

# Integration of Depth Modules: Stereopsis and Texture

E. B. JOHNSTON,\*† B. G. CUMMING,\* A. J. PARKER\*

Received 28 January 1992; in revised form 8 September 1992

---

**Global shape judgements were employed to examine the combination of stereopsis and shape-from-texture in the determination of three-dimensional shape. Adding textural variations to stereograms increased perceived depth. Thus, texture was not simply vetoed by the strong stereo cue. In experiments where the depth specified by texture was incongruent with that specified by stereo, the data were well described by a weighted linear combination rule. Although only a small weight was assigned to texture, this weight was somewhat greater at a farther viewing distance. This could be a consequence of the decreased reliability of stereopsis at far viewing distances.**

Stereopsis Shape-from-texture Three-dimensional shape perception Integration of depth cues

---

## INTRODUCTION

A central problem in human visual processing is to understand how three-dimensional percepts are derived from the two-dimensional retinal images. There are many cues potentially available for specifying three-dimensional structure, such as stereopsis, structure-from-motion, texture gradients and shading. These cues have been investigated extensively in isolation (Kaufman, 1974; Marr, 1982) under the assumption that depth is processed in separate modules corresponding to the different sources of three-dimensional information. The “2½D sketch” (Marr & Nishihara, 1978) is one explicit proposal for the form of representation that could result from the combination of independent depth modules. It consists of a map of the distance and orientation of surface points with respect to the viewpoint. The different sources of depth information are processed individually to extract local measures of either the distance of a surface point from the viewer, or the local orientation of the surface with respect to the viewpoint (Ullman, 1979; Marr & Poggio, 1976; Ikeuchi & Horn, 1981; Witkin, 1981). However, the combination rule for the independent measures yielded by the different depth modules was not considered.

Recently, there has been interest in how these modules, if they are independent, could be integrated to provide a unified three-dimensional shape percept (Doshier, Sperling & Wurst, 1986; Bülthoff & Mallot, 1988; Bruno & Cutting, 1988; Rogers & Collett, 1989; Maloney & Landy, 1989; Landy, Maloney & Young, 1990; Young, Landy & Maloney, 1992). Three possibilities for cue interaction are generally considered.

The first is *vetoing*, where one “strong” cue completely overrides another “weaker” cue. This type of interaction is found where cues are strongly inconsistent. An old example is provided by stereo photographs of complex scenes. When the pairing between the viewing eye and photographs is switched, the depth fails to reverse although the sign of the disparity changes (Schriever, 1925). This demonstrates that the combination of a number of consistent pictorial cues, such as perspective, shading and texture, can veto stereopsis. A more recent example is found in the work of Bülthoff and Mallot (1988): when stereo indicates a flat surface but shading indicates an ellipsoid, no significant depth is perceived, showing that stereo can veto shape-from-shading. However, in that particular study little depth was signaled by shading alone, so it is hard to exclude some of the other combinations described below.

A second means of interaction is *weighted linear combination*. In this scheme, depth cues are first processed in separate modules. The independent depth estimates from each module are then linearly combined, with differential weights assigned to each cue. In a recent book on sensor fusion Clark and Yuille (1990) term this combination rule *weak fusion*, because the cues do not interact prior to the independent extraction of depth measures. Linear addition accounted for the data of Doshier *et al.* (1986) which examined the combination of stereo, perspective and proximity luminance covariance in disambiguating kinetic depth. In a task where subjects made judgements about the depth relations of three planes, Bruno and Cutting (1988) also found support for linear cue combination in describing the interactions of (1) motion parallax, (2) occlusion, (3) height in the picture plane, and (4) familiar size. Rogers and Collett (1989) specified a linear combination rule which described their data collected with differing combinations of motion parallax and stereo. Landy *et al.* (1990) also

---

\*University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, England.

†Present address: Sarah Lawrence College, Bronxville, NY 10708, U.S.A.

proposed linear combination in their study of kinetic depth and texture integration. They addressed the question of the weighting of different cues, and found evidence for modification of the weights depending on the reliability of the information available from each source. It is also possible that the interaction of two cues could increase the reliability of a depth estimate, without changing the mean value.

The third means of cue interaction is more complex. It is termed *strong fusion* by Clark and Yuille (1990), and involves a cooperation between cues prior to obtaining depth estimates. This may be implemented by analysing some interaction in the image (such as monitoring the rate of change of disparity in combining stereo with motion), or may take place between two relatively independent modules. Maloney and Landy (1989) discuss a specific instance of this which could aid in providing commensurate depth estimates from separate cues. This is *promotion* where the incompleteness of one depth cue is compensated for by another cue which provides information needed to yield independent depth estimates from the incomplete cue.

Another type of strong interaction is disambiguation. The information available from some depth modules (e.g. kinetic depth) is inherently ambiguous. In some situations information from another depth module can be used to determine which of two potential interpretations is correct (Doshier *et al.*, 1986; Blake & Bühlhoff, 1990).

The focus of this paper is the means by which texture and stereopsis are combined. These two cues have been extensively studied individually, but the rules that govern their combination have not been analysed in detail. Interactions between linear perspective and stereopsis in specifying receding planar surfaces have been studied (Gillam, 1968; Youngs, 1976; Stevens & Brookes, 1988). When perspective and stereopsis conflict, some compromise is reached which depends upon the effectiveness of each cue in isolation. However, the perspective component of the texture cue—that is the uniform changes in size of the texture elements with distance—is small for the size of many objects commonly available for grasping and manipulation. In this study we portrayed curved objects of ~5 cm radius to examine the interaction between texture and binocular stereopsis.

The effectiveness of stereopsis as an independent source of depth information is well demonstrated by the Julesz random-dot stereogram (Julesz, 1971). Stereopsis is a potentially powerful depth cue because it is very precise and it does not require knowledge of surface properties. Provided eye position is known, binocular disparities can be converted into measures of absolute distances of surface features from the observer. However, previous work on perceived depth from disparity demonstrates that the necessary scaling for viewing distance is not correctly performed (Foley, 1980; Johnston, 1991; Cumming, Johnston & Parker, 1991). Shape perception from surfaces defined by binocular disparity in random-dot stereograms is only veridical at

intermediate viewing distances close to 1 m (Johnston, 1991). At far distances, depth from stereo is underestimated, while at near distances it is overestimated. These distortions are best explained by suggesting that observers scale disparities with an inaccurate measure of viewing distance.

This scaling problem evident with stereopsis as a single cue to depth could be ameliorated by the addition of other cues which do not depend on viewing distance. One way this might occur is if the second cue provided a good estimate of the viewing distance, which could be used for scaling disparity data—this would be an example of promotion. Alternatively, a weighted linear combination with a cue which is more veridical could also produce some improvement. Linear combination is not possible between depth modules that yield incommensurate shape measures—if for example stereo yielded only a measure of peak depth, and texture yielded a measure of surface curvature, these two measures could not simply be added together.

Gibson (1950) first drew attention to the gradient of texture density as a cue to surface slant. If texture elements (texels) are uniformly distributed on a plane, changes in density are directly correlated with distance of the surface from the observer. Cutting and Millard (1984) discussed other texture gradients: the compression gradient, or foreshortening, which corresponds to changes in the width/height ratio of texels, and the perspective gradient, which is available from the uniform decrease in texel size as the distance of the surface from the observer increases. They found that compression was the most effective component for curved surfaces. Todd and Akerstrom (1987) also found the compression cue to be most effective for curved surfaces, and found good depth perception from monocularly viewed textured ellipsoids. In a study investigating which component of texture is effective in *binocular* viewing, we also found that texel compression is most important in specifying the shape of curved surfaces (Cumming, Johnston & Parker, 1993). Extracting a measure of surface orientation from the texture compression does not require a measure of the viewing distance, so texture potentially provides a cue that is free of the distance scaling problem inherent in stereopsis. In a conventional random-dot stereogram the texture is isotropic and thus specifies a fronto-parallel planar surface, at variance with the depths specified by any disparities present. It may be that stereopsis is a sufficiently powerful source of depth information that this inappropriate texture cue has no effect on the perceived depth from random-dot stereograms (i.e. stereo vetoes texture), but it is also possible that the texture cue modifies the depth percept. The experiments here attempted to discover which of these interactions occurs by generating stereograms with a shape-from-texture cue. The shape specified by texture was manipulated independently of the shape specified by stereopsis, to allow the exploration of the rules governing the combination of texture and stereo information in human vision.

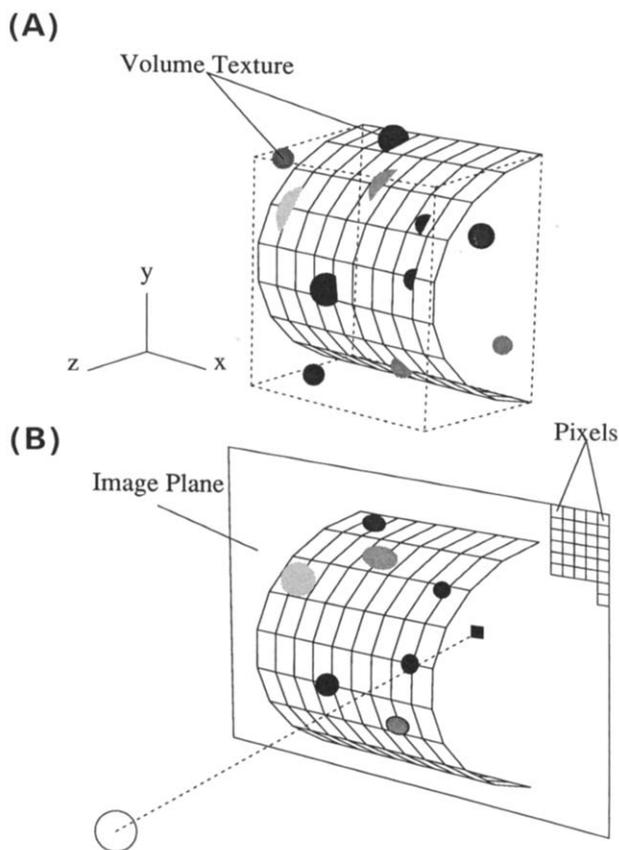


FIGURE 1. Diagram illustrating the method of ray-casting with volumetric texture. (A) Illustrates how three-dimensional space is filled with volume texture (spheres in this diagram), so that the texture on the cylinder's surface is defined by its intersection with this volume texture. (B) Shows how a ray is "cast" to the viewpoint, from each pixel in the image. The point at which this ray intersects the three-dimensional surface (the cylinder here) is then calculated. The ray here (dotted line) is shown passing through a black sphere at this point of intersection. The illustrated pixel would therefore be rendered as black. This also illustrates the geometry used, with the projection plane coinciding with the back edge of the cylinder.

**METHODS**

*Stimulus generation*

*Volumetric representation and ray-tracing.* In this study, textured images were generated by volumetric texture rendering (Watt, 1989; first used in psychophysics by Bülhoff & Mallot, 1988), a computer graphics technique which is equivalent to carving an object out of a solid textured material (such as granite). The three-dimensional space in which the object lay was divided up into small volume elements (*voxels*). This was programmed on a TAAC-1 graphics accelerator board and a Sun 3/160 workstation, using the volumetric representation software supplied with the TAAC-1 accelerator. The volume was then filled with texture by assigning different grey-levels to individual voxels. Two types of textured blocks were used. To produce the "random" block, the voxels were set to a randomly generated grey-level between 1 and 254. Random numbers were generated using a non-linear additive feedback routine supplied with the graphics accelerator. The block was then convolved with a three-dimensional Gaussian of space constant 0.42 mm (1 voxel) to provide

smooth variations in grey-level. The "spheres" block was constructed by randomly positioning non-overlapping spheres of grey-levels between 1 to 117 and 137 to 254 in a block of background grey-level 127. The spheres varied randomly in radius from 0.17 to 0.42 cm. Voxels which lay only partially inside the spheres were assigned a grey-level based upon the proportion of the voxel that was contained within the sphere, thus providing the correct information for anti-aliasing the final images.

To render images from the textured blocks, the three-dimensional coordinates of the surface to be portrayed were first calculated. Cylinders, ellipsoids and roof-like shapes were used. For each pixel on the image plane, a ray was cast to each eye's viewpoint, and the intersection of this ray with the three-dimensional textured surface was calculated. The grey-level of the block at that point was then displayed at the pixel from which the ray originated.

Figure 1 illustrates the method and the geometry used. The image plane is positioned at the back of the hemicylinder, which sits within the textured block. This method produces an exact perspective projection for each eye's viewpoint, generating binocular disparities that portray surfaces protruding from the screen.

Although this representation of three-dimensional space is sampled and quantized, we were able to overcome the aliasing problems this would have introduced by means of interpolation: for any arbitrary location in three-dimensional space, the grey-level was determined using a trilinear interpolation routine supplied with the TAAC-1 accelerator. Trilinear interpolation involves linear interpolation in each dimension between nearest voxel centres. This method represents three-dimensional space with smoothly varying grey-levels specified to sub-voxel accuracy. As the size of the voxels was only slightly larger than that of the pixels on the

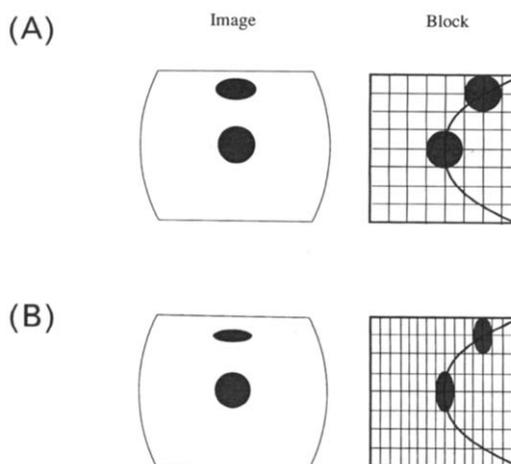


FIGURE 2. Schematic diagram illustrating how the depth portrayed by texture is manipulated independently of the depth portrayed by stereo. In the upper half of the figure, the volume is scaled isotropically, so that the texture variation is commensurate with the surface described by stereo. In the lower half of the figure the scale of the block is anisotropic, so that the same surface passes through a larger number of voxels. This produces proportional increases in the compression and density of texels in the image, for the same displayed stereo slant.

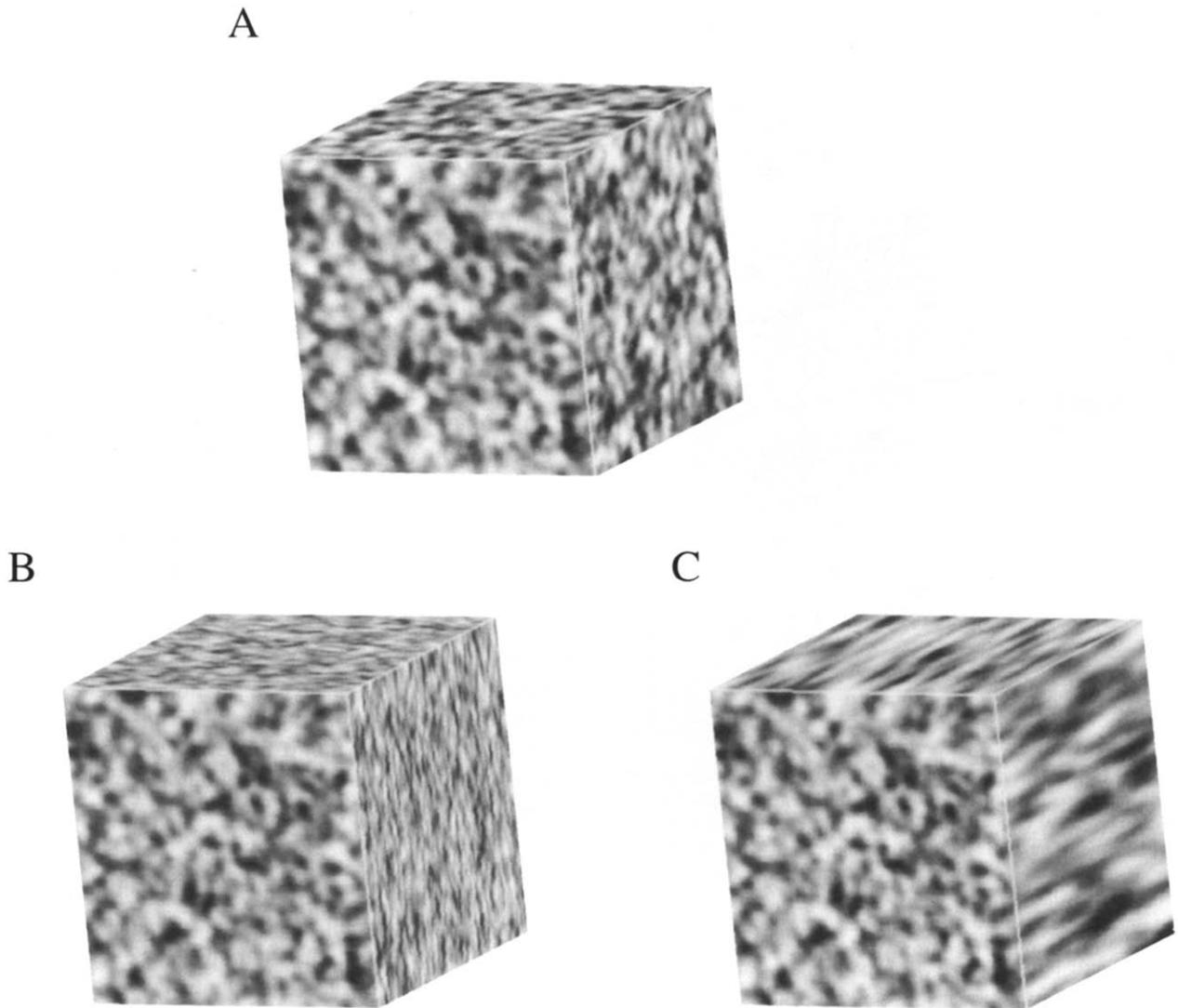


FIGURE 3. Example random textured blocks. (A) Isotropic block—surfaces cut from this block have a T/S ratio of 1, the two cues are perfectly consistent. (B) Compressed block, where the voxels are half as deep as they are high and wide—surfaces generated from this block have a T/S ratio of 2. (C) Expanded block, where the voxels are twice as deep as they are high and wide—surfaces cut from this block have a T/S ratio of 0.5.

monitor (each pixel measured 0.27 mm, the width of each voxel was 0.42 mm), our representation of three-dimensional space did not limit image quality any more than the limitations imposed by the pixel array of the monitor.

*Manipulation of the texture cue.* When an isotropic texture block is used, the textural variations with surface orientation are consistent with the depth specified by stereo disparity. That is, if subjects reconstruct shape-from-texture using the assumption that the surface markings are isotropic, stereo and texture will specify the same shape. By varying the properties of the textured block the texture cue can be varied independently from the stereo cue. Figure 2 illustrates the consequences of compressing the texture block along the  $z$ -axis. Figure 2(A) shows a  $y$ - $z$  cross-section of a horizontal cylinder in a block of isotropic texture elements. Figure 2(B) illustrates that using the same surface generated from a

block of texture elements half as deep as they are wide generates a surface whose textural variations specify a surface of twice the depth (if the assumption of texture isotropy is still used). However, the stereo disparities still specify the original surface. Thus, a surface with a texture/stereo (T/S) depth ratio of 2 is created. It is important to note that the stimuli are presented protruding from a fronto-parallel plane, giving a sample of the texture from which the surface is rendered. Since the modification of the texture cue affected only the scaling in the depth dimension, the statistical properties of the texture in this fronto-parallel region did not change as the T/S ratio was changed. Consequently, the texture in this region of the image was always isotropic, supporting the use of the assumption that the texture on the surface of the cylinder is isotropic when extracting shape-from-texture. This is important because shape-from-texture is only made possible by the use of assumptions about the

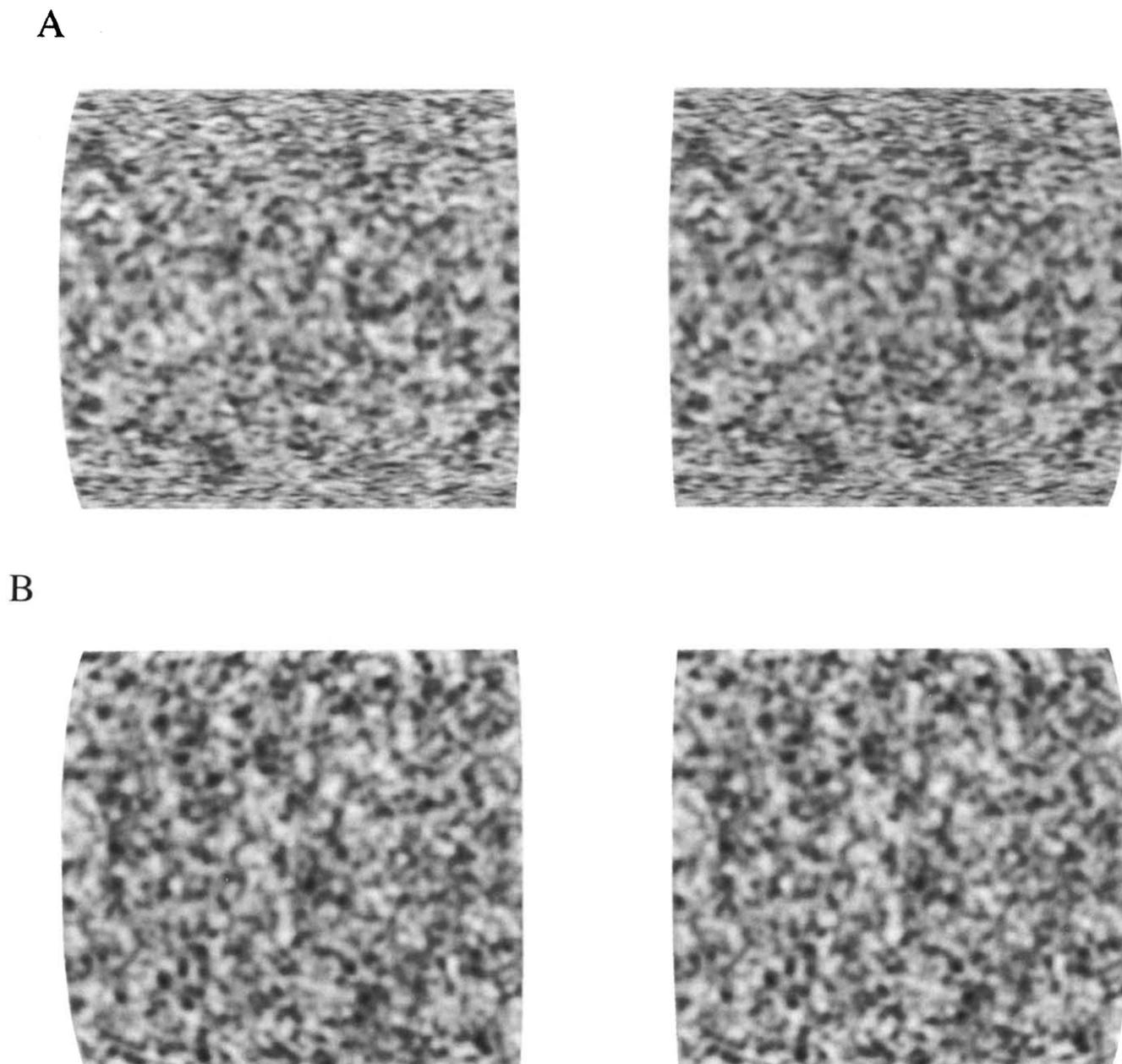


FIGURE 4. Example stimuli. Two stereo pairs portraying cylinders, made from the random block. (A) A cylinder with an appropriate texture cue. (B) The texture cue corresponds to a fronto-parallel plane. The stimuli are shown against a white background for clarity, but during experiments they were shown against a random-dot background.

properties of unprojected texture. So a T/S ratio of 2 means that the shape specified by texture is twice the depth of the shape specified by stereo, *on the assumption that the unprojected texture is isotropic*.

Figure 3 shows some examples of textured blocks. Surfaces cut from the isotropic block [Fig. 3(A)] have a T/S ratio of 1, surfaces generated from the compressed block [Fig. 3(B)] have a T/S ratio of  $>1$ , and surfaces made from the expanded block [Fig. 3(C)] have a T/S ratio of  $<1$ . Any ratio of texture to stereo depth can be generated simply by scaling the  $z$ -axis of the textured block. To produce a T/S ratio of 0, corresponding to a conventional random dot stereogram (where the texture does not vary with depth), the block is expanded until all of the  $x$ - $y$  planes within the block are identical. In

this case, whatever the  $z$ -value (or depth) required, the grey-level assigned will be the same, as in a conventional random-dot stereogram.

Figure 4 shows example stereo stimuli cut from the random block. Figure 4(A) shows a cylinder with a texture cue: as the surface curves away from the viewer the texture becomes increasingly foreshortened. Figure 4(B) is an example of random-dot stereogram with no texture variation—the texture specifies a fronto-parallel plane, while the stereo depicts the same cylinder as Fig. 4(B).

#### *Stimulus dimensions*

The stimulus depicted was always 10 cm in both height and width. At the two viewing distances of 50 and

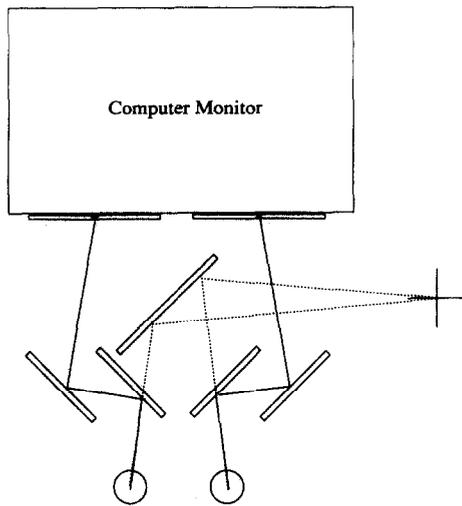


FIGURE 5. Schematic diagram of the modified Wheatstone stereoscope. The two proximal mirrors were half-silvered and mounted on adjustable tilt and rotation stages, and the two outer mirrors were front-silvered and fixed at  $45^\circ$  to the interocular axis. The fifth mirror, which was also front-silvered, allowed viewing of a physical fixation cross (indicated by the +). The physical fixation cross was placed at the same distance as the path from the subject's eyes to the computer monitor, so the four light paths shown are all of equal length. The images of the physical fixation cross were monocularly aligned with identical fixation crosses on the monitor centred on the positions marked by  $\times$  on the computer monitor.

200 cm, the stimulus subtended  $11.3$  and  $2.8^\circ$  respectively. The individual texels in the spheres block varied in size from  $0.17$  to  $0.42$  cm, subtending visual angles of  $11.68$ – $28.88'$  at  $50$  cm and  $2.92$ – $7.22'$  at  $200$  cm. For the random-dot stimuli, the voxel size was scaled with the viewing distance such that one voxel always subtended  $1.5'$ . This ensured that the spatial frequency content of the stereograms did not alter with viewing distance. The stimuli were always presented on a background made from slices through the appropriate textured block, so a comparison of the textured three-dimensional surface with a flat surface specified by the same texture was always available. The flat region of the image allows for calibration of the assumptions made about the texture (isotropy, homogeneity, texture element shape).

#### Apparatus

The stimuli were presented on a Manitron VLR2044 white phosphor monitor (phosphor P4) corrected for linearity in the luminance domain by using signals from a coupled pair of video DACs, according to the method of Watson, Neilson, Poirson, Fitzhugh, Bilson, Nguyen and Ahumada (1986). The screen was  $29 \times 25$  cm, and  $1192 \times 900$  display pixels were available. Stimuli were presented in the central portion of the screen, whose  $x$ - $y$  geometry was linear measured by the technique described by Maloney and Koh (1988). The mean luminance was  $34.3$   $\text{cd m}^{-2}$ .

A modified Wheatstone stereoscope was used to achieve independent presentation of each eye's image. Figure 5 shows a schematic diagram of the apparatus. The outer mirrors were fixed at  $45^\circ$ . The two proximal mirrors were mounted on high precision Ealing

rotational and tilt stages which allowed adjustment of the mirror positions to within  $1'$ . A separate physical fixation cross was viewed through the half-silvered proximal mirrors. The physical fixation cross was placed at the viewing distance required and the proximal mirrors were then adjusted under monocular viewing to align the physical fixation cross with each eye's image of the stereo fixation cross displayed on the monitor. This procedure ensured that the vergence position was correct for the viewing distance used, and that the viewing apparatus replicated exactly the geometry used in stimulus generation. Separate stimuli were generated for each subject (see above) in order to account for variations in interocular separation, and the separation of the two proximal mirrors was set to the interocular separation of the subject. The subject's head was fixed by means of a chinrest, as well as a forehead rest. However, a bite bar was not used so head and eye movements were possible, although subjects were instructed to keep their heads still and fixate the centre of the stimulus.

#### Procedure

A global three-dimensional shape judgement task was used, rather than a local judgement of depth, orientation or curvature. A major advantage of using global shape judgements when investigating the combination of different sources of three-dimensional shape information is that they allow information to be integrated from a variety of sources. This may be especially important if individual cues result in different types of information about the surface, such as a local depth map from stereo, and surface orientation from texture (Bülthoff, 1991).

The experimental task was identical to the task used previously to assess the veridicality of three-dimensional shape perception from cylindrical surfaces displayed as random-dot stereograms (Johnston, 1991). Subjects were presented with a series of horizontally oriented elliptical cylinders which differed in their elongation in depth from trial to trial. On each trial, observers decided if the cylinder was more or less extended in depth than a cylinder of circular cross section. Expressed differently, they were asked to determine if the ratio of depth to half-height was greater or less than one. The point of subjective equality was determined, which corresponds to a cylinder which *appears* to be a circular cylinder.

In order to prevent subjects from discriminating stimuli on the basis of portrayed depth alone, a small degree of random variation ( $\pm 15\%$ ) was added to the overall size of each cylinder portrayed. Thus subjects were forced to make judgements about the *shape* (depth/height) of each individual cylinder, rather than purely making a comparison of depth between cylinders.

In this experiment we also extended the shape judgement task, that was originally devised for cylinders (Johnston, 1991), to other canonical shapes. In the case of ellipsoids, subjects decided if the depicted surface was more or less extended in depth than a sphere. In the case of roof stimuli, composed of two sides of a triangular prism, observers judged whether the sides of the roof were inclined at more than or less than  $45^\circ$  (or

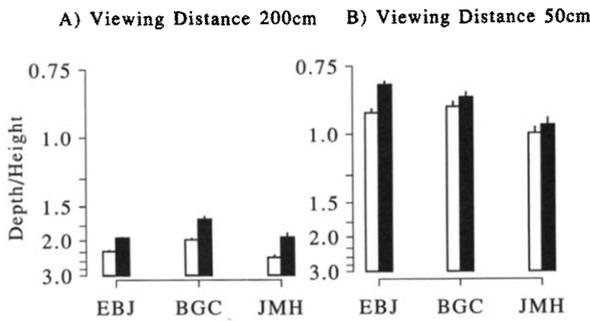


FIGURE 6. The portrayed depth/height of the cylinders which appeared circular to the three subjects (indicated on the x-axis) using the random texture. In all conditions the cylinder half-height was 5 cm. The error bars show one standard error. Since increasing ratios of portrayed depth/height correspond to decreasing perceived depth, the ordinate is plotted with a reciprocal scale. (A) Data collected at a 200 cm viewing distance. Hollow bars signify the 0 T/S ratio condition (contradictory stereo and texture) and solid bars signify a T/S ratio of 1 (congruent stereo and texture). When no texture variation was present (texture specifies a flat plane), all subjects required exaggerated depth to be portrayed by disparity in order to perceive the cylinder as circular. Thus, the depth/height ratio is considerably greater than 1—depth is underestimated. When a commensurate texture cue was introduced (1 T/S ratio) the depth/height ratio decreased towards more veridical perception. (B) Data collected at a 50 cm viewing distance. At this distance the depth/height values are <1—depth from stereo is overestimated. Adding the correct textural variations again increases the perceived depth (compare the solid and hollow bars) but at this close distance the result is a less veridical percept than in the 0 T/S ratio condition.

equivalently, whether the angle formed between the planes was greater than or less than 90°). All of these tasks can be summarized as judgements of whether an object's depth was greater or less than its half-height. This allows us to quantify the shape distortions in units which are commensurate across different shapes.

The results are plotted in terms of the *depth/height* ratio of the cylinder which appears circular, or the ellipsoid which appears spherical, or the roof which appears to have sides of 45° slope. (As the shapes depicted are opaque the ratio is strictly the depth/half height, so a ratio of one corresponds to veridical per-

ception.) Note that larger depth/height ratios correspond to shapes in which larger depths were perceived equal to the portrayed height. So depth/height ratios > 1 indicate that depth is *underestimated*. Since larger ratios describe less effective depth stimuli, the ratio will be plotted on a reciprocal scale when used to quantify perceived depth.

A modified staircase method was employed. The stimuli were rank-ordered by stereo depth value. The stepsize between consecutively presented stimuli was initially set to six times the final value. After each stimulus presentation the stepsize was decreased to  $6/N$ , where  $N$  is the trial number in that staircase, until a reversal with a stepsize of 1 occurred (Levitt, 1970). As measures of response variability were of some importance for this study, we were particularly concerned that reversals were independent. To this end, after a reversal occurred with a stepsize of 1, a completely new staircase was started with a stepsize of 6, beginning with one of the original, widely spaced, starting values. Up and down staircases were randomly inter-leaved (Cornsweet, 1962). Each experimental run consisted of three up staircases and three down staircases. Each run was performed twice, so that each data point presented here results from a total of twelve staircase determinations, each with one reversal at the smallest step size.

*Subjects*

Four subjects completed each experiment. Three subjects took part in all experiments: the first two authors and one other observer, who was naive concerning the aims of the experiment and the stimulus manipulations performed. Two further naive subjects participated, one in each experiment. All subjects wore the appropriate optical correction.

**EXPERIMENT 1: THE EFFECT OF ADDING A CONGRUENT TEXTURE CUE**

This experiment was designed to investigate whether there is any effect of texture in the presence of binocular

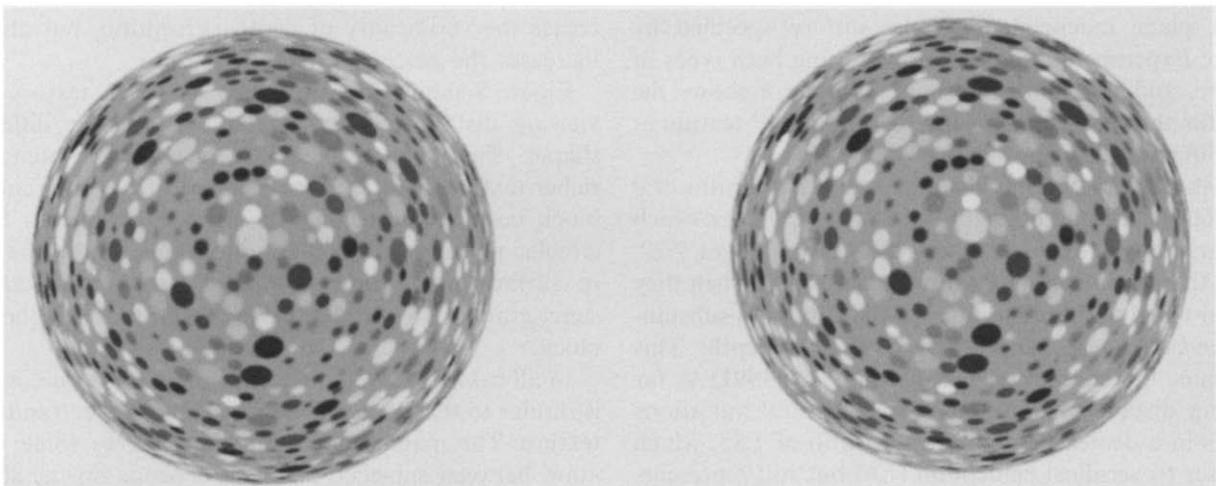


FIGURE 7. Example stereogram of an ellipsoid generated from the "spheres" texture block. This stereogram is designed for crossed fusion. Note the changes in size, density and ellipticity of the texture elements towards the edges of the surface where the surface is more steeply angled with respect to the viewer.

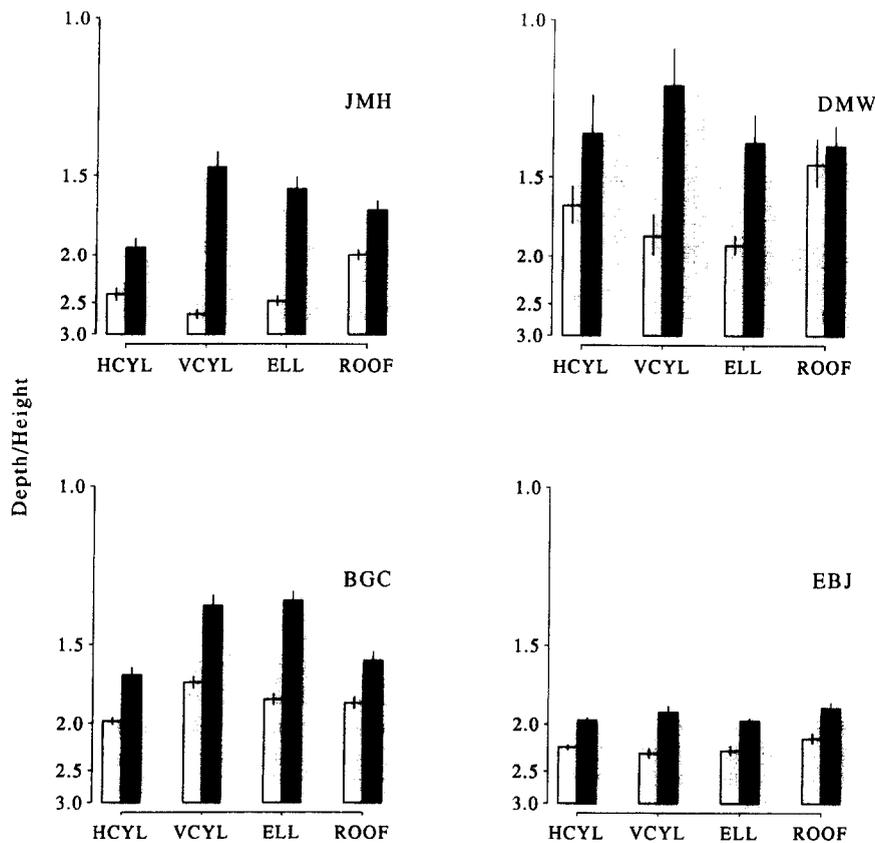


FIGURE 8. Portrayed depth/height of the surfaces which appeared to have equal height and depth to the three subjects (plotted individually) collected with the spheres texture. The viewing distance was 200 cm. The cylinder half-height was 5 cm. Since increasing ratios of portrayed depth/height correspond to decreasing perceived depth, the ordinate is plotted with a reciprocal scale. Four types of surfaces were depicted—horizontally oriented cylinders (HCYL), vertically oriented cylinders (VCYL), ellipsoids (ELL) and surfaces composed of two joined planes of opposite slant (ROOF). Hollow bars represent the 0 T/S condition and solid bars represent the 1 T/S condition. For all four shapes the perceived depth in the 1 T/S condition is larger than in the 0 T/S condition. That is, adding the correct texture cue increases the perceived depth.

disparity, or whether stereo is such a strong cue that it vetoes texture. The random textured block was used, so that the effect of texture in conventional random-dot stereograms could be assessed. The two conditions were a T/S ratio of 1, where the variations in size, density and orientation of the texture elements were consistent with the cylindrical surface specified by stereo, and a T/S ratio of 0, where the texture was homogeneous and specified a flat plane independent of the surface specified by stereo. Experiments were performed using both types of texture, and four different shapes. Figure 6 shows the data for three observers, using the "random" texture at two different viewing distances.

At the 200 cm viewing distance, with a T/S ratio of 0 [Fig. 6(A)], the depth/height ratio of the cylinder which appears to be circular to the observers averages 2.22. Since the depth is more than twice the height when they appear equal to the subject, perceived depth is a substantial *underestimate* of portrayed physical depth. This replicates the effect reported in Johnston (1991) at far viewing distances. Addition of the textural variations results in a decreased depth/height ratio of 1.85, which is closer to veridical perception (1.0) but still represents a large underestimate of depth. At the short viewing distance of 50 cm [Fig. 6(B)] where the depth is *overestimated* in conventional random-dot stereograms

(Johnston, 1991), adding the texture cue further increases perceived depth, from an average depth/height ratio of 0.95 where  $T/S = 0$ , to 0.87 where  $T/S = 1$ . It is interesting that although perception is close to veridical without the texture cue (especially for subject JMH), the increase in perceived depth produced by the texture cue makes shape judgements *less* veridical in all subjects. In summary, adding the texture cue does not always increase the veridicality of depth perception, but always increases the perceived depth.

Figure 8 shows data using the "spheres" texture at a viewing distance of 200 cm, and using four different shapes. This "spheres" texture provides a potentially richer texture cue than that produced from the random block because the change in shape of the texels from circular to ellipsoidal is a good local cue to the changes in surface orientation. Figure 7 shows an example stereogram of an ellipsoid generated from the "spheres" block.

In all cases there is an effect of the texture cue, which is similar to that shown in Fig. 6(A) using the "random" texture. The magnitude of the effect shows some variation between subjects, and also depends on the shape used. Since there is considerable inter-subject variation in the perceived depth-from-stereo alone (as was previously reported by Johnston, 1991), further analysis is

TABLE 1. Weights of texture cue using a variety of shapes, textures, and viewing distances

Condition	EBJ	BGC	JMH	DMW	RBC
<i>"Random" texture, horizontal cylinders</i>					
Viewing distance 200 cm	7 (1)	10 (2)	10 (3)		
Viewing distance 50 cm	14 (4)	5 (5)	4 (6)		
<i>Different shapes, "spheres" texture, at 200 cm</i>					
Horizontal cylinders	6 (1)	15 (4)	9 (2)	15 (11)	19 (2)
Vertical cylinders	8 (2)	16 (3)	31 (4)	32 (11)	
Ellipsoids	6 (1)	21 (3)	23 (4)	22 (8)	
Roofs	6 (2)	9 (3)	9 (3)	4 (9)	

The texture weight is shown as a percentage of the sum of the weights for stereo and texture. The numbers in brackets give the standard errors.

required in order to compare the effectiveness of the texture cue across subjects. To do this we have calculated the relative weightings of the stereo and texture cues from the data shown in Figs 6 and 8 (the method for calculating the weights is described below). The values for the percentage weight given to texture under each condition are shown in Table 1.

For horizontal cylinders, the weight assigned to texture is between 5 and 20% for all subjects. The texture cue has a stronger effect for vertical cylinders than for horizontal cylinders, the difference being greatest for the two naive subjects. This could be related to the reported stereo anisotropy (Rogers & Graham, 1983; Gillam, Chambers & Russo, 1989)—surfaces generating a sheared disparity field (e.g. horizontal cylinders) are more easily seen and have more apparent depth than surfaces which generate either a horizontally expanded or compressed disparity field (e.g. vertical cylinders). It has also been reported that this anisotropy does give rise to differences in the effectiveness of texture cues on stereoscopically presented surfaces (Buckley & Frisby, 1993). The fact that the magnitude of this effect varies between subjects is not surprising in light of the large individual differences found in the existence and size of the stereo anisotropy (Mitchison & McKee, 1990; Cagenello, 1990).

It is clear that a vetoing form of interaction does *not* take place between stereo and texture—across all subjects and all conditions the addition of a texture cue increases perceived depth. The result with the "random" texture demonstrates that conventional random-dot stereograms cannot be considered devoid of a texture cue to depth, since the flatness specified by the homogeneously textured plane can influence perceived depth. Many studies use stereograms composed solely of fronto-parallel planes, in which a correct texture would contain only a small change in the scale of the texture pattern (as a consequence of perspective projection) with no changes in texture compression. Thus, the observation that texture affects the perception of random-dot stereograms may only be important for studies using stereograms to depict curved or slanted surfaces.

With the spheres texture, and a T/S ratio of 0, the image is composed solely of circular texture elements. In this condition two subjects (JMH and EBJ) reported that texture elements did not always appear to lie flat on the

surface depicted by stereo, so the surface no longer appeared smooth. Thus the texture cue may have quite strong local effects (making the slant of the texel appear quite different from the slant defined by stereo), whilst having a relatively small influence on the global shape judgement. A related point was noted by Ninio (1981) using stereograms constructed from random curved lines. He described seeing lines that did not appear to lie on the surface that is specified by stereo, and attributed this to assumptions made about the straightness of the lines in three-dimensional space.

#### EXPERIMENT 2: VARYING THE TEXTURE CUE

Experiment 1 showed that the form of interaction between texture and stereo is not simply a vetoing operation. Experiment 2 was designed to assess whether a model based on weighted linear combination can account for the interaction, and, if so, to quantify the relative weighting of the texture and stereo cues. In Expt 1, using a T/S ratio of 1 meant that, as the portrayed depth varied, both the texture cue and the stereo depth were varied together. We were concerned that this might allow subjects to use monocular cues (based on the different texture patterns) to solve the task, without having to reconstruct a three-dimensional representation from combining stereo with texture. In order to avoid this, the procedure was slightly modified here: the experiment was run in blocks of trials, within which only one depth was portrayed by texture for all stimuli, while the depth portrayed by stereo was varied systematically. This was achieved by varying the T/S ratio for each stimulus to produce the required stereo depth in conjunction with the selected texture depth. This modified procedure was not used with one of the naive subjects (RBC), who was presented with stimuli in which both cues varied together. Seven different blocks were run, with texture depths ranging from 0 to 15 cm. For each block, the stimulus that appeared circular to the subject was determined. The logic of this experimental manipulation is that as the depth specified by texture is increased less stereo disparity should be necessary for subjects to perceive the stimulus as circular. If the interaction is linear, specific predictions can be made about the combinations of stereo and texture that will cause cylindrical stimuli to be perceived as circular.

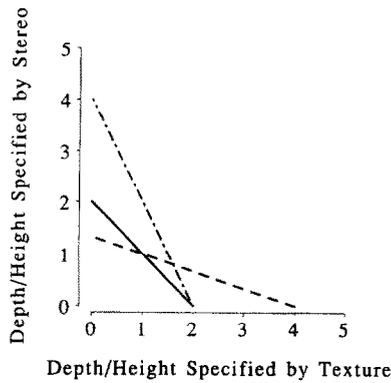


FIGURE 9. Schematic plot of possible linear cue integration functions. The solid line shows a linear combination when stereo and texture have an equal weighting of 0.5 each, and both provide independent veridical depth estimates. The dashed line shows a stereo weighting of 0.75, and a texture weighting of 0.25, again assuming veridicality of depth perception from both cues. The dot-dashed line shows the effect of a non-veridical shape from stereo mechanism, which underestimates depth by 50%, when both cues have weights of 0.5.

In order to illustrate the various predictions of a linear combination model, hypothetical data lines are plotted in Fig. 9. The axes show ratios of depth/height specified by texture (shown on the  $x$ -axis) and depth/height specified by stereo (shown on the  $y$ -axis). The lines show hypothetical points at which the combined effect of texture and stereo would cause stimuli to appear circular in cross-section.

The simulated data lines are calculated from a simple linear model for combining cues, assuming that texture and stereo are the only available cues.

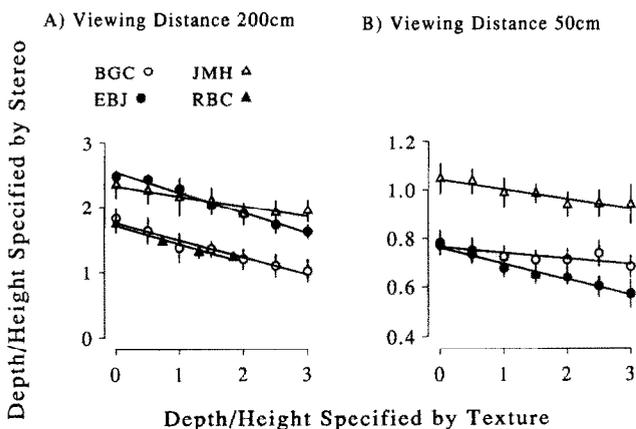


FIGURE 10. Texture/stereo integration data, showing the combination of cues resulting in the portrayed cylinder appearing circular, for three subjects (indicated by different symbol types). In one experimental run the depth portrayed by texture was held constant and the depth portrayed by stereo was varied to determine the combination of texture and stereo which produced a stimulus which appeared circular to the subject. The lines joining the data are calculated by linear regression. The error bars show standard deviations. (A) 200 cm viewing distance. The effect of texture can be seen from the linearly decreasing depth from stereo required to judge the cylinder as circular as the depth portrayed by texture is increased. (B) 50 cm viewing distance. The effect of texture is small at this viewing distance—the slopes of the regression lines are close to 0. Note the expanded  $y$ -axis, done in order that the slopes of the regression lines and their fit to the data could be better appreciated.

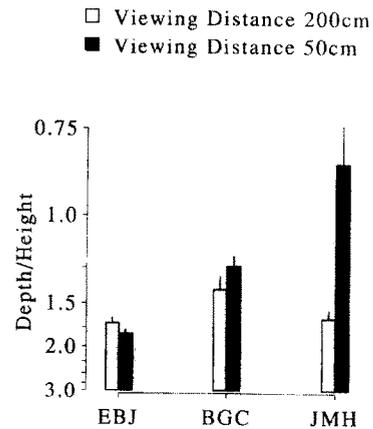


FIGURE 11. The effectiveness of the texture cue alone. The stimuli were monocularly viewed ellipsoids of half-height 5 cm. The ratio of portrayed depth/height is plotted on a reciprocal scale, as in Figs 6 and 8. Open bars—viewing distance 200 cm. Depth from texture alone is underestimated by an average of 59%. Solid bars—viewing distance 50 cm. Depth from texture alone is underestimated by subjects BGC and EBJ, and slightly overestimated by subject JMH.

The linear addition of stereo and texture can be described as:

$$d = \alpha_s S + \alpha_t T, \quad \alpha_s + \alpha_t = 1 \quad (1)$$

where  $d$  is perceived depth,  $S$  and  $T$  are the depth estimates produced from stereo and texture (which are each initially assumed to be veridical), and  $\alpha_s$  and  $\alpha_t$  signify the weights assigned to the stereo and texture depth measures. If texture and stereo were equally weighted and those weights summed to 1, the data would fall on the solid line shown in Fig. 9. If the weighting of texture was only 0.25 then the data would fall on a line of shallower slope indicated by the dashed line in Fig. 9. Thus, from data plotted on the axes shown in Fig. 9, it should be possible to calculate the relative weighting of texture and stereo from the slope of the regression line fit to the data.

A number of assumptions are made in the simple linear combination model described by equation (1). At least one of these, that the perceived depth from each cue alone is veridical, cannot be maintained. Our previous work on shape from stereopsis (Johnston, 1991; Cumming *et al.*, 1991), and the data for textured stereograms shown in Figs 6 and 8, demonstrated that at a viewing distance of 200 cm perceived depth from binocular disparity is underestimated. Equation (1) can be modified to take this into account:

$$d = \alpha_s f_s(S) + \alpha_t f_t(T) \quad (2)$$

where the depth measures supplied by stereo and texture are some unknown function of the geometrical depths specified by the information available. However, if the functions  $f_s$  and  $f_t$  are themselves linear, then linear data plots are still predicted. The pattern of shape judgement data at a variety of portrayed depths and viewing distances (Johnston, 1991) suggests that the non-veridicality of depth perception from stereo is a consequence of misestimating the viewing distance parameter. An error in this single scaling parameter gives a

geometrically incorrect, but still linear function,  $f_s(S) = k_s S$ . The consequence of equation (2) for the data plots is that the intercepts with the  $x$  (texture) and  $y$  (stereo) axes will not necessarily be at the positions expected from equation (1). Further, the *slope* of the lines is a function not only of the relative weighting of the two cues, but also of the functions mapping portrayed depth to perceived depth. The dot-dashed line in Fig. 9 is calculated with equal weighting of texture and stereo, a function  $f_s$  which underestimates depth by 50%, and a veridical function  $f_t$ . The slope of this line, and the  $y$ -intercept, are both twice as large as those of the line generated with equation (1) using equal weights assuming that  $f_s$  and  $f_t$  are veridical (the solid line in Fig. 9).

The experimental data are plotted in Fig. 10 for three subjects at two viewing distances, 50 and 200 cm, and for a fourth subject at 200 cm. Regression lines provide a good fit to the data, indicating a linear combination of texture and stereo. All seven lines pass an ANOVA test for linearity (Armitage, 1971, p. 271). The difficulty with calculating the relative weighting of texture and stereo from the data presented in Fig. 10 is that the functions  $f_s$  or  $f_t$  in equation (2) are unknown, in addition to  $\alpha_t$  and  $\alpha_s$ .

The  $y$ -intercepts in Fig. 10 show the familiar pattern of the function  $f_s$  providing an underestimate of depth at the far viewing distance, and a small overestimate at the near distance. It should be understood that the  $y$ -intercept cannot be used directly to estimate  $f_s(S)$ : when the  $T/S$  ratio is 0, the texture cue specifies a fronto-parallel plane, so the contribution from texture will make cylinder appear flatter than if there were no effect of texture. When the  $T/S$  ratio is 0,  $f_t(T) = 0$ , so the perceived depth [from equation (2)] will be  $\alpha_s f_s(S)$ . However, when  $\alpha_s \gg \alpha_t$  (so  $\alpha_s \approx 1$ ), then  $\alpha_s f_s(S) \approx f_s(S)$ , so the misestimate will be small. Similarly, the large  $x$  intercepts do not necessarily indicate that the function  $f_t$  (the mapping between portrayed and perceived depth from texture) produces underestimates of depth. These intercepts indicate how great a texture cue would be required to counteract a set of binocular disparities portraying a fronto-parallel plane.

The data in Fig. 10 alone are not sufficient to solve equation (2) for  $\alpha_s$ ,  $\alpha_t$ ,  $f_s(S)$  and  $f_t(T)$ . The equation can be solved if an independent estimate of one of the unknowns is obtained, and the assumption made that the weights sum to 1. We therefore attempted to estimate  $f_t(T)$  by performing the shape-judgement task monocularly.

The data, plotted in Fig. 11, show that there is an underestimation of depth from texture alone, with the exception of subject JMH at the close distance. The other two subjects underestimate depth from texture by similar amounts at both viewing distances.

While this experiment provides data consistent with our findings in texture-stereo integration, we do not wish to emphasize the underestimation of depth from texture. The monocular stimuli with depth defined only by texture do not provide the compelling depth sensation yielded by the stereo stimuli. They are representations of

depth that allow subjects to make some judgement about the depth depicted, or at least to distinguish between representations of different depths, but this is subjectively different from the depth percept produced by stereopsis. Thus, it is quite possible that observers fall back on strategies involving monitoring properties of the two-dimensional images, without having to create a three-dimensional representation. The objective here was to use stimuli equivalent to those used in the binocular study, to provide an estimate, however crude, of  $f_t(T)$ , allowing us to calculate values for the coefficients  $\alpha_t$  and  $\alpha_s$ . In equation (2) the term  $d$  refers to the perceived height/depth. Our experimental data yields pairs of stimulus values ( $S_i, T_i$ ) all of which give rise to a perceived height/depth of 1. Substituting  $d = 1$  into equation (2), and using the constraint that  $\alpha_t + \alpha_s = 1$  (Maloney & Landy, 1989), equation (2) can be re-written

$$1 = (1 - \alpha_t) f_s(S) + \alpha_t f_t(T) \tag{3}$$

rearranging:

$$f_s(S) = \frac{1 - \alpha_t f_t(T)}{(1 - \alpha_t)} \tag{4}$$

Therefore plotting  $S$  against  $T$  should yield a straight line, which intercepts the  $x$ -axis at  $1/\alpha_t f_t(T)$ .

Table 1 shows the weights calculated for texture by this method. Two sets of coefficients are shown—one assuming that  $f_t(T)$  is veridical, the other calculating  $f_t$  from the data in Fig. 11. Note that when  $f_t(T)$  is assumed to be veridical, lower weights are assigned to the texture cue. This is because the measured contribution of the texture cue is  $\alpha_t f_t(T)$  in equation (2), so an increase in  $f_t(T)$  must result in a decrease in  $\alpha_t$ . Stereo is weighted more heavily than texture—an average ratio of 4.5:1 at the 200 cm viewing distance, and an average ratio of 11.5:1 at the 50 cm distance. The difference in weighting of texture at the two viewing distances could reflect the relative reliability of the two cues. At 200 cm, perception of depth from stereo alone is not veridical, and small disparities necessarily represent relatively large depth values. Small errors in estimating stereo disparities can lead to larger errors in evaluating portrayed depth, making the stereo cue inherently less reliable at far distances. Thus, the texture cue is assigned greater weight as the stereo cue becomes less dependable.

### DISCUSSION

When appropriate variations of texture are added to a stereoscopically displayed three-dimensional shape, the perceived depth is altered in accordance with the depth portrayed by the shape-from-texture cue. Clearly, stereo is not such a strong three-dimensional cue that the shape specified by texture has no effect. The alteration of perceived depth was also found for a texture which resembled that of a random-dot stereogram, indicating that the textural cue to "flatness" in conventional random-dot stereograms influences perceived depth. The shape specified by texture also influenced perceived depth when the volume texture was composed of

TABLE 2. Weighting of the texture cue

Subject	Distance	% Texture ( $\alpha_t/(\alpha_t + \alpha_s)$ )	
		(A) If veridical	(B) As measured
EBJ	200 cm	12	21
BGC		15	22
JMH		7	11
RBC		16	
EBJ	50 cm	9	17
BGC		3	4
JMH		4	3

The texture weight is shown as a percentage of the sum of the weights for texture and stereo ( $\alpha_t/(\alpha_t + \alpha_s)$ ), so the stereo weighting can be calculated by subtraction from 100%.

Column (A) shows the texture weight calculated when it is assumed that shape-from-texture is veridical. Column (B) shows the weight calculated when the non-veridicality of shape-from-texture (shown in Fig. 11) is taken into account. In each case the standard deviation for the estimate of the weight was <1%.

spheres. The apparently circular cylinder task we used in previous work on shape from stereo was extended to include ellipsoidal and roof stimuli, where a similar effect of texture was found.

One alternative explanation for this result is that the changes in texture somehow changed the way in which stereo information allows the visual system to reconstruct the surface. For example, it might be that the stereo correspondence problem is solved in a different way for circular image features than for elliptical ones. There are two reasons why such an explanation is unlikely. First, since the apical region of the cylinder is more or less fronto-parallel, the apex of the cylinder and the background plane have the same appearance, regardless of the texture used (see Fig. 4). This means that the peak depth of the cylinder is defined by features which do not change their shape with changes in the T/S ratio. Second, for a horizontally oriented cylinder, the effect of increasing the T/S ratio is to compress the image features in a vertical direction. Consequently along any horizontal scan line the statistics of point to point matches are unaltered, the scan lines are effectively just moved closer together.

An important point is that although the effect of texture is reliable it is quite small. At the far viewing distance (200 cm), adding the correct textural variations produced an average change of 17.2% in perceived depth, and at the close viewing distance (50 cm), a change of 7.7% was found. One possible reason for the differential strengths of texture and stereo is that the stereo cue does not depend on the form of the surface markings (provided sufficient texture markings are present), whereas interpretation of the texture cue depends upon making assumptions about the form of the object texture in order to interpret changes in the image texture. The assumptions discussed in the computational literature (Blake & Marinos, 1989) are homogeneity, that the texture covering the surface is uniformly distributed, and isotropy, that the orientations of texture elements are uniformly distributed. The use of these assumptions in human vision is discussed in more detail in the accompa-

nying paper (Cumming *et al.*, 1993). There are many natural textures that do not obey these assumptions, and other textures which provide poor shape-from-texture because the texture elements are not distinct enough to allow shape-from-texture to proceed robustly. This contrasts markedly with stereopsis, when binocular views of any scene contain some robust information about three-dimensional layout. It may be that stereo is assigned a greater weight because it is more generally applicable.

Another possibility is that the procedure adopted in Expt 2, in which images differed from one another only in their stereo depth, biased subjects towards a strategy that gave little weight to the texture cue. Two pieces of evidence argue against this interpretation. First, in Expt 1 the stereo and texture cues varied together, and the magnitude of the effect of texture was similar in the two experiments (compare Tables 1 and 2). Second, for one subject (RBC) Expt 2 was performed with stimuli in which both stereo and texture changed, and his data are similar to those of the other subjects.

As discussed in the Introduction, many studies on depth cue integration have found data that is well described as a linear additive interaction. This study extends this general finding to the interaction of stereopsis and shape-from-texture, for the class of surface (cylinder) and texture (spheres) used. Since the data in Fig. 10 falls closely on straight lines, it is unnecessary to invoke any form of strong fusion or multiplicative interaction to account for them.

The mounting evidence for linear cue combination suggests that it is a very general and common means of integrating depth cues. This supports the modular view of depth cues promoted by Marr and Nishihara (1978). As Bruno and Cutting (1988) discuss, linear combination of independent modules makes sense from a developmental point of view, as the ability to make use of the different sources of depth information emerges at different times, with the ability to use motion parallax preceding the ability to use stereo, which in turn precedes the ability to use pictorial cues such as relative size and occlusion (Yonas & Granrud, 1985). Further, independent processing of individual cues is a good strategy to adopt in light of the variation in the number of cues available to the observer at any one time or in any particular portion of the scene, depending upon the observer's motion, the contents of the scene, and the viewing optics. A strong dependence of one cue upon another could be damaging if those cues do not always occur together. Parallel processing of different features seems to be an organizational principle in the physiology of the primate visual system (Lennie, Trevarthen, van Essen & Wässle, 1990). However, we currently have very little knowledge concerning the physiological processing of the pictorial cues to depth (such as shape-from-texture and shape-from-shading), so it is not possible to say whether separate brain areas are involved in processing stereo and texture.

The results of Expt 2 suggest that there are problems in producing correct depth estimates from either stereo

or texture independently. The nonveridicality of stereo is likely to be a consequence of misestimating the viewing distance, but this problem is not alleviated by the addition of texture. This is clear from the finding that adding texture improves the veridicality of depth perception at the far distance but has the reverse effect at the near viewing distance. A more accurate viewing distance estimate should improve the depth estimate at both viewing distances. Nonveridicality of shape-from-texture is suggested by the results collected with monocularly presented stimuli containing only the texture cue (Fig. 11), but this is not free of confounding two-dimensional cues to the task. This underestimation of depth from texture may simply reflect that shape-from-texture does not provide a compelling depth sensation in isolation.

The data presented here are consistent with the statistical framework introduced by Maloney and Landy (1989) which emphasized some interesting aspects of cue interaction within a scheme of linear combination. Firstly, the weight assigned to a cue depends upon some measure of its reliability. This reliability measure could be partly derived from an ancillary cue: some other piece of information which does not in itself yield a depth measure but indicates the appropriateness of a given cue. For example, a densely textured surface with an object texture which obeys the homogeneity and isotropy assumptions is a good candidate for extracting shape-from-texture. Some support for the context-dependent weighting of cues comes from the finding that texture is weighted more heavily at the far viewing distance, where stereo is less veridical and disparities are small, thus more likely to be misestimated. A second idea Maloney and Landy put forward is promotion, where information provided by one cue makes up for a deficit in another cue. In the case of stereo the deficit is the need for information about viewing distance. Shape-from-texture does not depend upon viewing distance, and does not appear to be able to provide an estimate for stereo. This is clear from the fact that while texture improves veridicality at the far distance it decreases veridicality at the close distance. The third concept introduced by Maloney and Landy is that of robustness, which is not tested here as it only applies when more than two cues are considered. They propose that consistency between the cues can be used to determine if one cue should be selectively down-weighted due to the discrepancy between its depth estimate and those provided by the other cues. However, a slightly different form of robustness (Poggio, 1989) can be applied when there are only two cues. In that cue combination algorithm, the weight assigned to the weaker cue depends upon the similarity of its depth estimate to that of the stronger cue. If this sort of interaction occurred between stereo and texture, the effect of texture should have been greater when nearly congruent combinations were used, compared with when texture and stereo portray very different surfaces. This would have produced a sigmoid shape to the lines shown in Figs 9 and 10. However, the psychophysical data shown in Fig. 10 show no sign of such a

nonlinearity, suggesting that no adjustment is made when the two depth modules produce markedly discrepant shape estimates.

Together the present experiments suggest that stereopsis and shape-from-texture are independent processes in their early stages. In spite of (or perhaps as a result of) inaccuracies in the processing of both cues, we find that the stereo and texture cues portrayed in our stimuli seem to interact simply by means of weighted linear combination, irrespective of whether the surfaces portrayed by the two cues are congruent. In this combination, information from stereopsis is weighted much more heavily than that provided by shape-from-texture, presumably reflecting the more reliable nature of the information provided.

## REFERENCES

- Armitage, P. (1971). *Statistical methods in medical research*. Oxford: Blackwell Scientific.
- Blake, A. & Bülthoff, H. (1990). Does the brain know the physics of specular reflection? *Nature*, *343*, 165–168.
- Blake, A. & Marinos, C. (1990). Shape from texture: Estimation, isotropy and moments. *Artificial Intelligence*, *45*, 323–280.
- Bruno, N. & Cutting, J. E. (1988). Minimodularity and the perception of layout. *Journal of Experimental Psychology: General*, *117*, 161–170.
- Buckley, D. & Frisby, J. P. (1993). The integration of stereo, texture, and outline cues in the shape perception of three-dimensional ridges. *Vision Research*. In press.
- Bülthoff, H. (1991). Shape from X: Psychophysics and computation. Landy, M. S. & Movshon, J. A. (Eds), *Computational models of visual processing* (pp. 305–330). Cambridge, Mass.: MIT Press.
- Bülthoff, H. & Mallot, H. (1988). Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America*, *5*, 1749–1758.
- Cagenello, R. B. (1990). Perception and representation of stereoscopic slant and curvature. Unpublished D.Phil. thesis, University of Oxford, Oxford, England.
- Clake, J. J. & Yuille, A. L. (1990). *Data fusion for sensory information processing systems*. Boston, Mass.: Kluwer.
- Cornsweet, T. N. (1962). The staircase method in psychophysics. *American Journal of Psychology*, *75*, 485–491.
- Cumming, B. G., Johnston, E. B. & Parker, A. J. (1991). Vertical disparities and 3-D shape perception. *Nature*, *349*, 411–413.
- Cumming, B. G., Johnston, E. B. & Parker, A. J. (1993). Effects of different texture cues on curved surfaces viewed stereoscopically. *Vision Research*, *33*, 827–838.
- Cutting, J. E. & Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General*, *113*, 198–216.
- Dosher, B. A., Sperling, G. & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research*, *26*, 973–990.
- Foley, J. (1977). Effect of distance information and range on two indices of visually perceived distance. *Perception*, *6*, 449–460.
- Foley, J. (1980). Binocular distance perception. *Psychological Review*, *87*, 411–434.
- Gibson, J. J. (1950). *The perception of the visual world*. Boston, Mass.: Houghton Mifflin.
- Gillam, B. J. (1968). Perception of slant when perspective and stereopsis conflict: Experiments with aniseikonic lenses. *Journal of Experimental Psychology*, *78*, 299–305.
- Gillam, B. J., Chambers, D. & Russo, R. (1988). Postfusional latency in stereoscopic slant perception and the primitives of stereopsis. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 163–175.
- Ikeuchi, K. & Horn, B. K. P. (1981). Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, *17*, 141–184.

- Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research*, *31*, 1351–1360.
- Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago, Ill.: University of Chicago Press.
- Kaufman, L. (1974). *Sight and mind*. New York: Oxford University Press.
- Landy, M. S., Maloney, L. T. & Young, M. J. (1990). Psychophysical estimation of the human depth combination rule. In Schenker, P. S. (Ed.), *Sensor fusion III: 3-D Perception and recognition, proceedings of the SPIE* (1383, pp. 247–254).
- Lennie, P., Trevarthen, C., van Essen, D. & Wässle, H. (1990). Parallel processing of visual information. In Spillman, L. & Werner, J. S. (Eds.), *Visual perception: The neurophysiological foundations*. New York: Academic Press.
- Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*, 467–477.
- Maloney, L. T. & Koh, K. (1988). A method for calibrating the spatial coordinates of a visual display to high accuracy. *Behavior Research Methods, Instruments and Computers*, *20*, 372–389.
- Maloney, L. T. & Landy, M. S. (1989). A statistical framework for robust fusion of depth information. *Proceedings of the SPIE: Visual Communications and Image Processing* (Part 2, pp. 1154–1163).
- Marr, D. (1982). *Vision*. San Francisco, Calif.: Freeman.
- Marr, D. & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B*, *200*, 269–294.
- Marr, D. & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science*, *194*, 283–287.
- Mitchison, G. J. & McKee, S. P. (1990). Mechanisms underlying the anisotropy of stereoscopic tilt perception. *Vision Research*, *30*, 1781–1791.
- Ninio, J. (1981). Random-curve stereograms: A flexible tool for the study of binocular vision. *Perception*, *10*, 403–410.
- Poggio, T. (1989). A parallel vision machine that learns. In Cotterill, R. M. J. (Ed.), *Models of brain function*. Cambridge: Cambridge University Press.
- Rogers, B. J. & Collett, T. S. (1989). The appearance of surfaces specified by motion parallax and binocular disparity. *Quarterly Journal of Experimental Psychology*, *41A*, 697–717.
- Rogers, B. J. & Graham, M. E. (1983). Anisotropies in the perception of three-dimensional surfaces. *Science*, *221*, 1409–1411.
- Schriever, W. (1925). Experimentelle studien über stereoskopisches sehen. *Zeitschrift für Psychologie*, *96*, 113–170.
- Stevens, K. A. & Brookes, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research*, *28*, 371–386.
- Todd, J. T. & Akerstrom, R. A. (1987). Perception of three-dimensional form from patterns of optical texture. *Perception and Psychophysics*, *13*, 242–255.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, Mass.: MIT Press.
- Watson, A. B., Neilson, K. R., Poirson, A., Fitzhugh, A., Bilson, A., Nguyen, K. & Ahumada, A. J. (1986). Use of a raster framebuffer in vision research. *Behaviour Research Methods, Instruments and Computers*, *18*, 587–594.
- Watt, A. (1989). *Fundamentals of three-dimensional computer graphics*. New York: Addison-Wesley.
- Witkin, A. P. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence*, *17*, 17–47.
- Yonas, A. & Granrud, C. E. (1985). The development of sensitivity to kinetic, binocular, and pictorial depth information in human infants. In Ingle, D. J., Jeannerod, M. & Lee, D. N. (Eds.), *Brain mechanisms and spatial vision* (pp. 113–145). Dordrecht, The Netherlands: Martinus Nijhoff.
- Young, M. J., Landy, M. S. & Maloney, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research*. Submitted.
- Youngs, W. M. (1976). The influence of perspective and disparity cues on the perception of slant. *Vision Research*, *16*, 79–82.

---

*Acknowledgements*—This research was funded by the Wellcome Trust, the SERC and the MRC. Additional support was provided by the Oxford McDonnell-Pew Centre for Cognitive Neuroscience. We are grateful to Ron Cagenello, Mike Landy and Mark Young for helpful comments on earlier versions of the manuscript. We thank Julie Harris, Ron Cagenello, and Dan Wolpert for acting as observers.