

Cross-modal slant and curvature matching of stereo- and motion-induced surfaces

SJOERD C. DE VRIES and PETER WERKHOVEN
Utrecht Biophysics Research Institute, Utrecht, The Netherlands

In many laboratory setups and in many day-to-day situations, a unique solution of the structure-from-two-views problem is unobtainable. Yet, when the visual system is presented with two projections in a sequence, it nevertheless appears to generate a reasonably stable percept of structure. In the research reported here, we examined whether the same surface would be perceived when subjects were presented with a pair of views that alternated in time monocularly (two-frame motion) or were shown simultaneously to both eyes (stereo). In Experiment 1, we studied *slant perception*: human observers were asked to match the slant of a motion-induced planar surface with its stereo-induced counterpart. In Experiment 2, the perceived *curvature* of parabolic surfaces was matched in a similar way. The results show that motion-induced slant is matched with a *higher* value of the stereo-induced slant. However, the curvature experiment showed that motion-induced curvature is matched with a *lower* stereo-induced curvature. One possible explanation may be that the slant and curvature are internally inconsistent in at least one of the modalities.

Mathematically, the metric three-dimensional (3-D) structure of a rigid set of identifiable points is generally fully specified by only two perspective projections of the object from different viewpoints (see, e.g., Longuet-Higgins, 1981). Representations of such 3-D structures can be useful to systems interacting with their environment.

The human visual system appears to be able to extract certain (not necessarily Euclidean) 3-D properties of a rigid scene from two distinct projections shown either to both eyes simultaneously (i.e., stereo) or sequentially to one eye (i.e., motion parallax).

The displacement of a point in the scene between projections is called *disparity* in the stereo domain and *velocity* in the motion domain (given the time frame). The displacement fields in both domains, however, can in principle be processed identically in order to extract 3-D information about the scene. Indeed, ample psychophysical evidence indicates a close relationship between the processing in both modalities (Graham & Rogers, 1982; Nawrot & Blake, 1989; Norman, Todd, Tittle, & Perotti, 1993; Rogers & Collet, 1989; Rogers & Graham, 1982, 1983).

Our aim is to study the ability of the visual system to extract and match 3-D properties in these modalities. Our study focuses on the extraction of surface slant and curva-

ture specified by either stereo or motion parallax and puts some constraints on the stratification of processes in the two domains.

When we refer to *specified slant* or *specified curvature* in the following text, this should be read as the slant or curvature that can be calculated from the stimuli, given the information of the viewing geometry. Furthermore, if the specified slant or specified curvature of a motion stimulus equals, respectively, the specified slant or specified curvature of a stereo stimulus, their displacement fields will be identical.

Figure 1 is a scheme of possible processing streams for extracting slant and curvature in both domains. The internal representations S_m and S_s of slant S in, respectively, motion and stereo (and, similarly, the representations C_m and C_s of curvature C) are psychophysically not observable. However, we can measure the slants S_m and S_s (and similarly the curvatures C_m and C_s) that observers judge to be equivalent, by using matching experiments. That is, one can obtain the relations M_c and M_s : $S_s = M_s(S_m)$ and $C_s = M_c(C_m)$. In Experiments 1 and 2, we will measure these relations M_c and M_s . We also tried to compare depth in both modalities. During the pilot measurements, we found depth from motion to be basically nonexistent. The relative movement of two surfaces (a planar foreground surface and a planar background) with respect to each other looked in most cases just like two surfaces sliding closely over each other. This phenomenon was previously reported by Gibson, Gibson, Smith, and Flock (1959): "The two velocity case yielded consistent perceptions of the separation of one surface into two. The flow-gradient case (motion perspective) yielded consistent perceptions of slant, or rate of recession in depth." The problem is that the situation of two surfaces sliding over each other may be a correct alternative interpretation of the stimulus. It is probable

S. C. de V. was funded by the SPIN project "3D Computer Vision" of the Dutch Ministry of Economic Affairs; P. W. was funded by the In-Sight II project of the ESPRIT Basic Research Actions of the European Community. The authors would like to thank Harald Vogt for serving as an observer, Maarten Hogervorst and Hendrik-Jan van Veen for many helpful discussions, and André Noest for his useful comments on this manuscript in its draft form. Correspondence should be addressed to S. C. de Vries, TNO Human Factors Research Institute, Kampweg 5, 3769 DE Soesterberg, The Netherlands (e-mail: sdevries@fys.ruu.nl or sjoerd@tm.tno.nl).

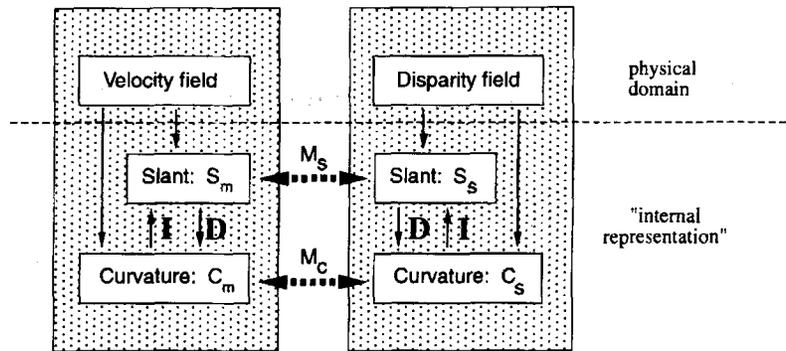


Figure 1. Scheme of possible processing streams for the extraction of slant and curvature from motion and stereo. Slant can be derived from curvature by integration and curvature can be derived by slant by means of derivation. Furthermore, both quantities may be derived independently. The dashed arrows indicate where our measurements fit in. M_s indicates the relation that is found by matching motion-defined slants with stereo-defined slants and M_c indicates the relation that is found by matching motion-defined curvatures with stereo-defined curvatures.

that a perception of depth (if it exists) is only obtained in situations in which at least a slant in depth is present. The consistency of depth and slant "modules" might then be tested by examining the perceived depth difference of two positions connected by a series of (differently) slanting surfaces. This, however, is beyond the scope of the present research.

Structure From Motion

Metzger (1934), as well as Wallach and O'Connell (1953), demonstrated the ability of the human visual system to acquire a 3-D structure from the projections of moving objects which, when viewed monocularly and statically, were purportedly devoid of 3-D depth cues. That observers obtain a unique solution is intriguing, since the computation of structure from motion is an ill-posed problem: an infinite number of moving structures project onto the same retinal images (the movement of images on a projection screen itself is an example of such a solution). To reduce the family of solutions, assumptions have to be made about the underlying structure; constraints such as connectivity, (piecewise) rigidity, smoothness of the objects, and/or knowledge of the common motion parameters have been proposed to deal with the ambiguities (see, e.g., Hoffman & Bennett, 1986; Koenderink, 1986; Longuet-Higgins, 1981; Longuet-Higgins & Prazdny, 1980; Roach & Aggarwal, 1979, 1980; Subbarao, 1988). And even then, the solutions might not be unique.

The literature cited above is concerned with perspective projection. Aloimonos and Brown (1989) have shown that in the case of orthographic or parallel projection, the rigidity constraint is not sufficient for two views with an arbitrary number of points to arrive at a unique solution. Several researchers (Bennett, Hoffman, Nicola, & Prakash, 1989; Koenderink & Van Doorn, 1991; Kontsevich, 1993) have demonstrated how a one-parameter family of solutions can be found from two parallel views. Because the quantitative difference between parallel projection and perspective is negligible in many cases (small objects at

about arm's length or farther away) and is often even objectively zero in typical laboratory setups (e.g., on discrete raster displays under the same viewing conditions), one cannot expect the visual system to do any better than to arrive at such a set of solutions (see also van Veen & Werkhoven, 1994). To arrive at a unique solution, extra assumptions have to be made about the specific surface transformation between projections or additional information has to be derived from other sources (e.g., from accommodation, familiar size, etc.). The solution may differ from setup to setup; a specific environment may force a certain response. An important question is whether the assumption depends only on the visual input itself. More specifically, is there a difference in the way we arrive at a unique solution for planar surfaces as opposed to curved surfaces?

Structure From Stereo

To derive structure from stereo, fewer assumptions have to be made about the outside world than have to be made in order to derive structure from motion in the case of the passive observer. The two frames of a stereo image are obtained simultaneously; therefore, they have to be connected by a rigid transformation. The direction of the displacement between the eyes is known, as is the direction of the axis of rotation (however, torsional movements of the eyes may change the axis of rotation). Using the same algorithms as mentioned above for motion, a reconstruction of the visual world may be obtained. Knowledge about the interocular distance (which could be gauged by experience of reaching or moving around in the world) and vergence angle (from eye muscle information) could provide additional proprioceptive information. However, although there is evidence that the vergence angle is used by the visual system, it appears that the angle is not registered very accurately (Cumming, Johnston, & Parker, 1991; Hofsten, 1976; Johnston, 1991; Ritter, 1977, 1979). There is, however, also evidence to the contrary. Stevens (1983) found that the slant of two planes at two different distances could

be matched quite accurately. One interpretation is that the distances in his experiments were known, but this claim should be taken with caution, since Stevens's planes were richly textured with reasonably large texture elements that may have served as additional, distance-independent, cues for slant.

Other Cues

We wanted to study mechanisms in the visual system that are tuned to specific *local* properties of the visual world presented in the form of disparity or velocity fields. For the surface patches studied in our experiments, such properties consist of slant (related to the first derivative of these fields) and curvature (related to the second derivative of these fields). We used a matching technique to study these properties. For example, observers were asked to match the slant of one surface patch with that of another. To discourage observers from using cues other than those in the velocity dimension, we had to take special precautions. When the two matched surface patches had identical shape and size in 3-D, the geometry of the *discontinuities* in the projected velocity field of the surface patches (at the edges of the patch) would be a sufficient cue to match their slant. Such a cue would not be sufficient to match the curvature of surface patches. Since we were interested in the consistency of *local* properties, we wanted to eliminate this cue. Therefore, we chose to randomize the shape of the motion- and disparity-specified surface patches. More specifically, we used elliptical surface patches, the size and aspect ratios of which were randomized.

A complication of the cue elimination described above is that it introduces a conflicting cue that may bias the extraction of slant based on local velocity or disparity fields: the foreshortening cue. Pilot experiments had shown that observers interpreted a projected surface patch with an aspect ratio that deviated from 1 as being a circular surface patch in space slanted according to the aspect ratio. Such an interpretation may interact with the slant estimation based on the disparity or velocity information of the surface patch. In particular, when the shape of the surface patch in space is not circular (but is assumed to be circular by the observer), this foreshortening cue provides a slant estimate that differs from the disparity-/velocity-based estimate. To our knowledge, the interaction between the slant induced by the shape of the surface patch and the slant induced by motion or disparity distributions has been studied only scantily (Youngs, 1976). We analyzed our data for the presence of such an interaction.

Another conflicting cue would be the accommodation of the eye lens. However, the accuracy of this cue is lower than the depth range used in our experiments (Helmholtz, 1910), so this cue should not pose much of a problem.

GENERAL METHOD

Procedure

We used a matching paradigm. Subjects were first shown a motion sequence with a duration of 3 sec (reference) with one eye cov-

ered and then a stereogram (test), also shown for 3 sec. The stimuli were separated by a static noise pattern which was shown for 0.5 sec. The values of the test (be it slant or curvature) could then be adjusted. After each adjustment, the resulting test image was shown. The minimum adjustment step was 0.5° in the slant experiment and 0.02 cm^{-1} in the curvature case, but larger steps could be used. Each test image had a random aspect ratio drawn from the range of the reference's aspect ratios.

Subjects could view the reference again at any time. There was no time limit. When the subjects indicated they had obtained a match, a new reference was presented.

The references were drawn in a random order. In the slant experiment, both the slant values (three) and aspect ratios (five) were mixed; each combination was presented three times, giving a total of 45 references for each tilt direction. In the curvature experiment, each of the 5 reference curvatures was presented five times in one session, with a new random texture surface for each new occurrence. Each curvature session was repeated three times, mostly on different days.

Stimuli

Both motion and the stereo stimuli were generated with the same program, which calculated the intersections of the lines from the viewpoints to the object points and the screen. In principle this is a perspective projection; more about this can be found in the General Discussion section. The motion frames are shown sequentially (each image was displayed for the duration of seven vertical retraces of the monitor [i.e. 1/10th of a second]) and the stereo frames (semi) simultaneously. Thus, the two-frame motion that is obtained in this way simulates the visual input of an observer who moves over the interocular distance (ca. 6.5 cm) between frames (see also the Appendix). The stimuli were presented at the same distance as that used for their calculation (60 cm). The center, which was to be judged, was indicated by means of a blinking dot.

Random dot stereograms were used to depict the stimuli. Each surface was surrounded by a background with about the same density (15%). We tested to ensure that no shape information whatsoever could be extracted, monocularly.

Owing to the discrete character of the screen, disparities on the screen could only change stepwise. In order to smooth the surface, the disparity of a point on the screen could be increased by one pixel, with a probability proportional to the rounding error that was made by truncating the calculated disparity of that point. In this way the average height (z) of a small neighborhood of a point was about the same as it would be in the continuous case. More details on the stimulus generation can be found in de Vries, Kappers, and Koenderink (1993, 1994).

Apparatus

We presented the random dot stereograms and the two-frame motion sequences on an Atari SM124 white phosphor monitor connected to an Atari mega ST4 computer. Left- and right-eye images were displayed on alternating vertical blanks of the monitor. The appropriate eye was selected by using LCD spectacles (Crystaleyes from StereoGraphics Corp.). This reduced the effective frame rate per eye to about 35 Hz, just above the flicker-fusion frequency for the lighting condition we used. Experiments were done in a dark room; average luminance of the screen was 40 cd/m^2 .

Screen dimensions were $20.75 \times 13.0 \text{ cm}$ ($19.6^\circ \times 12.4^\circ$), with 640×400 pixels. The screen was viewed from a distance of 60 cm.

Observers

Three observers (H.V., P.W., and S.V.) took part in the experiments. H.V. was naive with respect to the goals of the research. Observer P.W. was naive with respect to the use of the aspect ratios. Spatial acuity and stereo acuity were tested and found to be normal or above normal. H.V. and S.V. used correction lenses ($-1.5/+0.75 \text{ D}$

and $-5.5/-5D$, respectively). The right eye was used for the monocular task. This was the dominant eye for all subjects.

EXPERIMENT 1

Matching the Slant of Stereo- and Motion-Induced Planar Surface Patches

This section describes an experiment in which observers were asked to adjust the slant of a stereo-induced planar surface patch such that its slant was judged equal to the slant of a motion-induced surface patch (see the Method section for a detailed description). In the following subsection, we discuss how the matched slant of a surface patch depends on surface slant.

Method

In the first experiment, we used flat slanted surfaces with an elliptic rim (see Figure 2). The aspect ratio of the elliptic rim (defined in this case as the vertical extent divided by the horizontal extent) was randomized within the range of 0.8–1.2 in order to discourage observers from using the depth differences at the boundaries of the projection of the rim (the contour) as a cue to extract slant. The effect that the aspect ratio of the rim has on judged slant is discussed in a later subsection. The specific values of the aspect ratio were chosen in this way so that with a circularity assumption the range of apparent slants included the range of stereo-specified slants. Since the surfaces were surrounded by an equal-density background, the boundary cue was in the cyclopean domain. The surfaces subtended a circular disk (in the case of an aspect ratio of 1) with an average radius of 3.5 cm.

Originally, four slant values were used: 7.5° , 15° , 22.5° , and 30° . Slant was in the horizontal direction (a tilt of 0°) and in the vertical direction (a tilt of 90°). During the pilot phase of the experiments we

found that the 7.5° slant condition did appear more like a deforming patch, which dominated the slant percept. Therefore, we excluded this value from the experiments. The slant values were chosen so that the pixel displacements did not supersede D-max (Braddick, 1974).

Results

Dependence on stereo slant. Figure 3 shows which stereoscopically presented displacement field is matched to a particular dynamically presented displacement field. Of course, subjects were instructed to match the apparent slant only. The settings are quantified as the corresponding slants that can be calculated if the viewing geometry is known. The direction of the slant (the tilt) was horizontal in the case of the data presented in Figure 3a, and vertical in the case of Figure 3b. The reference slant (i.e., the motion-specified slant value) is given along the horizontal axis. We have shown the matched slant (i.e., the stereo-specified slant value) along the vertical axis. The three graphs correspond to the results for the 3 observers. The dashed line gives the results that would be expected if observers judged the slants of motion- and stereo-specified surface patches to be equal. Clearly, this was not the case.

For example (see Figure 3a), a slant (in horizontal direction) of 15° for a motion-specified surface patch was judged equal to a slant of 17° (averaged across observers) of a stereo-specified surface patch. Furthermore, motion-specified slant values of 22.5° and 30° were matched with (averaged) stereo-specified slant values of 32° and 35.5° , respectively. In general, observers matched a given motion-specified slant with a stereo-specified slant that was markedly higher.

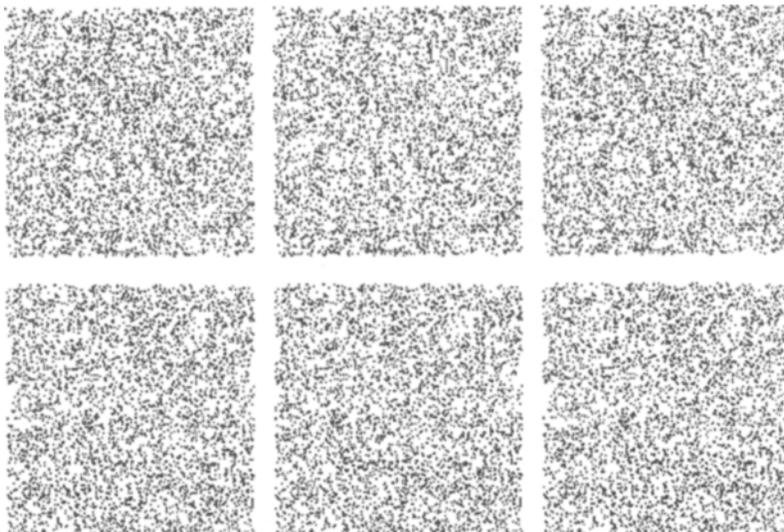


Figure 2. Examples of the presented stereograms. The upper panels show one of the stimuli of Experiment 1, a slanted planar surface; the lower panels show a horizontally oriented parabolic cylinder used in Experiment 2. Size, density and disparity levels differ from the actual setup. To view the stereograms the right-hand and middle panels should be cross-viewed or the left-hand and the middle panels should be viewed with lines of sight parallel.

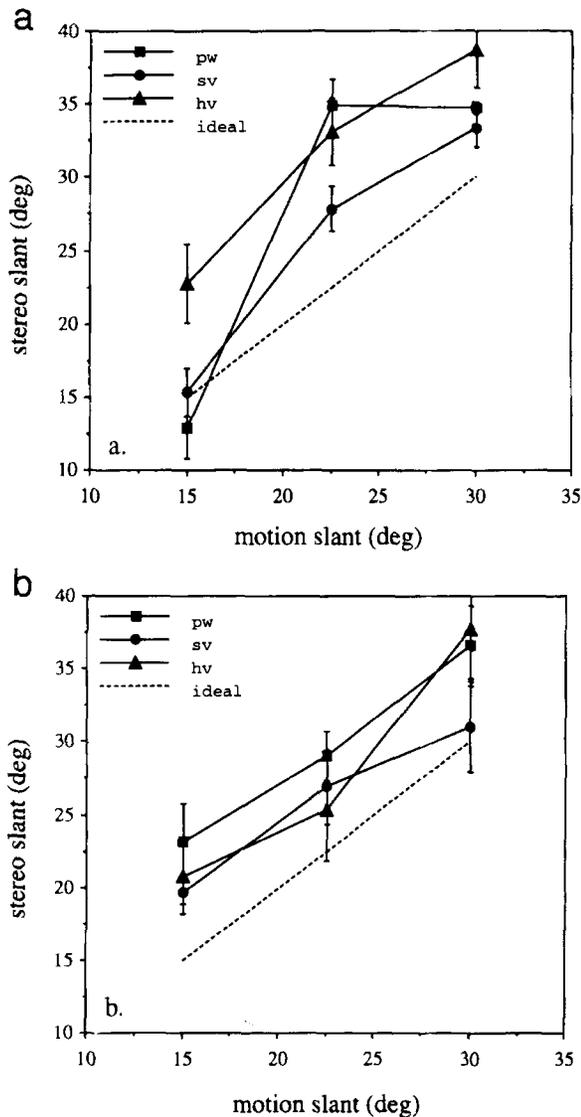


Figure 3. Matched stereo-specified slant as a function of motion-specified slant. Along the vertical axis we have shown the slant of a stereo-specified surface patch that is judged equivalent to the slant of a motion-specified surface patch (horizontal axis). The dashed line gives the results expected when observers judge the slants of motion- and stereo-specified surface patches equal. The three curves are for the Observers P.W., S.V., and H.V. (a) Motion- and stereo-specified surfaces were slanted in a horizontal direction. (b) Motion- and stereo-specified surfaces were slanted in a vertical direction. Error bars indicate the measurement error.

We observe similar results when the slant has a vertical direction (see Figure 3b). For the motion-specified slants within the range studied (15° , 22.5° , 30°), the matched stereo-specified slant values are markedly higher (21.1° , 27.0° , 35.1° , respectively).

Conclusion. The slant of motion-specified surface patches was generally matched to a higher value of the slant of stereo-specified surface patches. Thus, the same pair of images yielded different slant judgments when presented simultaneously and binocularly (stereo) than when presented sequentially and monocularly (motion).

However, we should restrict ourselves somewhat here: if the slant depends on the vergence angle (and there is experimental evidence which supports this), then, for different vergence angles, the stereo slant should be different. Such variability makes the comparison dependent on the particular value of the vergence angle, and the obtained relation (motion slant is stronger than stereo slant) cannot be used for the comparison between slant and curvature cases in general.

Dependence on aspect ratio. We have reorganized the data used for Figure 3 in order to reveal the effect that the shape of the motion-specified surface patch had on the matched slant of the stereo-specified surface patch. In Figure 4, we have averaged the matched slant values of the stereo-specified surface across all possible shapes of the motion-specified surface patches. In this section, we have singled out the data for the different aspect ratios of the motion-specified surfaces.

Figure 4 shows the matched slant of stereo-specified surfaces (vertical axis) as a function of the aspect ratio of the motion-specified surface (horizontal axis) for three slant values of the motion-specified surface: (1) 15° , (2) 22.5° , and (3) 30° . The results for the vertical slant direction (a tilt of 90°) are indicated with square symbols; the results for the horizontal slant (a tilt of 0°) condition are indicated with circular symbols.

For conditions in which the compared slant directions are vertical, the matched slant of the stereo-specified surface patch roughly *decreases* when the aspect ratio of the elliptic patch increases. That is, motion-specified patches that are elongated in a direction orthogonal to their slant direction yield higher matched slant values for the stereo-specified patches than for circularly shaped motion-specified patches.

For the condition in which the compared slant directions are horizontal, the matched slant value of the stereo-specified surfaces roughly *increases* with the aspect ratio of the motion-specified surface patch. A description in terms of the elongation of the surface patch for this "horizontal slant" condition is similar to that for the "vertical slant" condition: an increment of the elongation of a motion-specified patch orthogonal to the direction of its slant yields an increment of the matched slant of a stereo-specified patch.

Discussion

From Figure 3 we conclude that the matched slant of a motion-specified surface patch depends on the motion information contained in that surface patch (the velocity gradient, or the *motion cue*). Our visual system makes use of the fact that local velocities depend on local surface depth. Figure 4 reveals the dependence of matched slant on a second source of information: the contour of the projection of the surface patch. We will refer to the use of contour information as the *foreshortening cue*. Both cues interact in the observers' match of slant.

To discuss the use of the foreshortening cue, we use the term *edge* for the curve that "bounds" the surface patch in 3-D space. In the projection, the image of the edge appears

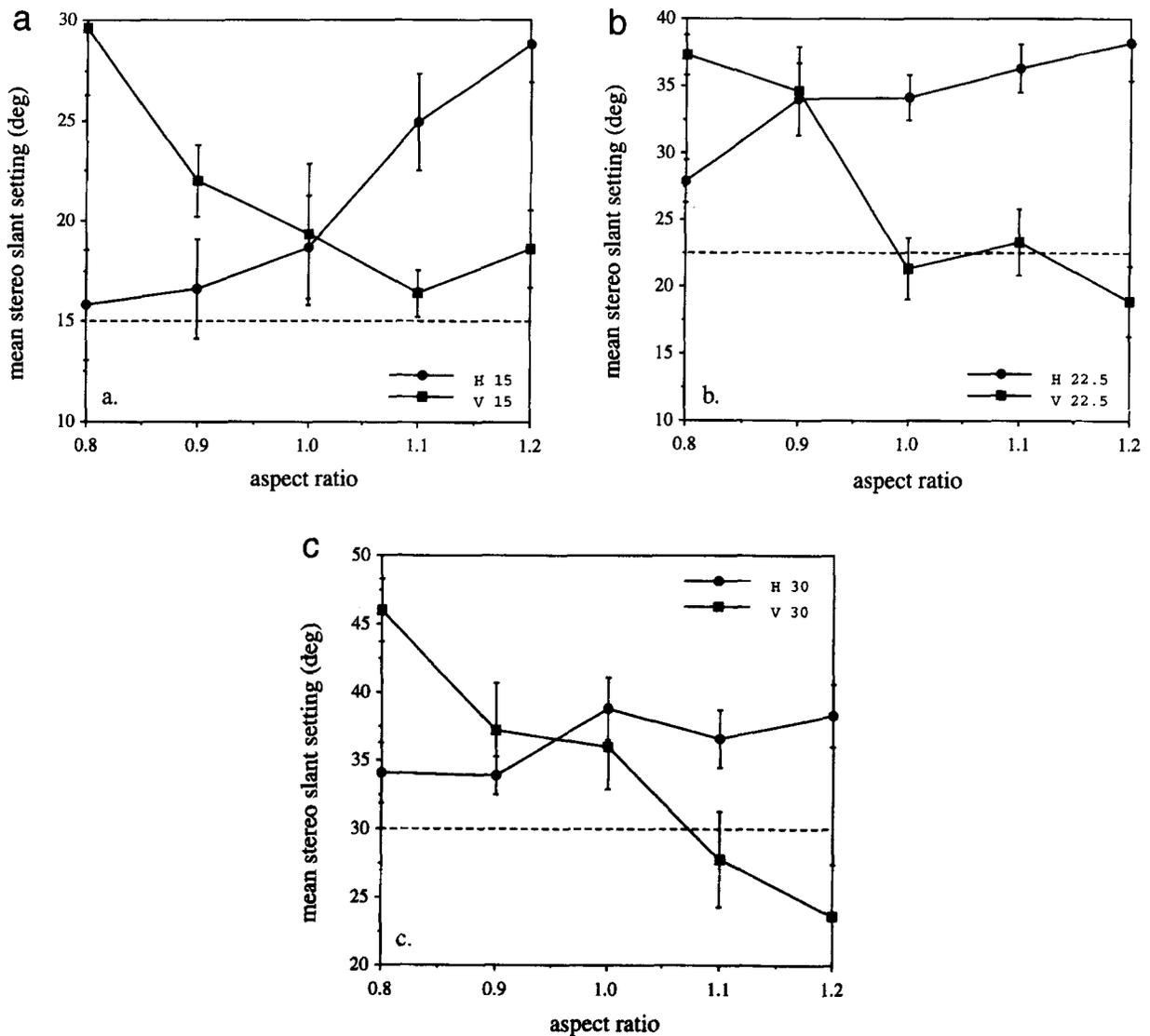


Figure 4. Effect of aspect ratio. The data used for this figure are the data presented in Figure 3, reorganized to reveal the effect of the shape of a surface patch. Plotted are the results averaged across observers. Note that observers matched the slant of a stereo-specified surface patch (of which the aspect ratio of the elliptical rim was randomized) with the slant of a motion-specified surface patch (with a rim with aspect ratio, say, α). The aspect ratio α of the motion-specified surface patch is shown along the horizontal axis. The slant of the motion-specified surface is different for each panel: (a) 15° slant; (b) 22.5° slant; (c) 30° slant. The matched slant of the stereo-specified surface is plotted along the vertical axis (averaged across Observers P.W., S.V., and H.V.). Each panel contains two curves: square symbols indicate the condition in which the stereo- and motion-specified surfaces to be compared had a vertical direction; circles indicate that compared surfaces were slanted in a horizontal direction.

as the contour. To make use of the foreshortening cue, our visual system has to make assumptions concerning the edge of the surface patch. For example, under the assumption of a circular edge, the aspect ratio of the contour is directly correlated with the slant of the surface patch. Under this “circularity” assumption, a manipulation of the edge (e.g., taking an elliptical edge) will yield a contour-based estimation of slant that deviates from the motion-based slant estimation. It is likely that the foreshortening cue biases the motion cue when the edge of the surface patch is not circular.

Under the circularity assumption, an aspect ratio of 1 for the edge of the surface patch is expected not to bias the

slant matches. Therefore, the motion-based slant matching is revealed best by the data for aspect ratio 1. From all the panels in Figure 4, we see that the matched slant of the stereo-specified surface patch is higher than the motion-specified value (except in panel b, when the surface slant is 22.5° in the vertical direction). This suggests that stereo-induced slant is indeed underestimated relative to motion-induced slant.

For aspect ratios of the edge higher than 1, the matched slant of the stereo-specified surface increases for a *horizontal* slant direction (circular symbols in Figure 4). This is consistent with the “circularity” assumption. The con-

tour of the motion-specified surface is elongated in the direction orthogonal to its slant direction. Thus, the foreshortening cue cooperates with the motion cue. For a particular aspect ratio smaller than 1 (dependent on the surface slant), the contour becomes circular, suggesting a flat surface. These values are 0.97 for a slant of 15° , and 0.92 and 0.87 for slants of 22.5° and 30° , respectively. Beneath these values, the judged contour-based slant of the surface changes in direction from horizontal to vertical. For such conditions, the motion-induced slant of the surface is smaller than the motion-induced slant of a surface patch with aspect ratio of 1, suggesting that contour-based and motion-based slant estimates interact even if the tilt values are orthogonal.

For the condition in which motion- and stereo-based surface slant is vertical (square symbols in Figure 4), the observations are similar when discussed in terms of reciprocal aspect ratios. For a purely isotropic visual system, the ascending curves (circular symbols) and descending curves (square symbols) in a panel should be each other's mirror images with the vertical through the aspect ratio 1 as a symmetry axis. Such symmetry is observed in Figure 4a (15° slant).

Obviously, however, such symmetry does not generally exist. The effect of contour-based information on the judgment motion-specified surface slant is clearly anisotropic for higher slant values (Figures 4b and 4c).

EXPERIMENT 2

Matching the Curvature of Stereo- and Motion-Specified Curved Surface Patches

Method

In this experiment, we used convex parabolic arches with a central curvature of 0.45, 0.68, 0.91, 1.13, and 1.36 cm^{-1} (see Figure 2). As in the slant case, the values were chosen so that all displacements were within D-max. In pilot experiments, we noticed that areas on the surfaces with a zero velocity (this is the case for positions whose simulated distance from the monitor was zero) seemed to separate from the adjoining areas. Since the observers were required to judge the center part of the arches, we raised the arches 2.5 cm so that at least these parts were always moving.

Results: How Matches Depend on the Curvature of the Surface Patch

In Experiment 1, observers matched the slant of a planar motion-specified surface patch with that of a stereo-defined planar surface patch. Here, we studied curvature and use parabolic cylinders. The procedure followed was similar to that of Experiment 1, except that observers were now asked to match the curvature of a motion-specified surface patch with that of a stereo-specified surface patch.

Figure 5a shows matched curvature values when the parabolic cylinder was oriented in the vertical direction—that is, the surface was curved in a horizontal direction. We call this the “curved-horizontal” condition. The matched curvature values of the stereo-specified patches are plotted along the vertical axis as a function of the curvature of the motion-specified patch (horizontal axis). The performance of 3 observers (three different curves) is very sim-

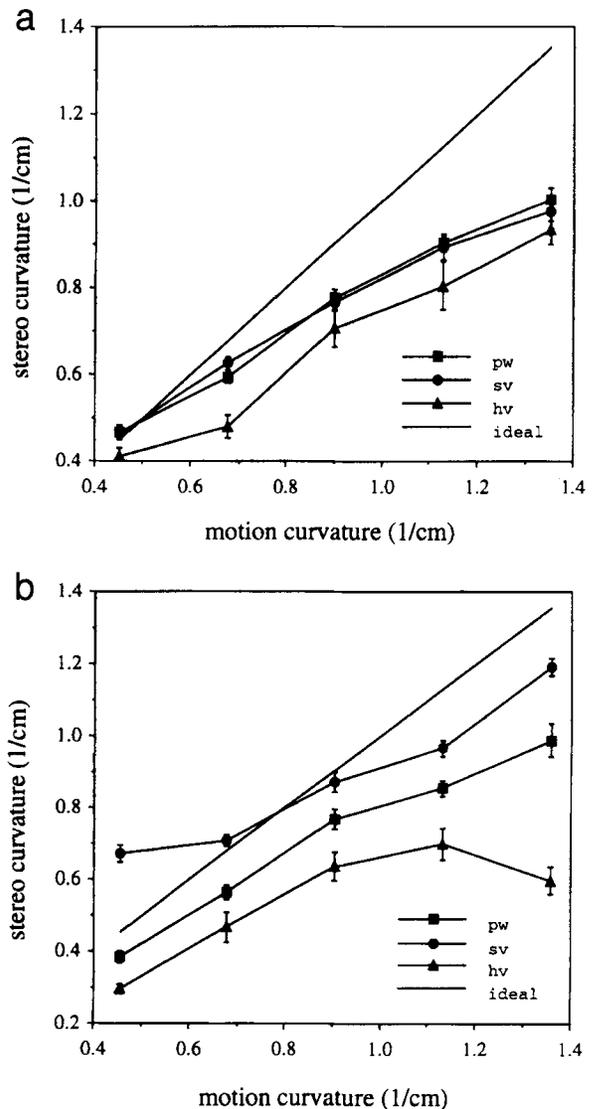


Figure 5. Matched stereo-specified curvature as a function of motion-specified curvature. Along the vertical axis, we have shown the curvature of a stereo-specified surface patch that is judged equivalent to the curvature of a motion-specified surface patch (horizontal axis). The straight line shows the results that would have been obtained had the observers matched dynamically presented displacement fields to the same displacement fields presented in stereo. The three different curves are for three different observers (P.W., S.V., and H.V.). (a) Motion- and stereo-specified surfaces were curved in a horizontal direction. (b) Motion- and stereo-specified surfaces were curved in a vertical direction.

ilar. The straight line gives the results that would be expected if observers judged the curvature of motion- and stereo-specified surface patches as equivalent.

In general, observers matched the curvature of a motion-specified surface with a lower curvature of a stereo-specified surface. This effect can be rather extreme. For example, a motion-specified curvature of 1.35 cm^{-1} was matched with a stereo-defined curvature of 0.97 cm^{-1} (averaged across observers). Matched curvature values

correspond to identically curved patches only for the lowest curvature value measured (0.45 cm^{-1}).

Figure 5b shows matched curvature values when the surface patches were curved in the vertical direction (horizontally oriented parabolic-cylindrical surface patches): the “curved-vertical” condition. As can be observed for the “curved-horizontal” condition, in the “curved-vertical” condition motion-specified surface patches are generally matched with “flatter” stereo-specified surface patches. Although observers performed very similarly in curvature matching when parabolic cylindrical surface patches were curved in the horizontal direction (Figure 5a), performance differed somewhat when the surface patches were curved in the vertical direction. For example, Observers H.V., P.W., and S.V. matched a motion-specified curvature of 1.35 cm^{-1} with stereo-specified curvature values of 0.59, 0.99, and 1.20 cm^{-1} , respectively. For 1 observer, the results for the lowest curvatures measured deviate from this general trend: Observer S.V. matched a motion-specified curvature of 0.45 and 0.68 cm^{-1} with a stereo-specified curvature of 0.68 and 0.71 cm^{-1} , respectively.

Discussion

Generally, the curvature of a motion-specified surface patch is matched with a *lower* curvature of a stereo-specified patch. Roughly speaking, matched stereo-specified curvature varies linearly with motion-specified curvature. The steepness of these approximately linear curves in Figure 5 can be interpreted as the “gain” with which curvature is extracted in the stereo domain relative to such gain in the motion domain. Averaged across observers, this relative gain of curvature extraction is approximately 60% for the “curved-horizontal” and 76% for the “curved-vertical” condition. However, it is clear that the linear relation should break down at lower slant values, since the nonzero intercept would imply a considerable curvature of the frontoparallel plane. Such a high curvature was never observed in the case of *apparent frontoparallel plane tasks* (Helmholtz, 1910), in which only small deviations of a real frontoparallel plane were found.

GENERAL DISCUSSION

Inconsistency of Slant and Curvature Extraction

We asked observers to match quantities of the 3-D environment specified by two images that were presented either simultaneously and binocularly (stereoscopic surface) or sequentially and monocularly (moving surface). We were particularly interested in comparing human performance for matching surface slant and surface curvature. In our experiments, matched stereo-specified *slant* was generally *higher* than motion-defined slant, whereas matched stereo-defined *curvature* was *lower* than motion-defined curvature. From this it follows that in at least one of the modalities the “internal representations” of slant and curvature are not related in a strict mathematical way, since an overestimation of the slant should necessarily lead to an overestimation of curvature if this were the case.

Did the Stimuli Contain Perspective Information?

The observer-object geometry and the metric object representation can be fully specified by two perspective projections of five surface markers in general positions given unlimited precision (Longuet-Higgins, 1981). The same information is specified for two views of an object moving rigidly relative to the observer (except for special movement conditions). A key notion here is the use of perspective information. When the contribution of perspective transformations becomes unmeasurable (small viewing angle, small depth range), we can consider the two images as two parallel projections of the object.

In order to assess the amount of perspective information that was present in our stimuli, we calculated two perspective projections $P_1(x_i, y_i, z_i)$ and $P_2(x_i, y_i, z_i)$ of a given set of points x_i, y_i, z_i ($i = 1 \dots N$, $N = 841$) uniformly distributed across the surface patch. P_1 and P_2 differ from each other, owing to the stochastic rounding process (see the Method section). We also calculated a parallel projection $P'(x_i, y_i, z_i)$ of the same set of points.

As a measure of the difference between two perspective disparity fields we use the averaged squared difference: $\sum_i^N [P_1(x_i, y_i, z_i) - P_2(x_i, y_i, z_i)]^2 / N$. We call this the *internal variance*. Similarly, the averaged squared difference between a perspective and a parallel projection is: $\sum_i^N [P_1(x_i, y_i, z_i) - P'(x_i, y_i, z_i)]^2 / N$, which we call the *external variance*.

If the external variance is not significantly higher than the internal variance, the difference between a parallel projection and a perspective projection will be negligible. For a set of five slanted surfaces with a slant of 30° , we found an average internal variance of 0.62 ± 0.02 and an average external variance of 0.65 ± 0.03 . Hence, our stimuli cannot be distinguished from parallel projections.

Computational Differences Between Motion and Stereo Processing

With only two parallel projections, metric information is lost, yielding a one-parameter family of possible object representations (Koenderink & van Doorn, 1991). Only affine properties are invariant within this one-parameter family (e.g., distance ratios along the same direction). Consequently, the slant and curvature of a surface patch are no longer uniquely determined. They are known only as a function of an unknown parameter: the magnitude of the surface rotation about an axis in the frontoparallel plane, or the *turn*. To reduce this family of solutions for the surface properties to a single surface representation, missing information has to be added or extra assumptions need to be made. For example, one can pick a solution with the smallest surface slant (Koenderink & van Doorn, 1991). Alternatively, one can guess a turn (in motion) or a vergence angle (in stereo) and complete the reconstruction. It is obvious that the extraction of surface properties via such assumptions may no longer be veridical.

When a static scene is viewed, the processing of motion parallax by a moving observer and of stereo information by a static observer are not *necessarily* different: the ob-

server may have access to ego-motion parameters (for the motion domain) or relative eye positions (for the stereo domain) through proprioceptive information.

For a dynamic scene, however, processing in the two domains *does* differ *fundamentally*. That is, in the stereo domain, the observer can still rely on proprioceptive information, whereas in the motion domain, the object-motion parameters (e.g., the turn) are unknown. Therefore, the extraction of 3-D information from the projections of a revolving object necessarily relies on assumptions of the turn (or, equivalently, the slant of the axis of rotation; see Bennett et al., 1989).

We have analyzed two possible strategies of an observer: (1) picking a solution for which the slant (attitude) of the surface patch is nearest frontoparallel, and (2) guessing a turn.

Affine Structure From Two Parallel Projections of a Surface Patch

Here, we describe a method of finding the complete family of consistent representations of a planar surface from two parallel projections. A detailed analysis is given in Koenderink and van Doorn (1991).

In our experiments, the surface transformation between projections is a rotation R around the vertical (y -)axis with a turn ρ . The horizontal axis is labeled the x -axis. The viewing direction is the z -axis. A plane is specified by three arbitrary points on that plane: O (taken to be the origin), $\vec{A} = (A_1, A_2, A_3)$ and $\vec{B} = (B_1, B_2, B_3)$. Let the equation describing this plane be: $z = G_1x + G_2y$. The tilt of the plane is defined as: $\tau = \arctan(G_2/G_1)$. The slant σ is defined as: $\sigma = \arctan(\sqrt{G_1^2 + G_2^2})$. Points \vec{A} and \vec{B} move to \vec{A}' and \vec{B}' under rotation R . The y -components are invariant under R , whereas the z -components are not observable. Thus the information about the slant and tilt of the plane is contained in the change of the projected x -components (which are the x -components themselves in parallel projection):

$$\begin{aligned} A'_1 &= \cos(\rho)A_1 - \sin(\rho)(A_1G_1 + A_2G_2) \\ B'_1 &= \cos(\rho)B_1 - \sin(\rho)(B_1G_1 + B_2G_2). \end{aligned} \quad (1)$$

The equations for other points on the plane are linear combinations of Equation 1 (more points do not add information!). The solution of G_1, G_2 is given by

$$\begin{pmatrix} G_1 \\ G_2 \end{pmatrix} = \begin{pmatrix} A_1 & A_2 \\ B_1 & B_2 \end{pmatrix}^{-1} \begin{pmatrix} A_1/\tan(\rho) - A'_1/\sin(\rho) \\ B_1/\tan(\rho) - B'_1/\sin(\rho) \end{pmatrix}. \quad (2)$$

G_1 and G_2 , and thus the slant and tilt of the surface, are a function of the unknown turn ρ . Generally, one obtains a range of possible orientations of the tilt and a lower bound on the slant. Two points are worth mentioning here: (1) the dependence of slant on the turn is independent of the particular choice of pair A and B as long as these pairs are in the same plane, and (2) the relation between slant and turn is approximately inverse. This predicted relation is supported by psychophysical evidence (van Veen & Werkhoven, 1994).

Minimum Slant Solutions

We have computed the minimum slant solutions for three stimuli presented in the motion domain: a vertically slanted patch with a slant value of 15° , 22.5° , or 30° revolving about the vertical axis with a turn of 6° between frames. Following Koenderink and van Doorn (1991), the minimum slant values for these stimuli are $\sigma = 14.0^\circ$, 11.9° , and 9.6° , respectively. The corresponding turns are $\rho = 19.7^\circ$, 16.8° , and 13.5° between frames, respectively. When the surface patches are slanted in a horizontal direction, the minimum slant values are all zero; that is, a frontoparallel patch is a possible solution.

Assuming that the differences between slant judgments in motion and stereo are due to the unknown turn in the motion domain, the minimum slant solution is clearly inconsistent with the results presented in Figure 3.

Guessing the Turn

The family of possible 3-D representations of a surface patch is reduced to a unique representation when the turn ρ is guessed. Conversely, given the judged 3-D properties (slant, curvature) of the surface patch, one can calculate the corresponding turn that was guessed by the observer.

Following Koenderink and van Doorn (1991), we have calculated the corresponding turn from the judged slant values of planar patches. That is, we have solved Equation 2 for ρ , given the σ' judged by the observers. For example, a patch with a vertical slant $\sigma = 30^\circ$ that is judged as $\sigma' = 35.1^\circ$ (see Figure 3a, averaged across observers) corresponds to a guessed turn $\rho' = 4.9^\circ$ between frames (note that the simulated turn was $\rho = 6^\circ$ between frames). We have averaged the judged slant values presented in Figures 3a (vertically slanted patches) for other slant values: the calculated guessed turns are $\rho' = 4.29^\circ$ and 4.9° between frames for $\sigma = 15^\circ$ and 22.5° , respectively. Furthermore, for horizontally slanted patches (see Figure 3b), the guessed turn is $\rho' = 4.5^\circ$, 3.7° , and 4.6° between frames for slant values $\sigma = 15^\circ$, 22.5° , and 30° , respectively.

We have also calculated the guessed turn for the judged curvature values of the curved patches. This calculation is based on a consistent relation between local slant and the curvature of a patch. When viewed along the y -axis, the vertical cylinders used in Experiment 2 are exact parabolas with a curvature c at the fixation point. Physical slant values σ vary along the position x along the x -axis: $\sigma(x; c) = \arctan(cx)$, where c is the physical curvature. We now assume a consistent relation between the extraction of slant and curvature: The judged slant values σ' vary along the position x along the x -axis: $\sigma' = \arctan(c'x)$, where c' is the judged curvature. We have calculated the guessed turn that maps σ to σ' as a function of x and c . The guessed turn varies only slightly for these conditions and is described by a distribution with a mean $\rho' = 6.32^\circ$ between frames with a standard deviation of 0.63° between frames.

In conclusion, our data are consistent with a model in which perspective information is not used (parallel pro-

jections) and in which the judgment of surface properties (where curvature is the spatial derivative of slant) is determined by a systematic underestimation by a factor of 0.6–0.8 of the turn for planar patches and a small and insignificant overestimation for cylindrical patches. However, this does not necessarily mean that in the case of curvature the turn is guessed correctly, since the turn was calculated on the basis of a stereo-specified slant that did not have to be veridical. It might be interesting to know what happens to the turn in the curvature case if, for instance, we double the viewing distance. This could show whether the turn estimate was fixed or whether it was correlated with vergence or other factors such as accommodation.

The preceding calculations suggest that the estimates of turn are fairly constant (but they do not exclude the possibility that they covary with the vergence estimate). This finding contrasts strongly with the results of Todd and Bresnan (1990). In their experiment, a metric task (comparing the length of two nonparallel line elements) was performed. The results for two-frame motion were best explained by assuming that the turn was selected *at random* upon each presentation. Recent research of LITER, BRAUNSTEIN, and HOFFMAN (1993), however, has shown that the turn estimate is not random, and that it depends on certain aspects of the stimulus, such as velocity differences after curl is removed.

In the analysis above, we have assumed a consistent relation between slant judgments and curvature judgments; that is, the curvature at the fixation point of our parabolic cylinder is extracted by taking the local spatial derivative of slant values across the surface. Under this assumption, the empirical inconsistency between matched slant and curvature values leads to the conclusion that the guessed turn or magnitude of rotation depends on the shape of the surface patch.

Alternatively, we can assume that the guessed turn is independent of the shape of the surface patch and that curvature extraction and slant extraction are simply inconsistent. That is, our visual system does not extract curvature by spatially differentiating slant values across the surface patch (or does not extract slant by spatially integrating curvature). Such a conclusion suggests different task-specific mechanisms for slant and curvature extraction. Since we have matched stereo- and motion-specified slant and curvature in our experiments—that is, we have made *relative* judgments—such inconsistency between the processes of slant and curvature extraction within a domain can occur in either the stereo-domain or the motion domain, or in both.

CONCLUSIONS

The results of the matching experiments concerning stereo- and motion-specified slant and curvature described in this paper lead to the following conclusions: (1) The judged slant of a surface patch is strongly biased by its projected contour. (2) Slant and curvature matches are not consistent: motion-specified slant is *overestimated* in terms of stereo-specified slant, whereas curvature is *underestimated*.

This may be due to (a) the dependence of the guessing of the missing information (the turn) on the shape of a motion-specified surface (planar vs. parabolic-cylindrical), or (b) a nonconsistent relation between the coding of mechanisms tuned to curvature and those tuned to slant.

REFERENCES

- ALOIMONOS, J., & BROWN, C. M. (1989). On the kinetic depth effect. *Biological Cybernetics*, **60**, 445-455.
- BENNETT, B. M., HOFFMAN, D. D., NICOLA, J. E., & PRAKASH, C. (1989). Structure from orthographic views of rigid motion. *Journal of the Optical Society of America A*, **6**, 1052-1069.
- BRADDICK, O. J. (1974). A short range process in apparent motion. *Vision Research*, **14**, 519-527.
- CUMMING, B. G., JOHNSTON, E. B., & PARKER, A. J. (1991). Vertical disparities and perception of three-dimensional shape. *Nature*, **349**, 411-413.
- DE VRIES, S. C., KAPPERS, A. M. L., & KOENDERINK, J. J. (1993). Shape from stereo: A systematic approach using quadratic surfaces. *Perception & Psychophysics*, **53**, 71-80.
- DE VRIES, S. C., KAPPERS, A. M. L., & KOENDERINK, J. J. (1994). Influence of surface attitude and curvature scaling on discrimination of binocularly presented quadratic shapes. *Vision Research*, **34**, 2409-2423.
- GIBSON, E. J., GIBSON, J. J., SMITH, O. W., & FLOCK, H. (1959). Motion parallax as a determinant of perceived depth. *Journal of Experimental Psychology*, **58**, 40-51.
- GRAHAM, M., & ROGERS, B. J. (1982). Simultaneous successive contrast effects in the perception of depth from motion-parallax and stereoscopic information. *Perception*, **11**, 247-262.
- HELMHOLTZ, H. VON (1910). *Handbuch der physiologischen Optik*. Hamburg: Voss. [(1962). *Treatise on physiological optics* (J. P. Southall, Trans.). New York: Dover.]
- HOFFMAN, D. D., & BENNETT, B. M. (1986). The computation of structure from fixed-axis motion: Rigid structures. *Biological Cybernetics*, **54**, 71-83.
- HOFSTEN, C. (1976). The role of convergence in visual space perception. *Vision Research*, **16**, 193-198.
- JOHNSTON, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research*, **31**, 1351-1360.
- KOENDERINK, J. J. (1986). Optic flow. *Vision Research*, **26**, 161-180.
- KOENDERINK, J. J., & VAN DOORN, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A*, **8**, 377-385.
- KONTSEVICH, L. L. (1993). Pairwise comparison technique: A simple solution for depth reconstruction. *Journal of the Optical Society of America A*, **10**, 1129-1135.
- LITER, J. C., BRAUNSTEIN, M. L., & HOFFMAN, D. D. (1993). Inferring structure from motion in two-view and multiview displays. *Perception*, **22**, 1441-1465.
- LONGUET-HIGGINS, H. C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature*, **293**, 133-135.
- LONGUET-HIGGINS, H. C., & PRAZDNY, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London: Series B*, **208**, 385-387.
- METZGER, W. (1934). Tiefenerscheinungen in optischen Bewegungsfeldern [Perception of depth in optical motion fields]. *Psychologische Forschung*, **20**, 195-260.
- NAWROT, M., & BLAKE, R. (1989, May 12). Neural integration of information specifying structure from stereopsis and motion. *Science*, **244**, 716-718.
- NORMAN, J. F., TODD, J. T., TITTLE, J. S., & PEROTTI, V. J. (1993). Directional anisotropies determine the perceived structure of contradictory motion and stereo disparity fields [Abstract]. *Investigative Ophthalmology & Visual Science*, **34**, 1082.
- RITTER, M. (1977). Effect of disparity and viewing distance on perceived depth. *Perception & Psychophysics*, **22**, 400-407.
- RITTER, M. (1979). Perception of depth: Processing of simple positional disparity as a function of viewing distance. *Perception & Psychophysics*, **25**, 209-214.

- ROACH, J. W., & AGGARWAL, J. K. (1979). Computer tracking of objects moving in space. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **1**, 127-135.
- ROACH, J. W., & AGGARWAL, J. K. (1980). Determining the movement of objects based on visual observations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **2**, 554-562.
- ROGERS, B. J., & COLLETT, T. S. (1989). The appearance of surfaces specified by motion parallax and binocular disparity. *Quarterly Journal of Experimental Psychology*, **41**, 697-717.
- ROGERS, B. J., & GRAHAM, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research*, **22**, 261-270.
- ROGERS, B. J., & GRAHAM, M. (1983, September 30). Anisotropies in the perception of three-dimensional surfaces. *Science*, **221**, 1409-1411.
- STEVENS, K. A. (1983). Slant-tilt: The visual encoding of surface orientation. *Biological Cybernetics*, **46**, 183-195.
- SUBBARAO, M. (1988). *Interpretation of visual motion: A computational study*. London: Pitman.
- TODD, J. T., & BRESSAN, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics*, **48**, 419-430.
- VAN VEEN, H. A. H. C., & WERKHOVEN, P. (1994). *Metamerisms in structure from motion*. Manuscript submitted for publication.
- WALLACH, H., & O'CONNELL, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, **45**, 205-217.
- YOUNGS, W. M. (1976). The influence of perspective and disparity cues on the perception of slant. *Vision Research*, **16**, 79-82.

APPENDIX

When we use a stereogram to generate a motion sequence, we simulate a monocular observer moving from the left-eye position to the right-eye position, or alternatively, a rotation and translation of the scene. During this movement, the angle between the surface normal of the projection screen and the line of sight changes somewhat. When we now project both images on the screen for the observer to view them monocularly and stationary from, for instance, the left-eye position, we introduce a small distortion. This is clear from Figure A1.

Figure A1 (panel a) shows the construction of the stereogram. The dots on the projection screen symbolize the two images that will be shown sequentially. Panel b shows the left eye viewing the right eye's image; the angles are slightly different when compared with the original angles from panel a, owing to small differences in distances. To calculate the influence of these differences approximately, we examine the reconstruction made by an observer with knowledge of the angle between both views. We

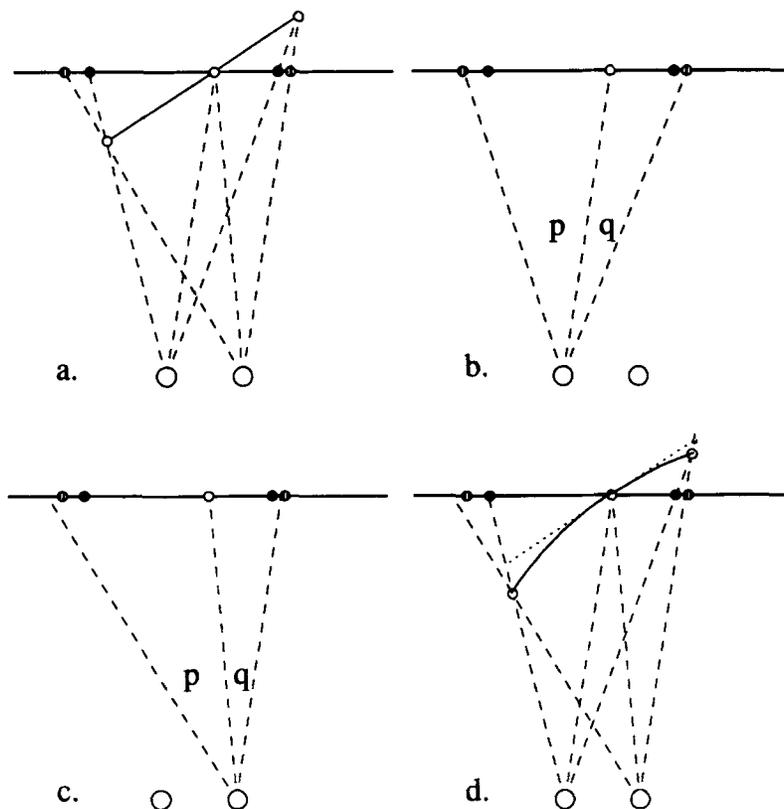


Figure A1. Possible reconstruction of the scene when a stereogram is viewed as an oscillatory two-frame motion sequence. Panel a shows the construction of the two images of the stereo pair. In a motion sequence the left eye views the right eye's image (panel b) as well as its own—left eye's—image. A system knowing the original angle between the two images could then transport angles p and q to the right eye (panel c) and make the reconstruction, as in (panel d). The dotted line shows the original slanted line. Interocular distance and size of the stimulus are both magnified by a factor of about three. The distortions in the real setup are about 10 times smaller. Notice that the slant at the fixation point is not distorted.

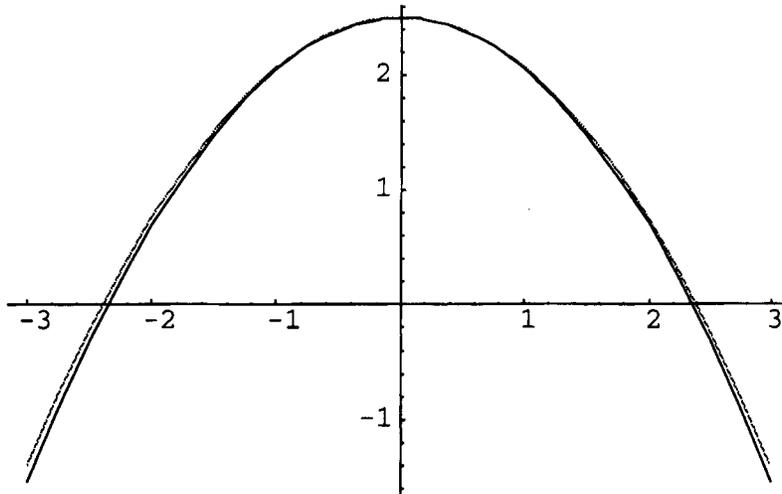


Figure A2. Influence of incorrect projection on a curved surface. Continuous line, original shape; dashed line, distorted shape.

transport the angles p and q from the left eye to the right eye (panel c) and obtain a new solution (panel d). The real distortion is smaller than that shown in the figure. If we divide the plane in half, we find slant to be decreased by about 2° in the left half and increased by about the same amount in the right half. The slant of a straight line fitted to the distorted line deviates only slightly (ca. 0.2°) from the original. In slant detection experiments, we have found average thresholds of about 4° . It is therefore reasonable to assume that the distortions are negligible.

The distortion of a curved object can be seen in Figure A2. This is a parabolic arch with curvature 0.9 cm^{-1} and a distorted curvature of 0.87 cm^{-1} . Again, the distortion is fairly small, very much smaller than the effect found in Experiment 2.

(Manuscript received May 12, 1994;
revision accepted for publication April 19, 1995.)