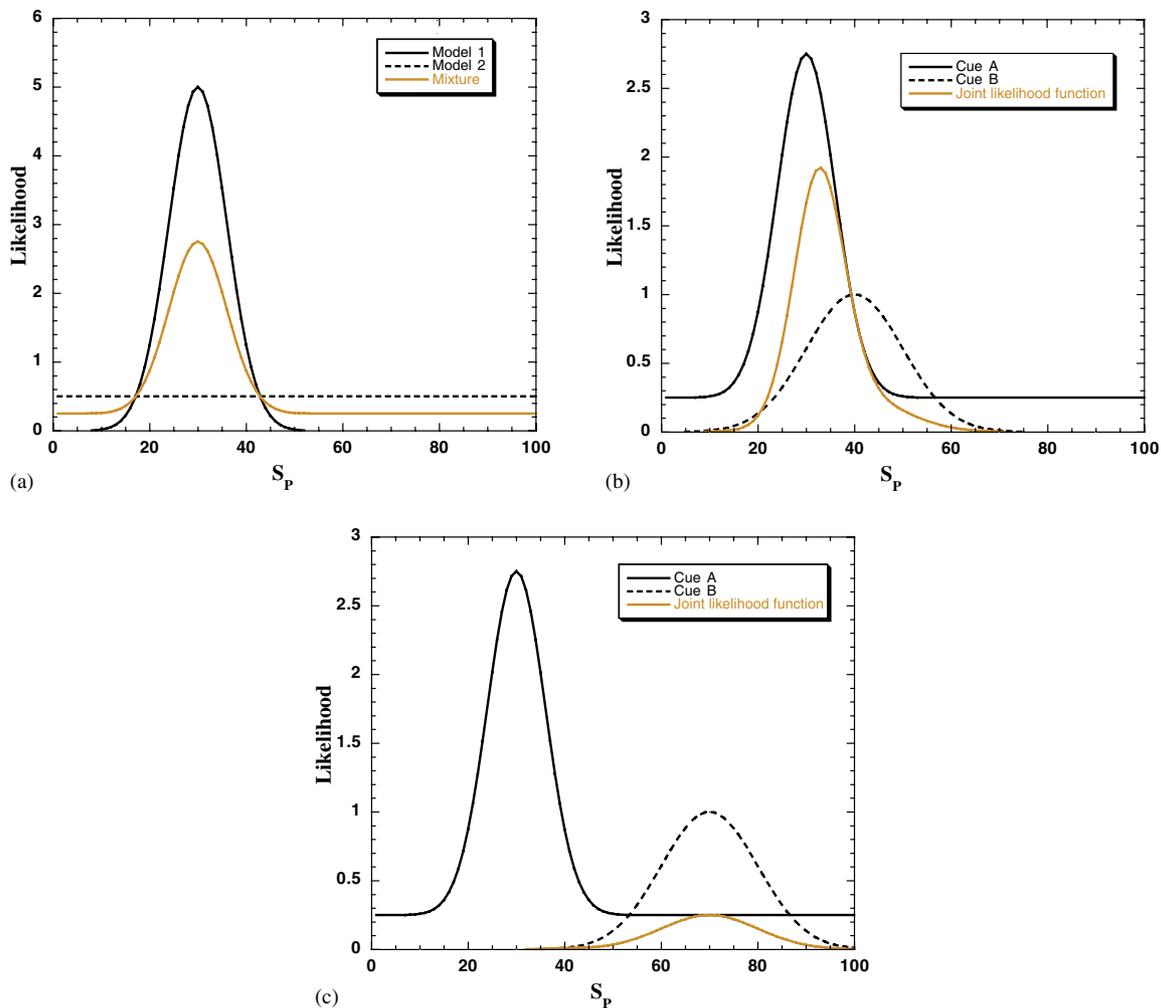


Fig. 5. Apparent cue vetoing can occur when the hidden prior contains both informative and uninformative sub-models (e.g., rigid vs. non-rigid motion). (a) The individual model likelihoods and the full mixture likelihood for cue A. (b) The results of integrating cue A with a cue that suggests an interpretation close to that of cue A under the informative model (model 1)—a joint likelihood function centered on a point that is approximately the weighted average of the means of the component likelihood function for cue A corresponding to model 1, and the likelihood function for cue B. (c) When cue B suggests a much different interpretation, its mode is far away from that of cue A under the informative model, and multiplying the likelihoods results in a joint likelihood function that is approximately equivalent to the likelihood function for cue B alone.

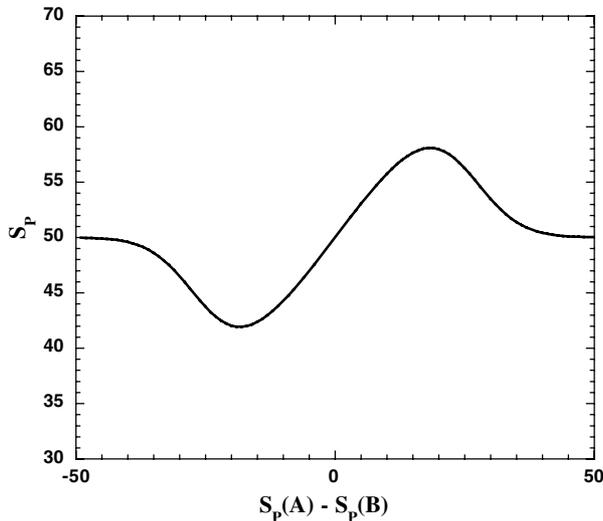


Fig. 6. Results of combining cues A and B from Fig. 5 as a function of the difference between the value of  $\vec{S}$  suggested by cue A under the informative model and the value suggested by cue B. The mode of the likelihood function for cue B is to be at  $\vec{S} = 50$ . The results are effectively the same regardless of whether an estimator uses the MSE cost function or the TMSE cost function.

#### 2.4. Implications for psychophysics

Our analysis of mixture models has several important consequences for understanding human visual perception. It points to a change in research emphasis from determining which *one* of several competing models of prior environmental constraints the visual system uses to interpret a 3D cue to determining the collection of constraints that the visual system employs. More deeply it points to a new aspect of visual processing that needs to be studied—how the visual system chooses from among several models to interpret the information provided by a cue. Researchers have recently begun to study how cue uncertainty determines relative cue weights in linear approximations of human cue integration strategies (Ernst & Banks, 2002; Gharamani, Wolpert, & Jordan, 1997; Jacobs, 2002). The multi-modality of cues suggests more complex, non-linear relationships between cue uncertainty and visual processing:

- Perceptual estimates of scene properties should be biased toward those consistent with strong constraints when images are projected from scenes that are close to matching those constraints.
- Perceptual biases toward constrained interpretations of a cue should disappear as scenes deviate more from the assumptions of constrained models.
- An otherwise unreliable cue can serve to cooperatively select the appropriate model to use when interpreting a more reliable cue, even while being given little apparent weight when fitting a linear model to the cue combination.

- Proportional biases toward more constrained interpretations of a cue should decrease predictably as image information is improved (see Appendix A).
- The space of scenes for which perceptual estimates show these biases should shrink as image information improves; that is, the visual system should be better able to choose a less constrained model (see Appendix A).

The following section described a series of experiments that apply the framework developed here to the problem of perceiving surface orientation from texture. The experiments serve as an example of how the mixture model framework can be effectively applied to psychophysics to develop a deeper understanding of human perceptual performance. The results show that human observers automatically switch between different models of surface textures to interpret image texture information and that they do so in a qualitatively optimal way; namely, they show partial biases consistent with a constrained model for surface textures, but these biases disappear for images projected from surface textures that deviate markedly from the model. Moreover, as the reliability of the image information increases, they are better able to reject the constrained model when it is not appropriate.

### 3. Estimating surface orientation from texture

That texture information provides a strong source of information about surface orientation and shape has been known for some time (Gibson, 1950). A large number of psychophysical studies have explored the cues that human observers use to estimate surface geometry from texture and to determine their relative importance to perceptual performance (Blake, Bulthoff, & Sheinberg, 1993; Buckley, Frisby, & Blake, 1996; Cutting & Millard, 1984; Knill, 1998a, 1998b; Li et al., 1997; Todd & Akerstrom, 1987). Less work has been done on what prior constraints on surface textures underlie human observers' use of the cue (Knill, 1998b; Rosenholtz & Malik, 1997). Since multiple, different constraints might apply, mixture models should play an important role in estimating surface geometry from texture. This section applies the framework of mixture models to the problem of estimating planar surface orientation from texture and describes a series of psychophysical experiments designed to test predictions that result from the theoretical analysis.

#### 3.1. The structure of texture information

Perspective projection distorts a texture pattern in two distinct ways: by scaling and compressing the pat-

tern. When surface textures have certain statistical properties, these distortions reflect themselves in specific patterns of optical texture that provide cues to the shape and orientation of a surface. The two statistical properties that can imbue optical texture patterns with reliable information about surface geometry are homogeneity and isotropy.

Homogeneous surface textures are ones whose statistical properties are invariant to position on a surface. Texture elements in images of homogeneous surface textures shrink, on average, as they recede from the viewer. This gives rise to a scaling cue in the image. Texture elements are also foreshortened by an amount determined by the local orientation of a surface relative to the line of sight. Since the viewpoint-relative orientation of a surface varies as a function of position on a surface, the pattern of local texture element shape and orientation in an image covaries systematically with the orientation and curvature of the surface in space. Thus, homogeneous textures also support a foreshortening cue in the form of texture shape gradients in an image.

Isotropic surface textures are ones whose statistical properties are invariant to orientation on a surface (they have no global orientation). Because isotropic textures have a specific average shape (circular), images of isotropic textures support much stronger inferences about surface geometry from the foreshortening cue than do images of anisotropic, homogeneous textures. In effect, when using an isotropic constraint, observers can use the local statistics of texture element shape (texture shape statistics) to make inferences about local surface orientation.

Previous studies have shown that foreshortening information is a dominant cue for judgments of surface orientation and shape (Buckley et al., 1996; Knill, 1998a, 1998b). Since this cue relies on prior assumptions about the shape statistics of surface textures, knowing what assumptions human observers use is central to understanding how humans estimate surface orientation from texture. On one hand, prior knowledge that a surface texture is isotropic renders a shape-from-texture estimator significantly more informative. Such knowledge can reduce the variance of an ideal estimator by one to two orders of magnitude (Knill, 1998c). On the other hand, mistakenly applying an isotropy constraint can lead to large biases in one's estimate of surface orientation. For a given stimulus, how can an observer determine whether or not to apply isotropy as a constraint on surface textures? The answer will fall out of a mixture-model formulation of texture information.

### 3.2. A mixture model for surface orientation from texture

This section describes an optimal model for estimating surface orientation from texture foreshortening that assumes that textures come in two classes—homoge-

neous or isotropic (a subclass of textures with homogeneous shape statistics). The model is limited to the shape properties of textures; thus, the definitions used here strictly apply only to those properties of textures; thus, for purposes of this discussion, isotropic textures that have inhomogeneous size statistics are considered to be homogeneous, since their shape statistics are trivially homogeneous. Without a general model for inhomogeneous textures, we cannot include them as a third class; however, intuition suggests that for any given image, the likelihood function for an inhomogeneous texture model will be a much broader version of that for a homogeneous model. For simplicity, we assume that were an inhomogeneous class included in the formulation, it would simply add a flat component to the mixed likelihood model.<sup>1</sup>

We model arbitrary homogeneous textures as resulting from a process that globally stretches a homogeneous, isotropic texture. This process can stretch textures by random amounts at random orientations. Thus, the global shape statistics of any given homogeneous surface texture are characterized by two free parameters—a stretch factor,  $\alpha$ , and the angle in the plane,  $\theta$  in which the texture is stretched. Within the framework of this paper,  $\alpha$  and  $\theta$  are secondary scene variables. The primary variables are the slant (angle away from the fronto-parallel) and tilt (direction of slant) of a surface.

Assuming that a non-zero proportion of textures are homogeneous (with arbitrary stretch factors) and that another non-zero proportion are isotropic, we can write the likelihood function for a given set of image texture measurements,  $\vec{T}$ , as a mixture of likelihoods for the two classes of texture

$$p(\vec{T}|\sigma, \tau) = \phi_i p_i(\vec{T}|\sigma, \tau) + (1 - \phi_i) p_h(\vec{T}|\sigma, \tau), \quad (10)$$

where  $\phi_i$  is the probability that a surface texture is isotropic and  $1 - \phi_i$  is the probability that it is homogeneous (but not necessarily isotropic). Knill (1998a) derived the likelihood function for the slant and tilt of a surface based on the pattern of texture element shapes in an image for a generic class of homogeneous surface textures. The general form of the likelihood function is given as a function of slant, tilt and the stretch parameters of a homogeneous texture ( $p(\vec{T}|\sigma, \tau, \alpha, \theta)$ ), since these determine the shape statistics of the texture pattern in the image. In order to derive the likelihood function for slant and tilt, we have to marginalize the full likelihood function over the stretch parameters  $\alpha$  and  $\theta$ ,

<sup>1</sup> Special subsets of homogeneous surface textures undoubtedly exist that would support estimating surface orientation from texture. Consideration of such texture ensembles is beyond the scope of this paper.

$$\begin{aligned}
p(\vec{T}_I|\sigma, \tau) &= \phi_i \int_0^{\pi/2} p(\vec{T}_I|\sigma, \tau, \alpha = 1)p(\theta) d\theta \\
&+ (1 - \phi_i) \int_0^{\pi/2} \int_0^\infty p(\vec{T}_I|\sigma, \tau, \alpha, \theta)p(\theta) \\
&\times p(\alpha) d\alpha d\theta, \tag{11}
\end{aligned}$$

where  $\alpha$  and  $\theta$  are assumed to be independent.<sup>2</sup> Assuming a uniform prior on  $\theta$  gives

$$\begin{aligned}
p(\vec{T}_I|\sigma, \tau) &= \phi_i p(\vec{T}_I|\sigma, \tau, \alpha = 1) \\
&+ (1 - \phi_i) \frac{2}{\pi} \int_0^{\pi/2} \int_0^\infty p(\vec{T}_I|\sigma, \tau, \alpha, \theta) \\
&\times p(\alpha) d\alpha d\theta. \tag{12}
\end{aligned}$$

Since stretching by a factor  $\alpha$  in a direction  $\theta + \pi/2$  is equivalent to stretching by a factor  $1/\alpha$  in the direction  $\theta$ , the prior on  $\alpha$  must satisfy the constraint that  $P(\alpha_1 < \alpha < \alpha_2) = P(1/\alpha_2 < \alpha < 1/\alpha_1)$ . Assuming that the prior on  $\alpha$  is flat for  $\alpha \leq 1$ , this constraint gives as the prior on  $\alpha$ ,

$$p(\alpha) = \begin{cases} 1/2; & 0 < \alpha \leq 1, \\ 1/2\alpha^2; & \alpha > 1. \end{cases} \tag{13}$$

### 3.3. Texture information for or against isotropy

As described in Section 2.3, marginalization over the stretch parameters imbues texture information with the power to discriminate between isotropic and anisotropic categories of surface texture. Fig. 7 illustrates the application of the mixture model to a specific texture image, using the likelihood function derived in Knill (1998c). Since all of the model likelihood functions are integrated over texture orientation,  $\theta$ , the figure shows the likelihood expressed as a function of surface slant,  $\sigma$  and the texture stretch factor,  $\alpha$  (for illustrative purposes, the tilt is assumed to be vertical). The averaging effect of marginalization shrinks the peak likelihood for the homogeneous model, which in this case is much lower than for the isotropic model. The likelihood function for the homogeneous model also has a larger spread, reflecting the greater uncertainty built into the model.

While the isotropic model dominates the mixture for an image created using an isotropic texture, Fig. 8 shows what happens for an image of a similar texture that has been stretched in the direction perpendicular to surface tilt prior to projection into the image (equivalently, compressed in the direction of tilt). The peak of the isotropic likelihood function shifts away from the true slant. This reflects the bias induced by assuming isotropy for an anisotropic texture. On the other hand, the

peak likelihood derived from the isotropic model shrinks as the stretch factor of a surface pulls it away from isotropy. The result is that the homogeneous model begins to dominate the mixture for images of textures generated using stretch factors that differ significantly from 1.0. This leads to the characteristic pattern of mixture models built from nested sets of constraints that was described in Section 2.3, here shown in Fig. 9. The plot shows the average biases of a set of ideal slant estimators for the class of textures used in Experiment 1 (with a field of view of 13°). The estimators differ in their assumptions about what proportion of textures are isotropic (the  $\phi_i$  parameter). When stimuli are generated from textures stretched by factors near 1.0, the estimators track the isotropy model, but as the stretch factor moves away from 1.0 and the homogeneous model begins to dominate the mixture, the bias levels off and eventually goes to zero.

### 3.4. Predictions and previous results

Section 2.4 listed a number of predictions that derive from optimal estimators using mixtures of priors. We can apply some of these to the problem of estimating planar surface orientation from texture. In particular, such estimators make specific predictions about subjective biases in the interpretation of images of planar surface textures that have been stretched by varying amounts:

1. For stretch factors near 1.0, subjects should show biased percepts of surface orientation, in accordance with the behavior of the isotropy model.
2. Making texture information more reliable should reduce the proportional biases for images of slightly stretched textures.
3. Subjective biases should weaken as surface textures are stretched by larger amounts away from isotropy. The pattern of absolute biases should curve back toward zero as the stretch factor increasingly deviates from 1.0 (see Fig. 9).
4. Making texture information more reliable should shift the transition zone of the bias function back toward a stretch factor of 1.0; that is, the range of stretch factors that lead to perceptual biases should decrease.

Two previous studies have reported conflicting results on the strength of subjective biases toward isotropic interpretations of surface textures. Using large field of view stimuli (36°) similar to those shown in Fig. 10, Rosenholtz and Malik (1997) found significant but weak biases toward isotropic interpretations for images of stretched surface textures. They used a direct orientation matching task in which subjects set a gauge figure to indicate the perceived orientation of stimulus surfaces.

<sup>2</sup> Independence of  $\alpha$  and  $\phi$  follows from an assumption that a texture may have any global orientation within the plane of a surface relative to the viewer.

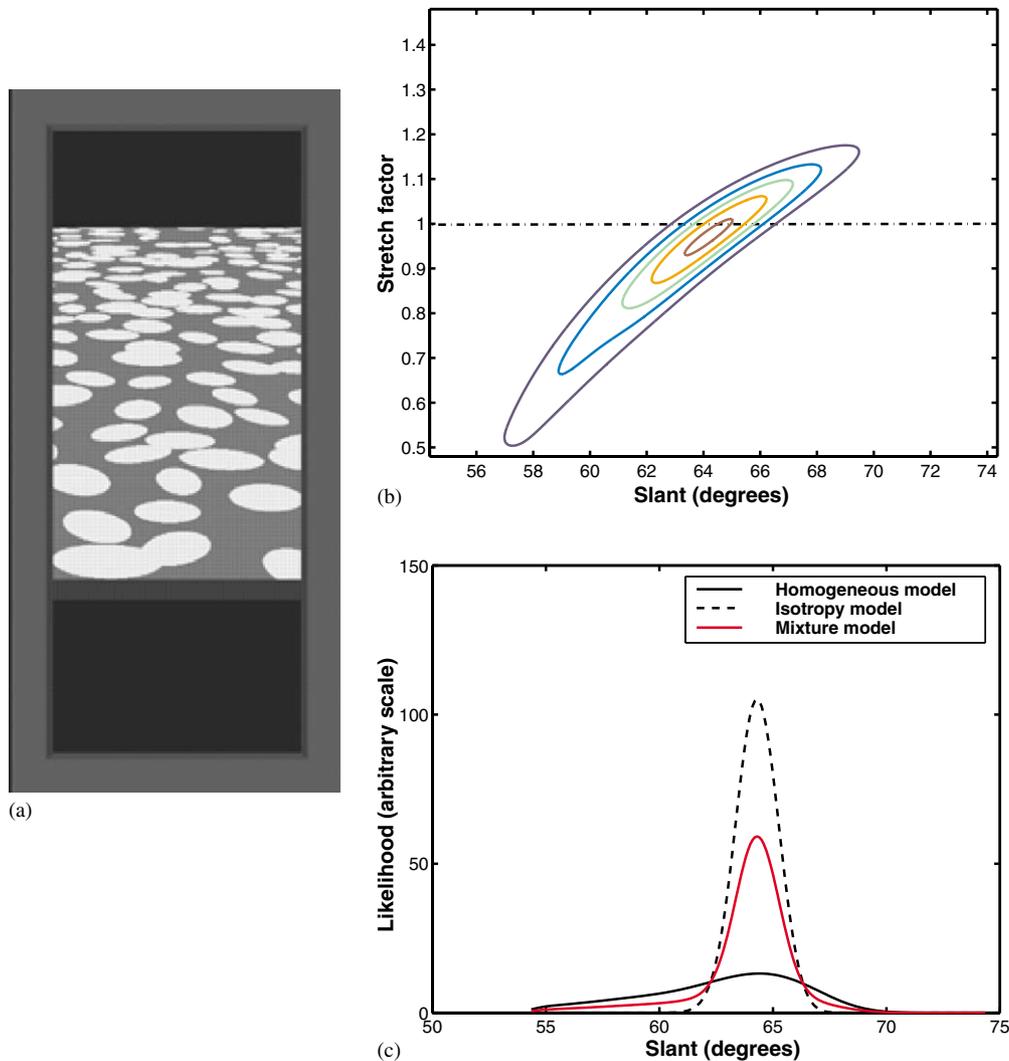


Fig. 7. (a) An isotropic texture projected from a slant of  $65^\circ$  with a simulated vertical field of view of  $13^\circ$ . (b) A contour plot of the joint likelihood function for slant and texture stretch factor, computed for the texture in (a). (c) The marginalized likelihood functions for slant for the homogeneous and isotropic models. The mixed likelihood function assumes an equal prior probability is assumed for each model.

Using similar texture patterns, but a different task (a discrimination task in which subjects made relative slant judgments between images of texture patterns stretched by random amounts), Knill (1998b) found significantly larger biases for small field of view ( $13^\circ$ ) images of surfaces. The difference in results between the two studies is consistent with the prediction that increasing the reliability of the texture information (by increasing the field of view<sup>3</sup>) should decrease the magnitude of the proportional bias toward the more constrained, isotropic model. However, differences in subjects' task and experimental conditions (e.g., Knill used a larger slant than

in any of those used in the Rosenholtz and Malik study) could as easily have led to the difference in results.

The experiments reported here more directly test the predictions of the mixture model. The four main experiments (1, 2, 4 and 5—3 is a control experiment) measure subjective biases in perceived surface slant for images of surface textures created using a range of stretch factors. Experiments 1, 2 and 3 test predictions 1 and 2 listed above. Experiment 1 provides a more direct comparison to the earlier Knill data by using similar stimuli and the same discrimination task to measure subjective biases for large field of view stimuli. Because of a few remaining methodological differences with the earlier study, we use the current methods to replicate the earlier results for small field of view stimuli in Experiment 2. Experiment 3 is a control experiment to insure that the differences that we find between large and small field of view stimuli are not an artifact of subjects giving more weight in the large

<sup>3</sup> Pilot studies have shown that, as with many other cues, simply adding more elements to a texture display, while keeping other parameters like the field of view constant, does not significantly impact human subjects' abilities to discriminate surface orientation from texture. Changing the field of view on a surface does (Knill, 1998a).

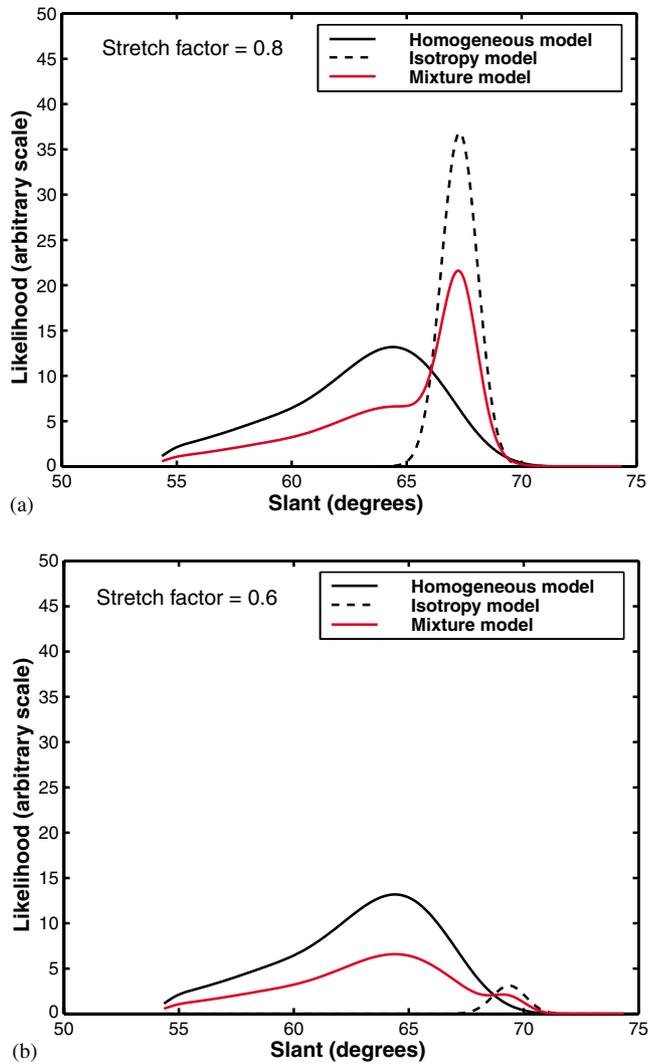


Fig. 8. Model likelihood functions for surface slant for stimuli like the one shown in Fig. 7(a), but using surface textures that have been stretched before projecting them into the image (see Fig. 10 for examples). (a) Surface textures stretched by a factor of 0.8 (compressed in the direction of surface tilt), (b) surface textures stretched by a factor of 0.6. The peak of the isotropic likelihood function shifts with the stretch factor because it is derived with the assumption that the stretch factor = 1 (isotropic textures). As the surface texture projected into the image becomes more compressed, the homogeneous likelihood function begins to dominate the mixture (as in (b)).

field of view stimuli to the texture scaling cue (which is unbiased by stretching surface textures). Experiments 4 and 5 test predictions 3 and 4. These experiments measure subjective biases for a large range of texture stretch factors for both small (Experiment 4) and large (Experiment 5) field of view stimuli.

3.5. Experiment 1: isotropy biases for large field of view stimuli

In the Knill (1998b) study, subjects were asked to judge which of two texture images (as in Fig. 11) de-

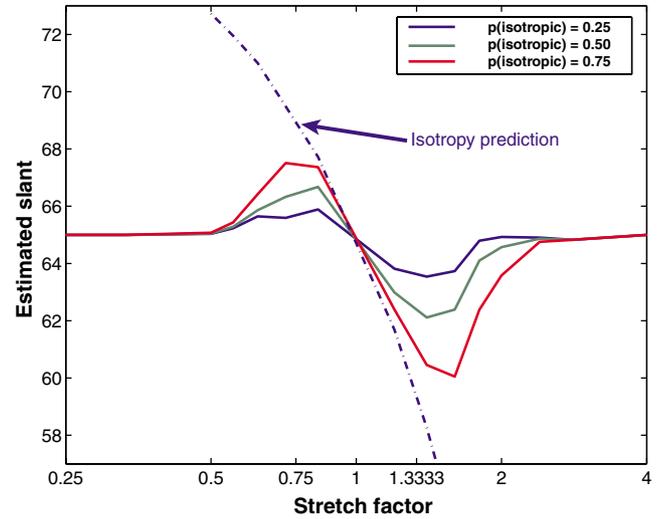


Fig. 9. The pattern of slant biases for an estimator that assumes a mixed prior on surface textures, expressed as a function of the stretch factor used to create surface textures projected into an image. The different curves reflect different priors on the relative frequencies of isotropic and homogeneous (and anisotropic) textures. The specific patterns shown here were derived from repeated simulations of the ideal observer for textures like those used in Experiment 1 at different levels of the stretch factor (see Fig. 11) (100 repetitions per stretch factor). The predicted biases decrease if the quality of the texture information is improved.

icted a surface with greater slant. The surface textures used to generate each of a pair of test stimuli were stretched in random directions by small random factors (stretch factors = 0.7–1.3). A psychometric model was fit to the discrimination data that was based on a weighted linear sum of the outputs of three ideal observers applied to the stimuli, one that used the foreshortening cue with an assumption of isotropy, one that used the scaling cue and one that used the density cue, and a putative unbiased observer that knew the slant of each surface. Using these weights, we derived a measure of each observer’s isotropy bias as a proportion of the theoretical bias induced by a pure assumption of isotropy. The bias reflected a combination of the degree to which subjects relied on the foreshortening cue and the degree to which their interpretation of that cue was biased by an isotropy assumption. Since subjects generally gave very little weight to scaling and density cues, the measure was dominated by the effect of the isotropy bias on subjects’ interpretations of foreshortening information. The results of the analysis showed that subjects’ slant biases were close to that predicted by a strong isotropy model—with a proportional bias of 0.7 for Voronoi textures (left image in Fig. 11) and 0.79 for elliptical textures (right image in 11).

Experiment 1 was designed to measure subjects’ isotropy biases using similar stimuli and the same discrimination task used in Knill (1998b), but for larger field of view stimuli. Fig. 10 illustrates the logic of the

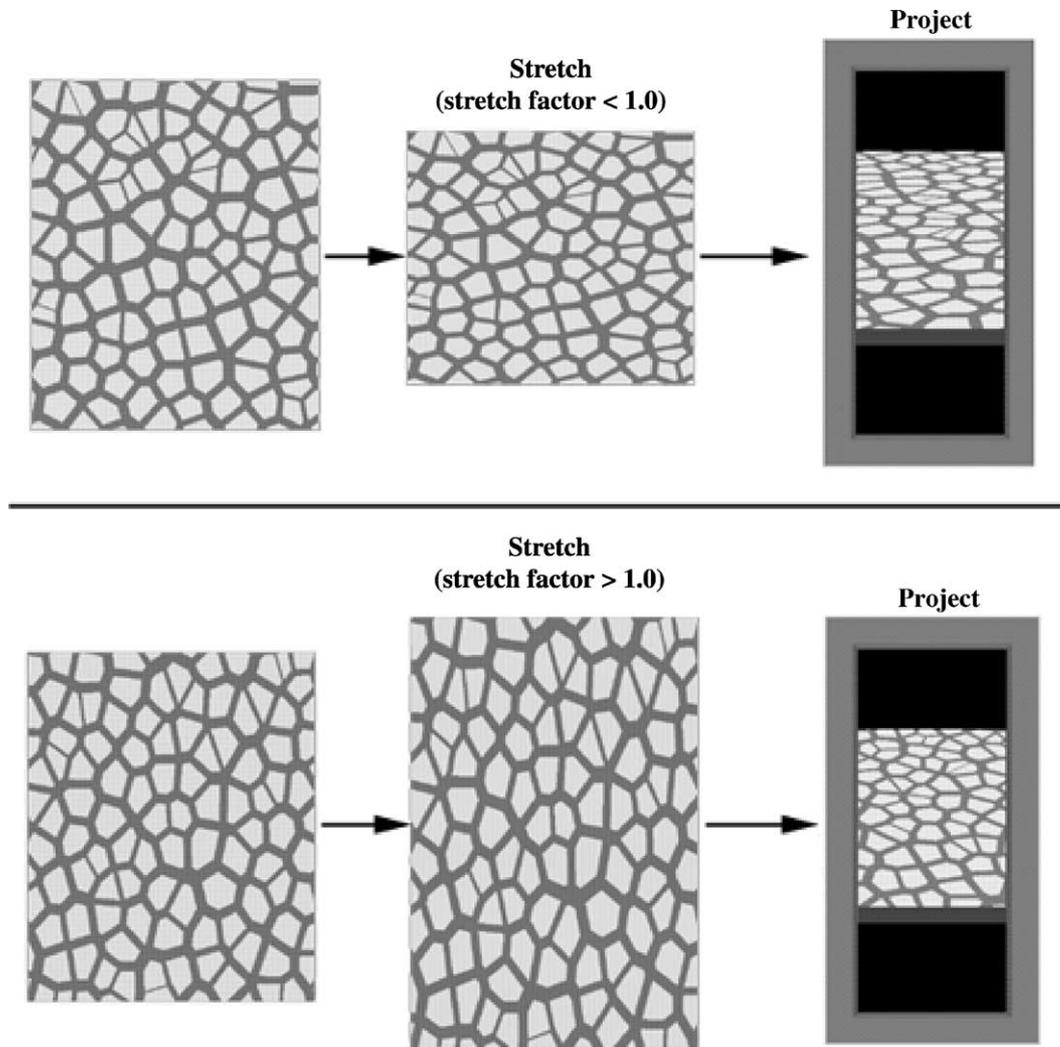


Fig. 10. Stimuli for the experiments were created in three stages. First, a random, isotropic texture pattern was generated. This was then stretched by some amount in the vertical direction. The resulting texture was projected into the image at a slant of  $65^\circ$  and a vertical tilt. A subject that assumes surface textures are isotropic would overestimate the slant of the top stimulus and underestimate the slant of the bottom one.

experiment (and of experiments 2, 3 and 5). Test stimuli were generated by stretching a surface texture by some factor,  $\alpha$ , in the direction of surface tilt prior to projecting it at a slant into the image. In each trial subjects were asked to judge whether the slant of a stretched test stimulus or the slant of an isotropic comparison stimulus was greater (see 11). For each value of the stretch factor,  $\alpha$ , the point of subjective equality (PSE) between test and comparison stimuli was found and used as a measure of a subject's slant bias. An observer who assumed that surface textures were isotropic would attribute all of the texture stretching in the test image to perspective foreshortening and thus overestimate (for  $\alpha < 1$ ) or underestimate (for  $\alpha > 1$ ) the slant of the surface. Slopes of regression lines fitted to the PSEs against the stretch factor,  $\alpha$ , provided measures of the strength of the isotropy bias for each subject.

### 3.5.1. Methods

**3.5.1.1. Apparatus.** Stimuli were presented on the display monitor of an SGI computer. The monitor was an SGI model TFS6705, 17 in., color display with a resolution of  $1280 \times 1024$  pixels. Stimuli were generated in gray-scale on the display. Since the stimuli did not contain smooth shading variations, we did not do gamma correction. Subjects viewed the stimuli presented on the monitor monocularly through a reduction screen, with their heads placed in a chin rest and resting on a front head-rest. Subjects' non-viewing eyes were covered with an eye-patch to eliminate any potential for binocular rivalry. Subjects were tested in a room painted matte black to minimize secondary reflections back onto the monitor. Finally, a matte black occluder was placed over the front of the monitor to obscure the physical screen boundaries. The monitor was calibrated using test patterns of dots viewed through a piece of

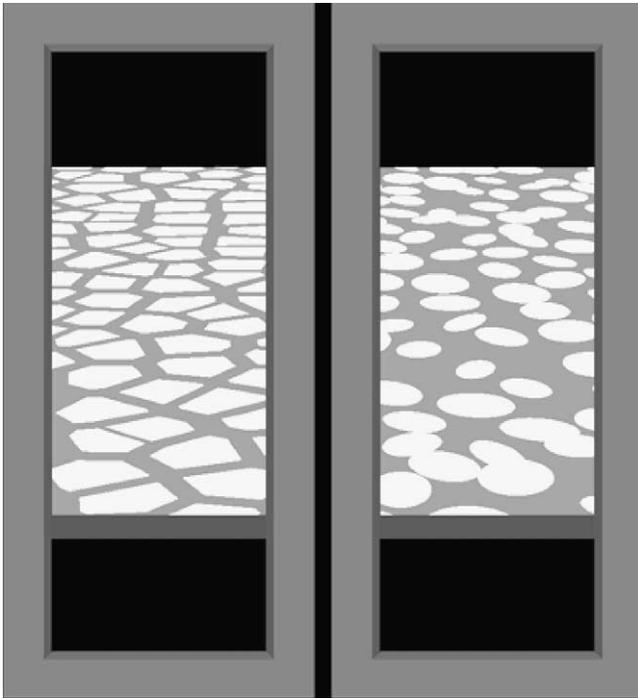


Fig. 11. Example stimuli from Experiment 2. These stimuli subtended  $10^\circ \times 13^\circ$  of visual angle. Stimuli for Experiment 1 were similar, but were expanded (keeping the texture density fixed) to fill  $20^\circ \times 25.4^\circ$  of visual angle.

metal with a square grid of holes to insure a square geometry.

Subjects viewed the display from a distance of 28 cm, giving a total angular extent of the display area on the screen of approximately  $48^\circ \times 40^\circ$  of visual angle.

**3.5.1.2. Stimuli.** Fig. 11 shows stimuli similar to the ones used for Experiment 1. Stimuli were created by perspective rendering of planar surface textures slanted away from the fronto-parallel plane around the horizontal axis (having a vertical tilt). One stimulus in a pair was a test stimulus, created by globally stretching a surface texture (as in Fig. 10) and projecting the texture into the stimulus image at a fixed slant of  $65^\circ$  (measured relative to the line of sight to the middle of the display). The other stimulus was a comparison stimulus, generated by projecting an isotropic texture pattern at a slant chosen by the adaptive procedure used in the experiment. Stimulus pairs were presented side by side in the experiment, with each stimulus image having its own simulated window frame. The innermost vertical boundaries of the two surface images were 70 pixels from the vertical mid-line of the screen (including the space taken up by the inner frames), which, for the viewing conditions used, gave a  $6^\circ$  separation between inner edges of the stimuli. For each condition in an experiment, the vertical positions of a surface's boundaries as they appeared in an image were the same for

both test and target stimuli, so that boundary height in the image plane did not provide a cue to surface slant.

Stimulus images subtended a width of 500 pixels and a height of 640 pixels (displayed side-by-side). At the viewing distance used of 28 cm, the images of surfaces subtended approximately  $20^\circ \times 25.4^\circ$  of visual angle. Test stimuli were created from random elliptical element surface textures. Each texture was generated by drawing randomly shaped ellipses at the points of a pre-generated, isotropic, random lattice. Random lattices were created using a constrained, stochastic reaction diffusion process designed to impose enough regularity on texture element spacing to insure that texture elements rarely overlapped. Prior to stretching, the distribution of ellipse orientations was uniform, consistent with the textures being isotropic. Prior to projection into the stimulus image, the elliptical textures were globally stretched by factors of 1.2, 1.1, 1.0, 0.9, or 0.8 in a direction aligned with the surface tilt. The density of the lattices used to generate the textures was set so that on average, 150 texels appeared in a stimulus image. Textures were scaled so that the average number of texels in a stimulus was constant as a function of surface orientation.

Comparison stimuli were created from constrained, random Voronoi textures. Voronoi polygons were generated from the same set of random lattices used to position elliptical texture elements on the surfaces used for test stimuli (see Knill (1992) for details of the construction). The polygons were then shrunk by a factor of 70% toward their centers of mass, creating the type of tiled pattern illustrated in Fig. 11.

The elliptical element textures were designed to have the same statistics as the Voronoi textures—we used the statistics of the second order moments of inertia of the polygons in the Voronoi textures (i.e., the distributions of sizes, aspect ratios and orientations of the fitted ellipses) to parameterize the distributions from which random ellipses were drawn on a surface. The ellipses were drawn at the same positions on a surface as the polygons in the Voronoi textures.

Test stimuli were projected under perspective projection from a fixed slant of  $65^\circ$  onto the computer screen. All surfaces were rotated away from the fronto-parallel around a horizontal axis, giving them a vertical tilt in the image. Comparison stimuli were rendered under perspective projection at whatever slant was selected by the adaptive procedure described below.

**3.5.1.3. Procedure.** We used a two-alternative forced choice procedure in which subjects judged which of two simultaneously presented texture images appeared to be more slanted. All conditions in an experiment were randomly interleaved, including the side of the display on which the correct stimulus appeared. The screen was blanked between trials, a period which lasted anywhere

from 1/2 to 1 s, depending on the time it took to generate stimuli for the next trial. Subjects were given unlimited time to view the displays on each trial, but were explicitly instructed to make judgments based on their immediate guess as to which surface was more slanted. They were told that on some trials the choice would be clear and on others it would be more ambiguous, but to stick with their first guess regardless of how uncertain it seemed. Feedback was given in the form of a summary score every 20 trials. The feedback was used simply to make the task more palatable for subjects, as pilot studies showed subjects found the experiment with no feedback extremely unpleasant and we suffered from many drop-outs. No trial-by-trial feedback was given, in order to minimize, as much as possible, the learning of simple 2D strategies for doing the task.

Test stimuli were generated to simulate a slant of 65°. Three non-parametric staircases (three-up/one-down, one-up/one-down and one-up/three-down), in which the slant of the comparison stimulus was increased or decreased, were interleaved for each condition. The staircases were used as sampling procedures; points of subjective equality (PSEs) and thresholds were estimated off-line using maximum likelihood fits to a standard psychometric function.

Before starting the main part of the experiment, subjects were run in a brief demonstration version of the experiment using textures generated from surfaces with very large differences in slant (65° and 73° for test and comparison stimuli respectively).

**3.5.1.4. Data analysis.** The raw data was organized into arrays specifying the proportion of trials on which subjects reported the comparison stimulus to be more slanted than the test stimulus, as a function of the slant of the comparison stimulus. In pilot experiments, we found that naive subjects like those used in these experiments have a high guessing rate (e.g., because of attentional lapses). This was reflected in psychometric functions that leveled off at points below 1.0 and above 0.0. In order to correct for guessing, we fit a modified cumulative Gaussian psychometric function to each subject's data in which the probability of selecting a comparison stimulus was assumed to be a mixture of an underlying Gaussian discrimination process and a random guessing process. Writing subjects' decision as

$$D = \begin{cases} 1; & \text{Comparison stimulus judged more slanted} \\ 0; & \text{Test stimulus judged more slanted} \end{cases} \quad (14)$$

the psychometric model was

$$p(D = 1|\Delta\sigma) = (1 - p)\Gamma(\Delta\sigma; m, s) + pq, \quad (15)$$

where  $\Delta\sigma$  is the difference in slant between comparison and test stimuli,  $m$  is the mean of the cumulative

Gaussian,  $s$  is the standard deviation of the cumulative Gaussian,  $p$  is the probability that a subject guessed on any given trial and  $q$  is the probability that subject guessed the comparison stimulus, given that he or she guessed at all. The mean parameter,  $m$ , provides a measure of the point of subjective equality between test and comparison stimuli.

Guessing parameters for each subject were assumed to be constant across conditions within an experiment. Parameters for the psychometric model were computed from maximum likelihood fits to the raw data. The inverse of the Hessian of the likelihood function derived from the psychometric model provided estimates of std. errors for the fitted parameters.

**3.5.1.5. Subjects.** Subjects were four undergraduates at the University of Rochester. Subjects were paid for their participation, had normal or corrected to normal vision and were naive to vision science and to the goals of the experiment.

### 3.5.2. Results

Fig. 12(a) shows the PSEs for each of the four subjects in Experiment 1. The averages of the fitted guessing parameters were  $p = 0.18$  and  $q = 0.72$ . The standard deviation parameters of the fitted cumulative Gaussians provided a measure of subjects' abilities to discriminate slant from texture, after correction for guessing. Using these to estimate 75% thresholds for subjects, we found an average discrimination threshold of 1.9°. PSEs for  $\alpha = 1.0$  showed that subjects judged the Voronoi textures to be slightly more slanted than equivalent elliptical element textures (average bias = 1.7°). This is consistent with the small differences found between the two types of textures in a previous study (Knill, 1998a).

In order to determine the biases predicted by a model in which subjects assume that surface textures are isotropic, we ran an ideal observer for estimating slant from texture for the stimuli used in the experiment. The ideal observer incorrectly assumed that surface textures were isotropic for all stimuli (it assumed a stretch factor equal to 1.0), but in all other ways was the statistically optimal estimator for the class of textures used (Knill, 1992). The dashed lines show the performance of the isotropic ideal observer. The bias function for the isotropic ideal observer is approximately linear in the range of stretch factors used in the experiment. Slopes derived from weighted linear regressions of PSE vs. the stretch factor, therefore, provide a summary measure of the proportion of isotropy bias shown by subjects. Fig. 13 (left panel) shows the fitted slopes for each of the subjects along with the fitted slope for the isotropic ideal observer. The average proportional bias, given by the

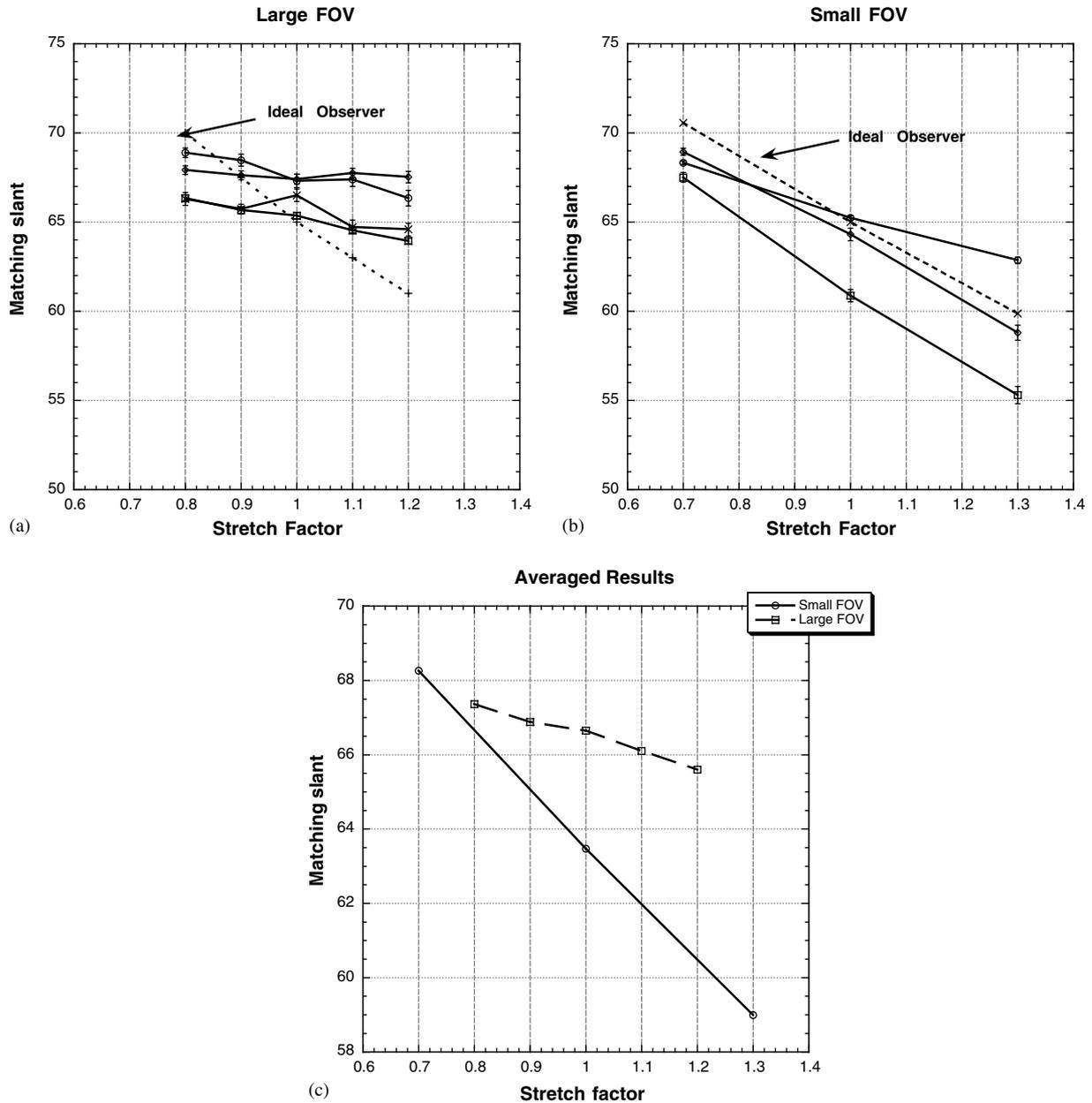


Fig. 12. Plots of subjects' points of subjective equality between test stimuli containing images of stretched textures and comparison stimuli containing images of isotropic textures, as a function of the stretch factor: (a) results from Experiment 1 (large FOV), (b) results of Experiment 2 (small FOV) and (c) the averaged PSEs of subjects in the two experiments.

ratio of subjects' regression slopes to the ideal observer's regression slope was 0.19.

3.6. Experiment 2: isotropy biases for small field of view stimuli

While the task and stimuli used in Experiment 1 were the same as the earlier experiment using small field of view stimuli, the stimulus set was different (surfaces were only stretched in the direction of surface tilt) and the method for estimating the strength of subjects' isotropy biases was different. A previously unreported experi-

ment using the same methodology as in Experiment 1, but with a smaller field of view addresses this issue. The experiment was equivalent to Experiment 1 with three differences. First the field of view was approximately half the size of the one used in Experiment 1 ( $10^\circ \times 13^\circ$ ), but with equal density of texture elements. This resulted in textures containing, on average, 60 texels per image. Second, only three stretch factors were used to create test stimuli (1.3, 1.0 and 0.7). Third, elliptical textures served as test stimuli and Voronoi textures served as comparison stimuli. All other methods were equivalent to Experiment 1.

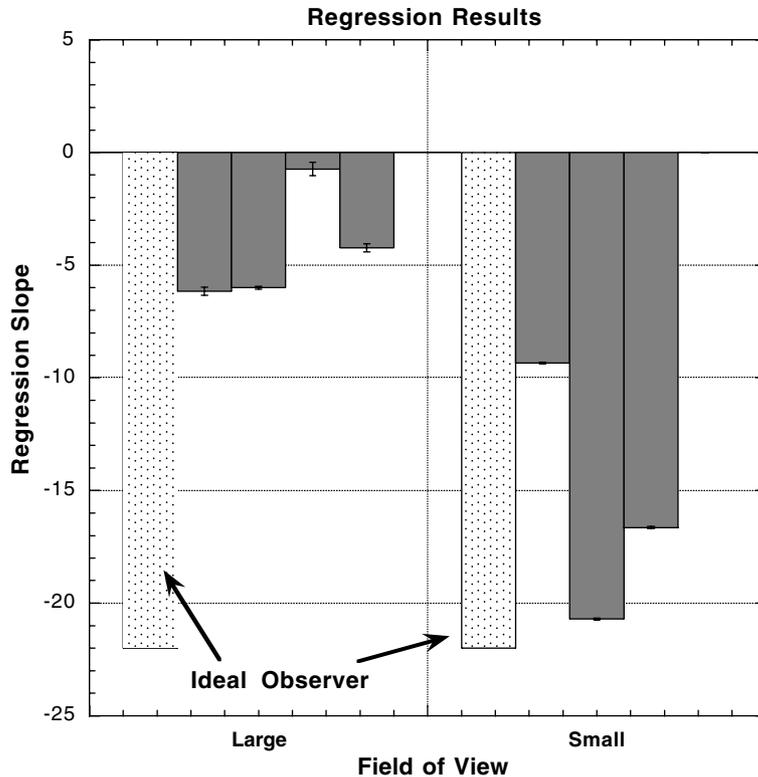


Fig. 13. Slopes of the best fitting lines through the curves in Fig. 12(a) and (b), derived from a weighted regression of PSE vs. stretch factor. The “ideal observer” shown here is one that assumes surface textures are isotropic, but is otherwise statistically optimal.

Three undergraduate subjects at the University of Pennsylvania served as subjects. All three subjects had corrected to normal vision and were naive to the goals of the experiment.

### 3.6.1. Results

Fig. 12(b) plots PSEs for each subject as a function of the stretch factor. The averages of the fitted guessing parameters were  $p = 0.17$  and  $q = 0.53$  and the average 75% threshold was  $2.3^\circ$ . For  $\alpha = 1.0$ , subjects showed a small bias, as in Experiment 1, to judge the Voronoi textures as slightly more slanted than the elliptical element textures (average bias =  $1.6^\circ$ ). Fig. 12(c) shows average PSEs as a function of  $\alpha$  for experiments 1 and 2. The curves clearly show a significantly larger effect for the small field of view stimuli than for the larger field of view stimuli. This is reflected in the regression slopes shown in Fig. 13. Subjects' average proportional bias in Experiment 2 was 0.7, as compared to 0.19 in Experiment 1. This difference was significant ( $T(5) = 3.58$ ;  $p < 0.01$ ).

### 3.6.2. Discussion—experiments 1 and 2

The results of Experiment 2 replicate the findings of the earlier study which served as the point of comparison with Experiment 1. The proportional isotropy biases found for Voronoi textures in Experiment 2 (small field

of view stimuli) were equivalent to those measured in that study (0.7, Voronoi textures). This supports comparing the results of Experiment 1, which measured the strength of isotropy bias for stretched elliptical textures in large fields of view to similar estimates for elliptical textures in small fields of view derived from the earlier study. While subjects in Experiment 1 showed some bias (proportional bias = 0.19), it was significantly smaller than the bias measured in the earlier study using small field of view stimuli (proportional bias = 0.79, elliptical element textures). It was also significantly smaller than the bias found for the small field of view stimuli used in Experiment 2. Strictly speaking, we cannot make a direct comparison between the results of experiments 1 and 2, because different types of textures served as test stimuli in the two. The equivalence of estimates derived from Experiment 2 and the earlier study, however, combined with the earlier results showing near equal biasing effects for elliptical and Voronoi textures, strongly argues that the large difference found here between experiments 1 and 2 is due to the change in field of view size. The combined results of the current experiment and previous experiments (Knill, 1998b; Rosenholtz & Malik, 1997) are consistent with a model observer that employs a mixed model of surface textures, assuming that a fraction of surface textures are isotropic and a fraction are anisotropic.

Before drawing such a strong conclusion, we must consider an alternative explanation for the decreased isotropy bias found in Experiment 1—that texture cues other than foreshortening are weighted more heavily in large field of view images of textured surfaces than in small field of view images. According to this hypothesis, an observer could impose a hard assumption of isotropy for the interpretation of foreshortening information, but the biasing effects of that assumption would be mitigated by other texture cues, like scaling, that are not biased by stretching surface textures. If subjects were to give more weight to these cues in large field of view stimuli, subjects would show the observed decrease in the bias caused by texture stretching.

### 3.7. Experiment 3: measuring weights for foreshortening and scaling cues

The third experiment tests the alternative hypothesis by measuring the weights given by subjects to foreshortening and scaling cues when estimating surface slant from texture. In a previously reported study using stimuli similar to those in Experiment 2 (with the same, small field of view, slant, etc.), subjects were found to weight foreshortening and scaling information in approximately a 4:1 ratio (Knill, 1998b). This experiment, therefore, was designed to measure cue weights for large field of view texture stimuli. Since studies have consistently shown that density information is ineffective for human observers (Buckley et al., 1996; Knill, 1992), the experiment focused only on the relative weighting of foreshortening and scaling cues.

A standard perturbation technique was used to measure the relative weights given to the foreshortening and scaling cues. The technique prescribes creating stimuli in which the cues suggest slightly different values of slant, measuring subjects' estimates of slant and correlating their estimates across different cue conflict conditions with the values suggested by each cue. In this experiment, test stimuli were created with five different combinations of foreshortening and scaling cues:  $\{62^\circ, 65^\circ\}$ ,  $\{67^\circ, 65^\circ\}$ ,  $\{65^\circ, 65^\circ\}$ ,  $\{65^\circ, 62^\circ\}$ , and  $\{65^\circ, 67^\circ\}$ . Points of subjective equality between the test stimuli and stimuli created with consistent foreshortening and scaling cues provided a measure of subjects' slant percepts for the test stimuli.

#### 3.7.1. Methods

**3.7.1.1. Stimuli.** Test stimuli were created from surface textures composed of random arrays of ellipses. For all test stimuli, the positions of texels in the image were computed by perspective projection of surface texel positions on a surface slanted away from the fronto-parallel at an angle of  $65^\circ$  around the horizontal. The aspect ratios and orientations of the projected ellipses were determined by perspective projection of the surface

texels at the slant specified for the foreshortening cue ( $62^\circ$ ,  $65^\circ$ ,  $67^\circ$ ,  $65^\circ$  and  $65^\circ$  for the five test stimuli respectively). The lengths of the projected ellipses were determined by perspective projection of the surface texels at the slant specified for the scaling cue ( $65^\circ$ ,  $65^\circ$ ,  $65^\circ$ ,  $62^\circ$  and  $67^\circ$  for the five test stimuli respectively). Comparison stimuli for each trial were generated by perspective projection of Voronoi textures at the slants specified by the adaptive procedure for each trial, just as they were for Experiment 1. Stimulus dimensions ( $20^\circ \times 25.4^\circ$ ) and average texture density (150 texels per stimulus) were the same as in Experiment 1.

**3.7.1.2. Procedure and data analysis.** The procedure and method of data analysis used in experiments 1 and 2 were repeated in this experiment to find the point of subjective equality between the cue conflict (test) stimuli and cue consistent (comparison) stimuli.

**3.7.1.3. Subjects.** Three subjects for the experiment were drawn from the student body at the University of Pennsylvania. Subjects were paid for their participation, had normal or corrected to normal vision and were naive to vision science and to the goals of the experiment.

#### 3.7.2. Results

Fig. 14 plots each subject's point of subjective equality between cue-conflict test stimuli and the cue consistent comparison stimuli for the five conditions in the experiment. The solid curves show PSEs for test stimuli with the foreshortening cue fixed at a  $65^\circ$  slant and with a variable scaling cue. The dashed curves show PSEs for test stimuli with the scaling cue fixed at a  $65^\circ$  slant and with a variable foreshortening cue. The relative slopes of the two curves reflects the relative weights given by subjects to the two cues. Fig. 15 plots the weights computed from a linear regression of the PSEs against the slants suggested by each cue, both as raw weights and as relative weights. On average, subjects weighted foreshortening over scaling information in a ratio of 3.4:1.

#### 3.7.3. Discussion

Subjects gave approximately the same relative weights to foreshortening and scaling information in the large field of view stimuli used here as was found for small field of view stimuli of a similar type (average ratio of 4:1) (Knill, 1992). The weights found in the current experiment did not sum to one, however. In theory, this could reflect an increased weight being given to density information; however, given the large body of results in the literature showing the inefficacy of the density cue, this seems unlikely. More likely is the possibility that

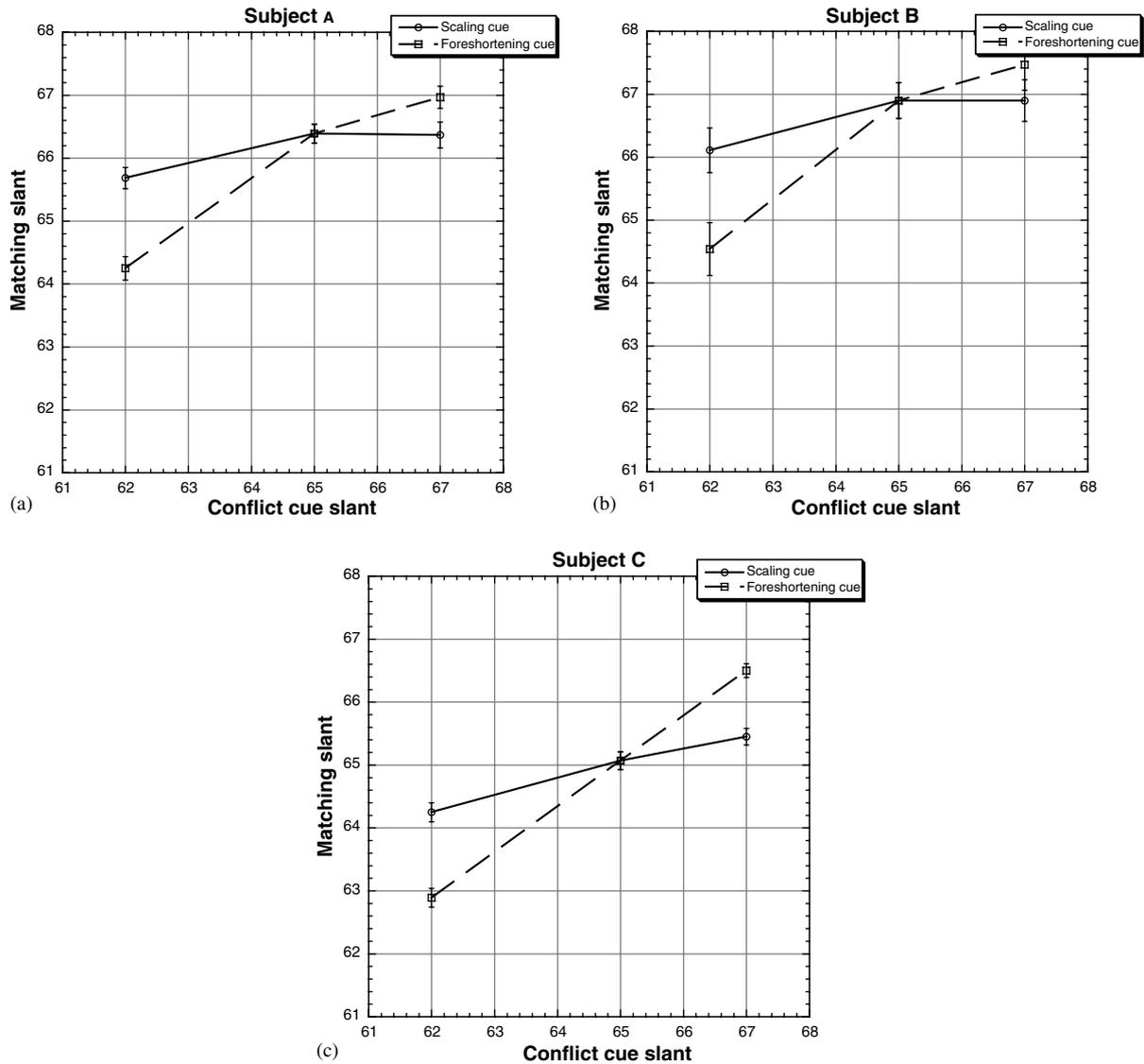


Fig. 14. Plots of subjects' points of subjective equality between test stimuli containing conflicts between scaling and foreshortening cues and cue-consistent, comparison stimuli. Solid lines plot PSEs for cue conflict stimuli in which the foreshortening cue was fixed at 65° and the scaling cue was varied. The dashed lines plot PSEs for stimuli with the reverse cue-conflicts.

scaling and foreshortening information on a coarser scale than individual texels contributed to subjects' judgments. The arrays of texels used to generate surface textures were semi-regular; therefore, the relative spacing between texels provided both scaling and foreshortening information. Since texel positions were projected from a fixed slant of 65° in all stimuli, the coarser scale information provided by texel spacing consistently specified a constant slant. Any weight given by subjects to scaling and foreshortening information at a larger scale than individual texels would diminish the effects of perturbations in the scaling and foreshortening information carried by the individual texels. Unfortunately, independently perturbing the scaling and foreshortening cues on a large scale is mathematically impossible, as it is impossible to isolate those cues in the

perspective mapping of texel positions from the surface into the image.<sup>4</sup>

The results of the experiment support the conclusion that the differential biases found in experiments 1 and 2 resulted from differences in the way the visual system interpreted the texture foreshortening cue, rather than from changes in the relative weights given to foreshortening and scaling information. Thus, the initial results appear to reflect the mixed nature of the visual system's model of texture isotropy.

<sup>4</sup> One can isolate them in the differential of the mapping, hence, assuming that the differential is approximately constant at the scale of individual texels, one can independently manipulate the scaling and foreshortening information provided by the shapes and sizes of texels, which is what we did.

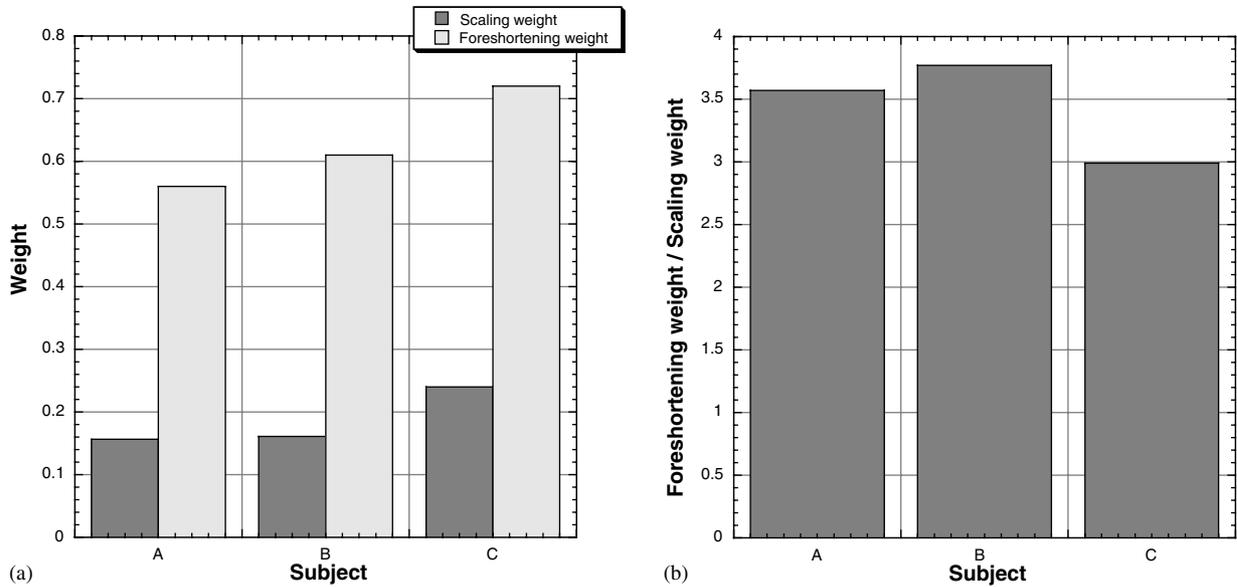


Fig. 15. (a) Raw weights derived from linear regression of subjects' PSEs against the slants suggested by scaling and foreshortening cues. (b) Ratios of foreshortening to scaling weights for the three subjects.

3.8. Experiments 4 and 5: turning off isotropy

The mixture model hypothesis predicts that the anisotropy model will begin to dominate subjects' estimates of surface orientation when a surface texture deviates markedly from being isotropic. More specifically, the model predicts that observers' isotropy bias

will decrease as a surface texture is stretched by larger amounts. In experiments 4 and 5, we tested this prediction by replicating experiments 1 and 2 using a larger range of stretch factors. Fig. 16 shows examples of the small field of view stimuli used in Experiment 4. Experiment 5 repeated the procedure for stimuli with larger fields of view.

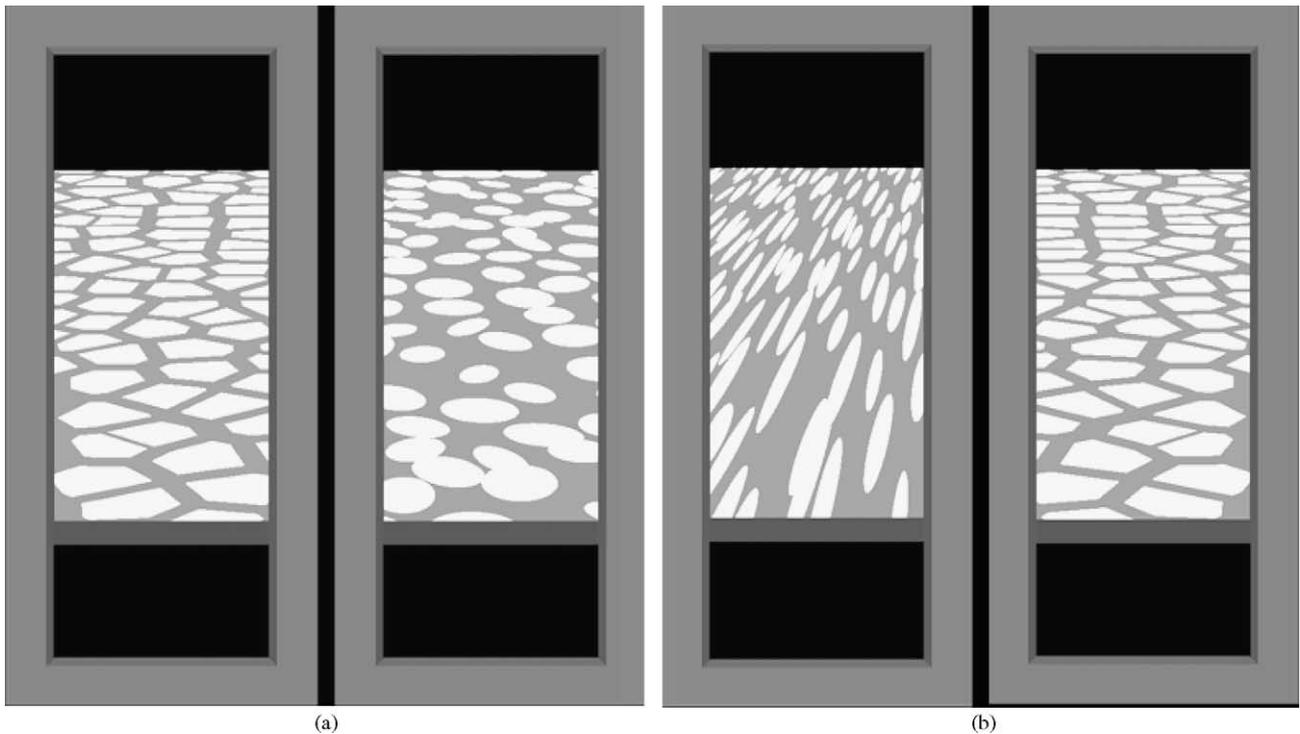


Fig. 16. Example stimuli from Experiment 4. These stimuli subtended  $10^\circ \times 13^\circ$  of visual angle. Stimuli for Experiment 5 were similar, but were expanded (keeping the texture density fixed) to fill  $20^\circ \times 25.4^\circ$  of visual angle.

### 3.8.1. Methods

The methods were equivalent to those used in experiments 1 and 2 with one exception: the stretch factors used in experiments 4 and 5 were {1, 2, 4, 8, 16}. Stimuli in Experiment 4 subtended  $10^\circ \times 13^\circ$  of visual angle. Stimuli in Experiment 5 subtended  $20^\circ \times 25.4^\circ$  of visual angle. The number of texels correspondingly averaged 60 in Experiment 4 and 150 in Experiment 5.

*3.8.1.1. Subjects.* Three subjects for the experiment were drawn from the student body at the University of Rochester. Subjects were paid for their participation, had normal or corrected to normal vision and were naive to vision science and to the goals of the experiment.

### 3.8.2. Results

Fig. 17 shows the results from both experiments. In Experiment 4, subjects showed a large bias at a stretch factor of 2.0, but the bias leveled off and even decreased for the largest stretch factor used of 16. In Experiment 5, subjects showed a smaller initial bias and the point at which the bias curve leveled off and changed direction was earlier than in Experiment 4.

### 3.8.3. Discussion

The results qualitatively follow the predictions of the mixture model. Subjects show an isotropy bias for small levels of surface texture stretching, but the bias decreases at high levels of the texture stretch factor. Were texture cues the only ones available in the experiment, the mixture model would predict that the bias would disappear completely for the highly stretched textures. Subjects' estimates, however, remain biased even for these highly anisotropic textures. In light of the presence in the stimuli of cues suggesting that the surfaces were fronto-parallel (e.g., accommodation and blur), this result is expected. The reliability of texture foreshortening information is significantly lower when isotropy does not apply than when it does. One would therefore expect that subjects' estimates would be biased toward the fronto-parallel for highly anisotropic textures, as the weight given to non-texture cues would correspondingly be increased in such cases. Moreover, the fronto-parallel bias should be larger for small field of view stimuli, with weaker texture information, than for large field of view stimuli, as was found here.

## 4. General discussion

The experimental results support the hypothesis that humans interpret surface orientation from texture using a mixed model of textures. The visual system, at least implicitly, assumes that a significant percentage of surface textures are isotropic, but that some are not. The

mixed assumptions lead to strong, but incomplete biases toward isotropy for surface textures that are close to being isotropic. The visual system still uses foreshortening information for texture images generated from highly anisotropic surface textures, but without treating them as isotropic. More concretely, the visual system appears to smoothly transition between using local foreshortening cues (deviations of local texture patterns away from isotropy) and global foreshortening gradients, which do not rely on an assumption of isotropy.

### 4.1. Implications for cue weighting

One of the notable features of the results in experiments 4 and 5 is that images of highly oriented textures appeared less slanted than images of isotropic textures. As noted above, this may have resulted from the relative weakness of anisotropic texture cues as compared to isotropic texture cues. This points to a complex interaction between cues when the interpretation of one of the cues is built on a mixture model—the apparent weight given to cues will depend on which component model contributes most to the interpretation. Generally, more constrained models will result in higher cue weights than less constrained models. This provides a further strong prediction of the mixture model. Further studies can test whether or not the greater fronto-parallel bias shown for highly stretched textures results from a decrease in the effective weight given by subjects to texture information in those conditions.

### 4.2. Alternative approaches to model selection

We have formulated a rigorous Bayesian approach to model selection, in which the structure of the information in the cue and how it interacts with other cues implicitly determines the model that dominates scene interpretation. An alternative approach to model selection is to rely on key features in the image (Richards et al., 1996) to disambiguate the appropriate model to use when interpreting the cue. The texture images in Fig. 16 suggest such a feature for textures—image textures that form a highly oriented “flow” in the image strongly suggest that the underlying surface texture is oriented. For planar surfaces, the highly oriented textures appear to converge at a vanishing point. For curved, developable surfaces, homogeneous, oriented textures project to texture flows that have a generalized form of parallelism in the image. Textures like these form what is akin to a special subclass of textures. These flow-like textures are conducive to qualitatively different interpretation strategies than are images of isotropic textures. In particular, they could efficiently be interpreted using a shape-from-contour like mechanism rather than a texture gradient-based method (Knill, 2001; Li & Zaidi, 2000). For

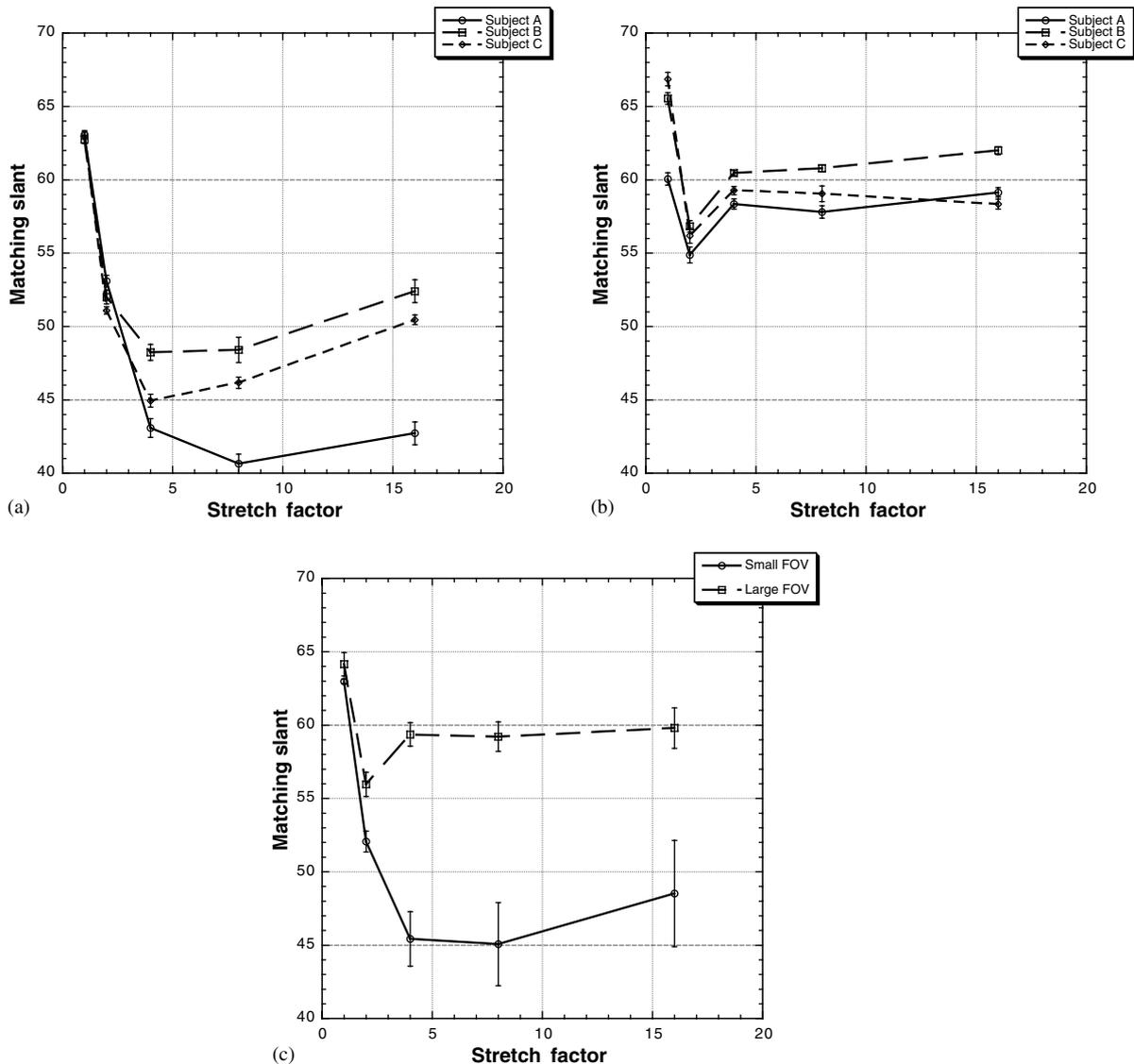


Fig. 17. Plots of subjects' points of subjective equality between test stimuli containing images of stretched textures and comparison stimuli containing images of isotropic textures, as a function of the stretch factor: (a) results from Experiment 4 (small FOV), (b) results of Experiment 5 (large FOV) and (c) the averaged PSEs of subjects in the two experiments.

curved surfaces, images of highly oriented texture flows are often much more informative than images of isotropic textures (Li & Zaidi, 2000).

### Acknowledgement

This work was supported by NIH grant NEI-EY09383 to the author.

### Appendix A

In this appendix, we derive the likelihood functions for a nested pair of hidden prior models when the joint likelihood function on primary and secondary scene

variables is Gaussian. For simplicity, we treat the case in which both the primary and secondary scene variables are scalar quantities (the  $n$ -dimensional case generalizes easily from the bivariate case). The joint likelihood function is a bivariate Gaussian,

$$p(\vec{I}|S_p, S_s) = \mathcal{L}_m \exp \left[ -\frac{1}{2} (\vec{S} - \vec{\mu})^T \Sigma^{-1} (\vec{S} - \vec{\mu}) \right], \quad (\text{A.1})$$

where  $\mathcal{L}_m$  is the maximum value of the joint likelihood function,  $\vec{S}$  is the vector containing both primary and secondary scene variables,  $\vec{S} = [S_p, S_s]^T$ , and  $\vec{\mu}$  is the mean of  $\vec{S}$ ,  $\vec{\mu} = [\mu_p, \mu_s]^T$ .  $\Sigma$  is the covariance matrix of the joint likelihood function, given by

$$\Sigma = \begin{bmatrix} \sigma_p^2 & \rho \sigma_p \sigma_s \\ \rho \sigma_p \sigma_s & \sigma_s^2 \end{bmatrix}, \quad (\text{A.2})$$

where  $\sigma_p^2$  is the variance of  $S_p$ ,  $\sigma_s^2$  is the variance of  $S_s$  and  $\rho^2$  is the covariance between the two variables.

We are interested in the likelihood function for the primary variable,  $S_p$ . This is given by marginalizing the joint likelihood function over  $S_s$ ,

$$p(\vec{I}|S_p) = \int p(\vec{I}|S_p, S_s)p(S_s) dS_s. \tag{A.3}$$

If the prior on  $S_s$  is a mixture of two models, the likelihood function becomes a mixture model,

$$p(\vec{I}|S_p) = \phi_1 p_1(\vec{I}|S_p) + \phi_2 p_2(\vec{I}|S_p) \tag{A.4}$$

$$p(\vec{I}|S_p) = \phi_1 \int p(\vec{I}|S_p, S_s)p_1(S_s) dS_s + \phi_2 \int p(\vec{I}|S_p, S_s)p_2(S_s) dS_s, \tag{A.5}$$

where  $\phi_i$  is the prior probability of model  $i$  and  $p_i(S_s)$  is the prior density function on  $S_s$  that is associated with model  $i$ . When  $S_s$  is a scalar, a constrained model nested within an unconstrained model has  $S_s$  fixed at a particular value. Without loss of generality, let  $S_s = 0$  for the constrained model. The likelihood function then becomes

$$p(\vec{I}|S_p) = \phi_1 p(\vec{I}|S_p, S_s = 0) + \phi_2 \int p(\vec{I}|S_p, S_s)p(S_s) dS_s. \tag{A.6}$$

The likelihood function for the constrained model is simply the slice through the joint likelihood function at  $S_s = 0$ , given by

$$p_1(\vec{I}|S_p) = p(\vec{I}|S_p, S_s = 0) = \mathcal{L}_m e^{-\mu_s^2/2\sigma_s^2} \exp \left[ \frac{-(S_p - (\mu_p - \rho\sigma_p\mu_s/\sigma_s))}{2(1 - \rho^2)\sigma_s^2\sigma_p^2} \right]. \tag{A.8}$$

In order to compute the likelihood function for model 2 (the unconstrained model), we require a prior density function for  $S_s$ . A uniform prior leads to the model likelihood function

$$p_2(\vec{I}|S_p) = \int p(\vec{I}|S_p, S_s)p_2(S_s) dS_s = \mathcal{L}_m \frac{1}{\Delta} \int_{S_0}^{S_0+\Delta} \exp \left[ -\frac{1}{2}(\vec{S} - \vec{\mu}_s)^T \Sigma^{-1}(\vec{S} - \vec{\mu}_s) \right] \times dS_s, \tag{A.9}$$

where  $S_0$  is the lower bound on  $S_s$ . Assuming that the joint likelihood function is well within the bounds,  $[S_0, S_0 + \Delta]$  (i.e.,  $\Delta \gg \sigma_s$ ), we can evaluate the integral as

$$p_2(\vec{I}|S_p) \approx \mathcal{L}_m \frac{1}{\Delta} \int_{-\infty}^{\infty} \exp \left[ -\frac{1}{2}(\vec{S} - \vec{\mu}_s)^T \Sigma^{-1}(\vec{S} - \vec{\mu}_s) \right] \times dS_s \tag{A.11}$$

$$\approx \mathcal{L}_m \frac{1}{\Delta} \sigma_s \sqrt{2\pi(1 - \rho^2)} \exp \left[ -\frac{(S_p - \mu_p)^2}{2\sigma_p^2} \right]. \tag{A.12}$$

Comparing Eqs. (A.8) and (A.12) clarifies a number of relationships between the constrained and unconstrained models. First, the constrained model is biased by an amount  $-\rho\sigma_p\mu_s/\sigma_s$ . The bias reflects the distance between the peak of the joint likelihood function and the assumed value of  $S_s$  under the constrained model ( $S_s = 0$ ). Second, the constrained model is penalized by a goodness of fit factor that decays exponentially from 1 as the joint likelihood function moves further away from the model's assumed value of  $S_s$ . Third, the likelihood function for the unconstrained model is penalized by a constant Occam's factor that depends only on the spread of the joint likelihood function and the prior uncertainty ( $\Delta$ ) in  $S_s$ . Unlike the goodness of fit factor, the Occam's factor does not depend on the mean of the joint likelihood function. Finally, the variance of the likelihood function for the constrained model is smaller by a factor of  $1 - \rho^2$  than the likelihood function for the unconstrained model. This factor determines how much information is gained by accepting the more constrained model.

The relative heights of the model likelihood functions depend on the relative values of the goodness of fit factor and the Occam's factor. Writing the peak values of the two likelihood functions as  $\mathcal{L}_1$  and  $\mathcal{L}_2$ , we have for the ratio of peak constrained likelihood vs. peak unconstrained likelihood

$$\frac{\mathcal{L}_1}{\mathcal{L}_2} = \frac{\Delta e^{-\mu_s^2/2\sigma_s^2}}{\sigma_s \sqrt{2\pi(1 - \rho^2)}}. \tag{A.13}$$

An estimator that computes the mean of the mixed likelihood function<sup>5</sup> computes a weighted average of the means of the component model likelihood functions. The ratio of the weights is proportional to the ratio of the areas under the two likelihood functions (Eq. (8)). The relative weights for the constrained and unconstrained interpretations in the combination rule are given by

$$\frac{w_1}{w_2} = \frac{\Delta e^{-\mu_s^2/2\sigma_s^2}}{\sqrt{2\pi}} \tag{A.14}$$

When the joint likelihood function peaks at a distance of  $0.1\sigma_s$  away from the value of  $S_s$  assumed in the constrained model, the ratio of weights is  $0.38\Delta$ . When the

<sup>5</sup> Alternative estimators will show similar qualitative properties to the mean, when considered on average.

peak deviates from the constrained value by a distance of  $4\sigma_s$ , the ratio shrinks to  $0.05\Delta$ .

## References

- Aggarwal, J. K., Cai, Q., Liao, W., & Sabata, B. (1998). Nonrigid motion analysis: articulated and elastic motion. *Computer Vision and Image Understanding*, *70*, 142–156.
- Bakshi, S., & Yang, Y. H. (1997). Towards developing a practical system to recover light, reflectance and shape. *International Journal of Pattern Recognition*, *11*(6), 991–1022.
- Blake, A., & Marinos, C. (1989). Shape from texture: estimation, isotropy and moments. *Oxford University Technical Report*, OUEL 1774/89.
- Blake, A., Bulthoff, H. H., & Sheinberg, A. (1993). Shape from texture: ideal observers and human psychophysics. *Vision Research*, *33*(12), 1723–1737.
- Brady, M., & Yuille, A. L. (1984). An extremum principle for shape from contour. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-6*(3), 288–301.
- Buckley, D., Frisby, J., & Blake, A. (1996). Does the human visual system implement an ideal observer theory of slant from texture? *Vision Research*, *36*(8), 1163–1176.
- Cutting, J., & Millard, R. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General*, *113*(2), 198–216.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature Neuroscience*, *4*(5), 429–433.
- Freeman, W. T. (1996). Exploiting the generic viewpoint assumption. *International Journal of Computer Vision*, *20*, 243–261.
- Garding, J. (1992). Shape from texture for smooth curved surfaces in perspective projection. *Journal of Mathematical Imaging and Vision*, *2*(4), 327–350.
- Garding, J. (1995). Surface orientation and curvature from differential texture distortion. In *Proceedings of 5th international conference on computer vision*, Cambridge, MA, pp. 733–739.
- Gharamani, Z., Wolpert, D. M., & Jordan, M. I. (1997). Computational models of sensorimotor integration. In P. S. Morasso, & V. Sanguineti (Eds.), *Self-organization computational maps and motor control*. Amsterdam: North-Holland Press.
- Gibson, J. J. (1950). The perception of visual surfaces. *American Journal of Psychology*, *63*, 367–384.
- Horaud, R., & Brady, M. (1988). On the geometric interpretation of image contours. *Artificial Intelligence*, *37*, 333–353.
- Ikeuchi, K., & Horn, B. K. P. (1981). Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, *17*, 141–184.
- Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences*, *6*, 345–350.
- Jepson, A., Richards, W., & Knill, D. C. (1996). Modal structure and reliable inference. In D. C. Knill, & W. Richards (Eds.), *Perceptio as Bayesian inference*. Cambridge, England: Cambridge University Press.
- Kanade, T. (1981). Recovery of the three-dimensional shape of an object from a single view. *Artificial Intelligence*, *17*, 409–460.
- Knill, D. C. (1992). Perception of surface contours and surface shape: from computation to psychophysics. *Journal of the Optical Society of America A*, *9*(4), 1449–1464.
- Knill, D. C. (1998a). Discriminating surface slant from texture: comparing human and ideal observers. *Vision Research*, *38*, 1683–1711.
- Knill, D. C. (1998b). Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Research*, *38*, 2635–2656.
- Knill, D. C. (1998c). Surface orientation from texture: Ideal observers, generic observers and the information content of texture cues. *Vision Research*, *38*, 1655–1682.
- Knill, D. C. (2001). Contour into texture: the information content of surface contours and texture flow. *Journal of the Optical Society A*, *18*(1), 12–36.
- Koenderink, J. J., & van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A*, *8*, 377–385.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, *35*(3), 389–412.
- Langer, M. S., & Zucker, S. W. (1994). Shape from shading on a cloudy day. *Journal of the Optical Society of America A*, *11*, 467–478.
- Li, A., & Zaidi, Q. (2000). Perception of three-dimensional shape from texture is based on patterns of oriented energy. *Vision Research*, *40*, 217–242.
- Li, R., Maloney, L. T., & Landy, M. S. (1997). Combination of consistent and inconsistent depth cues. *Investigative Ophthalmology and Visual Science*, *38*(4), 4229–4229.
- Mackay, D. J. C. (1992). Bayesian interpolation. *Neural Computing*, *4*, 415–447.
- Malik, J., & Rosenholtz, R. (1995). Recovering surface curvature and orientation from texture distortion: a least squares algorithm and sensitivity analysis. In *Lecture notes in computer science: Vol. 800. Proceedings of 3rd European conference on computer vision* (pp. 353–364). Springer-Verlag.
- Pentland, A. P. (1990). Linear shape from shading. *International Journal of Computer Vision*, *4*, 153–163.
- Richards, W., Jepson, A., & Feldman, J. (1996). Priors, preferences and categorical percepts. In D. C. Knill, & W. Richards (Eds.), *Perceptio as Bayesian inference*. Cambridge England: Cambridge University Press.
- Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge, England: Cambridge University Press.
- Rosenholtz, R., & Malik, J. (1997). Shape from texture: isotropy or homogeneity (or both)? *Vision Research*, *37*(16), 2283–2294.
- Stevens, K. A. (1981). The visual interpretation of surface contours. *Artificial Intelligence*, *17*, 47–73.
- Todd, J., & Akerstrom, R. (1987). Perception of three-dimensional form from patterns of optical texture. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 242–255.
- Ullman (1979). The interpretation of structure form motion. *Proceedings of the Royal Society of London Series B*, *203*, 405–426.
- Weiss, I. (1988). 3D shape representation by contours. *Computer Vision, Graphics and Image Processing*, *41*, 80–100.
- Witkin, A. P. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence*, *17*(1), 17–45.
- Yuille, A., & Bulthoff, H. (1996). In D. C. Knill, & W. Richards (Eds.), *Perceptio as Bayesian inference*. Cambridge, England: Cambridge University Press.
- Yuille, A. L., & Clark, J. J. (1993). Bayesian models, deformable templates and competitive priors. In L. Harris, & M. Jenkin (Eds.), *Spatial vision in humans and robots*. Cambridge, England: Cambridge University Press.