Temporal integration of motion and stereo cues to depth

FULVIO DOMINI
Brown University, Providence, Rhode Island

CORRADO CAUDEK
University of Trieste, Trieste, Italy

and

PETER SKIRKO
Brown University, Providence, Rhode Island

In three experiments, we investigated the integration of three-dimensional information provided over time by different depth cues. In the first experiment, we found that the perceptual derivation of surface orientation from the optic flow was affected by the prior presentation of static stereo information in the same spatial location. This bias weakened as the length of the motion sequence increased, but it was still present after 800 msec. In the second experiment, conversely, we found that the perceived orientation of a stereo-specified surface was not influenced by the prior presentation of a static stereo surface. In a third experiment, we found that two surfaces defined by identical disparity fields did not elicit the same perceived depth if, previously, one of them had been specified by a conjunction of stereo and motion information. This effect was found to last for at least 400 msec. Taken together, these findings indicate that interactions exist among different sources of depth information, even when they are provided at different moments of time.

Human observers obtain information about three-dimensional (3-D) shape from a large number of depth cues provided by the visual environment. In recent years, many studies have been carried out to investigate the strategies that human observers may follow to combine the information specified by the different sources of depth information (Clark & Yuille, 1990; Landy, Maloney, Johnston, & Young, 1995; Richards, 1985). These investigations have tried to establish the mathematical rule by which the different depth cues are combined, proposing both linear (Bruno & Cutting, 1988) and multiplicative (Massaro, 1988) combination rules and proposing weights that change according to the reliability of the different cues (Landy et al., 1995). Whereas these studies have differed from one another with respect to the specific depth cues that have been considered and the conclusions that have been reached, these investigations have shared a common assumption—that is, they have considered the problem of depth cue combination only for *cues that are present in the same spatial location and at the same moment of time* (for the learning effects on

the reliability of the depth cues, see, e.g., Jacobs & Fine, 1999). Much less attention has been provided, conversely, to the problem of the *temporal integration of the depth cues*. Such a form of integration arises, for example, when the same 3-D object is successively defined by different sources of information (e.g., a disparity gradient first and a motion gradient later). We can, therefore, ask the following question. Can the perceptual interpretation of a depth cue, made available at one moment of time, be affected by a different depth cue that had defined the same target object at a previous moment of time? For example, let us suppose that a stationary 3-D object is viewed binocularly by a static observer (disparity information) and, at a successive moment of time, the same object is viewed monocularly by the same observer while he or she is moving (motion parallax). What we want to know is whether the perceptual interpretation of motion parallax changes according to the properties of the disparity information previously presented.

One technique that has been used to study temporal interactions among depth cues is selective adaptation. Nawrot and Blake (1989), for example, have shown that "extended exposure to structure from stereopsis subsequently influences the perception of structure from motion" (p. 718). In their experiment, observers were shown an unambiguous 3-D stereoscopic sphere rotating about the vertical axis, and successively, they were asked to report the perceived direction of rotation of an ambiguous sphere defined under an orthographic projection by motion

information alone. Their results indicate that the perceived direction of rotation of the ambiguous structure-from-motion (SFM) sphere was biased by the previous exposure to the unambiguous motion from stereopsis (see, also, Nawrot & Blake, 1991).

In another study, Poom and Börjesson (1999) asked observers to adapt to a stationary slanted planar surface viewed binocularly or to motion parallax generated by the translation of the observers' heads. In the test phase, the observers were exposed to the same stimuli as in the adaptation phase, and their task was to determine the orientation in which the test surface had to be placed in order to appear frontoparallel. The results of their experiments revealed that *aftereffects were generated even when the surfaces used in the adaptation and test phases were defined by different depth cues*. Poom and Börjesson also found transfer effects among texture gradients, motion parallax, and binocular disparity.

Domini, Adams, and Banks (2001) studied whether the mechanisms underlying the stereoscopic curvature aftereffect (Graham & Rogers, 1982) are tuned to particular disparity patterns or to some other property, such as surface curvature. Their results clearly support the second hypothesis, thus showing that 3-D aftereffects are caused by adaptation among mechanisms specifying surface shape, rather than among mechanisms signaling the disparity pattern. If aftereffects are caused by adaptation to higher level mechanisms tuned to surface representation, it may be expected that the 3-D representation specified up to a certain moment of time by a certain depth cue may influence the interpretation of another depth cue that is provided at a second moment of time.

In the present study, the temporal interactions among different depth cues were not investigated by means of an adaptation procedure (as in the experiments described above), but rather, by using a method previously employed by Domini et al. (2002). In one of their experiments, Domini et al. presented in succession two optic-flow fields in the same spatial location, and observers were asked to judge, at the end of the stimulus sequence, the orientation of the perceived 3-D surface. Domini et al. (2002) found an effect of long-term temporal integration for the perceptual interpretation of the optic flow: The 3-D orientation perceived at the end of the stimulus sequence was strongly affected by the properties of the first of the two optic flow fields that had been presented in succession. Such temporal effects were found to last for over 1,200 msec.

In the present investigation, our intention was to extend those findings by studying the *interactions of different depth cues over time*. In particular, we tried to establish whether the perceived 3-D orientation of a planar surface defined by motion information (Experiments 1 and 2) or stereo information (Experiment 2) can be affected by the properties of a stationary disparity gradient that was presented in the same spatial location at a previous moment of time. Moreover, we studied whether the perceived depth of a sinusoidally corrugated surface defined by a stationary disparity field can be affected by the dynamic properties of a previously presented stereo-specified surface (Experiment 3).

EXPERIMENT 1

In Experiment 1, observers viewed two adjacent planar surfaces slanted in depth (see Figure 1). In the first segment of the stimulus sequence (*history*), the planar surfaces were defined by two static disparity fields. In the
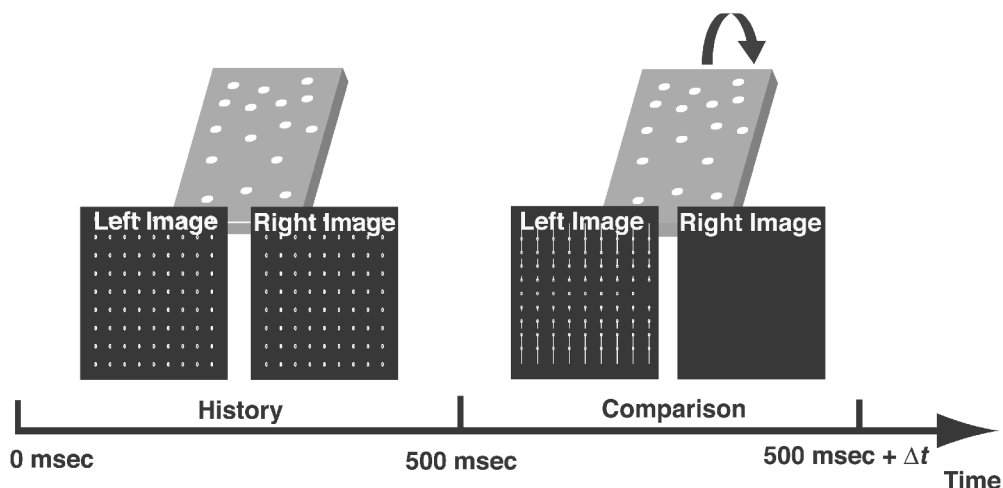

Figure 1. Schematic representation of the stimulus sequence used in Experiment 1. During the first 500 msec of the stimulus display (history phase), each of two adjacent surfaces (standard and test) was specified by stereo disparity (in this figure, only one surface is shown). During the comparison phase, the right image of each surface was replaced by a blank field, and the dots present on the left image were animated so as to produce a linear constant optic flow specifying a rotation in depth of a planar surface.

second segment of the stimulus sequence (*comparison*), the two planar surfaces were defined by two monocularly viewed velocity fields. We hypothesized that the disparity field shown during the first segment of the stimulus sequence would bias the perceptual interpretation of motion information presented in the same spatial location during the second segment of the stimulus sequence. According to this hypothesis, observers should perceive a larger slant when a given velocity gradient follows a large disparity gradient than when it follows a small disparity gradient. We propose, in fact, that perceived 3-D surface orientation builds up in time, by taking into account the orientation perceived in the previous moment of time (Domini et al., 2002). Alternatively, no integration of stereo and motion information occurs over time. In this second case, the judgment of the optic flow would not be biased by the prior presentation of disparity information.

Method

Participants. Three volunteers, recruited from the Brown University community and naive as to the purpose of the experiment, and one of the authors (P.S.) participated in this experiment. All the participants had normal or corrected-to-normal vision.

Apparatus. The displays were presented on a high-resolution color monitor (1,280 × 1,024 addressable locations), under the control of a Hewlett-Packard Visualize X550 Workstation. The screen had a refresh rate of 120 Hz. The stimuli were presented stereoscopically using liquid crystal shuttered glasses that were synchronized with the monitor refresh rate. The refresh rate of the stereo image, therefore, was equal to 60 Hz. An anti-aliasing procedure was used: For point locations falling on a pixel boundary, the screen luminance was proportionally adjusted in the relevant addressable locations. The observer sat 100 cm from the screen and viewed the displays through a reduction screen placed approximately 10 cm away from the monitor. The reduction screen limited the vertical viewing area to a 9 × 11 cm rectangular aperture. A chinrest was used to restrict head movement. The experiment was run in a dark room.

Stimuli. The stimuli were high-luminance random dots presented on a low-luminance background. Two rectangular regions located 1 cm to the left or to the right of a central fixation point depicted, at different instants of time, stereo-specified or motion-specified random-dot planar surfaces. At the beginning of each trial, a stereo image specifying two planar surfaces was presented (*history* phase). Each simulated planar surface could be either frontoparallel or slanted 30° about the horizontal axis. For the slanted surfaces, the maximum disparity was 6.3 arcmin. After 500 msec, the right image was replaced by a blank field, and the dots on the left image were animated (*comparison* phase). The velocity fields that were created in this manner simulated two linear velocity fields (all velocity vectors were parallel to the vertical axis). The horizontal gradient of each velocity field was always nil. The magnitude of the vertical gradient of one of the two motion fields (*standard*) was equal to 0.103 sec^{-1} (which produced a maximum velocity at the top and bottom of the stimulus of 37.8 arcmin/sec). The vertical gradient of the other motion field (*test*) took on, in different trials, the value of 80%, 90%, 100%, 110%, or 120% of the magnitude of the gradient of the standard. The duration of the comparison phase ranged from 200 to 800 msec, in 200-msec steps. The two random-dot fields shown on each trial were contained in a region approximately 9 cm wide and 11 cm tall. They were separated by a blank region approximately 2 cm wide. Each frame of the stimulus sequence displayed 500 dots, and dot density was kept con-

stant (Sperling, Landy, Dosher, & Perkins, 1989). The position (left or right) of the standard and test sequences was randomly selected on each trial.

Design. Three within-subjects variables were studied: the condition (during the history phase, the test and standard disparity gradients were [30°, 0°], [0°, 0°], [30°, 30°], or [0°, 30°]), the ratio between the velocity gradient of the test and the velocity gradient of the standard (0.8, 0.9, 1.0, 1.1, or 1.2), and the duration of the motion sequences (200, 400, 600, or 800 msec).

Procedure. The participants were asked to report whether the velocity field supporting the largest perceived slant at the end of the stimulus sequence was located on the left or on the right. The participants responded by pressing the corresponding mouse buttons connected to the workstation. No feedback was provided for "correct" responses.

Results and Discussion

The data from the naive observers and 1 expert observer were analyzed together, since a similar pattern was exhibited. A repeated measures analysis of variance (ANOVA) was conducted on the frequencies of "test more slanted" responses, with condition (test/standard disparity gradient: [30°, 0°], [0°, 0°], [30°, 30°], or [0°, 30°]), test/standard motion gradient ratio (0.8, 0.9, 1, 1.1, or 1.2), and duration of the SFM sequence (200, 400, 600, or 800 msec) as the within-subjects independent variables. As was expected, the effect of the condition variable was significant [$F(3,9) = 31.101, p < .001$] and interacted with the effect of duration [$F(9,27) = 3.604, p < .01$]. The effect of the motion gradient ratio was significant [$F(4,12) = 10.103, p < .001$] and interacted with the effect of the duration of the SFM sequence [$F(12,36) = 6.516, p < .001$]. Finally, the interaction between condition and motion gradient ratio was significant [$F(12,36) = 2.171, p < .05$]. No other effects or interactions were significant.

The percentages of "test more slanted" responses as a function of condition, ratio between the velocity gradient of the test and the velocity gradient of the standard, and duration of the motion sequences are shown in Figure 2 (the average standard error was 6%). In the conditions in which, during the history phase, the disparity gradients for the test and the standard were equal ([0°, 0°] and [30°, 30°]), the responses of the observers did not reveal any systematic bias. This is shown in the figure by the gray and black squares. Conversely, if different disparity gradients were used for the test and the standard during the history phase ([0°, 30°] and [30°, 0°]), the observers' responses were systematically biased (open and filled circles). If the standard was preceded by a 0° disparity field and the test was preceded by a 30° disparity field, the observers were more likely to judge the test to be more slanted; if the standard was preceded by a 30° disparity field and the test was preceded by a 0° disparity field, the observers were less likely to judge the test to be more slanted. From Figure 2, it is also easy to see that this bias decreases as the length of the SFM sequence increases. For the present stimuli, however, the influence of the prior presentation of stereo on the perceptual interpretation of motion lasted for up to 800 msec.
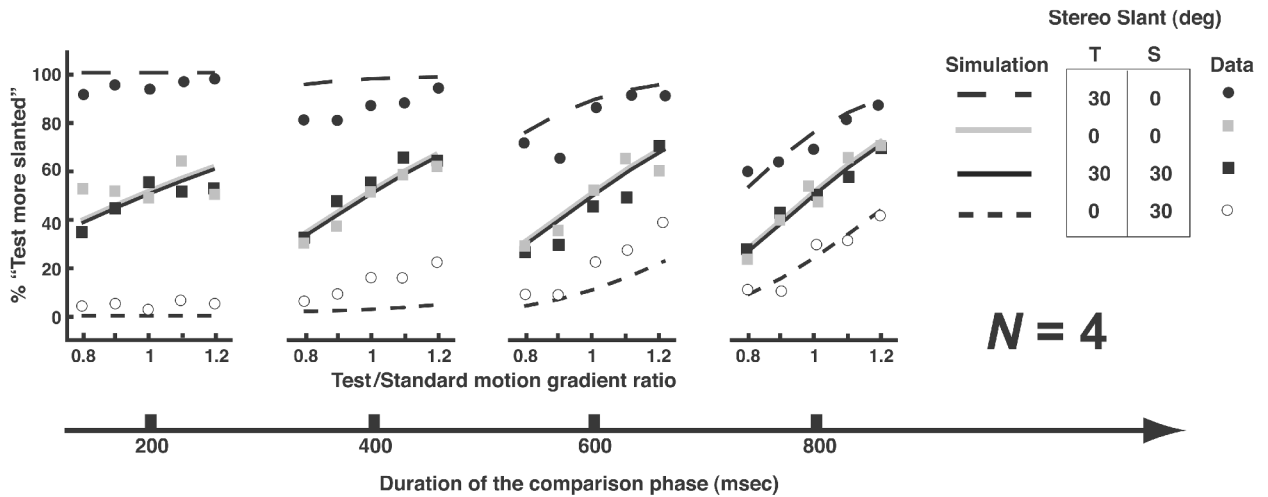
Figure 2. The mean percentages of "test more slanted" responses in Experiment 1 as a function of condition, test/standard motion gradient ratio, and duration of the comparison phase. Squares and circles represent the empirical data. Lines represent the predictions of the model proposed by Domini et al. (2002); for a discussion of this simulation, see the General Discussion section.

EXPERIMENT 2

The results of the previous experiment indicated that the perceptual interpretation of the optic flow is affected by the prior presentation of stereo information, thus indicating the combination through time of these two depth cues. The purpose of Experiment 2 was to determine whether temporal integration occurs also when stereo information specifies both the history and the test phases of a stimulus sequence otherwise identical to that used in Experiment 1. In other words, we investigated whether temporal integration occurs over time for two separate episodes providing disparity information only.

Method

Participants. Five volunteers, recruited from the Brown University community, participated in this experiment. They had not participated in the previous experiment and were naive as to the purpose of the present empirical investigation. All the participants had normal or corrected-to-normal vision.

Apparatus. The apparatus was the same as that in the previous experiment.

Stimuli. The stimuli in this experiment were identical to those used in the previous experiment, except that the simulated 3-D surfaces during the comparison phase could be specified either by motion or by stereo information. When the simulated 3-D surfaces were specified in the comparison phase by disparity gradients, the simulated slant for the standard surface was 41°. The simulated slant for the test surface was determined as specified below in the Procedure section. In the control condition, in the comparison phase both standard and test surfaces were defined by disparity information. The history condition was the same as that used in Experiment 1.

Design. Four within-subjects variables were studied: the comparison-phase type (stereo vs. motion), the condition [during the history phase, the test and standard disparity gradients were (30°, 0°), (0°, 0°), (30°, 30°), or (0°, 30°)], the ratio between the motion (stereo) gradient of the test and the motion (stereo) gradient of the standard (five levels), and the duration of the motion sequences (400 or 800 msec).

Procedure. The procedure was identical to that of the previous experiment, except for the fact that, in Experiment 2, a calibration procedure was used in order to equate the difficulty of the discrimination between test and standard, in both the stereo and the motion comparison types. Prior to the experiment, a preliminary session was run so that, for each subject, the ratios between the motion (stereo) gradient of the test surface and the motion (stereo) gradient of the standard surface were such that the percentage of "test more slanted" responses varied (as a function of the motion [stereo] test gradient) in the range between 10% and 90%. In this way, the difficulty of the stereo and the motion comparisons was equated.

Results and Discussion

A repeated measures ANOVA was conducted on the frequencies of "test more slanted" responses, with condition (test/standard disparity gradient: [30°, 0°], [0°, 0°], [30°, 30°], or [0°, 30°]), test/standard motion gradient ratio (0.8, 0.9, 1, 1.1, or 1.2), duration of the SFM sequence (400 or 800 msec), and comparison type (stereo or motion) as the within-subjects independent variables. The most important result was the interaction between the condition variable and the comparison variable type [$F(3,12) = 53.830, p < .001$]. The condition [$F(3,12) = 11.803, p < .001$] and the motion gradient ratio [$F(4,16) = 154.961, p < .001$] variables were significant. Also significant were the two-way interactions between duration of the SFM sequence and condition [$F(3,12) = 6.426, p < .01$], between comparison type and motion gradient ratio [$F(4,16) = 13.217, p < .001$], between duration of the SFM sequence and motion gradient ratio [$F(4,16) = 5.650, p < .01$], and between condition and motion gradient ratio [$F(12, 48) = 4.934, p < .001$] and the three-way interactions between comparison type, duration of the SFM sequence, and condition [$F(3,12) = 5.785, p < .05$], between comparison type, duration of the SFM sequence, and motion gradient ratio [$F(4,16) = 6.778, p < .01$], and between comparison type, condition, and motion gradient ratio [$F(12, 48) = 2.639, p < .01$]. No other effects or interactions were significant.

The percentages of "test more slanted" responses as a function of condition, test/standard motion gradient ratio, duration of the SFM sequence, and comparison type are shown in Figure 3 (the average standard error was 5.19%). The first thing that should be noticed is that, in all control conditions (i.e., when the test and the standard disparity gradients were [0°, 0°] or [30°, 30°] and for both stereo and motion comparisons), all the psychometric functions are parallel, thus indicating equal sensitivity for the stereo and the motion tasks. The important thing, however, is that the stereo history phase had no effect on the stereo comparison (Figure 3, left panel) whereas, in the motion comparison, the results obtained in Experiment 1 were replicated (Figure 3, right panel). The size of the bias for the motion stimuli, moreover, is comparable across experiments.

These results indicate that a disparity gradient presented at a certain instant of time does not influence the perceived 3-D orientation of a subsequently presented disparity gradient. This finding is apparently inconsistent with the results of Domini et al. (2002), which clearly indicate that perceived 3-D orientation specified by motion information is the result of a process of temporal integration. This apparent inconsistency, however, can be easily explained by noting that two velocity gradients presented in the same spatial location can, in principle, be produced by *the same* 3-D surface orientation, as the consequence of the acceleration or deceleration of the surface. Information about both episodes, therefore, can be accrued to better recover the orientation of the projected surface. A change in the disparity gradient, conversely, *must necessarily* correspond to a change in the 3-D orientation of the surface. In this second case, therefore, the recovery of 3-D surface orientation from the two dimensional (2-D) projection would not benefit from the accrual of information provided at different moments of time. It should thus be expected that tempo-

ral integration plays a lesser role in the perceptual interpretation of successive stereo episodes than it does in the perceptual interpretation of the optic flow.

EXPERIMENT 3

In a pilot experiment, we devised a stimulus display providing both stereo and motion information, so that the depth perceived from the simultaneous presentation of the two cues was larger than the depth perceived from stereo alone. The purpose of Experiment 3 was to determine whether the depth arising from the combination of stereo and motion information persists over time also when the dynamic stereoscopic display reaches a stop, thus specifying 3-D shape through disparity alone. Such a situation arises, for example, when a moving observer binocularly views a 3-D object. If the object is distant enough (more than 1 m; see Johnston, 1991; Johnston, Cumming, & Landy, 1994; Tittle & Braunstein, 1993), the 3-D shape that is supported by the combination of stereo and motion information can be deeper than the 3-D shape that is specified by stereo information alone. When the observer stops, however, 3-D shape is indeed specified by stereo information only. Here, we ask whether, in this case, perceived 3-D shape remains constant for a certain time interval or suddenly changes as a consequence of the disappearance of the dynamic information. In order to increase the realism of the stimulus displays, moreover, in this third experiment we modified the stimulus displays so that, unlike Experiment 1, they provided second-order temporal information. In other words, rather than using constant optic flow fields, we simulated a large-angle rotation (40°) of the projected surfaces.

Method

Participants. Three volunteers, recruited from the Brown University community, participated in this experiment. They had not
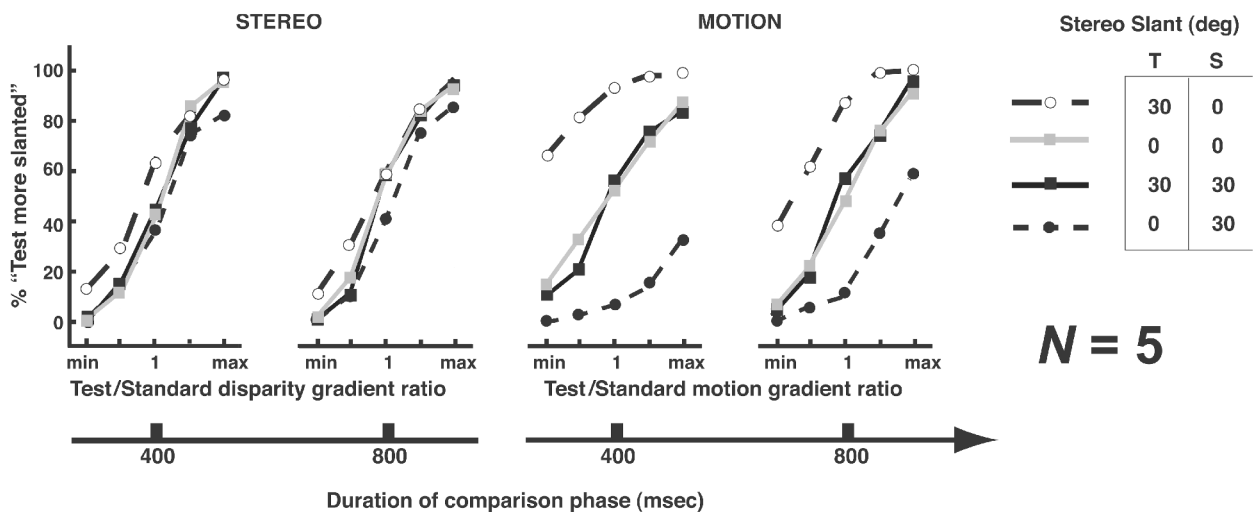

Figure 3. The mean percentages of "test more slanted" responses in Experiment 2 as a function of comparison type (stereo or motion), condition, test/standard disparity or motion gradient ratio, and duration of the comparison phase.

participated in the previous experiments and were naive as to the purpose of the present empirical investigation. One of the authors (C.C.) also participated in this experiment. All the participants had normal or corrected-to-normal vision.

Apparatus. The apparatus was the same as that in the previous experiments.

Stimuli. The stimuli were random-dot stereograms. At the beginning of each trial (see Figure 4), a stereo image specifying one sinusoidal surface was presented either at the top or at the bottom of the screen for 1,000 msec (the history phase). The stimulus display (standard) contained exactly two cycles of the sinusoidal corrugation. The peak to through depth was 2 cm, corresponding to a maximum disparity of 2 arcmin. In the stereo–motion condition, during the whole history segment of the stimulus sequence, a horizontal velocity field was added to the standard. The 3-D angular rotation varied linearly from a maximum of 80 deg/sec to 0 deg/sec over 60 frames (the history phase). At the beginning of the rotation, the simulated surface was rotated in depth by 40°. At the end of the rotation, the surface was parallel to the image plane. In the stereo-only condition, the sinusoidal surface during the history segment was defined by disparity alone. The history phase was followed by time interval Δt, during which the disparity field was the same as that during the history phase, but motion was no longer present. Δt took on the value of 0, 200, or 400 msec. After the time interval Δt, the comparison phase began with no interstimulus interval between them. In the comparison phase, the sinusoidal surface that had been presented previously was left on the screen for another 200 msec. During this time, a second sinusoidal surface (test) was presented, below or above the standard. The test simulated a horizontal sinusoidally corrugated surface with the same frequency as the standard. In different trials, the test/standard amplitude ratio was equal to 0.30, 0.65, 1.00, 1.35, 1.70, 2.05, or 2.40. During the comparison phase, both the standard and the test were defined by disparity

information alone, in both the stereo–motion and the stereo-only conditions. The two random-dot fields shown on each trial were contained in a region with a size of 252 × 252 pixels (5.4° × 5.4°). The standard and the test were separated by a blank region of 100 pixels. Both the standard and the test were defined by 600 dots. The position (top or bottom) of the standard and the test surfaces was randomly selected on each trial.

Design. Four within-subjects variables were studied: the condition (stereo–motion vs. stereo only), position of the standard (top or bottom), the ratio between the amplitude of the stereo-defined test and standard surfaces (0.30, 0.65, 1.00, 1.35, 1.70, 2.05, or 2.40), and the duration of the time interval Δt between the history and the comparison phases (0, 200, or 400 msec).

Procedure. The experimental task was to compare the depth of the two sinusoidal surfaces when they were both present on the screen (i.e., during the comparison phase). In each trial, the participants were asked to indicate whether the deeper surface was located at the top or at the bottom of the screen. The participants responded by pressing the corresponding mouse buttons connected to the workstation. No feedback was provided for "correct" responses.

Results and Discussion

As in Experiment 1, the data of the naive observers and of 1 expert observer were analyzed together, since they exhibited a similar pattern. A repeated measures ANOVA was conducted on the frequencies of "test deeper" responses with condition (stereo–motion vs. stereo only), test/standard stereo amplitude ratio (0.30, 0.65, 1.00, 1.35, 1.70, 2.05, or 2.40), duration of the time interval Δt (0, 200, or 400 msec), and location of the standard (top vs. bottom) as the within-subjects independent variables.
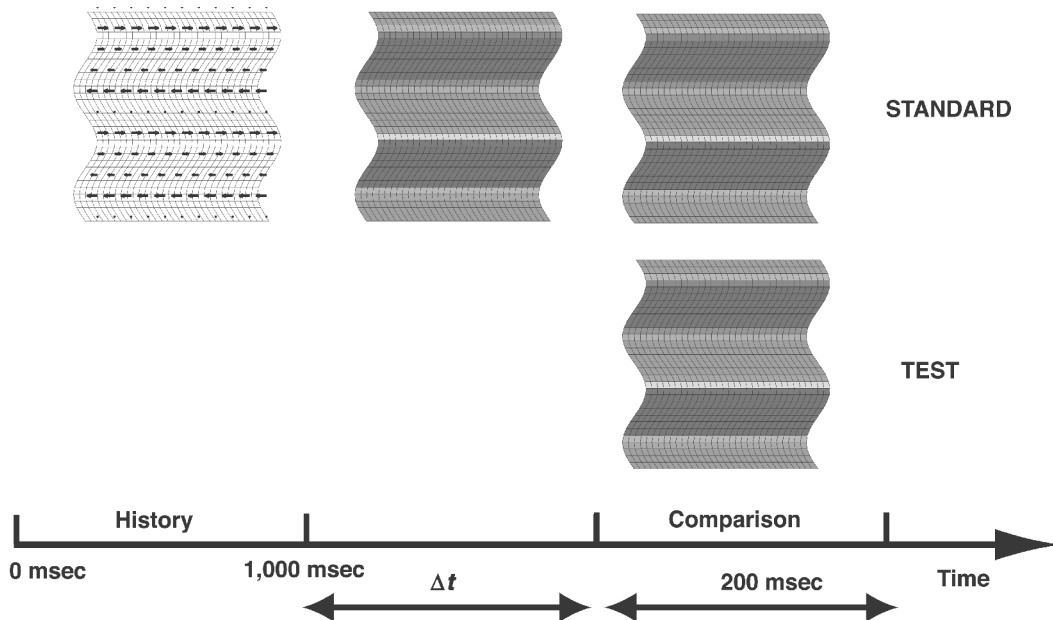

Figure 4. Schematic representation of the stimulus display used in Experiment 3. During the first 1,000 msec of the stimulus sequence (history), only one corrugated surface (standard) was shown. The three-dimensional shape of the standard during the history phase was specified by stereo and motion information (stereo–motion condition) or by stereo information only (stereo-only condition). In the stereo–motion condition, motion was present only during the history phase of the stimulus sequence. After a time interval Δt (0, 200, or 400 msec), another corrugated surface (test) was presented either above or below the standard. During the comparison phase, the standard and test surfaces were shown for 200 msec.

The most important result was the interaction between the condition variable (stereo–motion and stereo only) and the test/standard stereo amplitude ratio variable [$F(6,18) = 6.407, p < .001$]. The condition [$F(1,3) = 22.056, p < .05$] and test/standard stereo amplitude ratio [$F(6,18) = 56.275, p < .001$] variables were significant. Also significant were the three-way interaction between condition, location of the standard, and duration of Δt [$F(2,6) = 9.587, p < .05$].

The percentages of "test deeper" responses as a function of condition, test/standard stereo amplitude ratio, and duration of the time interval Δt are shown in Figure 5 (the average standard error was 5.19%). In the control condition (stereo only), the psychometric function representing the percentages of "test deeper" responses was unbiased. This is in sharp contrast with the psychometric function relative to the stereo–motion condition. In this second case, a large bias was present, indicating that the observers were much more likely to report that the standard was deeper (in the history phase of the stimulus sequence, in fact, the standard was preceded by a display providing both stereo and motion information). It is important to remember that, during the comparison phase, the standard and the test stimuli were identical across conditions (stereo–motion and stereo only). These results, therefore, indicate that *a larger depth is likely to be perceived when a surface that, previously, had been defined by stereo and motion is then specified by disparity alone.* In the present experiment, this effect was found to last for at least 400 msec.

GENERAL DISCUSSION

In three experiments, the participants were asked to indicate which of two adjacent displays (standard and test) defined by motion (Experiments 1 and 2) or stereo (Experiments 2 and 3) information appeared at the end of the stimulus sequence to be more slanted in depth, or deeper. In Experiments 1 and 2, we found that the amount of slant that was recovered from motion information was affected by the stereo information that had been shown in the same spatial location at a previous moment of time. In Experiment 3, we found that the increase of perceived depth that resulted by adding motion to a disparity field persisted over time when motion was no longer present. Taken together, these findings indicate that the perceptual processing of motion and disparity information may interact with each other, even when these two sources of depth information are not present in the same spatial location at the same moment of time.

In the first experiment, we investigated whether stereo information shown at the beginning of a stimulus sequence (history) would affect the perceptual interpretation of the motion information shown in a later moment of the stimulus sequence (comparison). When the two adjacent velocity gradients that were shown in each trial followed two identical disparity gradients (both specifying either a 0° or a 30° slant), the psychophysical curves representing the percentages of "test more slanted" responses as a function of the *standard/test* motion gradient ratios were not biased, for all the durations of the comparison segments. When the standard and the test velocity gradients followed two different disparity gradients (one specifying a 0° and the other specifying a 30° slant), the psychophysical curves were biased. If the standard followed a 0° disparity field and the test followed a 30° disparity field, the probability of judging the test to be more slanted increased; if the standard followed a 30° disparity field and the test followed a 0° disparity field, the probability of judging the test to be more slanted decreased. This perceptual bias weakened as the length of the comparison sequence increased, but it was still present after 800 msec.
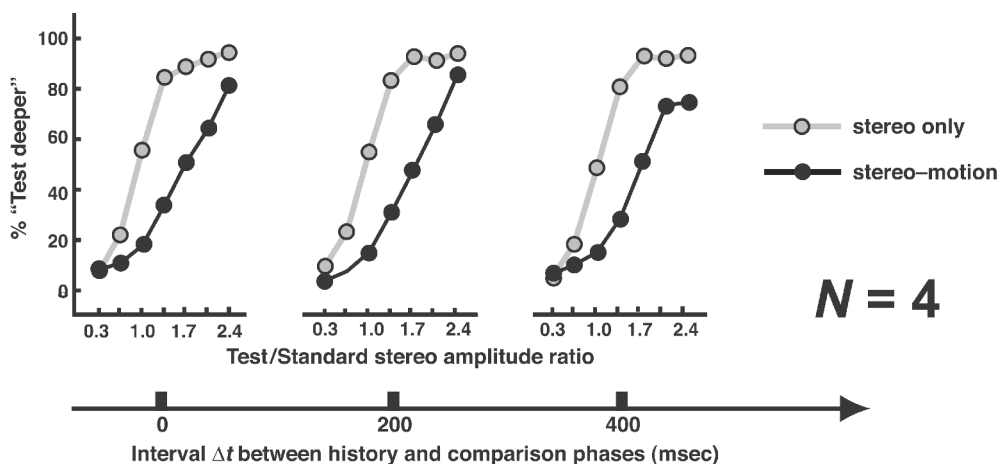

Figure 5. The mean percentages of "test deeper" responses in Experiment 3 as a function of condition (stereo only and stereo–motion), test/standard stereo amplitude ratio, and interval Δt between the history and the comparison phases.

In the second experiment, we compared perceived 3-D surface orientation in trials in which a stereo history segment was followed by a motion comparison segment and in trials in which a stereo history segment was followed by a stereo comparison segment. The results of Experiment 2 showed no bias in the stereo–stereo trials, whereas the responses of the stereo–motion trials revealed the same biases as those in Experiment 1. The absence of a perceptual bias in the stereo–stereo condition may be attributed to the fact that, unlike for stimuli providing motion alone, a change in the magnitude of the disparity gradient always corresponds to a change in the 3-D orientation of the viewed surface. In the case of the optic flow, on the other hand, a change in the magnitude of the velocity gradients may be produced, not only by a change of surface orientation, but also by the acceleration or deceleration of a surface that is maintaining a constant 3-D orientation. Temporal recruitment of information, therefore, is appropriate to better recover 3-D surface orientation from a 2-D projection in the case of successive motion episodes (Domini et al., 2002) or successive stereo and motion episodes (Experiment 1), but not in the case of successive stereo–stereo episodes (Experiment 2).

The ambiguity of motion information stems from the fact that, in deriving 3-D information from the optic flow, the human visual system makes use only of the information provided by two views (first order), which can be produced by the orthographic projection of an infinite number of different 3-D rigid structures. Evidence that the perceptual analysis of the optic flow is restricted to the first-order velocity field has been provided by several investigations (Caudek & Domini, 1998; Domini & Caudek, 1999; Domini, Caudek, & Proffitt, 1997; Liter & Braunstein, 1998; Liter, Braunstein, & Hoffman, 1993; Norman & Todd, 1993, 1995; Todd & Bressan, 1990; Todd & Norman, 1991) and, in particular, by the finding that performance in SFM tasks does not improve when the number of views is increased from two to many (Todd & Bressan, 1990; Todd & Norman, 1991). As a consequence of the necessity of fixing one of the infinite solutions that, in principle, are compatible with the first-order information (for a discussion, see Domini & Caudek, 1999), it is therefore to be expected that *the perceptual interpretation of motion information is likely to be affected by the 3-D interpretation that has been assigned by the unambiguous stereo information to the same target object at a previous moment of time* (see Experiments 1 and 2).

The fact that the perceptual interpretation of a static disparity gradient is not influenced by previously presented static stereo information, however, does not necessarily mean that a disparity field is only processed instantaneously. In the third experiment, in fact, we found that two surfaces defined by identical disparity fields did not give rise to the same perceived depth if, previously, one of them had been specified by a conjunction of stereo and motion information. Earlier investigations have shown that the addition of motion to a stereo display can increase the amount of perceived depth (e.g.,

Johnston, 1991; Johnston et al., 1994; Tittle & Braunstein, 1993). The original contribution of the present experiment is to show that this depth increment persisted over time when the surface that had been previously defined by both stereo and motion information was then defined by stereo information alone. For our stimuli, this motion-induced depth increment was found to persist for at least 400 msec. The present results suggest, therefore, that a combination of stereo and motion information can take place also over time *between the disparity information that is actually provided by the stimulus display and the persistence of a "memory trace" of motion information that had been presented at a previous moment of time*. Evidence that a 3-D representation defined by motion information persists over time when motion is no longer present had been previously provided by Domini and Braunstein (1998). In that study, the participants were shown a random-dot planar surface rotating in depth, and at the end of the SFM sequence, a pair of dots was flashed on the screen for a brief interval. The task of the observer was to judge the tilt of the imaginary line connecting the dots. Even if the dots were shown when motion was no longer present, the perceived 2-D orientation of the imaginary line connecting the two dots was found to depend on the perceived 3-D orientation of the surface that had been previously defined by motion information.

The fact that the perceptual interpretation of the optic flow may be affected by the prior presentation of stereo information is a novel result, but it should not come as an unexpected one. Several reports, in fact, have revealed the existence of interactions between stereo and motion information when these two depth cues are presented simultaneously (e.g., Cornilleau-Pérès & Droulez, 1993; Johnston et al., 1994; Tittle & Braunstein, 1993; Tittle, Perotti, & Norman, 1997; Tittle, Todd, Perotti, & Norman, 1995; see, also, Richards, 1985). Moreover, it is well known that motion information is inherently ambiguous. This ambiguity stems from the fact that, in deriving 3-D information from the optic flow, the human visual system makes use only of the information provided by two views (e.g., Caudek & Domini, 1998; Domini & Caudek, 1999; Domini et al., 1997; Liter & Braunstein, 1998; Liter et al., 1993; Norman & Todd, 1993, 1995; Todd & Bressan, 1990; Todd & Norman, 1991), which can be produced by the orthographic projection of an infinite number of different 3-D rigid structures. As a consequence of the necessity of fixing one of the infinite solutions that, in principle, are compatible with the first-order information (for a discussion, see Domini & Caudek, 1999), it is therefore to be expected that *the perceptual interpretation of motion information is likely to be affected by the 3-D interpretation that has been assigned by the unambiguous stereo information to the same target object at a previous moment of time*.

More unexpected is the finding that, also, the perceptual interpretation of disparity information may be modulated by a process of temporal integration. This finding, in fact, contrasts with the current view that the

perceptual analysis of disparity information is *uniquely determined*, provided that the viewing distance is fixed. Being unambiguous at each moment of time, therefore, the perceptual analysis of disparity is not expected to be biased by the 3-D information provided at a previous moment of time. The results of Experiment 2 are especially interesting, therefore, because of the contrast with such widespread opinion, and point to the limitation of any approach attempting to explain perceived 3-D shape solely on the basis of the information provided by a single depth source at any isolated moment of time.

The present results with regard to the integration of different depth cues over time, on the other hand, are consistent with the findings about long-term temporal integration in SFM reported by Domini et al. (2002). Domini et al. (2002) showed observers two optic flow sequences presented side by side. Each sequence was made up of two successive segments: the history and the comparison. The velocity gradients ($\phi_{x1} > \phi_{x2}$) used for the comparison segments were such that, when shown alone, the observers reliably associated (in at least 80% of the cases) the larger perceived slant (the angle between the planar surface and the frontoparallel plane) to the velocity field having the largest gradient (ϕ_{x1}). When ϕ_{x1} was preceded in the history segment by a very small velocity gradient and ϕ_{x2} was preceded by a very large velocity gradient, however, the opposite result was found: The observers consistently judged the velocity field with the smallest gradient (ϕ_{x2}) in the comparison phase as having the largest perceived slant. Such an effect was found to last up to 1,280 msec. Domini et al. (2002) explained these findings by means of a temporal integration model assuming that (1) a 3-D representation is derived heuristically from the first-order velocity field and (2) perceived local surface orientation is updated over time by averaging the slant magnitudes specified by the current optic flow, on the one hand, with the slant and angular rotation magnitudes perceived at previous moments of time, on the other. It is straightforward to extend this model to the case in which the 3-D information is specified over time by different depth cues.

CONCLUSION

The perceptual interpretation provided to a single depth cue shown in isolation (either motion or stereo) was affected by the prior presentation of a different depth cue. Such effects of cue combination over time have been found to last up to 800 msec. The results of the present investigation are consistent with the temporal integration model proposed by Domini et al. (2002).

REFERENCES

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

CAUDEK, C., & DOMINI, F. (1998). Perceived orientation of axis of rotation in structure-from-motion. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 609-621.

CLARK, J. J., & YUILLE, A. L. (1990). *Data fusion for sensory information processing systems*. Boston: Kluwer.

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.

DOMINI, F., ADAMS, W., & BANKS, M. S. (2001). 3-D aftereffects are due to shape and not disparity information. *Vision Research*, **41**, 2733-2734.

DOMINI, F., & BRAUNSTEIN, M. L. (1998, April). *Structure from motion displays can prime perspective distortions.* Paper presented at the meeting of the Association for Research in Vision and Ophthalmology (ARVO), Fort Lauderdale, FL.

DOMINI, F., & CAUDEK, C. (1999). Perceiving surface slant from deformation of optic flow. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 426-444.

DOMINI, F., CAUDEK, C., & PROFFITT, D. R. (1997). Misperceptions of angular velocities influence the perception of rigidity in the kinetic depth effect. *Journal of Experimental Psychology: Human Perception & Performance*, **23**, 1111-1129.

DOMINI, F., VUONG, Q. C., & CAUDEK, C. (2002). Temporal integration in structure from motion. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 816-838.

GRAHAM, M., & ROGERS, B. (1982). Simultaneous and successive contrast effects in the perception of depth from motion-parallax and stereoscopic information. *Perception*, **11**, 247-262.

JACOBS, R. A., & FINE, I. (1999). Experience-dependent integration of texture and motion cues to depth. *Vision Research*, **39**, 4062-4075.

JOHNSTON, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research*, **31**, 1351-1360.

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

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

LITER J. C., & BRAUNSTEIN, M. L. (1998). The relationship of vertical and horizontal velocity gradients in the perception of shape, rotation and rigidity. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 1257-1272.

LITER, J. C., BRAUNSTEIN, M. L., & HOFFMAN, D. D. (1993). Inferring structure from motion in two-view and multiview displays. *Perception*, **22**, 1441-1465.

MASSARO, D. W. (1988). Ambiguity in perception and experimentation. *Journal of Experimental Psychology: General*, **117**, 417-421.

NAWROT, M., & BLAKE, R. (1989). Neural integration of information specifying structure from stereopsis and motion. *Science*, **244**, 716-718.

NAWROT, M., & BLAKE, R. (1991). The interplay between stereopsis and structure from motion. *Perception & Psychophysics*, **49**, 230-244.

NORMAN, J. F., & TODD, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception & Psychophysics*, **53**, 279-291.

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

POOM, L., & BÖRJESSON, E. (1999). Perceptual depth synthesis in the visual system as revealed by selective adaptation. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 504-517.

PREVIC, F. H., BREITMEYER, B. G., & WEINSTEIN, L. F. (1995). Discriminability of random-dot stereograms in three-dimensional space. *International Journal of Neuroscience*, **80**, 247-253.

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

SPERLING, G., LANDY, M. S., DOSHER, B. A., & PERKINS, M. E. (1989). Kinetic depth effect and identification of shape. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 826-840.

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., & NORMAN, J. F. (1997). Integration of binocular stereopsis and structure from motion in the discrimination of noisy surfaces. *Journal of Experimental Psychology: Human Perception & Performance*, **23**, 1035-1049.

TITTLE, J. S., TODD, J. T., PEROTTI, V. J., & NORMAN, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 663-678.

TODD, J. T., & BRESSAN, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics*, **48**, 419-430.

TODD, J. T., & NORMAN, J. F. (1991). The visual perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception & Psychophysics*, **50**, 509-523.

NOTE

1. Results from studies on the discriminability of random-dot stereograms indicate that a fundamental anisotropy exists in the processing of visual targets in 3-D space (Previc, Breitmeyer, & Weinstein, 1995).