Psychophysical estimation of the human depth combination rule

Michael S. Landy

Laurence T. Maloney

Mark J. Young

Department of Psychology and Center for Neural Science
New York University, New York, NY 10003

ABSTRACT

We describe a series of experiments designed to test (1) whether human observers combine depth cues using a *weighted average* when depth estimates in different maps are nearly consistent, (2) whether human observers behave as *robust* estimators when depths become increasingly inconsistent, and (3) whether the weights used in the linear rule of combination change to reflect the *estimated reliability* of different depth cues. We report initial experiments concerning texture and motion. The data are clearly consistent with the notion that the depth percept is a linear combination of the individual depth values portrayed by each cue. By randomly varying the shapes of the texture elements, the texture cue is artificially made unreliable, and the data support the hypothesis that unreliable cues are given less weight. Finally, there is an indication that when cues are strongly inconsistent, the weight on one of the cues is lowered, consistent with the hypothesis of robustness.

1. INTRODUCTION

A visual observer has a variety of cues available to estimate the distance to objects in the environment and their shape. These cues include pictorial cues (linear perspective, texture perspective, shading, etc.), motion cues (the kinetic depth effect, motion parallax), and binocular stereopsis, among others. Recently, there has been interest in the problem of how multiple cues to depth are combined to form a single depth map or shape estimate (Braunstein, Andersen & Riefer, 1982; Bruno & Cutting, 1988; Bülthoff, in press; Bülthoff & Mallot, 1987; Cutting & Millard, 1984; Dosher, Sperling & Wurst, 1986; Epstein, 1968; Gillam, 1968; Goodale, Ellard & Booth, 1990; Proffitt, Bertenthal & Roberts, 1984; Schwartz & Sperling, 1983; Stevens & Brookes, 1987, 1988; Todd & Akerstrom, 1987; Todd & Mingolla, 1983; Wallach & Karsh, 1963; Youngs, 1976). In particular, we have proposed that depth combination be treated as a problem of statistical estimation (Maloney & Landy, 1989).

In this paper, we report initial results of psychophysical experiments designed to test this statistical paradigm of depth combination (Maloney & Landy, 1989). Subjects viewed stimuli which included both texture and motion cues to depth, and the depth portrayed by each cue was varied independently. By determining which combinations of portrayed depths appear to have the same combined depth, it was found that the linear combination rule assumed by Maloney and Landy (1989) does adequately describe observer performance, that the weights on individual cues vary with their individual reliabilities, and that there is some indication that the subject behaves in a robust fashion (described below). The statistical paradigm for depth combination is reviewed, and then the psychophysical results are presented.

2. THEORY

The first issue in analyzing the depth combination problem is to recognize that different cues provide qualitatively different information. For example, motion parallax (depth from self-motion) is potentially an *absolute depth cue* in that the precise depth (e.g. in meters) can be computed given knowledge of the self-motion and induced visual motion of each feature. Other cues may be specified only up to one or more unknown *parameters* which must be estimated in order to convert the raw depth information into depth estimates. Texture gradients and linear perspective provide estimates of depth at all feature locations in the scene up to an unknown multiplicative scale factor when polar perspective is used. The kinetic depth effect (depth from object motion) for displays which use parallel perspective provides depth information that is subject to an unknown additive depth factor and to *depth reversals* (where all relative depth values are inverted, and perceived rotation reverses as well). Thus, the depth estimate from the raw kinetic depth information $I_k(x,y)$ at location x,y is

$$d_k(x,y;\gamma_{k_1},\gamma_{k_2}) = \gamma_{k_1}\,(1+\gamma_{k_2} I_k(x,y)). \qquad (1)$$

γ_{k_1} is the additive parameter (i.e. the distance to the center of rotation), and parameter γ_{k_2} allows for depth reversals (and takes on values ± 1).

There are a number of ways in which different cues might interact and combine. Depth values from various cues can be averaged. One can decide to use one cue and ignore another. One cue can be used to disambiguate another, or otherwise fill in missing parameters required to *promote* the information of that cue to a more complete form (e.g. by determining γ_{k_1} and γ_{k_2} above – Braunstein et al., 1982; Maloney & Landy, 1989; Proffitt et al., 1984).

We are interested in whether the combination of depth estimates from multiple cues may be modeled as a weighted average:

$$\begin{aligned} d(x,y) = {} & \alpha_k\, d_k(x,y;\overline{\gamma_k}) + \alpha_t\, d_t(x,y;\overline{\gamma_t}) \\ & + \alpha_s\, d_s(x,y;\overline{\gamma_s}) + \alpha_p\, d_p(x,y;\overline{\gamma_p}). \end{aligned} \qquad (2)$$

In this case we have represented an average of depth estimates from four cues: kinetic depth (k), texture gradient (t), binocular stereopsis (s), and linear perspective (p). Note that we are averaging depth estimates (e.g. d_k), not the raw depth information (e.g. I_k). If two cues provide incommensurate information, they can not be averaged. For example, kinetic depth with parallel perspective provides a depth map up to an unknown *additive* constant (and up to depth reversals), whereas texture perspective (assuming polar perspective) provides a depth map up to an unknown *multiplicative* constant. Unless these two parameters are estimated, the two depth cues can not sensibly be combined. On the other hand, if two cues are already commensurate then they may be combined at an earlier stage. For example, if both cues are scaled by an estimate of the viewing distance to a particular feature, then the cues may be averaged without estimating that distance, and scaled later if needed.

This viewpoint on the depth combination problem formalizes the possible ways in which depth cues may interact. Depth cues may simply average, as we have suggested. The weights $\alpha_\bullet$ should reflect the relative reliabilities of each cue. If one cue is known to be far more reliable than another, then the presence of that cue should result in a high weight for it and correspondingly low weights for the others. Thus, the presence of a strong cue will cause it to *veto* the other cues (Bülthoff, in press; Bülthoff & Mallot, 1987). One cue may help determine the weight of another. For example, an occluding contour can make the depth from shading more salient (Ramachandran, 1988). Cues can interact by one cue determining the unknown parameters $\overline{\gamma_\bullet}$ of the other cue. For example, an occlusion cue (which merely provides an indication of depth order) may be used to specify which of the two depth-reversed interpretations are seen in a kinetic depth display (Braunstein et al., 1982; Proffitt et al., 1984), thus specifying γ_{k_2}.

When multiple cues are averaged, it is desirable that the resultant combined estimate be distributionally *robust*. Robust estimation theory (Hampel, 1974; Huber, 1981) can be taken as suggesting that the weights used to average multiple cues should depend not only on their estimated reliability, but on the depth estimates themselves, specifically the *consistency* of the estimates. For example, suppose that five different depth cues all estimated the distance to a particular object as about two meters, but one additional cue estimated that same distance as twenty meters. A robust estimator will correspondingly lower the weight of the discrepant cue.

Consider an experiment involving stimuli for which the depth portrayed by one (discrepant) cue is manipulated independently of those of five other cues. Suppose the five cues are consistent and specify the same amount of depth to the feature (say, 2 meters) and that the discrepant cue is varied so as to portray a wide range of depths. If we measure perceived depth from the ensemble of cues as the discrepant cue increased from two meters up to twenty, the combined estimate should rise at first (with a slope equal to the weight applied to that cue), and then drop back down to two meters as the discrepant cue is noted and ignored by lowering its weight. The curve which results (the robust average as a function of the value of a single datum) is called an *influence function* (Hampel, 1974; Huber, 1981).

In this paper, we test whether this statistical model of depth cue combination adequately represents cue combination in human vision. We manipulate displays containing two cues: kinetic depth and texture gradients. We will show that the cues are combined using a weighted average. We show further that the weights are determined by an estimate of the reliability of each cue. And, we have some indication that the estimator is robust in that the weight of a cue is lowered when its value becomes too discrepant.

The theory behind the psychophysical method is as follows (Maloney & Landy, 1989). Subjects are shown two displays containing both kinetic depth and texture information. In one display, both cues are consistent so that perceived depth should be equal to portrayed depth ($d' = d_k' = d_t'$). In the other display, the two depths are discrepant: $d_t = d_k + \Delta\text{cue}$. Using a forced-choice paradigm, the value of d' is determined which results in the two displays having subjectively equal depth. For this value of d' we have

$$\begin{aligned} d' &= \alpha_k d_k + \alpha_t d_t \\ &= \alpha_k d_k + \alpha_t (d_k + \Delta\text{cue}), \end{aligned}$$

and thus

$$\alpha_t = \frac{d' - d_k}{\Delta\text{cue}} = \frac{\Delta\text{depth}}{\Delta\text{cue}}. \tag{3}$$

In other words, the weight α_t may be estimated psychophysically. We may then determine the dependence of α_t on the reliability of the cue (experimenter-imposed) and on how discrepant d_t is from d_k.

3. METHODS

3.1. Displays

Subjects viewed displays which simulated the front half of a vertical elliptical cylinder (a stretched tin can) rotating back and forth about a horizontal axis. The cylinders were covered with a texture of dark spots on a lighter background. Sample stimuli are shown in Fig. 1A.

Stimuli included both a textural cue and a motion cue to depth. The portrayed depth from each cue was varied independently as follows. First, an elliptical cylinder surface was 'carved' from a simulated slab of a three-dimensionally textured material with the depth extent determined by the depth to be portrayed by the texture cue (Fig. 1B). This textured surface was projected onto a cylinder with the depth extent to be portrayed by the motion cue, and then rocked back and forth about a horizontal axis. These images were projected (using parallel projection throughout) onto the image plane.

The axis about which a cylinder rotated was one of the cylinder's ellipse axes, midway between the bottom and top of the display, perpendicular to the line of sight. The length of this axis (the *width axis*) corresponded to the width of the displayed surface, and was constant across surfaces in these experiments. The length of the other ellipse axis (the *depth axis*) was varied to give impressions of cylinders of variable depth extent, independently for the texture and motion cues. Thus, for a particular display there is a *texture depth* d_t and a *motion depth* d_k.

The method of generating the textured cylinders was modeled after the technique of Johnston, Cumming and Parker (1990). A virtual box is packed with randomly positioned, non-overlapping balls. The balls are black and the rest of the box is white. It is as if a solid block of marble has been simulated. The textured elliptical cylinder (of depth d_t) is 'carved' from this marble slab. The black balls which intersect the surface of the cylinder result in black texture elements.

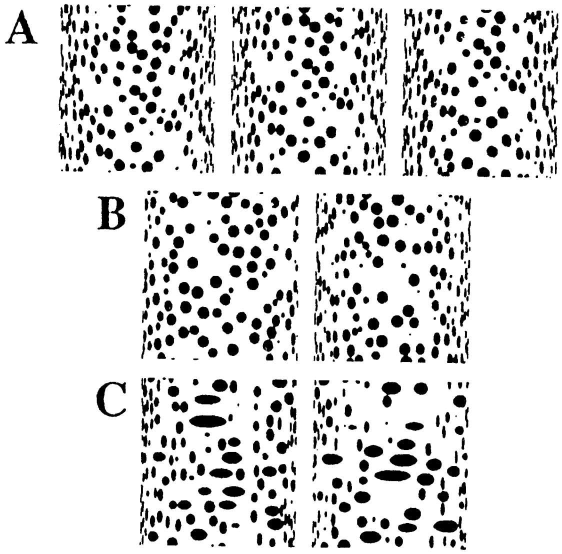

Figure 1. (A) Sample frames from a sequence combining both texture and motion cues to depth (10° downward, face-forward, and 10° upward). (B) Two stimuli with different values of d_t. (C) Stimuli with small (0.25) and large (0.5) amounts of variability imposed on the texture cue from Expt. 3.

Boxes were packed with balls to a density of approx. 0.25. For Expts. 1 and 2, the balls were spheres with radii of a fixed length (the varying texture element size derives from the different portions of the balls cut by the surface). For Expt. 3, the balls were ellipsoids with axes in the same directions as the cylinder axes. The vertical ellipsoid axis was of a fixed length, equal to that of the radii in Expts. 1 and 2. The other two ellipsoid axes were of equal lengths and were varied randomly in order to reduce the reliability of the texture cue (Fig. 1C). The varying axis lengths were randomly selected from a range and then tested for overlap with previously placed balls. The range was selected so as to yield a a distribution of lengths with mean equal to the fixed lengths of Expt. 1 and specific standard deviation (coefficient of variation 0.25 and 0.50). As the likelihood of fitting a larger ellipsoid decreases with increasing density, the resulting distributions of ellipsoid length are positively skewed.

The textured surface carved from the box of balls was spatially sampled. For each element of a grid of 564 × 240 pixels in the image plane (spanning the maximum cylinder height over all experimental conditions), it was noted whether a texture element on the cylinder surface overlay that pixel, and at what depth. (Note that parallel projection is used throughout.) Then, all of these depth values were scaled proportionally (by a factor of d_k/d_t, resulting in a textured cylinder with depth d_k. The resulting surface was rotated in steps of 1° from -10° to +10°, parallel projected to the image plane, and then windowed to a resolution of 240 × 240 pixels. This clipping prevented the observer from seeing the top and bottom edges of the cylinder. If visible, the curvature (after rotation) of these edges would have provided an additional cue to the shape and rotation extent.

We were concerned that subjects use perceived depth as their decision variable, rather than using a simpler strategy. Such strategies include choosing the stimulus with the faster speed, the shorter rotation time, the longer distance traveled by each texture element, etc. In order to make such strategies less useful, displayed image subsequences were chosen randomly from the 21 computed frames. Each time a stimulus was to be displayed, a subsequence of the available frames was chosen. First, the extent of rotation, which was always symmetric about the line of sight, was chosen randomly from ±3° to ±10°. Second, the rotation speed was chosen randomly from the range 1°/frame to 2.25°/frame. Given these two parameters, the subsequence was chosen. Note that the actual rotation amount on a given frame varied from that determined by the rotation extent and speed given the availability only of frames rotated in integer amounts from the face-front position.

A complete stimulus sequence was constructed by displaying two such surfaces side by side, with rotation magnitudes and velocities determined randomly and independently. A stimulus display was 60 frames in length. Each subdisplay began with the downward rotated frame, continuing upward to the opposite extreme, then back down. This cycle was repeated as many times as could fit fully within the total length of 60 frames. Extra frames not filled with a subdisplay were left blank (the intensity of the background). Three full rotations were displayed on the average. Frames were displayed at 12 Hz in Expts. 1 and 3 (five 60 Hz refreshes of each stimulus display frame). In Expt. 2, cylinders of greater depth were used. Stimuli were displayed at 7.5 Hz (eight 60 Hz refreshes per display frame) to maintain similar stimulus velocities. Each displayed surface window subtended 3.5 deg and was viewed from a distance of 1.7 m. The two windows were separated by 15 min.

The perceived depth of these displays depends on observer calibration of viewing distance. We took pains (using a reduction screen) to reduce the cues to absolute depth to the display. Nevertheless, in the sequel portrayed depth values are given in absolute distance terms. The 3.5 deg wide surface viewed from a distance of 1.7 m had a physical width of 10.5 cm. If a circular cylinder was portrayed, its depth will be reported as 5.25 cm (i.e., half the actual width). The observers may well have miscalibrated the viewing distance, but this should have an identical effect on the depth perceived for the two cues since parallel perspective was used for both. For example, the balls used to generate the textures had radii approximately equivalent to .44 cm (for Expts. 1 and 2, mean for Expt. 3).

Stimuli were displayed using an Adage RDS-3000 display system on a US Pixel PX15 monochrome monitor with a fast P4-like phosphor. The video format was 60 Hz, noninterlace, with 480 × 512 visible pixels. The background luminance (within the displayed textured cylinders, and elsewhere on the monitor screen as well) was 31 cd/m^2. The texture spots were black. Displays were viewed monocularly through a reduction screen approximately matched in luminance and color to the background.

3.2. Procedure

In each trial, two stimuli were displayed side-by-side. In one of the two stimuli $d_k = d_t$ (the *consistent cues* surface). In the other, the two depth values were generally inconsistent (the *mixed cues* surface). The subject's task was to indicate with a key press which surface appeared to extend further in depth. The mixed cues surface appeared equally often on each side of the display across trials. A response terminated the stimulus and the next trial commenced after a 1.0 sec pause. In the absence of a response, the entire 60 frames in the stimulus sequence were displayed repeatedly with a 0.5 sec pause between each redisplay until a response was made. If the subject viewed the entire 60 frame display, in most cases each surface completed several rotations and disappeared prior to the end the 60 frames due to the different ways stimuli were sampled (see above). Additional blank stimulus frames added up to a 20 frame delay (1.7 sec for Expts. 1 and 3, 2.5 sec for Expt. 2) to the 0.5 sec pause between repeats.

For each experiment, d_k for the mixed cues surface was fixed (6.6 cm for Expts. 1 and 3, 10.9 cm for Expt. 2). The mixed cues surface d_t varied between blocks of trials (the values were 4.4, 5.5, 6.6, 7.7, and 8.8 cm for Expts. 1 and 3, 4.4, 6.6, 8.8, 10.9, 13.1, 15.3, and 17.5 cm for Expt. 2). During a block, only the consistent cues surface depth d' varied. A block consisted of 50 comparisons with each of the consistent cues surfaces in random order (depths of 3.3, 4.4, 5.5, 6.6, 7.7, 8.8, and 9.8 cm in Expts. 1 and 3, and depths of 4.4, 6.6, 8.8, 10.9, 13.1, 15.3 and 17.5 cm in Expt. 2). In Expt. 3 the variance of the texture information in the mixed cues surface was fixed during a block, but across blocks took on values of 0.25 and 0.5 (coefficient of variation, see above). The consistent cues surfaces were the same as in Expt. 1 – there was no noise added to the texture information for depth in the consistent cues surfaces in any of the experiments.

In Expts. 1 and 2, two blocks were run at each level of d_t for the mixed cues surface, yielding 100 comparisons at each value of d' for each mixed cues surface (10 blocks for Expt. 1 and 14 for Expt. 2). In Expt. 3, two blocks were run at each level of texture variance, for a total of 20 blocks. For each experiment, all blocks were run in random order. The left-right positioning of the mixed and consistent cues surfaces was randomly determined for each trial. The rotational extent and velocity of each surface in a trial was separately determined by random sampling from the extent and velocity ranges.

3.3. Subjects

In all results shown here the subject was the third author. Pilot data on two additional subjects is consistent with these data but incomplete. All subjects had normal or corrected-to-normal vision.

4. RESULTS

The results of Expt. 1 are shown in Fig. 2. Fig. 2A gives the psychometric functions, and it is clear that increasing the depth portrayed by texture in the mixed cues stimulus (d_t) decreases the likelihood that a given consistent cues surface will be perceived as having more depth (or having greater curvature). These data were fit with a cumulative Weibull distribution, and from this fit function the point of subjective equality (PSE) was estimated from the 50% point of the function. These PSEs (the value of d' which matches a given mixed cues display) are plotted in Fig. 2B as a function of d_t. It is clear that the data fall on a straight line, consistent with the averaging model of Eq. 2. The slope may be used as an estimate of α_t. We estimated this value by fitting a regression line to the data of Fig. 2B, yielding a value of 0.62. Thus, the weight on the texture cue turns out to be higher than the weight on the motion cue under these experimental conditions, which was unexpected. One possible explanation is that the particular choice of a vertical cylinder rotating about a horizontal axis leads to a weak depth percept (from the motion cue) and that more compelling motion displays (e.g., using a vertical rotation axis with the current set of shapes) would yield a higher weight for the motion cue.

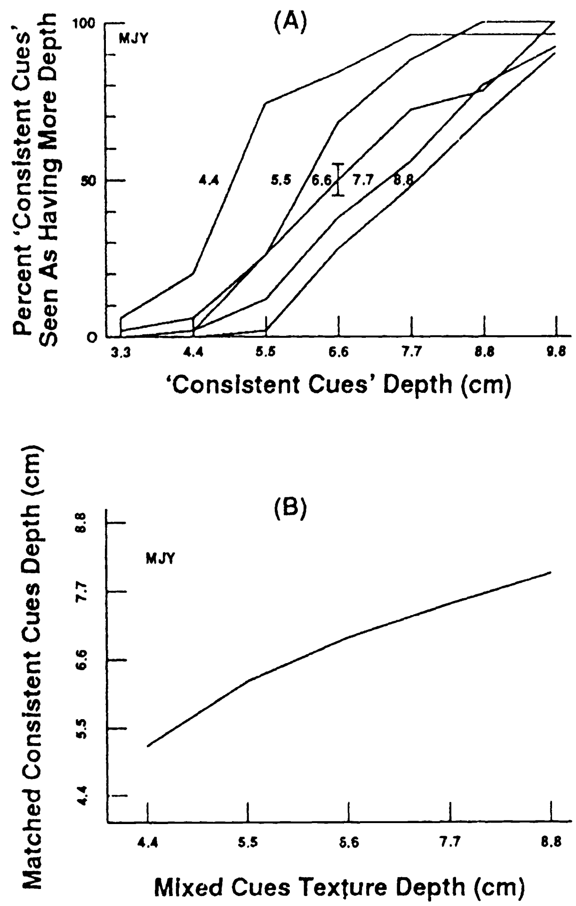

Figure 2. (A) Psychometric functions from Expt. 1. The amount of depth portrayed by motion in the mixed cues stimulus (d_k) was 6.6 cm throughout. The depth portrayed by texture in the mixed cues stimulus (d_t) varied from block to block and is indicated as the parameter of each curve. There were 100 trials for each point. A representative error bar is shown (plus or minus one sample error of the mean). (B) The matched consistent cues depth d' as a function of the displayed texture depth from the mixed cues stimulus d_t. Each psychometric function in part A was fit using a cumulative Weibull function and the point of subjective equality was derived as the 50% point of that function.

There is an orderly progression of the psychometric functions in Fig. 2A, so that an increase of d_t shifts the curve to the right without changing its shape. This is an important observation because it implies that the observer is not doing anything substantially different with the more inconsistent cues (leftmost and rightmost curves) as compared with trials in which the cues are consistent in both the consistent cues stimulus and the (nominally) mixed cues stimulus (the middle curve in Fig. 2A).

Consider an alternative to the averaging model of Eq. 1 which we call *strategy mixture* (see Sperling & Dosher, 1986). When two cues are inconsistent, one possible strategy is to 'combine' the two depth values by simply choosing to attend one depth cue to the exclusion of the other. With only two cues, both of which are reliable indicators of depth, this might result in a mixture strategy where on some trials the value of d_t was used, and on other trials d_k was used. If a subject employed this strategy, the psychometric functions for mixed cues stimuli should have increasingly shallow slope as the cues become more inconsistent. We see no evidence for this in our data, and conclude that subjects are truly averaging the depth values provided by the two cues.

It was mentioned that weight given to a particular cue should be decreased if the value of that cue is quite disparate from the values of several other cues if the depth estimator is to be robust. In Expt. 2, this issue was explored by repeating the experiment using a wider range of values of d_t. The results are shown in Fig. 3. Although the data are not compelling, it appears that the largest values of d_t are having less of an affect on the combined depth estimate for the mixed cues shape (the curve in Fig. 3B appears to be flattening for large values of d_t). In other words, as the depth value for texture becomes more disparate, the weight on the texture cue is decreasing. However, there is an important reason for questioning this implication. Theoretically, a cue should be downweighted if it is strongly inconsistent with *several* other cues all of which are mutually consistent themselves. In the experiments reported here only two cues were manipulated (and a reduction screen and monocular viewing was used to reduce the availability of other cues such as binocular stereopsis). Hence, as the texture cue became increasingly disparate from the motion cue, there is no theoretical impetus for choosing either cue over the other. A strategy mixture does not appear to have been employed either, since the slopes of the curves in Fig. 3A are reasonably consistent. Thus, the decreasing effectiveness of the texture cue with increasing values of d_t might well imply a lowered value of α_t, but this nominally robust strategy must be using side information (or outright assumption) in order to justify attending to the motion rather than the texture information.

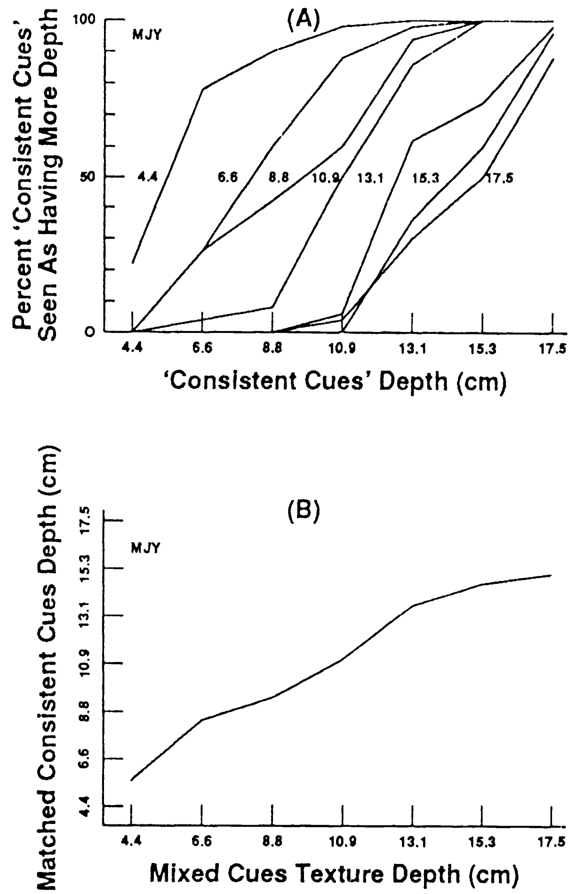

Figure 3. The results of Expt. 2, in which the range of depths portrayed by the texture cue was increased. (A) Psychometric functions. (B) The PSE as a function of the portrayed texture depth d_t.

In Expt. 3 the reliability of the texture cue was manipulated experimentally by varying the shapes of the texture elements randomly. The results are shown in Fig. 4. It is clear that increased texture variability decreases the effectiveness of the texture cue. Fitting a regression line to the PSE data results in estimates of α_t of .41 and .18 for the 0.25 and 0.5 texture size variability conditions, respectively.

5. SUMMARY

We have introduced a statistical paradigm for the combination of multiple depth cues. The theory suggests that the depth estimates from multiple cues should be combined using a weighted average, and that the weight of a cue should reflect its estimated

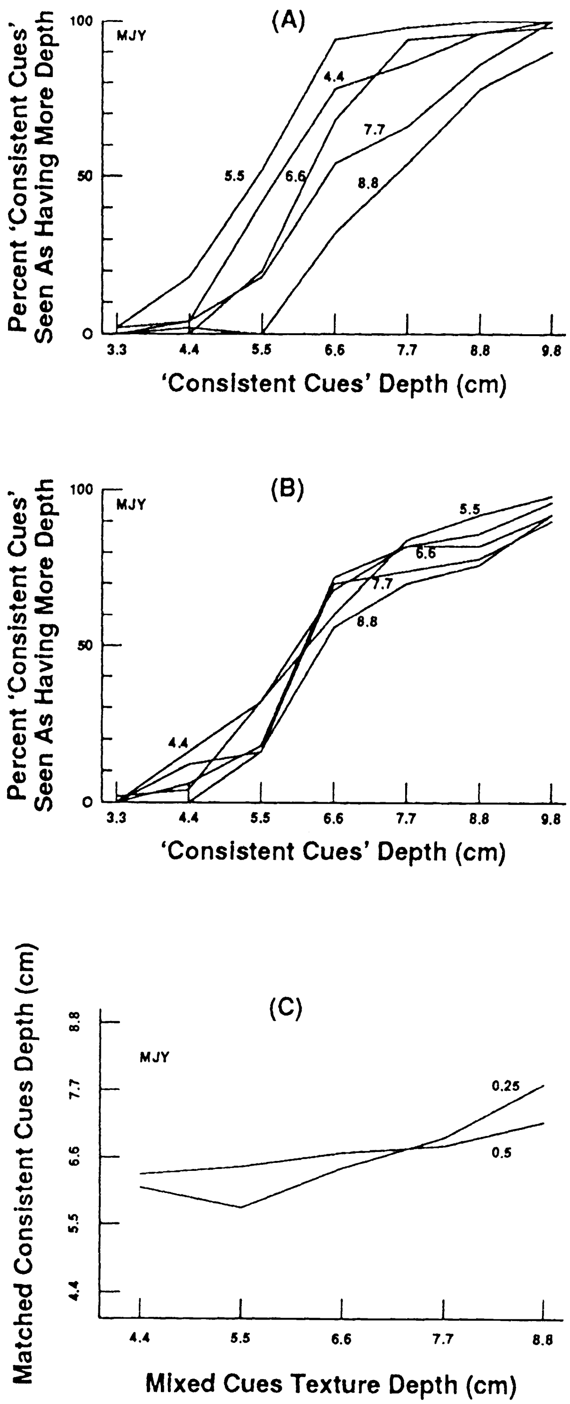

Figure 4. The results of Expt. 3, in which the reliability of the texture cue was reduced by adding noise to the texture element shapes. (A) Psychometric functions (coefficient of variation for texture of 0.25). (B) Psychometric functions (coefficient of variation for texture of 0.5). (C) The PSE as a function of d_t for the previous data.

reliability along with the consistency of the depth values estimated from that cue with those from other cues. In experiments combining texture and motion information all of these predictions received tentative confirmation. Further work will elaborate on a number of the issues raised in this paper, as well as examining other depth cues such as binocular stereopsis and linear perspective.

6. ACKNOWLEDGEMENTS

This work was supported in part by grant EY08266 from the National Eye Institute. We thank Robert Picardi, as usual, for helping to keep the equipment in working order. We also thank Elizabeth Johnston for helpful suggestions, both theoretical and practical.

7. REFERENCES

Braunstein, M. L., Andersen, G. J. & Riefer, D. M. (1982). The use of occlusion to resolve ambiguity in parallel projections. *Perception & Psychophysics*, *31*, 261-267.

Bruno, N. & Cutting, J. E. (1988). Minimodularity and the perception of layout. *Journal of Experimental Psychology: General*, *117*, 161-170.

Bülthoff, H. H. (in press). Shape from X: Psychophysics and computation. In Landy, M. S. & Movshon, J. A. (Eds.), *Computational Models of Visual Processing*. Cambridge, MA: MIT Press.

Bülthoff, H. H., and Mallot, H. A. (1987). Interaction of different modules in depth perception. *Proceedings of the IEEE First International Conference on Computer Vision*, 295-305.

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. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research*, *26*, 973-990.

Epstein, W. (1968). Modification of the disparity-depth relationship as a result of exposure to conflicting cues. *American Journal of Psychology*, *81*, 189-197.

Gillam, B. J. (1968). Perception of slant when perspective and stereopsis conflict: Experiments with aniseikonic lenses. *Journal of Experimental Psychology*, *78*, 299-305.

Goodale, M. A., Ellard, C. G. & Booth, L. (1990). The role of image size and retinal motion in the computation of absolute distance by the Mongolian gerbil (*Meriones unguiculatus*). *Vision Research, 30*, 399-413.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association, 69*, 383-393.

Huber, P. J. (1981). *Robust Statistics.* New York: Wiley.

Johnston, E. B., Cumming, B. G. & Parker, A. J. (1990). Stereo-texture interactions in 3D shape perception. *Investigative Ophthalmology & Visual Science (supp.), 31*, 304.

Maloney, L. T. & Landy, M. S. (1989). A statistical framework for robust fusion of depth information. *Proceedings of SPIE: Visual Communications and Image Processing, Part 2*, 1154-1163.

Proffitt, D. R., Bertenthal, B. I. & Roberts, R. J., Jr. (1984). The role of occlusion in reducing multistability in moving point-light displays. *Perception & Psychophysics, 36*, 315-323.

Ramachandran, V. S. (1988). Perceiving shape from shading. *Scientific American, 331*, 133-166.

Schwartz, B. J. & Sperling, G. (1983). Luminance controls the perceived 3-D structure of dynamic 2-D displays. *Bulletin of the Psychonomic Society, 21*, 456-458.

Sperling, G. & Dosher, B. A. (1986). Strategy and optimization in human information processing. In Boff, K. R., Kaufman, L. & Thomas, J. P. (Eds.), *Handbook of Perception and Human Performance, Vol. I* (pp. 2-1 - 2-65). New York: Wiley.

Stevens, K. A. & Brookes, A. (1987). Depth reconstruction in stereopsis. *Proceedings IEEE First International Conference on Computer Vision*, 682-686.

Stevens, K. A. & Brookes, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research, 28*, 371-386.

Todd, J. T. & Akerstrom, R. (1987). Perception of 3-dimensional form from patterns of optic texture. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 242-255.

Todd, J. T. & Mingolla, E. (1983). Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 583-595.

Wallach, H. & Karsh, E. B. (1963). The modification of stereoscopic depth-perception and the kinetic depth effect. *American Journal of Psychology, 76*, 429-435.

Youngs, W. M. (1976). The influence of perspective and disparity cues on the perception of slant. *Vision Research, 16*, 79-82.