Relationship between binocular disparity and motion parallax in surface detection

JESSICA TURNER and MYRON L. BRAUNSTEIN
University of California, Irvine, California

and

GEORGE J. ANDERSEN
University of California, Riverside, California

The ability to detect surfaces was studied in a multiple-cue condition in which binocular disparity and motion parallax could specify independent depth configurations. On trials on which binocular disparity and motion parallax were presented together, either binocular disparity or motion parallax could indicate a surface in one of two intervals; in the other interval, both sources indicated a volume of random points. Surface detection when the two sources of information were present and compatible was not better than detection in baseline conditions, in which only one source of information was present. When binocular disparity and motion specified incompatible depths, observers' ability to detect a surface was severely impaired if motion indicated a surface but binocular disparity did not. Performance was not as severely degraded when binocular disparity indicated a surface and motion did not. This dominance of binocular disparity persisted in the presence of foreknowledge about which source of information would be relevant.

The ability to detect three-dimensional (3-D) surfaces has been studied previously using static stereo displays (e.g., Uttal, 1985), motion parallax displays (Andersen, 1996; Andersen & Wuestefeld, 1993), and structure-from-motion (SFM) displays (Turner, Braunstein, & Andersen, 1995).[1] This research has examined the signal-to-noise ratio required to detect a surface in the presence of added "noise" points in stereo displays of smooth, quadratic surfaces (Uttal, 1985) and in motion parallax displays of sinusoidal gratings (Andersen & Wuestefeld, 1993). Other studies have considered the minimum number of points required to detect a smooth surface in motion parallax displays (Andersen, 1996) as a function of frequency and amplitude for sinusoidally corrugated surfaces and in SFM displays (Turner et al., 1995) as a function of frequency and amplitude or of shape index and curvedness (see Koenderink, 1990). This research has documented the ability of human observers to detect a variety of surfaces with a small number of feature points or with a small ratio of feature points to noise points, from either binocular disparity alone or motion alone. The interaction of binocular disparity and motion in surface detection, however, has not been previously investigated.

Although the interaction of binocular disparity and motion has not been investigated in detection tasks, the inter-

action of depth cues in determining the perceived shape of objects has been a subject of considerable attention. Bülthoff and Mallot (1988) discussed four ways in which depth information from different cues may be combined: *accumulation*, *veto*, *disambiguation*, and *cooperation*. In accumulation, the depths computed from the various cues are combined additively to provide an overall perceived depth. Veto implies that the depth computed from one cue determines the perceived depth, overriding the depths computed from other cues. Disambiguation refers to the use of information from one cue to disambiguate the depth information provided by another cue. In cooperation, the effectiveness of one cue can be enhanced by information from another cue, resulting in a combined depth that is greater than the sum of the depths provided by each cue in isolation. Accumulation and veto are consistent with a model of cue combination in which depths are computed separately by each cue—that is, a strict modularity or weak fusion model. Disambiguation and cooperation, on the other hand, involve interactions of the cues prior to the final depth computation and are not consistent with weak fusion. Disambiguation and cooperation can occur in the absence of modularity—that is, with strong fusion—but they can also occur with modules that interact prior to computing a final depth map but that still compute separate depth maps. Landy, Maloney, Johnston, and Young (1995) refer to this as *modified weak fusion*.

All four types of interactions discussed by Bülthoff and Mallot (1988) have been found in research concerned with recovering 3-D shape from combinations of binocular disparity and motion information. Johnston, Cumming, and Landy (1994) found that when binocular disparity and motion parallax indicated different curvatures, the per-

ceived shape was a linear combination of the curvatures simulated, with the weights constrained to sum to one. Norman and Todd (1995) found that one source of information would veto another when binocular disparity and motion indicated conflicting shapes, but the dominant source varied between subjects. Binocular disparity was the dominant source of information for determining surface depth when placed in conflict with motion parallax (Rogers & Collett, 1989). Motion information that was not compatible with the perceived shape was interpreted as relative motion within the surface. Tittle and Braunstein (1993) found a cooperative relationship between binocular disparity and SFM for transparent displays with high disparity. For some conditions, judged depth from combined binocular disparity and SFM was greater than the sum of the judged depths with either cue alone. Landy et al. (1995) interpreted this result in terms of changing the weights within a linear-combination approach. The role of binocular disparity in disambiguating SFM displays has been the subject of both theoretical analyses (Richards, 1985) and empirical investigations (Braunstein, Andersen, Rouse, & Tittle, 1986; Rouse, Tittle, & Braunstein, 1989).

In examining the relationship between binocular disparity and motion parallax in the detection of smooth surfaces, we will consider three combinations of information from these cues. Each of these combinations will be compared with detection with each cue presented in isolation. First, both cues can indicate the presence of a surface. In this case, the question of interest is how detection from the combined information compares with detection from each cue separately. The possible outcomes are that (1) detection is enhanced over that found with either cue alone, (2) detection is no better than that found with either cue separately (assuming that the effectiveness of the separate cues is matched), or (3) detection is inferior to that found with the separate cues. The first two outcomes could occur with any degree of modularity and would depend on the combination rules, with the second outcome consistent with a "winner-take-all" or veto rule. The third outcome, although unlikely, would suggest a form of strong fusion involving interference between the cues.

For the other two combinations of cues, only one cue indicates the presence of a surface. In the present experiments, the other cue indicated that the points were randomly distributed in a 3-D volume. It is unlikely that detection in these two cases would be enhanced over the single-cue case or over the combined-cue case. The more likely outcomes are that detection for each cue will be the same as in the single-cue case, it will be the same for one cue and inferior for the other cue, or it will be inferior for both. If detection is inferior for at least one cue, compared with performance in the single-cue case, it might be just slightly inferior or it could drop to chance levels. If detection is the same for one cue and drops to chance levels for the other cue, a veto or winner-take-all process is indicated. There is a fourth combination of cues, in which both indicate that the points are randomly

scattered in a volume. This combination provided the "noise" condition for our detection experiments.

In the three experiments presented here, the subject's task was to determine which of two sequentially presented displays contained information indicating the presence of a smooth surface. In the first experiment, detection was examined for binocular disparity and motion parallax separately and for the two sources of information in combination. The results of this experiment indicated a veto of motion information by binocular disparity when the two sources were present and only one indicated a smooth surface. The second and third experiments examined two methods of improving surface detection when motion carried the surface information: providing a pretrial cue indicating whether disparity or motion would carry the surface information on each trial, or presenting the trials in blocks and informing the subject of the relevant cue for each block.

EXPERIMENT 1

In Experiment 1, we examined the interaction of binocular disparity and motion information in determining the accuracy of surface detection both when the two cues provided disparate information about the presence of a smooth surface and when the two cues provided corresponding information. In a strict weak fusion model of cue interaction, the ability to register depth from one cue is independent of the information provided by other cues. If the two cues create independent depth maps, and the decision process can work independently on the two depth maps, then performance when the two cues agree is predicted by a probability-summation model. If the probability of detecting the surface using motion parallax information only is $P(M)$, and the probability of detecting the surface using disparity information only is $P(S)$, and the two are independent, then the probability of detecting the surface when both sources indicate the surface should be $P(M \& S) = P(M) + P(S) - P(M) * P(S)$. To determine whether probability summation occurs for surface detection from combined disparity and motion cues, we included conditions in which only disparity information was present (the points did not move) or only motion parallax information was present (the displays were viewed monocularly). These conditions provided estimates of $P(S)$ and $P(M)$. Probability-summation predictions based on these estimates were compared with the results in conditions in which both cues were present and provided corresponding information.

Method

Apparatus. All displays were shown at 40 frames per second using a point-plotting system with 4,096 × 4,096 resolution consisting of a 21-in. CRT monitor (Xytron AB2) with a P4 phosphor controlled by a VaxStation II. The actual display size was a square with 3,600 pixels (15.43 cm) on a side. The subjects viewed the display from a distance of 97.8 cm through a mirror stereoscope with a black cardboard septum extending from the viewer to the monitor. To collect the baseline data for motion parallax blocks,

one eyehole of the stereoscope was covered so that the corresponding half of the display was visible only to one eye.

Responses were collected using a two-pushbutton box attached to the VaxStation. The box was oriented so that one pushbutton was on the left and one was on the right as viewed by the seated observer. All trials were run in a completely dark room. The same apparatus was used in all the experiments.

Subjects. The subjects were 1 of the authors (J.T.) and 3 graduate students who were paid for their time. The graduate students were familiar with visual psychophysical experiments but were naive to the purposes of the present experiments until after the completion of Experiment 1. Each subject had monocular vision corrected to 20/40 (Snellen eye chart) and normal binocular vision (able to detect a disparity difference of 40 sec on the Stereo Optical Circles Test). The same subjects participated in all the experiments, which are presented here in chronological order.

Stimuli. The surface that the subjects were asked to detect is shown in Figure 1. The compound sinusoid was determined according to the following equation:

$$z = 0.354\, M \cos[1.7\pi(y + \phi)/M]$$
$$+ 0.532\, M \cos[2.3\pi(y + \phi)/M + \pi/4], \qquad (1)$$

in which ϕ was a random phase constant that could vary between plus and minus one quarter of the display height, and M was half the vertical extent of the display. The y value could vary between positive and negative M. A compound sinusoidal shape combined with the random phase offset was chosen to minimize any effects of periodicity on surface detection. The differences between maximum and minimum simulated depth was approximately equal to the height of the display. The random phase on each trial had the effect of repositioning the portion of the surface that was seen through the viewing window. In this way, the depth minima and maxima were not always at the same vertical location. The total area taken up by the surface was a square 9º on a side.

A surface configuration was created in the following way: Horizontal and vertical (x and y) locations for points were selected quasi-randomly, in that the same number of points were constrained to fall within each cell of a virtual 3×3 grid centered over the available screen locations. When there were fewer than nine points, the points were spread randomly in x and y. The method of back-projection to the surface under polar perspective was as follows: The depth (z) value of each point was computed according to Equation 1. Given this z value and the viewing distance, the y value in the image was adjusted to produce the correct polar projection. (The x values did not need to be changed since depth did not vary in the horizontal dimension. The points could thus be placed on the surface, while maintaining a uniform distribution in the im-

age, without altering the x coordinates. The resulting display did not simulate a uniform distribution on the 3-D surface.) To create a matched nonsurface configuration, the z values of the surface configuration were permuted randomly. The resulting configuration would have the same x and y values as the original surface configuration, but the depth values would no longer be in accordance with the original surface.

The velocities of the points depended on simulated depth according to the equation $dx = E * 10.0/(E + z)$, where dx is the projected displacement of a point in x in one frame transition and E is the eyepoint distance of the observer, both measured in pixels. The maximum and minimum possible velocities were 1.40º and 1.03º per second. The right and left stereo views of the points were calculated by rotating the configuration of points by a positive and a negative angle, respectively, which was adjusted according to the eye separation of the subject. Both the displacements and the disparities were constant across views. The motion parallax and binocular disparity information could specify different configurations since the depths used in calculating the velocities and those used in calculating the disparities were independent.[2]

In the *baseline* conditions, the depths in the displays were defined by either static disparity or monocular motion parallax. In a static stereo display, points had zero velocity. In a monocular motion parallax display, points were viewed monocularly and so had zero disparity. In the *combined* conditions, the depths in the displays were defined by both binocular disparity and motion parallax.

In a combined condition trial in which motion parallax carried the surface information, velocities of the points in a surface display were based on the surface; in a noise display, they were randomly permuted, representing points in a volume. If the disparity information was to be compatible with the motion parallax, then the same configurations of point depths were used to calculate the stereo views in the surface display as were used to calculate the point velocities in that display. If disparity was to be incompatible with motion parallax, then, in each display, the stereo views were calculated from different nonsurface configurations, in which the point depths were chosen randomly from the depth range of the surface. We will refer to this condition as *motion with disparity noise*.

Similarly, if binocular disparity carried the surface information, then, in one display, the disparities were based on surface depths, whereas, in the other display, they were not. If motion carried incompatible information, the depth configurations used to calculate the points' velocities were not based on a surface but were randomly assigned within the same range as the surface depth values. We will refer to this condition as *disparity with motion noise*. Thus, the 3-D configuration indicated by motion and the 3-D configuration specified by binocular disparity differed from one display to another in a trial, regardless of which source of information specified the surface. In one of the displays, either one cue or both cues specified a surface; in the other display, neither cue specified a surface.

For the noise (distractor) displays, both disparity and motion indicated that the points were randomly positioned in a volume. The depths were independently selected for the two cues. This was done so that a noise display could not be distinguished from a display in which one cue indicated a surface on the basis of the compatibility of the motion and stereo depths of the points.

Two rows of three dots were shown 2.25º above and below the display to aid fusion. The dots were separated horizontally by 0.24º and were shown in stereo at zero disparity and remained on throughout the trial. Each trial consisted of a 700-msec fusion target, a 3-sec (120-frame) display, an 800-msec interstimulus interval (ISI; with the fusion target still visible), and a second 3-sec display. A new trial began 3.5 sec after the subject's response to the previous trial. The subjects had 60 sec from the end of the second display to respond, and they had no difficulty responding within that time.

Design. For the baseline trials, the independent variables were type of display (static binocular disparity or monocular motion par-
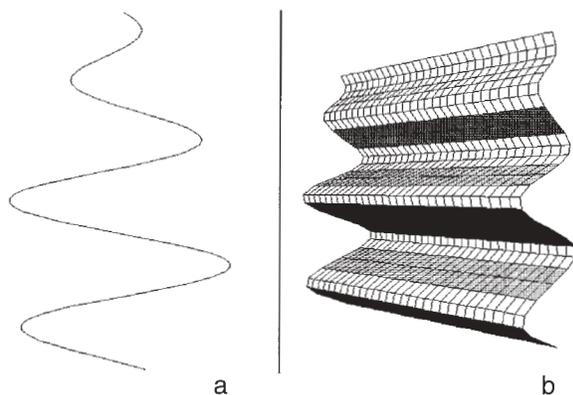

Figure 1. (a) A side view of the maximum extent of the surface. (b) A three-quarters view of the surface.

allax) and number of points (6, 8, 10, 12, 14, or 16). (Number of points was studied in previous research of the detection of smooth surfaces from motion parallax [Andersen, 1996] and SFM [Turner et al., 1995].) For the combined-cue trials, the independent variables were the relevant cue (binocular disparity, motion, or both) and number of points (same as for baseline trials).

The baseline and combined conditions were blocked and run in an ABBA design: Half of the baseline trials were run first, followed by the first half of the combined trials, the second half of the combined trials, and the remainder of the baseline trials. The baseline trials were blocked by number of points and type of display. Each block consisted of 25 trials, including 5 practice trials at the beginning. The combined trials were blocked only by number of points. The relevant cue conditions—disparity, motion, or both—were randomly ordered within each block. Each block consisted of 35 trials, including 5 practice trials at the beginning. There was a total of 40 trials in each of the 12 baseline and 18 combined conditions.

Procedure. The subjects were shown the diagrams in Figure 1 while the task was explained to them. They were told that, on each trial, they would see two displays of points, one after the other. In one display, the points would be on the surface. After seeing both displays, they were to push the left button if the surface was in the first display and the right button if the surface was in the second display. Feedback (one beep for a correct response and two for an incorrect response) was provided in the practice trials and throughout the experiment. The subjects were also told that there should be three fused points above and below the display and that, if they were unable to fuse the points, they should inform the experimenter immediately via the intercom system. After the instructions were read and the subjects indicated that they understood the procedure, they dark-adapted for 2 min before beginning the experiment. After any break in which they left the room, they dark-adapted for 2 min before beginning again. This procedure was used throughout all the experiments.

Four practice blocks were run at the beginning of the experiment. The first two blocks consisted of 25 static stereo and 25 motion parallax displays, with 50 points in each display. The second two blocks consisted of the same types of displays with 25 points. Two additional 50-point blocks were run at the end of the experiment. The subjects were required to meet a criterion of 90% correct in the 50-point blocks both at the beginning and at the end of the experiment. All subjects met this criterion.

Within the baseline condition, the order of the numerosity levels was randomly selected for 2 subjects, with the reverse of these random orders used for the other 2 subjects. The subjects saw a block of motion parallax trials and a block of disparity trials at each numerosity level. Within the combined condition, the order of the blocks was randomly selected for 2 subjects, with the constraint that one block at each numerosity level was shown before the next block at any numerosity level could be shown. The reverse of these random orders was used for the other 2 subjects. The subjects completed between three and six blocks of trials at every session and completed all sessions over the course of 1–3 weeks.

Results

The mean performance for the 4 subjects for the baseline conditions is shown in Figure 2. The percents correct for baseline binocular disparity and motion parallax for each subject were analyzed using a 2 (information source) $\times$ 6 (numerosity) within-subject analysis of variance (ANOVA). The difference between performance on the motion and disparity trials was not significant [$F(1,3) = 6.04$, $p > .05$]. (The mean percents correct for motion parallax and disparity were 76.97 and 74.79, respectively.) Performance improved significantly with increas-
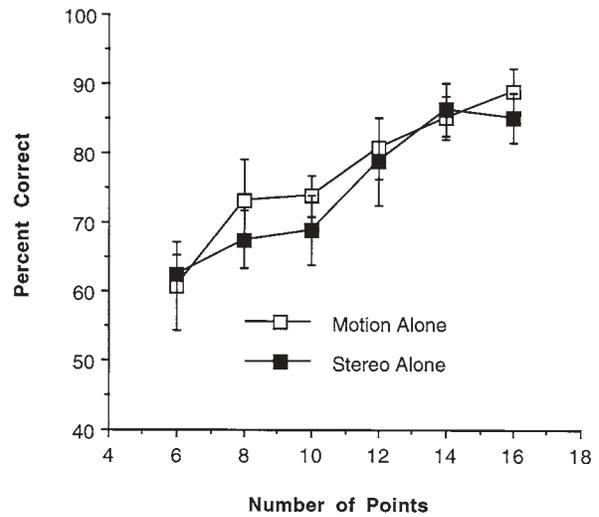

Figure 2. Mean performance of 4 observers in the baseline motion parallax and stereo conditions in Experiment 1, as a function of numerosity. Error bars denote ±1 *SE*.

ing numerosity [$F(5,15) = 36.44$, $p < .05$, $\omega^2 = .54$]. Post hoc tests (Tukey's HSD) showed that the mean percent correct at 6 points was significantly lower than at all other levels, and the means for 8 and 10 points were significantly lower than the means for 12, 14, and 16 points.

The mean performance in the combined conditions is shown in Figure 3. The percents correct for each subject were analyzed in a 3 (information source) $\times$ 6 (numerosity) within-subject ANOVA. The effect of information source was significant [$F(2,6) = 84.92$, $p < .05$, $\omega^2 = .45$], with all differences between conditions significant. The effect of numerosity was significant [$F(5,15) = 7.70$,
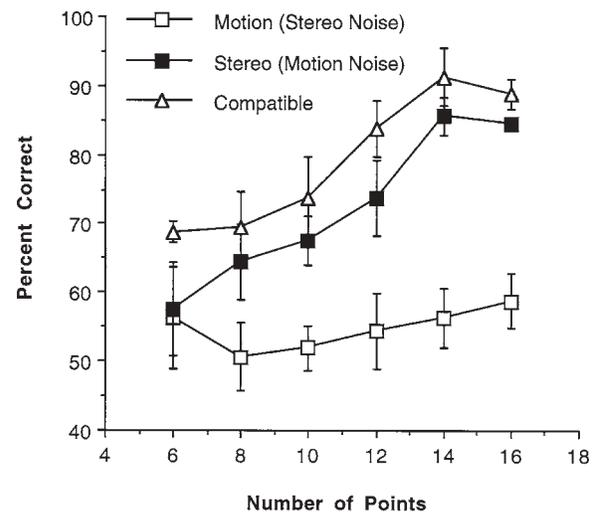

Figure 3. Mean performance of 4 observers in the motion-with-disparity-noise, disparity-with-motion-noise, and compatible conditions in Experiment 1, as a function of numerosity. Error bars denote ±1 *SE*.

$p < .05$, $\omega^2 = .18$], with performance at 14 and 16 points significantly better than performance at 6, 8, or 10 points. The interaction was also significant [$F(10,30) = 2.33$, $p < .05$, $\omega^2 = .04$].

Discussion

The baseline performance results show similar performance in detecting the surface for motion and disparity. For the number of trials in each condition, 66% correct is required to exceed chance with $p < .05$. In the motion blocks, the percent correct exceeded chance with 6 points for 1 subject, 8 points for 2 others, and 10 points for 1 subject. In the binocular disparity blocks, the percent correct exceeded chance at 6 points for 1 subject, 8 points for 2 others, and 12 points for 1 subject. In the combined condition in which the two sources agreed, performance did not improve beyond the baseline levels, as can be seen by comparing the two curves in Figure 2 with the top curve in Figure 3. Percent correct predicted from probability summation was higher than obtained performance for all subjects at all levels of numerosity. Figure 4 compares predicted and obtained performance averaged across subjects for the six levels of numerosity. These results fail to support the idea that the two sources of information drive two separate and independent processes of surface perception that have outputs combined through probability summation.

Performance on trials on which motion parallax carried the surface information and binocular disparity indicated points in a volume was much lower than on trials in which binocular disparity carried the surface information and motion indicated points in a volume. Performance when binocular disparity indicated a surface and motion indicated points in a volume was significantly worse than performance when the two sources were compatible.
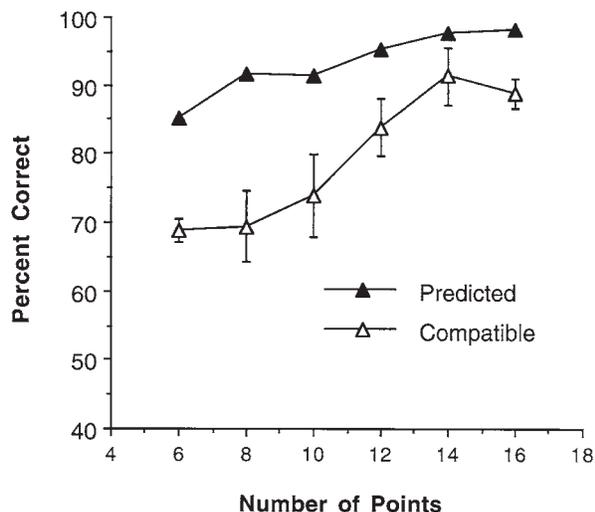

Figure 4. Mean performance of 4 observers in the compatible condition in Experiment 1, and the predicted performance from a probability-summation model using the mean baseline performances. Error bars denote ± 1 _SE_.

Thus, while the presence of inconsistent depth information from motion can degrade the ability to detect a surface using binocular disparity information, the presence of inconsistent binocular disparity information more seriously degrades the ability to detect a surface using motion information.

Although the presence of incompatible motion and binocular disparity information was quite apparent to all of the subjects, the binocular disparity information was weighted more heavily, approaching a veto relationship with respect to the motion information. Johnston et al. (1994) suggest that the weights placed on different sources in a linear combination are affected by which source is more reliable. A source of depth information providing depth values that are highly discrepant when compared with depth values from other sources would be considered unreliable. This might occur in natural settings when the viewing conditions for one source of depth information are degraded relative to other sources, as when viewing strongly shaded objects at a distance outside the range of effective binocular disparity. In Experiment 1, when motion and binocular disparity were not compatible, the cue that was informative about the presence of a smooth surface might be considered the more reliable cue. The subjects appeared to have used the binocular disparity cue almost exclusively, however, regardless of which cue was informative. In Experiments 2 and 3, we manipulated the information given to subjects about which source could reliably be used to distinguish between surface and volume displays. If the dominance of binocular disparity is at least partially under the subjects' control, performance on the motion with disparity noise trials should improve when subjects are informed that motion is the relevant source and disparity is noise. If the presence of highly incompatible binocular disparity information precludes the use of motion parallax information, informing the subjects which source is relevant should not improve performance.

EXPERIMENT 2

In Experiment 1, 3 of the 4 subjects were not informed beforehand that the sources of information would conflict, though they all reported rapidly becoming aware of the conflicting information. None of the subjects knew from trial to trial which source of information could be used to distinguish the surface and volume displays, when the two sources were incompatible. During the debriefing after Experiment 1, all subjects were explicitly made aware that the binocular disparity and motion parallax could conflict and that, in some cases, one source would indicate the surface display while the other cue indicated a volume. In Experiment 2, a pretrial cue informed the subject whether motion parallax or binocular disparity would carry the surface information for that trial. If the subject is able to control the weights given to binocular disparity and motion information (e.g., by attending to one or the other type of information), performance should be affected by this pretrial cue. If the weights are not af-

fected by such pretrial cues, the results should be similar to those in Experiment 1.

Method

Stimuli. The stimuli were the same as in Experiment 1, except that a visual cue was presented 900 msec before each trial. In the informative-cue (*cued*) blocks, the cue indicated whether binocular disparity or motion would carry the surface information; in the remaining (*uncued*) blocks, the cue was uninformative (i.e., the same cue was presented on every trial and did not indicate either source of depth information).

In the cued blocks, the cue consisted of a horizontal line and a vertical line. The horizontal line was always centered in the display, and the vertical line could be either above or below the horizontal line by 0.3º. For 2 of the subjects, the vertical line above the horizontal line indicated that binocular disparity was the relevant source of information; for the other 2 subjects, the vertical line above indicated that motion was the relevant source of information. This was done to balance any unexpected effect of the choice of cue. In the uncued blocks, the subjects saw a horizontal line and a broken vertical line centered on the horizontal line. The six-dot fusion target was presented with the cue (see Figure 5).

The horizontal line was presented to the left or right eye randomly from trial to trial, and the vertical line was presented to the other eye. Binocular presentation of the cue was used to ensure that the subjects were viewing the displays binocularly. The subjects were instructed to inform the experimenter if, on any trial, the horizontal and vertical lines did not line up to form either a T shape (upright or inverted) or a cross (depending on the particular block of trials). No such reports were made, and all subjects verbally confirmed afterward that they had seen a T shape or cross on each trial.

No baseline displays were shown during this experiment. The displays were calculated in the same way as in Experiment 1, though the number of points was limited to 6, 10, or 16. When motion and disparity information was compatible on a cued trial, the motion or disparity cue was presented with 50% probability.

Design. There were three levels of numerosity, two cuing conditions (cued or uncued), and three sources of information (disparity with motion noise, motion with disparity noise, or compatible). There were 60 trials in each of the 18 conditions.

The trials were blocked by numerosity and cuing condition. Each block had 35 trials including 5 practice trials at the beginning. The overall design was again an ABBA design: Each subject saw half of the uncued blocks, then half of the cued blocks, followed by the second half of the cued blocks, and the remainder of the uncued blocks.

Procedure. The subjects were shown the same static representation of the surface as in Experiment 1, and the task and the cuing were explained to them. As in Experiment 1, a block of 25 monocular motion parallax trials and a block of 25 static disparity trials with 50 points in each display were shown first. Two subjects saw the disparity block first, and the other 2 subjects saw the motion block first. The same was done with 25-point displays. All subjects maintained 90% correct or better on the 50-point displays, and they repeated that performance on another two blocks of 50-point displays at the end of the experiment.

The blocks were ordered in sets of three, one block at each numerosity level. For 2 subjects, the order was determined randomly within each set of three; for the other 2 subjects, the reversed order was used.

The subjects completed between three and six blocks of trials at every session, and they completed all sessions over the course of 1–3 weeks. One subject was unavailable for 2 weeks right after finishing the cued blocks. Before collecting the data from his final uncued blocks, he responded to new 50-point and 20-point blocks and to one cued block of 35 trials at each level of numerosity in a random order. Data from these additional practice trials were not included in the final analysis.

Results

The mean percent correct for 4 subjects is shown in Figure 6 for each condition. The percents correct were analyzed in a 3 (information source) $\times$ 2 (cued or uncued)
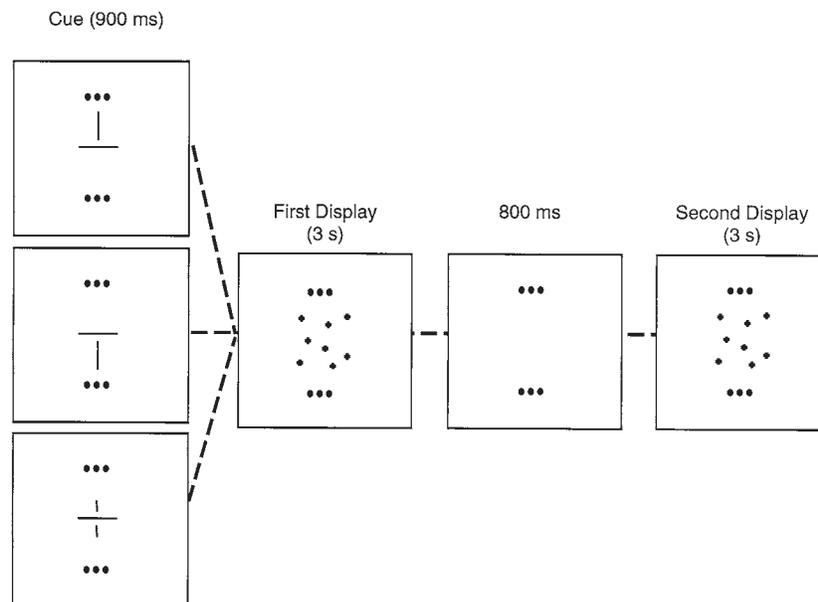
Cue (900 ms)

Figure 5. A trial in Experiment 2. Observers first saw one of the three possible cues, with the three fixation points above and below, then the first display, a blank interval, and the second display. All displays were presented binocularly.
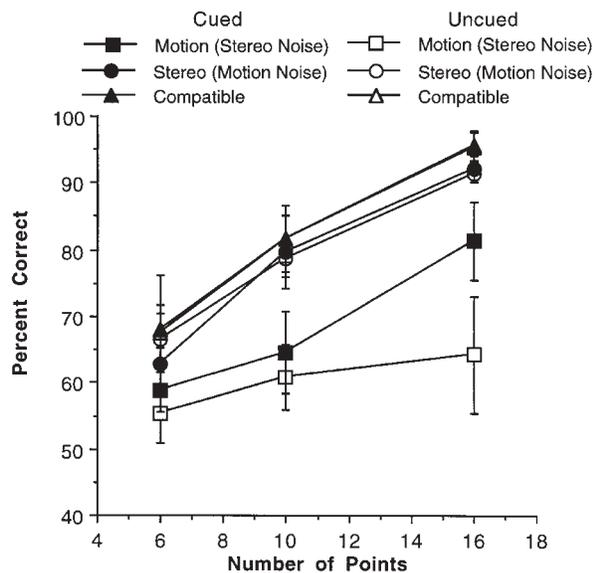
Figure 6. Mean performance of 4 observers in Experiment 2, as a function of numerosity. Error bars denote ±1 *SE*.

× 3 (numerosity level) within-subject ANOVA. Performance improved with numerosity [$F(2,6) = 67.37$, $p < .05$, $\omega^2 = .382$], with all differences between numerosity levels significant (Tukey's HSD, $p < .05$). Performance was affected by the information source [$F(2,6) = 20.24$, $p < .05$, $\omega^2 = .232$], with accuracy on the disparity-with-motion-noise trials and the compatible trials significantly better than accuracy on the motion-with-disparity-noise trials. The effect of cuing was significant [$F(1,3) = 12.48$, $p < .05$, $\omega^2 = .011$], as was the interaction between cuing and information source [$F(2,6) = 5.98$, $p < .05$, $\omega^2 = .010$]. Figure 6 shows that the difference between cued and uncued performance is attributable to the motion-with-disparity-noise conditions. No other interactions were significant.

Discussion

Knowledge of which source carries the surface information does not allow subjects to perform at the same level in the motion-with-disparity-noise condition as they do in the disparity-with-motion-noise condition. Although the performance on the motion-with-disparity-noise trials was better when cued, it was still worse than performance in the other conditions.[3]

EXPERIMENT 3

In Experiment 2, the subjects were informed on each trial which cue would be relevant, with the relevant cue varying from trial to trial. The effect of cuing was to improve performance on the motion trials, but performance was still markedly worse than performance even on the uncued disparity trials. It is possible that performance cannot be adjusted effectively according to information about the relevant cue on a trial-by-trial basis but could be adjusted over longer sequences of trials. In Experiment 3, only one source of information was relevant for an entire block of trials, allowing subjects to benefit as much as possible from cuing.

Prior to beginning Experiment 3, we collected a second set of baseline data to determine whether performance on motion-only and disparity-only trials remained comparable. The stimuli for these baseline trials were either motion parallax displays or static stereo displays. The motion parallax displays were presented stereoscopically but had no disparity. The displays were equivalent to the baseline displays of Experiment 1, except that the pretrial cuing symbols were retained from Experiment 2. Also, in the Experiment 1 baseline trials, viewing of the motion parallax display had been monocular. (The motion parallax trials were thus not completely comparable to those in Experiment 1.) Following the baseline trials, the subjects responded to one block of twenty-five 50-point displays and one block of twenty-five 20-point displays. Two subjects then were shown a set of disparity blocks—one block of disparity trials at each numerosity level in a random order—followed by two sets of blocks of motion trials, and the remaining set of disparity blocks. The other 2 subjects saw the reversed ordering. A total of 40 trials in each condition were shown. The subjects were informed before each block whether it would be a motion block or a disparity block. They also saw the appropriate pretrial cue before each trial.

Method

Stimuli. All the stimuli were produced in the same manner as the uncued displays in Experiment 2, with the exception that no trials were shown in which the two sources of information were compatible. On every trial, only one source of information indicated a surface.

Design. As in Experiment 2, there were three levels of numerosity and two cuing conditions (cued or uncued), but the source of information variable had only two levels (disparity with motion noise and motion with disparity noise). There were 60 trials in each of the 12 conditions. The trials were blocked by numerosity (6, 10, or 16 points) and cuing condition (*mixed*, *disparity*, or *motion*). In the mixed condition, either disparity or motion parallax could carry the surface information from one trial to the next, and the subjects were not informed which was relevant; in the motion and disparity conditions, the entire block of trials would be motion with disparity noise or disparity with motion noise, and the subjects were informed at the beginning of each block which it would be. The three types of blocks (mixed, motion, and disparity) were run in an ABCCBA order, with the constraint that the mixed blocks were always shown first and last. For half of the subjects, the order was mixed–motion–disparity–disparity–motion–mixed; for the other half, the order of the disparity and motion blocks was reversed. Each block had 35 trials in it including 5 practice trials.

Procedure. The subjects were shown the static representation of the surface and informed that the task would be the same as previously: They would see two displays and were to indicate which display had the surface in it. Before beginning the main part of the experiment, the subjects responded to two blocks of 50-point displays first—one of static binocular disparity displays and one of monocular motion parallax displays—and all performed above 90% correct on both blocks.

The subjects next responded to two blocks at each numerosity level in the mixed conditions. The orders of the blocks were generated randomly for the first 2 subjects, with the constraint that one block at each numerosity level was shown before the second block of any numerosity level could be shown. The order of blocks for the 3rd and 4th subjects was the reverse of the orders used for the 1st and 2nd subjects, respectively. Once the subjects had responded to the two sets of mixed blocks, 2 subjects responded to two sets of disparity blocks, then four sets of motion blocks, followed by the remainder of the disparity blocks, and, finally, the remainder of the mixed blocks. The other 2 subjects responded to two sets of motion blocks, then the four sets of disparity blocks, followed by the remainder of the motion blocks, and the mixed blocks. The entire experiment was run over the course of 2 weeks for each subject. As in Experiment 2, the subjects were required to inform the experimenter if they did not see a T-shaped cue before each trial. No such reports were made.

Results

The mean performance for disparity alone and motion alone was higher than the performance for the first baseline data collected in Experiment 1: For disparity alone, the percents correct, averaged across the 4 subjects, were 63%, 84%, and 95% for the 6-, 10-, and 16-point conditions, respectively (relative to 63%, 69%, and 85% in Experiment 1). For motion alone, the percents correct were 71%, 84%, and 90% for the three numerosity conditions (relative to 61%, 74%, and 89% in Experiment 1). Although performance improved overall relative to Experiment 1, performance levels for disparity alone and motion alone remained similar.

The results for the combined-cue trials, averaged over 4 subjects, are shown in Figure 7. A 2 (mixed or blocked) $\times$ 2 (source of information) $\times$ 3 (numerosity level) within-subject ANOVA again showed a significant effect of numerosity level [$F(2,6) = 59.97$, $p < .05$, $\omega^2 = .32$]. Post hoc tests (Tukey's HSD) showed all differences to be significant. Performance on blocked trials was significantly better than performance in mixed blocks [$F(1,3) = 14.45$, $p < .05$, $\omega^2 = .05$], and performance on disparity trials

Figure 7. Mean performance of 4 observers in Experiment 3 by condition, as a function of numerosity. Error bars denote ±1 *SE*.

was significantly better than on motion trials [$F(1,3) = 13.52$, $p < .05$, $\omega^2 = .21$]. None of the interactions were significant.

Discussion

Giving information about which source is relevant and maintaining that source over a series of trials improves performance, not only in the motion-with-disparity-noise conditions but also in the disparity-with-motion-noise condition. But, even with this improvement, performance when motion is the relevant source is still significantly worse than performance when disparity is the relevant source. The performance in the blocked motion-with-disparity-noise condition fails to match the performance even in the mixed (i.e., uncued) disparity-with-motion-noise condition.

GENERAL DISCUSSION

The results of these three experiments provide a consistent picture of the relationship between binocular disparity and motion information in surface detection. When the cues provide different information about whether or not a smooth surface is represented in a display, the disparity information completely dominates, or vetoes, the motion information. Specifically, when binocular disparity indicates a smooth surface and motion parallax indicates points randomly distributed in a 3-D volume, surface detection is as accurate or almost as accurate as when both sources indicate the presence of a surface. When motion parallax indicates the presence of a surface, and binocular disparity does not, performance in detecting the surface is markedly degraded. This occurs even though baseline performance is similar for binocular disparity alone and motion parallax alone.

Some improvement can be achieved in surface detection when motion indicates the presence of a surface and binocular disparity indicates random placement of points in a volume by informing the subject that motion is the relevant cue. There are several possible explanations for this improvement. One possibility is that subjects are able to attend more closely to the motion information when informed that motion is the relevant cue. Another possibility, consistent with the use of a linear-combination rule, is that subjects are able to adjust the weights given to the two cues in combining the depth information so that motion now has a nonzero weight. The present results do not distinguish between these possibilities, but they do indicate that the benefit achieved by informing the subject about which cue is relevant is limited: Sensitivity in the presence of conflicting binocular disparity information does not approach the levels achieved with motion alone. This is true whether information specifying the relevant cue is provided before each trial or for an entire block of trials.

When both cues indicate the presence of a surface, performance is at about the same level as with either cue alone. There is no evidence of probability summation. Probability summation has been found with binocular dis-
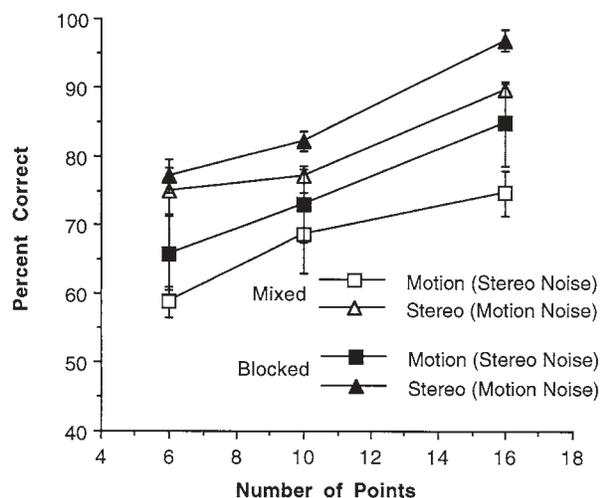
parity and motion in other tasks. Cornilleau-Pérès and Droulez (1993) found that a probability-summation model fit the increase in performance when these two cues provided compatible information in a curvature detection task. The lack of probability summation in surface detection is consistent with complete dominance of one cue over the other. Although it is not possible to determine which cue is dominant when the cues provide compatible information, the dominance of binocular disparity when the cues provided conflicting information strongly suggests that the lack of probability summation is also due to the dominance of binocular disparity.

The veto relationship between binocular disparity and motion in surface detection is consistent with a weak fusion model—that is, with strict modularity. Each module under this model would compute a depth map independently. The veto would occur at the stage at which the outputs of the modules are combined. The present results do not, however, rule out modified weak fusion (Landy et al., 1995). Finding certain types of interactions can support that model; failure to find interactions does not disconfirm it. The complete dominance of one cue would not be expected in a strong fusion model. Without some modularity, it would seem unlikely that motion information, which in isolation is as effective as binocular disparity information for surface detection, would have no effect on the computation of a depth map when binocular disparity information is present.

The present results suggest that separate surface interpolations based on the individual depth cues are not available to the observer. Consider the following two models, which differ in the level at which surface interpolations are first computed. In the first model, illustrated in Figure 8a, each module is used to compute a depth value for each point, and these depth values are then combined. In the second model as well, as illustrated in Figure 8b, each module computes a depth value for each point, but each module goes on to attempt to fit a smooth surface through the computed 3-D coordinates. (For discussions of processes that subjects might be using to decide whether a smooth surface fits through a set of points once the 3-D coordinates are recovered, see Andersen, 1996, and Turner et al., 1995.) The results of these surface computations are then combined. In the first model, if the combination rule always selected the disparity depth for a point when the depths computed by the two modules disagreed, the disparity depth would determine the perceived depth, with the depth computed by the motion module having no effect. The surface computation would depend entirely on the disparity depths. Consider, however, how the second model would operate when motion indicated a smooth surface and stereo indicated points randomly spread in a volume. If each module independently attempted to compute a smooth surface before the depth information from the modules was combined, the result for the "surface" display would be stereo indicating a volume and motion indicating a surface. The result for the "noise" display would be both modules indicating a volume. The motion module would thus be capable of distinguishing

between the surface and noise displays, and the binocular disparity model would not. For binocular disparity to veto motion under these conditions, the combination rule would have to disregard the only information available to distinguish between the two alternatives. It would seem more reasonable for any combination rule that was applied following separate computation of surfaces by each module to select the display for which the motion module computed a surface, no matter how little weight it gave to motion in the presence of conflicting disparity information. The veto of motion by binocular disparity, when the two modules provide conflicting information, does not provide conclusive evidence that a model in which surface interpolation follows combination of depth estimates from different cues is correct and that a model that computes a surface interpolation from each cue individually is not correct, but it is more consistent with the first alternative.

Why does binocular disparity veto motion parallax in a surface detection task? Cutting and Vishton (1995) recently suggested that a useful approach to understanding the relationship of different sources of depth information is to consider the constraints underlying each source. If one source of information can disconfirm the assumptions required by the other source, and not vice versa, a complete veto by the first source may result. Both binocular disparity and motion parallax require solution of the correspondence problem, and this does not appear to be a source of difficulty for either cue in these experiments. Disparity further requires that the two images are projections through two separate projection points. For absolute depths, this separation must be specified, but for relative depths it need not be. There does not appear to be any alternative interpretation of a stereo pair available to the visual system in which this constraint is violated.[4] Motion parallax involves a very different constraint. To obtain relative depths from a set of relative velocities, the points must be moving rigidly in 3-D space. This constraint is more easily violated than are the constraints underlying binocular disparity: If the points are not moving rigidly, velocity and depth need not be related. We propose that the reason for the veto of motion parallax by binocular disparity is that computing conflicting depths from disparity can falsify the rigidity assumption underlying motion parallax, whereas computing conflicting depths from motion parallax cannot overcome the constraints underlying disparity.

This interpretation is supported by Rogers and Collett's (1989) observation that when binocular disparity and motion parallax are placed in conflict in a shape judgment task, the judged shape is in accord with the disparity information, and nonrigid motion is perceived. Similarly, subjects viewing our displays in which disparity and motion information indicated different depths reported perceptions of nonrigid motion. It might be possible to reinforce the constraint of rigid motion by linking the motion of the dots to motion of the subject's head (Rogers & Graham, 1979).[5] However, this would not necessarily prevent subjects from perceiving an additional, nonrigid
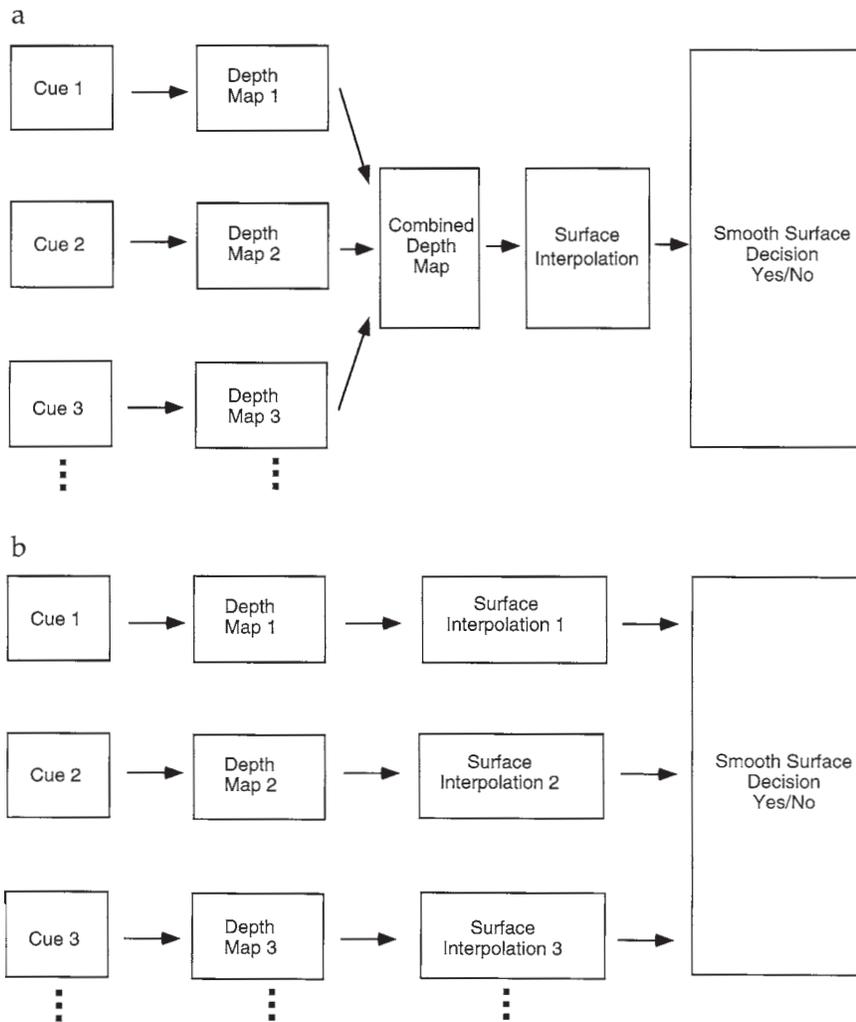
a

b

Figure 8. Diagrams of two possible levels of interaction between the different sources of depth information. (a) Each source of information determines its own depth map. The depth maps are then combined into a single depth map to which a surface interpolation routine is applied, and the decision about whether a smooth surface is present or absent is based on the composite depth map. (b) Each source of information determines its own depth map, to which a surface interpolation routine is applied. Each source of information thus has a vote in whether a surface is present or absent in any given visual stimulus.

component to the motion. The possibility that active head movements will increase the effectiveness of motion parallax in the presence of conflicting binocular disparity information will have to be resolved in future research.

The present analysis in terms of constraints is related to the normative analysis discussed by Landy et al. (1995). The rigidity constraint in motion parallax can be regarded as being more easily violated than are the constraints underlying depth recovery from binocular disparity because there is a reasonable interpretation of a stimulus in which binocular disparity indicates a surface and motion parallax indicates a volume—points moving at different speeds on a surface—but there is no equally reasonable interpretation in the opposite case. A normative analysis would thus accept a violation of the rigidity con-

straint in motion parallax more readily than it would accept a violation of the constraints underlying binocular disparity.

It is difficult to say at this point why some research examining the combined effects of binocular disparity and motion finds a veto effect and other research finds a compromise. Tittle, Perotti, and Phillips (1995) have suggested that differences in the way cues interact may relate to whether there are scale-independent differences or only scale-dependent differences between the surfaces indicated by the two cues, when inconsistent information is presented. For example, using Koenderink's (1990) method of classifying shapes, differences in shape index would be scale-independent, whereas differences in curvedness would be scale-dependent. Tittle et al. (1995)

found that dominance of one cue over another is more likely with scale-independent differences and that compromise solutions are more likely with scale-dependent differences. The difference between a smooth surface and points randomly arranged in a volume in the present experiments can be regarded as scale-independent (i.e., qualitative rather than quantitative). It may be that when the differences are qualitative, the disparity information indicates that the rigidity constraint required to recover depth from motion is violated and the surface structure is not recovered from the motion. When the differences are quantitative, the rigidity constraint may not be rejected but may be applied with some tolerance. The display may be perceived as quasi-rigid; however, as long as the rigidity constraint is not fully rejected, the motion may contribute to the recovery of the 3-D structure. This explanation must be considered speculative until considerably more is known about the interaction of cues in the detection of surfaces and in the judgment of surface shape and about the relationships among these tasks.

In summary, in a surface detection paradigm with conflicting binocular disparity and motion information in which one source indicated a random distribution of points in a volume (rather than an alternative surface interpretation), disparity information dominated performance and conflicting disparity information degraded the ability to respond to the motion, even when subjects were informed that motion provided the relevant information. We have suggested a possible reason for the dominance of binocular disparity and what type of processing would be consistent with a veto relationship in surface detection.

REFERENCES

ANDERSEN, G. J. (1996). Detection of smooth three-dimensional surfaces from optic flow. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 945-957.

ANDERSEN, G. J., & WUESTEFELD, A. P. (1993). Detection of three-dimensional surfaces from optic flow: The effects of noise. *Perception & Psychophysics*, **54**, 321-333.

BRAUNSTEIN, M. L. (1994). Structure from motion. In A. T. Smith & R. J. Snowden (Eds.), *Visual detection of motion* (pp. 367-393). New York: Academic Press.

BRAUNSTEIN, M. L., ANDERSEN, G. J., ROUSE, M. W., & TITTLE, J. S. (1986). Recovering viewer-centered depth from disparity, occlusion, and velocity gradients. *Perception & Psychophysics*, **40**, 216-224.

BÜLTHOFF, H., & MALLOT, H. (1988). Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America A*, **5**, 1749-1758.

CORNILLEAU-PÉRÈS, V., & DROULEZ, J. (1993). Stereo-motion cooperation and the use of motion disparity in the visual perception of 3-D structure. *Perception & Psychophysics*, **54**, 223-239.

CUTTING, J. E., & VISHTON, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. J. Rogers (Eds.), *Handbook of perception and cognition: Vol. 5. Perception of space and motion* (2nd ed., pp. 69-117). San Diego: Academic Press.

JOHNSTON, E. B., CUMMING, B. G., & LANDY, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research*, **34**, 2259-2275.

KOENDERINK, J. J. (1990). The brain as a geometry engine [In special issue, Domains of mental functioning: Attempts at a synthesis]. *Psychological Research*, **52**, 122-127.

LANDY, M. S., MALONEY, L. T., JOHNSTON, E. B., & YOUNG, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, **35**, 389-412.

NORMAN, J. F., & TODD, J. T. (1995). The perception of 3-D structure from contradictory optical patterns. *Perception & Psychophysics*, **57**, 826-834.

RICHARDS, W. (1985). Structure from stereo and motion. *Journal of the Optical Society of America A*, **2**, 343-349.

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., & GRAHAM, M. (1979). Motion parallax as an independent cue for depth perception. *Perception*, **8**, 125-134.

ROUSE, M. W., TITTLE, J. S., & BRAUNSTEIN, M. L. (1989). Stereoscopic depth perception by static stereo-deficient observers in dynamic displays with constant and changing disparity. *Optometry & Vision Science*, **66**, 355-362.

TITTLE, J. S., & BRAUNSTEIN, M. L. (1993). Recovery of 3-D shape from binocular disparity and structure from motion. *Perception & Psychophysics*, **54**, 157-169.

TITTLE, J. S., PEROTTI, V. J., & PHILLIPS, F. (1995). The integration of surface shape and curvedness from motion and binocular stereopsis. *Investigative Ophthalmology & Visual Science*, **36**, S847.

TURNER, J., BRAUNSTEIN, M. L., & ANDERSEN, G. J. (1995). Detection of surfaces in structure from motion. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 809-821.

UTTAL, W. R. (1985). *The detection of nonplanar surfaces in visual space*. Hillsdale, NJ: Erlbaum.

NOTES

1. We use motion parallax to refer to perspective effects (as in a polar projection of a translation) and SFM to refer to depth from motion that does not require perspective (as in a parallel projection of a rotation). See Braunstein (1994).

2. In the present experiments, the corrugations were oriented horizontally, and the motion and disparities were horizontal. Although it would be interesting to examine surfaces with vertical corrugations and motion in the vertical direction, the present restrictions were necessary to produce displays in which one cue indicated a smooth surface but the other cue indicated points in a volume, while keeping the 3-D configurations indicated by each cue rigid and constant from frame to frame.

3. Which cue was shown on a trial in which stereo and motion were compatible was randomized. An analysis of the observers' performance on these compatible trials, broken down by the particular cue, showed no consistent effects of the cue across observers.

4. Stimuli can be produced that violate this constraint, as when a stereoscope is used or when noncorresponding images are fused, but these stimuli are interpreted by the visual system as if the constraint applied.

5. This was suggested by Alan Gilchrist during a visit to our laboratory in April 1995.