Combining multiple cues for texture edge localization

Michael S. Landy

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

ABSTRACT

A visual feature such as an edge may be signalled by one of any number of visual cues such as modulation of brightness, color, stereo disparity, motion, texture and so on. Here, a theory is reviewed which suggests that when more than one cue is available, separate location estimates are made using each cue, and the resulting estimates are averaged. Cues which are more reliable are given a correspondingly larger weight in the average. A psychophysical paradigm (*perturbation analysis*) is described for testing whether the theory applies and estimating its parameters. Preliminary results are described for localization of texture edges signalled by changes in contrast, local texture orientation and local texture scale. The data are consistent with the theory: In the data, individual cues appear to be averaged and more reliable cues are given more weight.

1. INTRODUCTION: MODIFIED WEAK FUSION

In many cases there are multiple cues which may be used to solve a given visual task. The obvious example is depth perception. Any perception textbook (e.g. Kaufman's[6]) provides a long list of cues which may be used to estimate (or at least to constrain) depth. These include binocular stereopsis, texture gradient, structure-from-motion, motion parallax (such as that generated by observer motion), accommodation, convergence, and so on. It is important to understand whether and how any given cue is used by the visual system, and much work has been done on each of these cues in isolation. However, it is equally important to analyze how the visual system works in the natural environment, in which cues are rarely available in isolation.

In our previous work[4-5,9-12,17-18], it was suggested that perceived depth from multiple cues is computed by human observers as a weighted average of the depth estimated using each cue alone,

$$\text{depth} = \sum_i \alpha_i \, d_i \ , \tag{1}$$

where d_i is the depth estimate from cue i, and α_i is the weight given to that cue. In general, it is assumed that the weights sum to one so that this is truly a weighted average of the individual estimates.

The weight applied to a cue depends on several factors. For example, a cue which is perceived as less *reliable* than another cue should be given less weight. Also, there is good reason for an observer to use a *robust* estimator.[2-3] For example, if several estimates agree reasonably well, and one estimate is extremely inconsistent, it makes sense to downweight (or even completely ignore) the discrepant estimate, and give more credence to the consistent estimates. We call this statistical method of combining cues (robust averaging with weights based on cue reliability) *Modified Weak Fusion*[10] (as opposed to weak fusion, as defined by Clark and Yuille[1]).

Although these ideas about combining multiple cues were first described as a model for the visual perception of depth, it is clear that the aspects of the modified weak fusion model just described are applicable to any situation in which multiple cues are available and these cues are all useful in estimating a numerical (or at least 'averageable') quantity. In this paper, the modified weak fusion theory is shown to be equally applicable to an entirely different visual task: spatial localization.

 0-8194-1146-9/93/$6.00

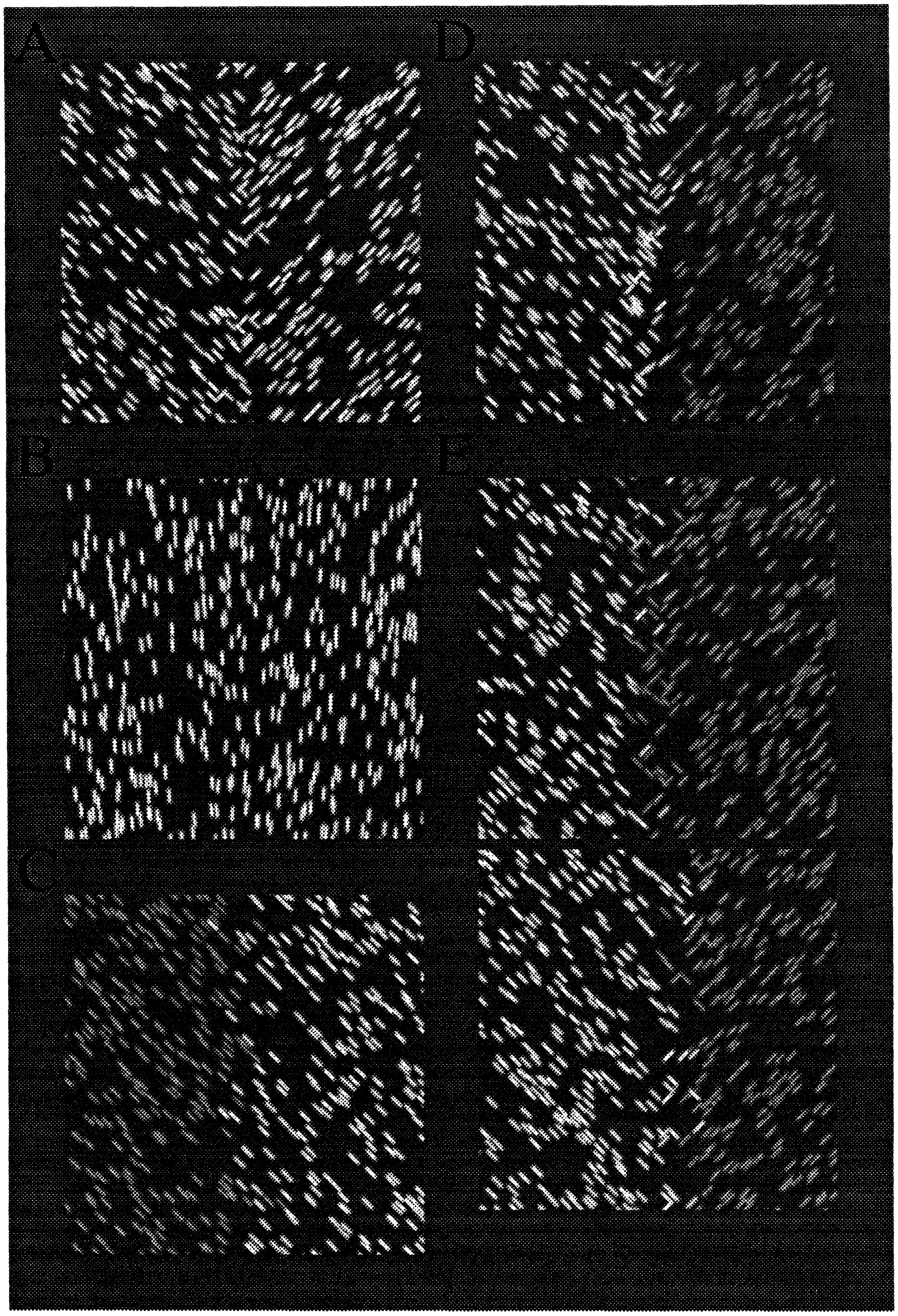

Fig. 1. Example stimuli from Expt. 1. (A) Orientation cue alone, Δ*orientation* = ±45°. (B) Δ*orientation* = ±5°. (C) Contrast cue alone, Δ*contrast* = ±25%. (D) Both cues. The contrast cue edge is shifted .18 deg to the right of the orientation cue edge. (E) A typical stimulus display. The upper texture edge is an inconsistent cues edge. The lower edge is a consistent cues edge and its location has been shifted rightward.

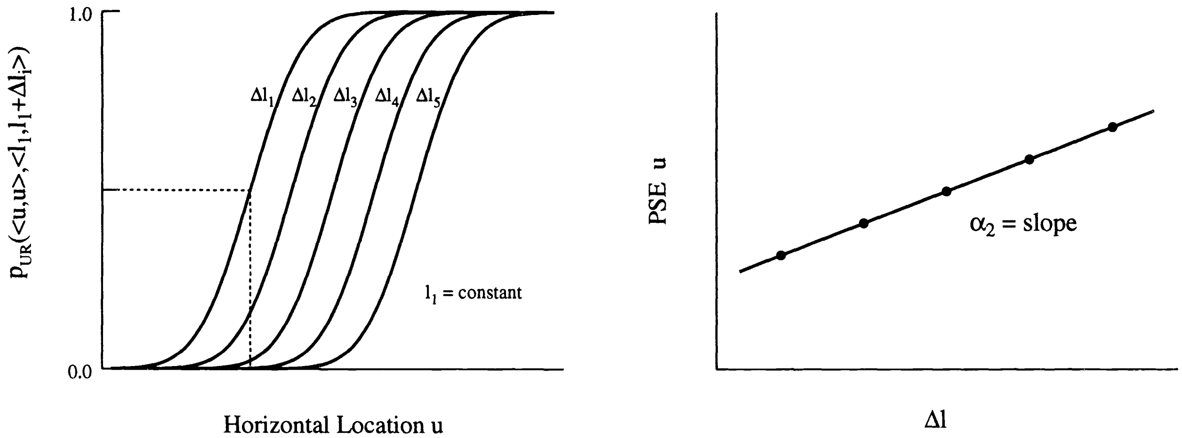

Fig. 2. (A) A series of psychometric functions predicted for an experiment in which an upper edge with two consistent cues which specify the edge location to be u is compared to a lower inconsistent cues edge where one cue is fixed at location l_1 and the other cue is shifted to location $l_2 = l_1 + \Delta l$. Each curve corresponds to a fixed choice of Δl. The ordinate indicates the probability that the observer perceives the upper edge to be to the right of the lower edge as a function of the position u of the upper edge. The dashed lines indicate the determination of the PSE for the leftmost curve. (B) If cues are averaged, this graph of the PSE as a function of Δl should be a straight line with slope equal to α_2 (the weight applied to the 2nd, shifted cue).

Consider the task of localizing texture borders such as those shown in Fig. 1. In this paper we describe experiments involving a vernier alignment task. Each stimulus display consists of two texture patches, one above the other, each of which contains a vertical edge (Fig. 1E). The subject's task is to indicate whether the upper edge is to the right or left of the lower edge. If each patch of texture contains only a single cue to localization (such as in Fig. 1A), then the data provide an estimate of $p_{UR}(u,l)$: the probability that the subject responds that the upper edge is to the right when the upper edge is at location u and the lower edge is at location l. If l is kept fixed through a block of trials and u is varied, one might expect the data to fall near a typical ogive curve such as a cumulative normal distribution.

When two cues are available to signal edge position, it is possible to create stimuli in which the two cues are inconsistent (such as the orientation and contrast cues in Fig. 1D) and signal different locations. Clearly, if these locations are sufficiently different then the subject will see two distinct edges and the task becomes ill-defined. This is how robustness considerations come into play with localization judgments. However, for two reasonably consistent cues, there are now four cue locations (two cues in each of two patches), and the data provide estimates of $p_{UR}(<u_1,u_2>,<l_1,l_2>)$ where the subscripts indicate the individual cues signalling edge location.

In our previous work[10-12,18] we have described an empirical method (called *perturbation analysis*) for testing the modified weak fusion theory and estimating individual cue weights. This straightforward paradigm works as follows. For a fixed *inconsistent cues* stimulus $<l_1,l_2>$ (where $l_2 = l_1 + \Delta l$) we measure the observer's performance using a range of *consistent cues* stimuli $<u,u>$ while varying u as before. This will again trace out an ogive psychometric function (Fig. 2A). We define the 50% point or point of subjective equality (PSE) as an estimate of where $<l_1,l_2>$ is localized (relative to the probe edge u). The experiment is repeated for a variety of values of Δl and the PSEs are then plotted as a function of Δl (Fig. 2B). If the theory applies to this task, then this graph of PSEs should be a straight line. A trivial bit of algebra reveals that the slope of this line provides an estimate of α_2. It is called perturbation analysis because cues are only perturbed away from complete consistency (the largest value of Δl should be kept relatively small). Otherwise, the stimulus may trigger the downweighting of a cue required by robust estimation (and the stimulus may appear to have two separate edges). In other words, this procedure is used to analyze what the subject does when cues are taken as being sufficiently consistent to simply average them, and any significant nonlinearity in the PSE plot indicates that averaging is not the method used by the observer.

Below, we report two preliminary experiments. Both experiments use the perturbation analysis procedure to examine vernier localization performance. In the first, the textures involved randomly placed oriented line segments, and in the second the textures involve spatially filtered noise. In both cases the results are consistent with the modified weak fusion theory.

2. EXPERIMENT 1: ORIENTED LINE TEXTURES

2.1 Methods

Stimuli. The stimuli consisted of randomly placed line segments on a grey background. There were three kinds of edges distinguished by the cues which indicated an edge was present. In one set of control experiments, line contrast was constant across the display and the edge was signalled by line orientation (Figs. 1A-B). In a second set of control experiments, line orientation was constant and line contrast signalled the edge (Fig. 1C). In the main experiment, both cues were used (e.g. Fig. 1D). In the control experiments, for each condition a stimulus was computed which was 1632×400 pixels containing a total of 4700 line segments (and a subset of a given strip was displayed in each trial). In the main experiments each stimulus strip was half as long and had half as many line segments. Each line segment had a length of 10 pixels (.15 deg as viewed from 1 m) and a Gaussian blurred shape (σ = 1 pixel = .015 deg). (Note that the term 'deg' is used here for degrees of visual angle, a measurement of stimulus size, and the symbol '°' is used to indicate degrees of rotation of a line segment or oriented texture.)

The stimulus generation worked as follows. For each line segment to be drawn, a random x,y location was chosen uniformly over the stimulus, and these determined the center of the line segment. Its contrast and orientation were then determined. For the control stimulus displayed in Figure 1A, the contrast was fixed at 100%, and the orientation was determined randomly based on the horizontal location of the edge (the x value). The curve shown in Figure 3A was used to determine the probability that a given line segment would be oriented up-and-rightward. As you can see, line segments placed in the far left portion of the stimulus were nearly certain to be oriented up-and-leftward, and those in the far right were nearly certain to be oriented up-and-rightward. In the neighborhood of the edge the probability of the up-and-rightward orientation passed smoothly from 0 to 1. The curve which determines the probability has a cumulative Gaussian shape with a value of σ = 12 pixels = .18 deg. This *probability blur* was used to render precise localization difficult. The edge displayed in Fig. 1B was made more difficult to localize by reducing the change in orientation across the edge, thus reducing cue reliability. The texture edge shown in Figure 1C was constructed similarly, although here the line segment orientation is fixed and the probabilistic edge is used to control the contrast of each line segment. In the main experiments, both contrast and orientation were controlled using separate probability blur curves (Fig. 3B), allowing separate control of the edge location signalled by each cue (Fig. 1D).

On a given trial 256×256 pixel patches were randomly extracted from two such stimulus strips and displayed one above the other (Fig. 1E). The subject's task was to indicate with a button press whether the edge in the upper stimulus patch was to the right or to the left of the edge in the lower patch (i.e. a texture vernier discrimination). Each patch was 3.8×3.8 deg, with a space of .12 deg between the two patches. The stimuli were viewed from 1 m distance. The stimuli were displayed on a Sony GDM-1950 monitor (1280×1024 pixels displayed at 60 Hz noninterlace) controlled by a Pixar II Image Computer, and computed using the HIPS software.[8] The background luminance was 35 cdm^2 and luminance was linearized using uniform patches as measured using a Minolta Chromometer CS-100.

All experiments (orientation control, contrast control and main experiment) were parameterized by the variations in contrast and orientation used in the stimuli (that is, Δcontrast and Δorientation). In some

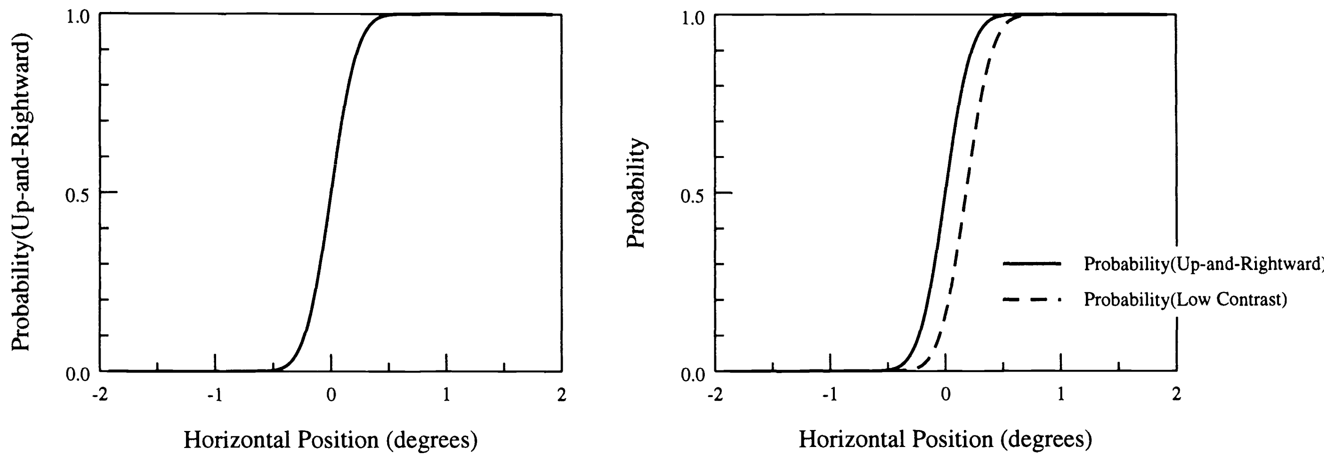

Fig. 3. Stimulus construction in Expt. 1. (A) In a one-cue stimulus, the probability of a line segment taking on a particular feature value (in this case, orientation) is a function of horizontal location. (B) In a two-cue stimulus, different probability curves are used for each of the two cues, resulting in an inconsistent cues stimulus.

experiments contrast varied ±25% (dim lines had 50% contrast and bright lines had 100% contrast; Fig. 1C); in others contrast varied ±10% (i.e. 65% and 85% contrast lines). In some experiments orientation varied ±45° (Fig. 1A), and in others it varied ±5° (as in Fig. 1B).

In a given orientation control experiment, 6 stimulus strips were precomputed: 3 with all low contrast line segments and 3 with all high contrast line segments, each with an edge determined by change in line segment orientation. In a given contrast control experiment, there were 3 strips with line segments oriented up-and-leftward and 3 strips with up-and-rightward orientation, and the edge was determined by change in line segment contrast. In the main experiment, there were 10 precomputed stimulus strips. In 5 strips the bright line segments were on the left, and in 5 strips the bright line segments were on the right. In all strips the orientation cue signalled an edge located in the center of the strip (at pixel 200 of the 400 pixel wide precomputed stimulus strip). The 5 strips in each group differed by the location signalled by the contrast cue which was shifted relative to the orientation cue edge (±12, ±6 and 0 pixels, or ±.18 deg, ±.09 deg and 0 deg).

Procedure. The method of constant stimuli was used throughout. In all experiments (orientation control, contrast control and main experiment) two blocks of trials were run for each of three combinations of Δorientation and Δcontrast: 45° and 25%, 45° and 10%, and 5° and 25%.

In the contrast control experiments there were 280 trials in a block. In a given trial, both edges would have the same fixed orientation. One edge (upper or lower) would be centered in the frame (the 'test' edge) and the other would be shifted right or left (to generate a function of probability of choosing the upper edge as being to the right as a function of the amount of actual shift). One edge would be displayed as is, and the other (top or bottom) would be reflected about a vertical axis. Thus, one edge had the high contrast on the left and the other had the high contrast on the right. The upper edge line segments all had one orientation and the lower edge line segments all had the opposite (reflected) orientation. A block consisted of all combinations of 2 fixed contrasts, 7 shifts of the upper edge location relative to the lower edge location (±24, ±16, ±8 or 0 pixels of shift, corresponding to ±.36, ±.24, ±.12 or 0 deg of shift), 2 locations of the 'test' edge (upper or lower), 2 locations of the reflected edge (upper or lower), and 5 replications of each. The entire stimulus was then jittered randomly left or right by as much as 8 pixels. This shifted patch was then extracted from one of the 3 precomputed stimulus strips chosen randomly, from a randomly chosen vertical portion of the strip (the horizontal portion was determined by the shift and jitter). Averaging over the irrelevant variables (fixed contrast, which had no effect on the results, and the choices of 'test' patch and reflected patch) resulted in 80 trials per shift over the two blocks of trials. The orientation controls were similar with the roles of contrast and orientation swapped. In the orientation controls the left-right reflection was irrelevant – all stimuli had up-and-leftward line segments in the left portion and up-and-rightward line segments in the right portion (Fig. 1A).

In the main experiments, a block consisted of 1120 trials which resulted from all combinations of the following: 7 shifts of the upper edge relative to the lower edge (as in the control experiments), 5 shifts of the contrast cue edge location relative to the orientation cue edge location in the inconsistent cues stimulus, 2 possible locations of the consistent cues stimulus (top or bottom), 2 types of top stimulus (high contrast on the left or right), 2 types of bottom stimulus (high contrast on the left or right), 2 stimuli to which the shift might be applied (top or bottom), and 2 replications of each combination. Again, the whole thing was then jittered left or right by as much as 8 pixels. Averaging over the various randomizations, this resulted in 64 trials per combination of edge shift and cue shift.

Each trial consisted of a 1 sec display of a fixation cue cross, .5 sec blank field at mean luminance, 200 msec stimulus flash, and a blank field which remained displayed until the subject's response, which in turn began the subsequent trial. No feedback was supplied. So far, only one subject (the author) has been run. He has corrected-to-normal vision.

2.2 Results

Figs. 4A-B show psychometric functions for the control experiments. It is clear that with a single cue the task was reasonably difficult, and with the low reliability cues ($\pm 5°$ or $\pm 10\%$), the task was nearly impossible. In the orientation control for a fixed orientation cue reliability ($\pm 45°$), the value of the constant contrast has no effect on performance. Likewise, in the contrast control for a fixed reliability ($\pm 25\%$), the value of the constant orientation has no effect.

Figure 4C shows psychometric functions for the main experiment for the highest quality displays (Δorientation = $\pm 45°$ and Δcontrast = $\pm 25\%$). As the edge indicated by the contrast cue in the inconsistent cues stimulus is shifted rightward relative to the edge indicated by the orientation cue, the entire psychometric function shifts rightward (compare these results to Fig. 2A). These curves were separately fit with cumulative normal distributions parameterized by mean, variance and probability of detecting the edge (i.e. asymptotic performance) using a maximum likelihood estimation procedure, and the 50% point (or PSE) was used as an estimate of perceived edge location. (Note that we could just as well have fit each reliability condition with a set of parallel psychometric functions, even omitting the detection parameter. This fits the data just as well; a nested hypothesis test confirms this, and the resulting fits affect none of the conclusions which follow.) These PSEs are displayed as a function of the cue shift in Fig. 4D along with the best-fitting line (determined by weighted least squares regression). It is clear that moving one cue relative to the other shifts edge location, although for this set of data linearity is rejected using a Bonferroni test on the five estimates relative to the prediction line ($p < .005$; linearity is accepted in the other two reliability conditions). The PSEs from all three cue reliability conditions are shown in Fig. 4E and the best fitting lines in Fig. 4F. As relative cue reliability is changed, the line rotates about the (0,0) point, indicating a change in the relative weights of the two cues as predicted.

3. EXPERIMENT 2: FILTERED NOISE TEXTURES

The oriented line experiments clearly show that the weights on cues in localization shift as relative cue reliabilities are changed. However, the task and its analysis were complicated by the fact that the reliability manipulations were too effective, often rendering the edge impossible to detect. In the next set of experiments filtered noise textures are used both to demonstrate the generality of the results and to produce a lowered reliability (that is, localizability) while retaining a detectable texture edge.

3.1 Methods

Stimuli. A number of aspects of this experiment are identical to Expt. 1 including the hardware, software, mean luminance, subject, viewing conditions, display timing and luminance calibration. A few aspects change

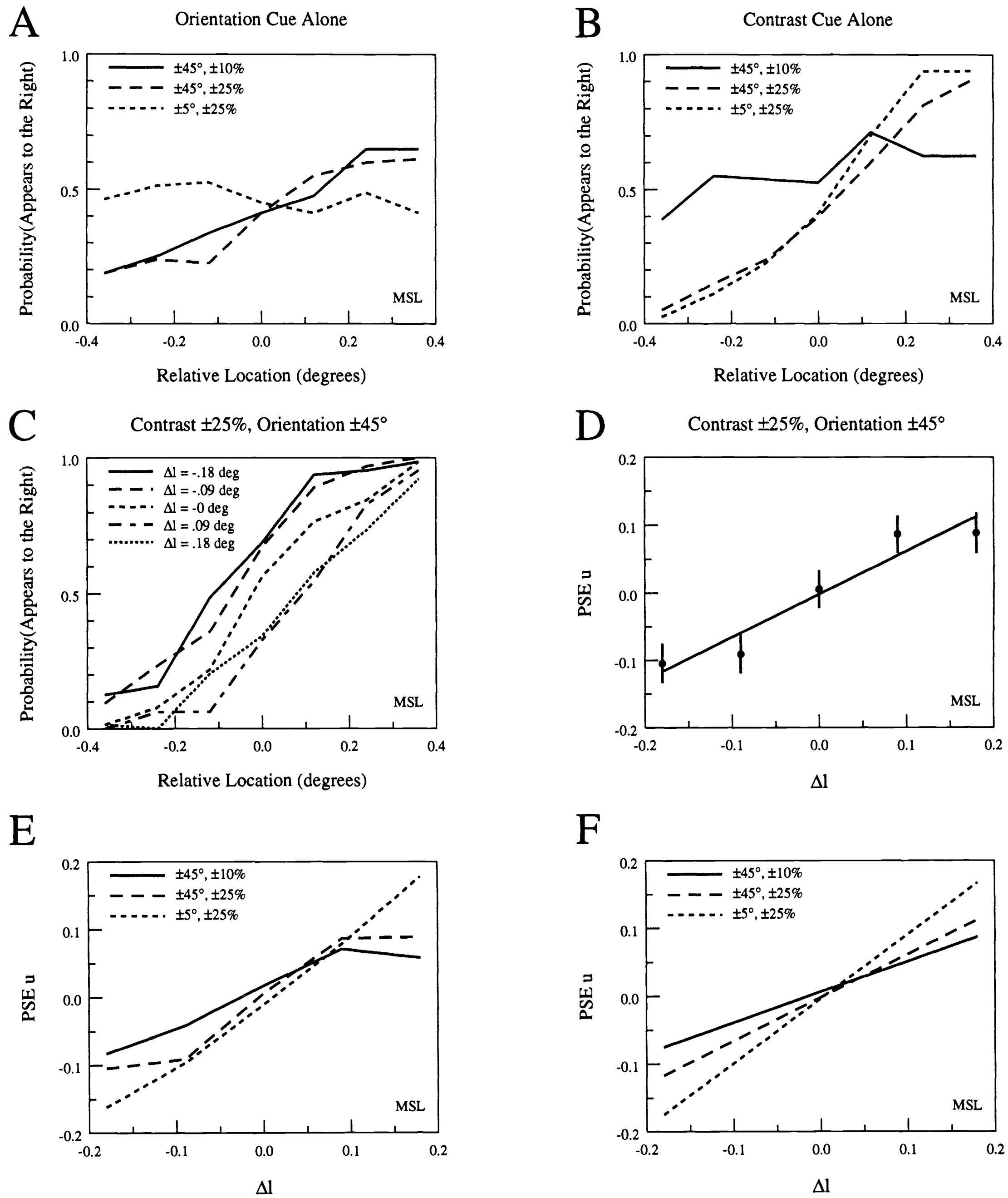

Fig. 4. Data for Expt. 1. (A) Psychometric functions for the orientation cue-only control experiments. Different curves represent different values of Δ*orientation* and Δ*contrast*. (B) Psychometric functions for the contrast cue-only control experiments. (C) Psychometric functions for the main experiments. Different curves represent different values of cue shift in the inconsistent cues stimulus. (D) PSEs from (C) as a function of cue shift. Error bars indicate two standard deviations of the estimated PSE parameter. The solid line is a best-fitting regression line (E) PSEs from all three reliability conditions. (F) Best-fitting regression lines to the data in (E). Note that slope (i.e. cue weight) varies with relative cue reliability in the predicted fashion (the three estimated values of slope are .451, .637 and .949).

only slightly from Expt. 1: stimulus strips are 2048 × 512 pixels, and stimulus patches are 400 × 400 pixels (or 6 × 6 deg) with .12 deg between the two patches. All stimuli are scaled to be 100% nominal peak contrast.

The stimuli consist of filtered noise. First, a Gaussian noise image is computed for which each pixel is an independent sample from a normally distributed random variable. This noise image is then filtered using a Gabor-shaped kernel, but for which the kernel used is a function of location across the image (see Ref. (7) for details). Two cues were investigated: spatial frequency and orientation. The spatial frequency cue involved a change from 3 to 6 cpd (Figs. 5A-B, which also show constant orientation values of vertical and horizontal, respectively). The orientation cue involved a rotation from horizontal to vertically oriented noise (Figs. 5C-D, which have constant spatial frequency values of 6 and 3 cpd, respectively). The orientation and spatial frequency at each horizontal location was determined by a cumulative Gaussian function (as in Fig. 3A, but with the y-axis determining either orientation of frequency). Cue reliability was varied by varying the σ of the cumulative Gaussian function. We refer to the σ of this function as the *feature blur*. For spatial frequency, feature blurs of .3 and .6 deg were used (Figs. 5A-B). For orientation, feature blurs of .3 and 1.2 deg were used (Figs. 5C-D). The low frequency and the horizontal orientation were always in the left portion of each patch. The lower patch was always reflected about a horizontal axis (see Fig. 5E) so that the subject would have no cause to have a bias for the upper edge patch location relative to the lower one. Without this reflection, the rotating texture of the lower patch visual "leads into" the texture of the upper patch, causing a strong bias toward the right for the upper patch with a magnitude which depends on other stimulus conditions. The orientation control used one 2048 × 512 pixel stimulus for each value of feature blur and each value of the constant frequency (3 or 6 cpd). Similarly, the contrast control used one such stimulus strip for each value of feature blur and constant orientation (vertical or horizontal). The main experiment used one stimulus strip for each value of the cue shift of the location of the edge defined by spatial frequency relative to that defined by orientation (±15, ±10, ±5 and 0 pixels, or ±.225, ± .15, ±.075 and 0 deg).

Procedure. Staircase procedures were used. Each block of trials consisted of a number of interleaved staircases which controlled the shift of one patch relative to another. Each staircase began with a step size of 16 pixels which dropped after 2 reversal pairs to a step size of 4 pixels.

In the orientation control experiments one block was run. The block consisted of eight interleaved staircases: two values of feature blur, two fixed spatial frequencies, and two staircases (converging on probabilities .707 and .293). There were 50 trials per staircase. On a given trial, the current staircase determined from which stimulus strip the patches were chosen (based on spatial frequency and feature blur), and the staircase value determined how far the upper edge was shifted relative to the lower edge. These shifts were performed symmetrically (e.g. upper edge rightward and lower edge leftward from a centered position) with an additional randomly chosen shift applied to both edges (as much as 12 pixels left or right). Averaging over the two spatial frequencies, this resulted in 200 trials per feature blur condition. The spatial frequency controls were similar with the roles of spatial frequency and orientation swapped. Initially a number of trial blocks were run with feedback in both control experiments to familiarize the subject with the materials and to eliminate any biases the subject had. Then, feedback was withheld and one block of each control experiment was run. Only the results from the no-feedback blocks are discussed below. All staircase trials were used in the subsequent analysis.

In the main experiment, three cue reliability conditions were tested: $\sigma_{orientation}$ and $\sigma_{frequency}$ combinations of .3 deg and .3 deg, 1.2 deg and .3 deg, and .3 deg and .6 deg. For each cue reliability condition, 3 blocks of 700 trials were run. Each block consisted of 14 interleaved staircases: 7 values of the shift of the spatial frequency cue relative to the orientation cue in the inconsistent cues stimulus, and two staircase criteria (of .707 and .293), with 50 trials per staircase. Averaged over the different staircase types, this resulted in 300 trials for each cue shift and reliability condition. On each trial the inconsistent cues stimulus was either the upper or lower edge (chosen randomly). The staircase determined where the consistent cues edge was located relative to the inconsistent cues edge. The bottom edge was always reflected about a horizontal axis (as in Fig. 5E). The pair of edge patches was randomly shifted right or left as much as 12 pixels.

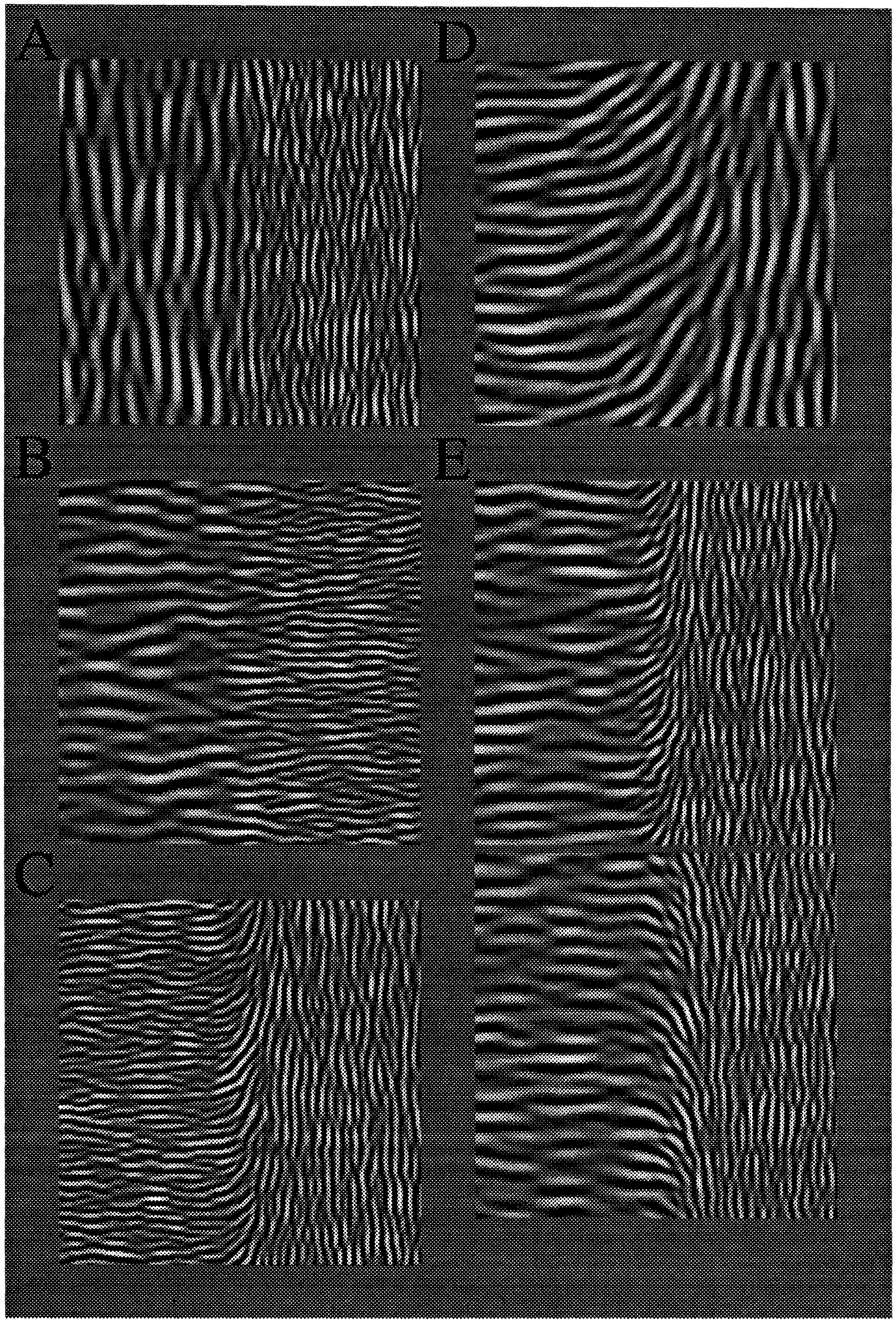

Fig. 5. Example stimuli from Expt. 2. (A) Frequency cue alone, $\sigma_{frequency}$ = .3 deg. (B) $\sigma_{frequency}$ = .6 deg. (C) Orientation cue alone, $\sigma_{orientation}$ = .3 deg. (D) $\sigma_{orientation}$ = 1.2 deg. (E) A typical stimulus display. The upper texture edge is an inconsistent cues edge ($\Delta l_{frequency}$ = -.225 deg). The lower edge is a consistent cues edge and its location has been shifted rightward .24 deg. $\sigma_{orientation}$ = $\sigma_{frequency}$ = .3 deg.

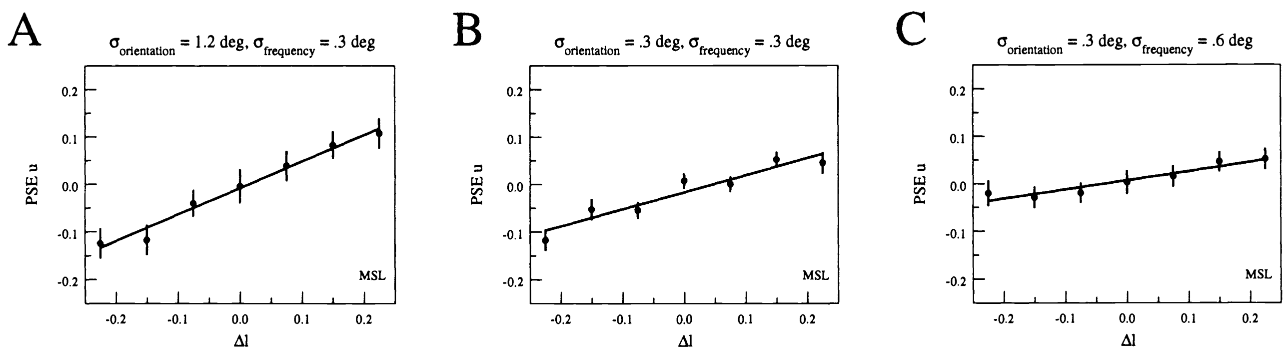

Fig. 6. Data for Expt. 2. Each graph shows PSE as a function of cue shift as in Fig. 4D for the three reliability conditions. Error bars indicate two standard deviations of the estimated PSE. Again, slopes change in a manner consistent with relative cue reliability (the three estimated values of slope are .191, .359 and .558).

3.2 Results

The data from the control and main experiments were each fitted separately by cumulative normal distributions using a maximum likelihood estimator. There was no evidence that the subject had any difficulty detecting the edges, and only a 2 parameter fit was required (mean and variance). In the single cue control experiments, an increase in feature blur resulted in an increase in variance (i.e. a decrease in slope) of the psychometric functions. For the main experiments, the values of the mean from these fits were taken as the PSE. Figure 6 shows the PSEs plotted as a function of the cue shift in the three reliability conditions. Several points may be noted from these data. First, there is no longer any problem in detecting the edges in either the main experiments or control experiments (the psychometric functions asymptote at 0% and 100% within the range of values tested; these raw data are not shown here). Second, in all three main experiments linearity obtains. In the $\sigma_{orientation}$ = .3 deg and $\sigma_{frequency}$ = .3 deg condition linearity is barely accepted (p = .1 with the Bonferroni correction), although it is not clear whether the possible nonlinearity in the data is meaningful (Fig. 6B). Again, the slopes of the three lines change with the relative reliability of the two cues. This is all consistent with a weighted average of location with weights determined by relative cue reliability. In addition, the psychometric function slopes in the two-cue conditions are steeper than those in single-cue trials, suggestive of the increase in reliability predicted when two independent random variables are averaged. We have had moderate success in fitting these data (both main experiments and controls) with a model of optimal cue combination where weights are chosen inversely proportional to cue variabilities. This modeling will be reported elsewhere.

Finally, we are compelled to point out one difficulty with this set of experiments. Across a block of trials in the main experiment, 700 trials were shown. In each trial the subject fixated about midway between the two stimulus patches for the 200 msec stimulus flash. Thus, every 3 sec or so the subject viewed a 200 msec flash of mostly 3 cpd horizontal noise to the left of fixation, mostly 6 cpd vertical noise to the right of fixation, and a texture edge located close to fixation. It turns out that this is a potent stimulus and timing for adaptation to occur. In fact, although the subject only views 700 flashes of 200 msec (for a total exposure of 2.33 minutes) over a 40 minute period, by about halfway through the sessions the nominally 100% contrast stimuli become difficult to see. The edges don't become impossible to detect, and the staircases do not wander off aimlessly. However, the task does become more difficult, and this may account for the noisiness in the data of Fig. 6B.

4. DISCUSSION

We have outlined a theory for the combination of multiple visual cues and seen its applicability to texture localization judgments. How does this theory add to existing theories of cue combination? In contrast detection experiments there is a history of discussion of multiple cues. In a high threshold theory of detection, the cue combination is often referred to as *probability summation*. Thus, if cue A is detected with probability p_A and cue B is detected with probability p_B, then a stimulus containing both cues will be detected whenever either cue alone is detected with probability $1-(1-p_A)(1-p_B)$ (assuming independence). Vector length models of detection also predict this improvement in detectability without the untenable high threshold assumption.[13,15,] In either analysis, a two cue stimulus will be more detectable than a single cue stimulus. Rivest[16] extended this logic to texture edge localization with two cues, and found that localizability was improved when some cues were added to a stimulus (the increase in psychometric function slope that was found here as well). What is new here is an analysis of what occurs when the two cues do not necessarily concur as to location, and the dependence of that cue combination on the relative reliability of individual cues.

An alternative explanation of these results might seek to explain them by reference to conscious strategies on the part of subjects. Consider a subject who, when presented with a 2-cue texture edge stimulus, simply attended to a single cue, ignoring the other. Obviously, if the subject were capable of doing this with no loss due to the other cue, then the 2-cue experiment results would be identical to the 1-cue control experiments (for the attended cue). This clearly isn't the case. But, consider an alternative strategy where the subject attends to only a single cue in each trial, but shifts from one cue to the other in alternate trials. This strategy would not produce the data shown here. First, in the consistent cues trials (the middle psychometric function in Fig. 4C) this would predict a psychometric function whose slope was no greater than the slope in either control condition. Second, this would predict a gradual decrease in slope as the cues become increasingly inconsistent (as these conditions would involve averaging two displaced single cue curves). Neither of these predictions accord with our results. Thus, we reject these conscious subject strategies as an explanation for our results.

As presented so far, the theory for cue combination in localization is a one-dimensional theory. This is clearly inadequate for describing visual performance which involves 2-dimensional displays. Pavel[14] has been investigating the cue combination problem. He suggests that similar considerations may be brought to bear on edge location and orientation judgments, thus generalizing these results to a true 2-dimensional edge identification. He also suggests that combination of cues involves combination across multiple spatial scales.

Finally, one might object to our analysis as being logically circular. How can a subject improve localization reliability by changing the weights on an average over multiple cues, when those weights must in turn be based on an estimate of an individual cue's reliability? We have suggested that this determination may be made by using *ancillary cues*, that is, cues which do not in and of themselves aid in the estimation of location, but are relevant to cue reliability.[10,12] For example, in the current task, for a given cue the subject may make estimates of edge location using receptive fields located at various points along the displayed vertical edge. If these individual estimates vary (for our vertical edges) or are highly variable and inconsistent with any smooth edge interpretation (in the general case), this may be used as an indication of low reliability. Thus, it should in principle be possible for an observer to use multiple cues and to simultaneously estimate their relative reliabilities in any given stimulus.

5. ACKNOWLEDGMENTS

The work described here was supported by NIH grant EY08266 and an NRC-NASA Ames Research Associateship. We would like to acknowledge the helpful comments and technical assistance of Jeff Mulligan, Misha Pavel, Mark Young, Carlo Tiana and, of course, Larry Maloney.

6. REFERENCES

1. J.J. Clark and A.L. Yuille, *Data Fusion for Sensory Information Processing Systems*, Kluwer, Boston, 1990.

2. F.R. Hampel, ''The influence curve and its role in robust estimation,'' *Jour. Amer. Stat. Assoc.*, Vol. 69, pp. 383-393, 1974.

3. P.J. Huber, P. J. *Robust Statistics*, Wiley, New York, 1981.

4. E.B. Johnston, B.G. Cumming and M.S. Landy, ''Integration of stereopsis and motion shape cues,'' *Mathematical Studies in Perception and Cognition*, Technical Report 92-1, New York University, New York, 1992.

5. E.B. Johnston, M.S. Landy, B.G. Cumming and L.T. Maloney, ''Integration of stereo and motion shape cues,'' *Invest. Ophth. Vis. Sci.*, Vol. 32, pg. 1180, 1991.

6. L. Kaufman, *Sight and Mind*, Oxford University Press, New York, 1974.

7. M.S. Landy and J.R. Bergen, ''Texture segregation and orientation gradient,'' *Vision Research*, Vol. 31, pp. 679-691, 1991.

8. M.S. Landy, Y. Cohen and G. Sperling, ''HIPS: A Unix-based image processing system,'' *Computer Vision, Graphics, and Image Processing*, Vol. 25, pp. 331-347, 1984.

9. M.S. Landy and L.T. Maloney, ''A statistical framework for combination of consonant depth cues,'' *Invest. Ophth. Vis. Sci.*, Vol. 31, pg. 173, 1990.

10. M.S. Landy, L.T. Maloney, E.B. Johnston and M.J. Young, ''Measurement and modeling of depth cue combination: In defense of weak fusion,'' *Mathematical Studies in Perception and Cognition*, Technical Report 91-3, New York University, New York, 1991.

11. M.S. Landy, L.T. Maloney and M.J. Young, ''Psychophysical estimation of the human depth combination rule.'' In P.S. Schenker (Ed.), *Sensor Fusion III: 3-D Perception and Recognition*, *Proc. SPIE*, Vol. 1383, pp. 247-254, 1991.

12. L.T. Maloney and M.S. Landy, ''A statistical framework for robust fusion of depth information.'' In W.A. Pearlman (Ed.), *Visual Communications and Image Processing IV*, *Proc. SPIE*, Vol. 1199, pp. 1154-1163, 1989.

13. J. Nachmias, ''A template matching model of subthreshold summation.'' In M.S. Landy and J.A. Movshon (Eds.), *Computational Models of Visual Processing*, pp. 135-145, MIT Press, Cambridge, MA, 1991.

14. M. Pavel, ''Information fusion over apertures and spatial frequencies,'' *OSA Annual Meeting Technical Digest*, Vol. 23, pg. 130, Optical Society of America, Washington, DC, 1992.

15. D. Pelli, ''Uncertainty explains many aspects of visual contrast detection and discrimination,'' *Journal of the Optical Society of America A*, Vol. 2, pp. 1508-1532, 1985.

16. J. Rivest, ''Multiple measurements in localization of contours.'' Unpublished doctoral dissertation, Harvard University, Cambridge, MA, 1992.

17. M.J. Young, M.S. Landy and L.T. Maloney, ''Depth from texture and motion,'' *Invest. Ophth. Vis. Sci.*, Vol. 32, pg. 1180, 1991.

18. M.J. Young, M.S. Landy and L.T. Maloney, ''Perturbation analysis of depth perception from combination of cues: texture and motion,'' *Mathematical Studies in Perception and Cognition*, Technical Report 92-2, New York University, New York, 1992. Accepted for publication in *Vision Research*.