Shape from Texture: Ideal Observers and Human Psychophysics

ANDREW BLAKE,* HEINRICH H. BÜLTHOFF,†‡ DAVID SHEINBERG†

Received 30 October 1991; in revised form 14 July 1992; in final form 8 January 1993

We describe an ideal observer model for estimating "shape from texture" which is derived from the principles of statistical information. For a given family of surface shapes, measures of statistical information can be computed for two different texture cues—density and orientation of texels. These measures can be used to predict lower bounds on the variance of shape judgements of "ideal" and human observers. They can also predict optimal weights for cue integration for the inference of shape from texture. These weights are directly proportional to the information carried by each cue. The ideal observer model therefore predicts that the variance of subjects' responses in a psychophysical shape judgement task should reflect the statistical importance of individual texture cues. Our results show that human performance in shape judgements for a one-parameter family of parabolic cylinders is often better than what an ideal observer achieves using a density cue alone. Therefore other information, for example the compression cue, must be used by human observers. For the first time, such results have been obtained without recourse to the unnatural cue conflict paradigms used in previous experiments. The model makes further predictions for the perception of planar slanted surfaces in the case of wide field of view.

1. INTRODUCTION

Texture cues in the image plane are a potentially rich source of surface information available to the human observer. A photograph of a cornfield, for example, can give a compelling impression of the orientation of the ground plane relative to the observer. Gibson (1950) designed the first experiments to test the ability of humans to use this texture information in their estimation of surface orientation. Since that time, various authors have proposed and tested hypotheses concerning the relative importance of different visual cues in human judgements of shape from texture (Cutting & Millard, 1984; Todd & Akerstrom, 1987). This work has generally relied on a cue conflict paradigm in which one cue is varied while the other is held constant. This is potentially problematic, since surfaces with conflicting texture cues do not occur in nature. It is possible that in a psychological experiment our visual system might employ a different mechanism to resolve the cue conflict condition. We show in this paper that the strength of individual texture cues can be measured and compared with an ideal observer model without resorting to a cue conflict paradigm.

Ideal observer models for estimation of shape from texture have been described by Witkin (1981), Kanatani and Chou (1989), Davis, Janos and Dunn (1983), Blake and Marinos (1990) and Marinos and Blake (1990). Given basic assumptions about the distribution and orientation of texture elements, an estimate of surface orientation can be obtained, together with crucial information about reliability of the estimate. The estimated reliability constitutes a theoretical best performance from the given visual information. This is the sense in which the model is a statistical "ideal observer" and is similar to the approach used by Barlow (1980) in the context of detecting symmetry in random-dot patterns and by Kersten (1990) for the computation of scene descriptions.

1.1. Texture cues for shape

In this paper, we primarily address the perception of statistical textures, as opposed to the regular patterns used in some previous studies (Stevens, 1981; Stevens & Brookes, 1988; Buckley, Frisby & Mayhew, 1989; Buckley & Frisby, 1993). We assume a texture model in which independent elements—texels—can be isolated. Voorhees and Poggio (1988) have shown that such an assumption is computationally plausible. Furthermore we assume oriented texels, elements with an in-built direction that is a material property, not an artifact of projection. Thin line elements, for instance, meet this requirement because the direction of the projected line in the image is a viewpoint-invariant feature. Whatever
the vantage point, the image line-element direction is the projection of a fixed direction on the surface, namely the direction of the surface line-element. Ellipses, however, do not meet this requirement. They have no obvious viewpoint-invariant features in the image-plane. For instance, the direction of the major axis of the projected ellipse is not viewpoint-invariant, since it is a function both of the shape and orientation of the original ellipse and also of the viewpoint.

Following Cutting and Millard (1984) we acknowledge three textural cues to shape: compression, density, and perspective (see also Table 1). In the following, we describe these three cues; more formal definitions are deferred until Section 2.

1.1.1. Compression. Compression has traditionally referred to the ratio of the width to the length of individual texture elements. This could, in principle, be a local cue for surface orientation if the shape of a texel were known a priori. Recent psychophysical evidence, however, has shown that human observers can estimate shape from texture without prior knowledge of texel shape. Nevertheless, such judgements were accurate only when the orientation of the compressed elements was consistent with projective geometry (Todd & Akerstrom, 1987).

In our model, compression refers solely to the tendency of texel orientations to align perpendicular to surface orientation (e.g. the horizon for planar surfaces). More precisely, the compression cue is defined to be the spatial distribution of observed texel orientations in the image plane. This definition can therefore apply equally well to surfaces textured by elements with no measurable area (e.g. line elements). The only necessary assumption for this model is that the distribution of element orientations is isotropic. Based on this assumption, we can compute the maximum information available in the projected texture pattern resulting from this cue, without needing to specify how it is extracted by the human observer.

1.1.2. Density. We define the density cue as the spatial distribution of observed texel centres in the image plane. The density cue is observable under perspective projection and also, generally, under orthographic projection [see Horn (1986) for definitions of perspective and orthographic projection]. The exception in the latter case is that, on a planar surface, the density cue provides no information about shape (i.e. the orientation of the planar surface). In order to use the density as a cue for surface orientation, it is necessary to make some assumption about the prior, statistical distribution of texture elements over the object surface. The natural default is a priori assumption of homogeneity: that the probability density function for the positions of texel centres on the surface is uniform, but of unknown magnitude. Any observed gradients of image density (Gibson, 1979) must then be due to projective effects and hence be dependent on local surface orientation relative to the line of sight. In the case of a plane under orthographic projection, the homogeneity assumption also leads to a uniform probability density for the positions of texel centres in the image. The orientation of the plane affects only the magnitude of the density, which does not serve as a cue given the assumption that the magnitude of the surface density is a priori unknown.

In contrast, when compression is spatially uniform, it is an observable cue. Even in the absence of any gradients of compression, the degree of compression is a valid cue to shape which has however been neglected (Cutting & Millard, 1984). It is a valid cue because, unlike density, for which the prior assumption is that it has some uniform but unknown value, compression is assumed, under the isotropy assumption, to be uniform and of known value: there is zero compression a priori, on the surface. Any texel compression observed in the image therefore indicates that the surface is slanted relative to the image plane, even when that compression is spatially uniform.

1.1.3. Perspective. Texture gradients due to perspective are a result of the scaling of texels inversely with
distance from the viewer. This scaling effect is additional to the effect of compression. In the limit of orthographic projection this cue disappears because distance from the viewer to points on a viewed surface is approximately constant. However, the perspective cue is difficult to model statistically because of the lack of any uniquely natural prior assumption about texel size (unlike the other two cues for which homogeneity and isotropy were the natural default assumptions).

Given that we cannot model the perspective cue statistically, how can this cue be eliminated experimentally? There are two ways to do this. First we can use a small field of view so that, approaching the limit of orthographic projection, inverse depth scaling is minimal. Second we can use thin line elements as texels, randomly orientated, for which perspective scaling cannot directly be measured.*

The properties of the three texture cues are summarized in Table 1.

1.2. Integration of texture cues

Each texture cue present in a particular image offers an “opinion” about surface shape. These opinions, each with their own reliability measure, must somehow be merged to form a joint opinion and associated reliability. Interactions between cues may either be consistent or contradictory. For example, consider viewing a golf ball with both eyes. There will be consistent depth information from stereo, shading and texture cues. Viewing an image of the same golf ball in a photograph, however, puts the stereo cues (which give constant depth for the entire photograph) into conflict with shading and texture, hence the cues are now contradictory. Psychophysicists have attempted to deal with the first case by taking weighted linear combinations with some success (Dosher, Sperling & Wurtz, 1986; Bruno & Cutting, 1988). The case of conflicting cues seems to require significant nonlinearity and is usually assumed to require a different, and independent, mechanism. For example, this case is explicitly excluded in the statistical framework for fusion of depth information proposed by Maloney and Landy (1989).

2. THE IDEAL OBSERVER MODEL

In this section we outline the framework for constructing ideal observer models for inference of shape from statistical texture. The following assumptions are made.

**Parametric surfaces:**

Visible surfaces are of the form

\[ z = f(X, Y; a) \]

where \( a = (a_1, a_2, \ldots, a_m) \) is a set of continuous parameters specifying a particular member of a certain family of surfaces and \( X, Y \) are image coordinates. An example is the two-parameter family of paraboloids of the form \( z = a_0X^2 + a_1Y^2 \). A world coordinate frame \((x, y, z)\) is defined with \((x, y)\) parallel to \((X, Y)\).

**Imaging projection:**

The surface is observed in a monocular view. In this paper, for the purposes of modelling, we assume that the field of view is restricted sufficiently that image projection is well approximated by the orthographic limit. This will simplify the mathematics of the ideal observer model.

**Surface texture:**

The texture is assumed to consist of a set of discrete, statistically independent texels, each generated in the tangent plane to the surface at a particular point. It is assumed that each texel is sufficiently small that it can be considered planar.

**Texel observation:**

The subject makes a set of observations

\[ w = (w_1, w_2, \ldots, w_N) \]

of \( N \) visible texels. Each observation \( w_s \) consists of a number of measurements of the texel in the projection of the surface onto the image. In this paper we will consider \( w_s = (X_s, Y_s, z_s) \) where \( X, Y \) denote image coordinates of the texel centre and \( z \) denotes a texel orientation defined relative to the \( X\)-axis.

**Prior distribution of texels:**

We assume that a priori distributions for the \( w_s \) are known. In principle, any prior knowledge, either general or specific, is usable. Specific distributions might be known in experiments where subjects were trained on a particular texture. Having observed numerous fronto-parallel planar instances of the texture they might learn prior distributions for texture position, orientation and shape. In this paper we consider general assumptions: first, that texel position is uniformly distributed over the area of the surface (homogeneity), and second, that texel orientation is uniformly distributed on the surface (isotropy). Note that the homogeneity assumption implies that the probability density for texel position is uniform over the surface, but that the magnitude of that density is a priori unknown.

**Posterior distributions of texels:**

The posterior (image) distributions of texel position, \( X, Y \), and orientation, \( z \), depend on the prior distributions and also on the biasing effects of projection. The posterior distribution itself is not, of course, observable. Instead the subject sees a sample, \( w \), of texels from that distribution. The posterior distribution of position is the density cue. The posterior distribution of orientation is the compression cue.
The ideal observer model uses the sample, \( w \), of image texels to estimate which surface from the allowed family of surfaces generated that particular sample of texels in the image. In the related psychophysical task, the subject observes \( w \) and is asked to estimate, by means of an adjustment task, the surface parameter \( a \).

2.1. The maximum likelihood principle

The applicability of Bayesian (MAP) estimation as a model for “shape from texture” was first noted by Witkin (1981). He applied it to the problem of estimating the orientation of an isotropically textured plane under orthographic projection.* Maximum likelihood estimation (MLE) differs from Bayesian estimation in that it depends only on texel distributions (compression, density) not on the prior distribution of surface orientation. In the planar case, for instance, this means that it is no longer necessary to attempt the difficult task of modelling the probability distribution for the orientation of planes (including possible bias towards ground planes etc.). Asymptotically, in the limit that many texels are visible, the two kinds of estimator are identical. That is, provided sufficiently many texels are visible, the prior distribution of surface orientation becomes negligible in comparison to the prior distribution (e.g. isotropy and homogeneity) of texel observables.

An MLE estimator can be regarded in either of the following two senses. It can be treated as an ideal observer model, appealing to standard results (Rao, 1973) showing that asymptotically the MLE is a best estimator in the sense of achieving the minimum possible variance—the so-called “Cramer–Rao lower bound”. This is the principle sense in which we use MLE models in this paper, as a laboratory standard of statistical efficiency against which to compare experimental results. Alternatively the MLE can be regarded as a candidate biological model for visual perception of shape. It has been shown (Blake & Marinos, 1990) that, in the case of orthographic views of planes, the MLE for planar orientation can be reduced to the calculation of a certain second moment of texel orientations. This can be represented as a linear network with oriented edge-detectors at the input and units value-coded for tilt at the output. Biological embodiment of the MLE need not therefore be ruled out on grounds of excessive complexity.

2.2. Maximum likelihood estimator for shape

An MLE for shape is constructed by means of a log-likelihood function \( L(w|a) \) which is the logarithm of the posterior probability of having observed a particular set of texels \( w \) given that the surface parameter was \( a \). The value \( a \) that maximizes \( L \) is the estimated shape. In turn \( L \) is constructed as a sum over all visible texels:

\[
L(w|a) = \sum_{n=1}^{N} \log p(w_n|a),
\]

where \( p(w_n|a) \) is the posterior probability of having observed a particular individual texel \( w_n \) given that the surface parameter had the value \( a \).

What is of particular interest is the case that each texel observation \( w_n \) contains more than one independent component. How are the individual components combined? We model a texel observable as \( w = (X, Y, z) \) which includes the density cue in terms of texel position \((X, Y)\) and the compression cue in terms of orientation \(z\). The effect of the two cues can be factorized:

\[
p(w|a) = p_D(X, Y|a)p_C(z|X, Y, a),
\]

where \( p_D \) and \( p_C \) are posterior distributions for single cues, density and compression respectively. Now the log-likelihood \( L \) splits additively into two components

\[
L(w|a) = L_D(w|a) + L_C(w|a),
\]

where

\[
L_D(w|a) = \sum_{n=1}^{N} \log p_D(X_n, Y_n|a)
\]

and

\[
L_C(w|a) = \sum_{n=1}^{N} \log p_C(z_n|X_n, Y_n, a).
\]

The two components \( L_C \) and \( L_D \) are log-likelihoods for the two cues treated individually over the visible texture.

2.3. Fisher information

The Fisher information contained in the observable texture \( w \) on the parameter \( a \) is defined to be (Rao, 1973)

\[
I(a) = -E[\delta^2 L(w|a)/\delta a^2]
\]

where \( E[\cdot] \) denotes expected value.† It measures information in the sense that \((1/N)^{-1} I(a)\) is a lower bound (Cramer–Rao) on the variance‡ of any unbiased estimator of \( a \). Asymptotically, the maximum likelihood estimator \( a \) reaches this bound and hence a “best” estimator.

In the two-cues case, Fisher information \( I \) like the log-likelihood \( L \), is additive so that:

\[
I(a) = I_D(a) + I_C(a).
\]

Since Fisher information is an expected value, it is calculated by integrating over all possible values of observable features. For instance, for the density cue,

\[
I_D(a) = -\int_{\text{image}} p_D(X, Y|a) \times (\delta^2 \log p_D(X, Y|a)/\delta a^2) \ dX \ dY.
\]

Similarly, for the compression cue,

\[
I_C(a) = -\int_{\text{image}} \int_0^{2\pi} p_C(z|X, Y, a) \times (\delta^2 \log p_C(z|X, Y, a)/\delta a^2) \ dz \ dX \ dY.
\]

* A good review of MAP, MLE and other relevant statistical concepts including the “Cramer–Rao lower bound” is given by Papoulis (1990).
† \( \delta / \delta a \) denotes differentiation with respect to each of the components of the vector \( a \). For instance, given a scalar function \( f \), its derivative is defined to be the following vector: \( \delta f / \delta a = (\delta f / \delta a_1, \ldots , \delta f / \delta a_n) \).
‡ When \( a \) has more than one component, \( I \) is a matrix and \( I^{-1} \) denotes the matrix inverse. In this paper we consider cases in which \( a \) is a single parameter and hence \( I \) is a scalar.
In each case, as expected, Fisher information is proportional to the number \( N \) of visible texels. Hence the variance bound is inversely proportional to \( N \), a familiar statistical property.

Later, in the experimental section of the paper, we compare variances of subjects' responses against theoretical lower bounds for density alone, compression alone, and both together. These bounds are based on \( I_D \), \( I_C \) and \( I \) respectively. For instance, a subject whose variance is significantly less than that computed from \( I_D \) cannot have been using the density cue alone. A subject whose variance is significantly below the one computed from \( I \) must have used extraneous information other than what was contained in the texture cues.

### 2.4. The role of statistical information in cue combination

Given that \( I_D \) and \( I_C \) are natural measures of the information present in density and compression cues it seems reasonable that they should be used when cues are combined to indicate relative reliability. In fact the optimal combination—in the sense of having the lowest overall variance—does indeed mix estimates based on each of the two cues individually, in the ratio of their respective information measures (Rao, 1973). Thus, given estimates \( \hat{a}_D, \hat{a}_C \) from individual cues, assumed unbiased, the best combined estimate from both cues is

\[
\hat{a} = \frac{\hat{a}_C I_C(w|\hat{a}_C) + \hat{a}_D I_D(w|\hat{a}_D)}{I_C(w|\hat{a}_C) + I_D(w|\hat{a}_D)}.
\]

If we hypothesise that human vision approaches this optimal strategy for cue combination we can make certain predictions about psychophysical performance. For instance, we can predict theoretically when one cue should be stronger than another and look for corresponding experimental results.

### 3. THEORETICAL PREDICTIONS

The previous section reviewed the calculation of Fisher information and the consequent bounds on estimator variance given the posterior probability density functions \( p_C \) and \( p_D \). These functions depend on the family of textured surfaces being viewed. In this section we show first how to construct the functions given a family of surfaces. Then we can compute Fisher information functions for the cases of parabolic cylinders viewed orthographically and planes viewed in perspective. These are used to predict performance in terms of variance in subjects' responses in estimation tasks and to predict the relative dominance of the compression and density cues under various conditions.

#### 3.1. Slant and tilt

It transpires that the probability density functions \( p_C \) and \( p_D \) can be defined concisely in terms of slant \( \sigma \) and tilt \( \tau \). Slant and tilt are used commonly to parameterize the orientation of a planar surface (see e.g. Witkin, 1981). They are simply polar coordinates for the direction of the plane normal, relative to the line of sight. Slant is the angle by which the surface dips away from the frontal plane and tilt is the direction in which the dip takes place (Marr, 1982). For a non-planar surface \( z = f(X, Y, a) \), the tangent plane at each point on the surface has its own slant and tilt. They can be evaluated from the local surface gradient

\[
\nabla f = (\partial f/\partial X, \partial f/\partial Y).
\]

Slant is given by:

\[
\sec \sigma = \sqrt{1 + (\partial f/\partial X)^2 + (\partial f/\partial Y)^2}
\]

and tilt is defined by:

\[
\tan \tau = \frac{\partial f/\partial Y}{\partial f/\partial X}.
\]

#### 3.2. Posterior probability function for the density cue

The probability density function \( p_D(X, Y, a) \) for the density cue is relatively simply derived. Applying the homogeneity assumption, the probability of finding a texel within some small patch of the image is proportional to the backprojected area of patch on the surface \( z = f(X, Y, a) \). Assuming orthographic projection

\[
p_D(X, Y, a) = \frac{\sec \sigma}{A}
\]

where the normalizing constant is

\[
A = \iint_{\text{image}} \sec \sigma \, dX \, dY,
\]

the total visible surface area. This expression for \( p_D \), for a particular family of surfaces, can now be inserted into equation (3) to compute the Fisher information function \( I_D(a) \) associated with the density cue.

#### 3.3. Posterior probability function for the compression cue

The compression cue is a little more complex, involving the biasing of the texel direction away from the gradient direction of the surface. Witkin (1981) shows that the biased posterior probability distribution \( p_C \) is then given by:

\[
p_C(\sigma|\sigma, \tau) = \frac{\sec \sigma}{2\pi \sin(\sigma - \tau) + \sec \sigma \cos(\sigma - \tau)}.
\]

Now \( \sigma, \tau \), representing local surface gradient, are functions of the surface parameter, \( a \), and of position on the surface which is a function of image location \( X, Y \). Thus \( p_C \) can be used in equation (4) to compute the Fisher information function \( I_C(a) \) associated with the compression cue.

#### 3.4. Parabolic cylinders in a narrow field of view

The preceding equations (3) and (4) form, together with equations (9) and (11), a procedure to evaluate the Fisher information functions \( I_C \) and \( I_D \). Using MATHEMATICA, a functional programming language, a surface family is defined by defining the form of the function \( f(.,.) \) which is then passed as a parameter to the procedure that evaluates \( I_C, I_D \). Because the language...
FIGURE 1(a, b). Caption on facing page.
SHAPE FROM TEXTURE

FIGURE 1. Textured parabolic cylinders (a), viewed perpendicularly to the cylinder axis (b), are used to test subjects' variance of response and cue-integration performance. The family is indexed by the parameter $e$ which specifies the "elongation" of the parabolic cylinder (c)—also see text.

includes symbolic calculus, the formulae (3) and (4) for cue-information can be programmed directly, with the integrals and partial derivatives left in symbolic form.

Experiments in the following section relate to a family of textured parabolic cylinders (Fig. 1) with vertical axes. These are surfaces of the form $z = f(x, y; e)$ where

$$f(x, y; e) = e(1 - x^2), \quad -1 \leq x \leq 1. \quad (12)$$

The $z$-axis points towards the viewer so that the surfaces appear convex. The parameter $e$ is the "elongation", controlling surface relief. Values of the parameter fall, for practical reasons, in the range $0 < e < 2.5$ as in Fig. 1(b).

Having defined a surface family, we can finally compute Fisher information functions for the two texture cues, both separately and jointly. The results are shown in Fig. 2, plotted as lower bounds on the standard deviations of subjects' responses (in units of elongation), as a function of the true elongation $e$ of the presented stimulus. Following the definition of the Cramer–Rao bound in the previous section, the plotted lower-bounds on standard deviation are $1/\sqrt{I_D(e)}$, $1/\sqrt{I_C(e)}$, and $1/\sqrt{I_D(e) + I_C(e)}$ for density, compression and both, respectively.

A number of predictions can be made on the basis of the graphs of Fig. 2.

1. Under the earlier hypothesis that cues might be integrated linearly, weighted in proportion to their statistical information, it is natural to define "dominance" of one cue over another to be the ratio of respective information measures (or, equivalently, the inverse square of the C.R. lower bounds on standard deviation). Figure 2(c) shows that compression dominates density ($I_C/I_D > 1$) for the family of parabolic cylinders.

2. The dominance of compression over density is least pronounced for zero elongation (planar), becoming more pronounced as elongation $e$ increases.

3. Since predicted standard deviations are significantly higher for the density cue alone than for the other two conditions, it should be possible to

*Note, that $e$ (elongation) is a single real number in this specially simple case of the family of parabolic cylinders. Thus the parameter vector $a$ mentioned earlier reduces now to a vector with one component $e$ only. From here on we refer to $e$ in place of $a$. 

FIGURE 2. The theory of maximum likelihood estimation and Fisher information is used to generate lower bounds on subjects' standard deviation in estimating the "elongation" of textured parabolic cylinders. (a, b) Predictions are shown for density cue only, compression cue only and both cues together. This is done for two texture densities, $\rho = 300, 450$, the difference between them being simply due to the $1/\sqrt{\rho}$ scaling of the standard deviations. (c) The information ratio $I_C/I_D$ measures the strength of the compression cue relative to the density cue. Compression is always stronger, its dominance over density becoming more pronounced as elongation $e$ increases.
test experimentally whether or not the compression cue may have been used. If standard deviations of responses fall below the lower bound for density alone some further information must have been used.

4. One might not expect to be able to distinguish from standard deviations of responses whether or not the density cue is being used. This is because the "compression only" and "both cues" curves in Fig. 2(a, b) are relatively close throughout the experimentally practicable range of elongations.

4. PSYCHOPHYSICAL RESULTS

Two types of experiments have been performed to test hypotheses arising from the ideal observer model for perception of textured parabolic cylinders. First, we have measured the standard deviations of observers' judgements of shape. These have been compared with ideal observer predictions. Second, following Cutting and Millard (1984) and Todd and Akerstrom (1987) we have measured responses to stimuli in which the compression and density cues are in conflict. We compare our results (Sheinberg, Bülthoff & Blake, 1990) with predicted cue-dominance based on statistical information for the ideal observer. In each case stimuli were generated using real-time graphics and their shape was estimated by the method of adjustment, after Bülthoff and Mallot (1988, 1990).

4.1. Stimulus generation

A Stardent GS1000 graphics mini-supercomputer was used for stimulus generation and display, together with a pair of electro-optic shutter glasses (StereoGraphics Corp.) for stereoscopic viewing. Subjects viewed a pair of computer generated surfaces (Fig. 1), arranged side by side. Each surface was defined in terms of vertically oriented parabolic cylinders, resting against a similarly textured, planar background. One surface was the textured stimulus, and was seen monocularly by blocking the view of one eye with a black occluder. The other surface, serving as a probe, was textured and shaded, and viewed stereoscopically. In order to achieve a visual match with the stimulus, the subject varied the shape of the probe by clicking the mouse. The probe was constructed in a square field of view approx. 120 x 120 mm. A sequence of approx. 30 surfaces, spanning 0.1 ≤ e ≤ 2.5, was precomputed for use as the subject-controlled probe. The stimulus occupied either a square field, similar to the probe's, or a slit approx. 120 mm horizontally by 50 mm vertically. In all experiments the viewing distance was 95 cm.

Discrete texels were generated at randomly sampled points on a parabolic surface with elongation e. This was done such that the probability of finding a texel centre in a particular small patch was proportional to dA, the area of the patch on the surface. This is equivalent to forming the parabolic cylinder surface from a flat sheet that has been exposed to a uniform "rain" of texels. Exposure was controlled to reach a specified mean density, ρ, which was the mean number of texels that would appear in the square field of view for a fronto-parallel planar surface.

Various texel shapes could be generated, the standard one being a line element whose length varied randomly, according to a uniform distribution, over a 2:1 range. This was done to remove the possibility of the projected element length acting as a deterministic shape cue.* The line element was defined in three dimensions to lie on the surface of a parabolic cylinder with elongation e. Element orientation was selected randomly, distributed uniformly in the tangent plane to the parabolic cylinder. Note that, in the case of the stimulus, e_c and e_p could be set independently. Choosing them to be unequal was the mechanism for generation of conflicting density and compression cues. For the probe, however, texture cues never conflicted and also were consistent with shading and stereoscopic cues.

Once the three dimensional texels were selected, by random sampling as above, they were projected onto the stereoscopic image. This is done in such a way (Bülthoff & Mallot, 1988) that the projection remains correct regardless of fixation point, given that the subject's head is located in a fixture (Fig. 3).

4.2. Experiment 1

The aim of the first experiment was to compare measured standard deviations in adjusted elongation e against the ideal observer model presented in the theoretical section. Stimuli were generated at random from the family of parabolic cylinders described earlier, with elongations lying in the range 0 ≤ e ≤ 2. Texture cues were consistent, so that e_c = e_p, and there was no shading cue—texels were superimposed on a plain-coloured background. Shape judgements are harder to make in the absence of shading but with the benefit that

*There remains however a possibility that the uniform distribution might be used by the observer as a prior distribution from which a statistical shape cue could be obtained. This possibility is yet to be investigated theoretically and experimentally.

FIGURE 3. Imaging geometry. Projection onto the x-z plane. Viewing distance is 95 cm. Nodal points of left and right eyes are e_l, e_r respectively, separated by 6.5 cm. A point p is imaged onto the screen at p_l for the view from the left eye and at p_r for the view from the right eye.
Results are shown in Fig. 4 for five subjects: ISA and GFB were naive, AB, HHB, DLS were not.* A reduced field horizontal slit was used for the stimuli since statistical efficiency is likely to be impaired by an excessively large field of view. Measured standard deviations with confidence intervals are plotted against the predictions of the ideal observer model. The predictions are essentially those based on the Cramer–Rao lower bounds shown in Fig. 2. There is one important difference however. Cramer–Rao bounds apply to unbiased estimators. Subjects’ responses prove not to be wholly unbiased however. Means often show some regression towards fronto-parallel, a phenomenon that was demonstrated by Büllthoff and Mallot (1988, 1990). Thus data analysis includes a linear fit for mean adjusted elongation vs true elongation, whose slope is defined to be the bias b. Lower bounds on standard deviations must then be multiplied by b. Note that the linear fit was checked by a χ² test and sessions which failed the test were rejected (11 sessions out of 45, and never more than two per subject, failed the linearity test). Statistical error in the linear fit is also taken into account in estimating b and hence in deriving lower bounds on standard deviation of elongation.

The standard deviations (Fig. 4) for all subjects range approximately from 1 to 2 times the Cramer–Rao lower bounds for the combined cues. This corresponds to an efficiency of use of statistical information of between 25 and 100% which is comparable with the value of 25% reported by Barlow (1980) for symmetry detection.

In some cases standard deviations were significantly below the lower bound for density information alone. In particular, it is at the largest elongation e = 2.0, that the difference between SD lower bounds for compression and density is greatest. In that case, three out of five subjects (AB, ISA, DLS) register data points that are significantly (at the 99% level) below the lower bound for the density cue. This is true for both texture densities ρ = 300, 450. The most likely conclusion is that at least some subjects are able to use the compression cue.

4.3. Experiment 2

In order to test subjects’ standard deviations as a psychophysical measure of the strength of shape cues we have to assume that our method is sensitive enough to record standard deviations below the Cramer–Rao lower bounds of the individual shape cues. Therefore we tested the sensitivity of our experimental paradigm in a control experiment. Subject HHB repeated the experiment with stereo and shading cues added back to the stimulus (Fig. 5). Under these circumstances we would predict that standard deviations could fall below the Cramer–Rao lower bounds for texture cues alone. From Fig. 5 it is apparent that standard deviations are greatly reduced relative to the standard experiment and fall below the Cramer–Rao lower bounds for texture cues. Recall that standard deviations below the Cramer–Rao lower bounds are only possible if cues other than the ones modelled in the ideal observer are employed. This control demonstrates the obvious fact that we can use other cues if they are available, and it also shows that the method is sensitive enough to record small standard deviations. Therefore possible limitations of measuring standard deviations close to the Cramer–Rao lower bounds are ruled out.

4.4. Experiment 3

Another possible methodological limitation was the use of line elements as texels—a limitation imposed by our ideal observer model. As an additional control we used circular texels instead of the standard line elements. The circles project to ellipses in the image so that, in principle, it is possible to estimate surface slant deterministically for each pixel. In other words the ideal observer for circular shapes should be exact. If the human visual system could fully use the elliptical shape cue, one would expect a great reduction in standard deviations of estimated elongations. Results of this control (Fig. 6) show that standard deviations are somewhat reduced but only by a factor of 2 or less, a lesser factor than in the previous experiment (Fig. 5).

4.5. Experiment 4

In order to compare our results with the cue conflict paradigm used in previous studies of shape from texture, we used our stimuli in a cue conflict experiment. Note that our technique of generating texture and shading on virtual surfaces allows us independent control of different shape cues by mapping texture compression or density on different surfaces. We presented monocularly viewed stimuli in a square field of view, with shading. All experiments were done with line element texels at a density of ρ = 300 with the following three conditions randomly interleaved: (1) consistent shading, density and compression, (2) consistent shading and density with “flat” compression, and (3) consistent shading and compression with “flat” density. “Flat” density or compression denote cues constructed with elongation e'e = 0, ec = 0 respectively (i.e. consistent with a plane). Each of nine conditions (three elongations each with flat density, flat compression or consistent cues) are presented six times, with random ordering, to each subject. The pooled results of seven subjects (Fig. 7) show that the compression cue dominates the density cue and that the relative dominance increases with increasing elongation.

*Naivety was not a major issue in our experiments because it is not possible to purposefully manipulate standard deviations without also affecting the mean. Systematic changes in standard deviation are then neutralized by our bias correction procedure (see text).
FIGURE 4. Caption on facing page.
FIGURE 4. Results of standard deviation measurements for judgements of the elongation of textured parabolic surfaces are shown by dots and error bars. They are plotted against theoretical predictions based on Fig. 2, which have been adjusted to take account of bias in subjects' mean responses. Results are shown for two densities $\rho = 300, 450$.

FIGURE 5. Results for a control in which the stimulus is shaded and viewed stereoscopically. Recorded standard deviations are considerably reduced compared with the same subject in Fig. 4.

FIGURE 6. The basic experiment of Fig. 4 is repeated here for circular shapes instead of line elements. Both subjects show a consistent but modest reduction in standard deviation.
5. DISCUSSION

Experiments 1 and 2 indicate that the human visual system is able to use texture compression as a cue to shape because the variance of subjects' shape judgements was often lower than that predicted by the ideal observer using the density cue alone. A significant virtue of our approach is that, unlike previous experiments (Cutting & Millard, 1984; Todd & Akerstrom, 1987), these results have been achieved without resorting to cue conflict.

Our experiment was not sufficiently sensitive however to demonstrate that both density and compression cues are being used. It seems unlikely that a sufficiently sensitive experiment could reasonably have been achieved with parabolic cylinders, because Cramer–Rao bounds for compression alone and for both cues together are so similar. Not all subjects gave conclusive results concerning compression. In the case of a subject like GFB, whose standard deviations never fall significantly below the Cramer–Rao bound for density, it is possible either that both cues are used but with relatively low efficiency, or (less plausibly) that one of the cues is used alone.

In the case of circular elements in Expt 3 standard deviations are reduced but only modestly. It is reasonable to infer that, rather than applying a deterministic procedure based on the compression of the projected circle, the visual system may be treating the circular element in a similar way to the line-element, avoiding any strong assumption of precise circularity of the native elements.

Experiment 4 tested for the relative strength of the compression and density cue when they are in conflict. Similar experiments have been performed by Todd and Akerstrom (1987) with paraboloids of rotation and square texels. They used informal measurement techniques in which subjects express parabolic elongation by comparison with a set of six templates which are effectively cross-sectional views of the paraboloids. In general, they found that the compression cue dominates the density cue over a range of elongations of $0 < \phi < 1.7$, similar to the range used in our experiments. Using a novel experimental technique (the “Apparently Circular Cylinder” test) Cumming, Johnston and Parker (1993) have also recently found dominance of compression over density. Our results show that:

1. As in Todd and Akerstrom's experiment, the compression cue dominates the density cue. This is consistent, of course, with the predictions of the ideal observer model in which (Fig. 2) the ordering of standard deviations for the two cues means that there is greater statistical information in compression than in density.

2. The ideal observer model further predicts (Fig. 2) that the relative dominance of the compression cue increases with elongation. This is reflected in the results (Fig. 7) in which the ratio of subjects' adjusted elongations for zero compression and zero density respectively is seen to increase with the elongation.
of the surface used to generate the stimulus. This increase is statistically significant at the 95% level.

3. Some considerable regression towards the fronto-parallel is evident. Significant regression towards fronto-parallel was also present in Expt 1, for which shading cues were absent. This phenomenon has been studied previously by Bülthoff and Mallot (1990).

5.1. Fisher information for planar surfaces

The experimental results for parabolic cylinders were for a small field of view and consequently predictions were based on the limiting case of orthographic projection. In that case it was found that the compression cue always dominates the density cue. However, when the field of view is larger so that perspective effects are significant, this need no longer be the case. In the case of planar surfaces we show below that, indeed, when the field of view is large, the density cue dominates.

The analysis of Fisher information can be extended, in principle, for perspective projection, to cover the case of a field of view that is wide enough that the orthographic limit does not apply. In the case of the density cue and a planar surface of variable slant $\sigma$, this has been done elsewhere (Marinos & Blake, 1990). The result is that the Fisher information is simply

$$I_{\theta}(\sigma) = 9 \frac{NM}{A} \sec^4 \sigma. \tag{13}$$

Here $M$ is a moment of inertia of the backprojected area of the image plane, about an axis through the backprojected image centre and orthogonal to the tilt direction; $A$ is the magnitude of the backprojected area. Approximating the backprojected area orthographically and assuming a square field of view of angular extent $\pm \theta$ gives

$$M \approx \frac{1}{2} A \tan^2 \theta$$

so that, approximately,

$$I_{\theta}(\sigma) = 3N \sec^4 \sigma \tan^2 \theta. \tag{14}$$

This approximation holds good roughly when $\pi/2 - 2\theta > \sigma > \theta$. An exact solution, free of any orthographic approximation, for $I_{\theta}(\sigma)$ has also been found using the symbolic algebra facilities in Mathematica. Owing to the length of the algebra* the formula is not reproduced here. However it has been used to confirm the accuracy of the approximation (14) and will also be used for comparison with $I_{\theta}(\sigma)$.

The Fisher information on slant $I_{\delta}(\delta)$ in the compression cue has been computed using a mixture of symbolic algebra and numerical integration (see below).
in compression and density cues respectively, as a function of field of view. The ideal observer weights the cues according to this ratio. A clear transition is predicted at a field of view of 10° (Blake & Marinos, 1990):

An orthographic approximation* is also available (Blake & Marinos, 1990):

\[ I_C = \frac{1}{2} \tan^2 \delta \]  

(15)

and this is also valid for \( \pi/2 - 2\theta > \delta > \theta \).

The exact forms of \( I_C(\sigma) \) and \( I_D(\sigma) \) are used to plot Cramer–Rao lower bounds on standard deviation, in Fig. 8. For \( N = 400 \) visible texels, the plots show bounds \( 1/\sqrt{\text{Ne}(\sigma)} \) and \( 1/\sqrt{\text{Nd}(\sigma)} \) on standard deviations for the compression and density cues respectively. Dependence on slant \( \sigma \) is shown for various sizes of field of view. The case of a small field of view approaches the orthographic limit in which the density cue vanishes [equation (14)] and hence the compression cue tends to dominate (its SD is lower). With larger fields of view, perspective effects contribute strongly and the density cue dominates uniformly, throughout the range of slants. Dominance of density over compression is predicted by the approximate equations (14) and (15) for fields of view exceeding about \( \pm 10^\circ \). This is consistent with the graphs (Fig. 8) which show that \( \theta = 10^\circ \) is close to the critical case—the respective information measures \( I_C \) and \( I_D \) are almost equal for slants between 20 and 70°.

5.2. Predictions for large field views

The predicted transfer of dominance from the density cue to the compression cue is a potentially testable, qualitative phenomenon. Strictly, experiments should be done with “needle” textures conforming to the homogeneity and isotropy assumptions. However the robustness of the prediction is such that one might also expect to observe it with textures which do not conform precisely, such as Fig. 9.

The graph (Fig. 8) for a \( \pm 5^\circ \) field of view shows a clear transition between dominance of density at low slant, and dominance of compression at high slant. However a complementary experiment, in which slant is varied, is probably impractical. This is because, in the region where the clearest transition of dominance occurs, the numerical value of the standard deviation of slant is too high. For a practicable number (400) of visible texels, with a field of view of \( \pm 5^\circ \), at the transitional slant of \( 10^\circ \), the standard deviation in the slant is almost 20°.

A more promising paradigm for experimentation is to keep slant fixed and vary the field of view between 0 and 20°. Equating the approximate formulae (15) and (14) indicates that, in a wide band of slants around 45°, a transition from density to compression dominance occurs at a field of view of about \( \pm 10^\circ \). This is shown clearly in exact plots of the information ratio \( I_C/I_D \) (Fig. 10). For \( \sigma = 45^\circ \), standard deviation of slant at the transitional point is small, just 5 , suggesting that experimental results would be reliable. On either side of the transition, the information ratio ranges from 10 to \( \frac{1}{10} \), indicating strong dominance by the respective cues. Again, this suggests a robust experimental effect.

6. CONCLUSION

The first conclusion is that experimental results are consistent with the ideal observer model, with typical statistical efficiencies of 25–100%. In most cases standard deviation increases with eccentricity of the cylinders, as predicted by computed Fisher information in the ideal observer model. The second conclusion is that, for a number of subjects, reliability is too good to result from density effects alone. It must be that other information, most likely information about compression, is used. For the first time (as far as we know), this has been shown without resorting to texture cue conflicts, that is, within the scope of “normal” surface viewing.

Having established plausibility for the idea that the visual system can estimate statistical information in texture cues, we return to the question of cue integration. Predictions for cue combination under orthographic projection agree with our own recent results. These show that the compression cue dominates the density cue, for parabolic cylinders in a narrow field of view. Furthermore, as predicted, the degree of dominance increases with increasing parabolic elongation. The predictions

*While \( I_C \) is approximated orthographically, in the case of \( I_D \) orthographic approximation simply yields \( I_D = 0 \) because planar density gradients vanish in the orthographic limit. Instead, only the trapezoidal backprojected area was approximated orthographically, as a rectangle, for the purposes of calculating \( M \).
also agree broadly with previous experimental results of Todd and Akerstrom (1987), using paraboloids and a qualitative rating technique.

REFERENCES


Acknowledgements—We gratefully acknowledge helpful discussions with David Buckley, Bruce Cumming, John Frisby, Elizabeth Johnston, Dan Kersten, Christopher Longuet-Higgins, Andrew Parker, Alan Yuille and Andrew Zisserman. We are grateful for support from the SERC Image Interpretation Initiative and the EEC Esprit programme. DLS was supported by an NFS predoctoral fellowship.