

Extraction of relief from visual motion

PETER WERKHOVEN and HENDRIK A. H. C. VAN VEEN
Utrecht University, Utrecht, the Netherlands

We quantified the ability of human subjects to discriminate the relative distance of two points from a slanted plane when viewing the projected velocities of this scene (orthographic projection). The relative distance from a plane (called relief) is a 3-D property that is invariant under linear (affine) transformations. As such, relief can *in principle* be extracted from the instantaneous projected velocity field; a metric representation, which requires the extraction of visual acceleration, is not required. The stimulus consisted of a slanted plane P (specified by three points) and two points Q_1 and Q_2 that are non-coplanar with P . This configuration of points oscillated rigidly around the vertical axis. We have measured the *systematic error* and *accuracy* with which human subjects estimate the relative distance of points Q_1 and Q_2 from plane P as a function of the slant of P . The systematic error varies with slant: it is low for small slant values, reaches a maximum for medium slant values, and drops again for high slant values. The accuracy covaries with the systematic error and is thus high for small and large slant values and low for medium slant values. These results are successfully modeled by a simple relief-from-motion computation based on local estimates of projected velocities. The data are well predicted by assuming (1) a measurement error in velocity estimation that varies proportionally to velocity (Weber's law) and (2) an eccentricity-dependent underestimation of velocity.

Structure-from-motion theories are concerned with the reconstruction of the three-dimensional (3-D) spatial structure of an object as specified by the information contained in its dynamic projection (projected positions and their temporal derivatives, such as velocities and accelerations, etc.). An analysis by Ullman (1979), for example, shows that *metric* properties of an object moving relative to the subject can generally be extracted from as few as two distinct *perspective* projections of five fiducial points of the object (for a detailed mathematical analysis and review, see Kanatani, 1993). The amount of structural information that can be extracted from two projections is reduced, however, when perspective distortions are small with respect to the subject's accuracy of processing the projected positions and displacements of points. For small viewing angles and small depth ranges, the condition of parallel projection is approached (Ullman, 1979).

Koenderink and van Doorn (1991) elegantly stratified the structure-from-motion problem for parallel projections. They showed that two successive projections of four points of an object—that is, their projected positions and velocities—allow for an affine representation of the object (even when the object is not rigid, but deforms linearly in 3-space while moving). An affine representation of a 3-D object does not specify the absolute distances between parts of the object, but specifies the 3-D properties that are invariant under linear transformations of

the object in space. More specifically, it specifies informative properties such as coplanarity and distance ratios along the same direction. A full metric representation of the object would require the rigidity assumption and the processing of a third projection of the object (Koenderink & van Doorn, 1991; Ullman, 1979). Three successive projections of the object specify the positions, the velocities, and the *acceleration* of the fiducial points.

These theories describe the minimal requirements for extracting certain 3-D properties from a dynamic scene. It is still an open question if and with what type of operations such algorithms are implemented in the human visual system. To what extent is the ability to judge metric properties from optic flow limited by the accuracy of position, velocity, and acceleration estimates? Recent literature has shown that human subjects perform poorly in visual motion tasks in which they have to judge metric properties of 3-D objects (Norman & Todd, 1993; Todd & Bressan, 1990; Todd & Norman, 1991; Todd, Norman, Perotti, & Tittle, 1993) from three or more projections, although human observers generally perform well in discriminating between nonrigid and fully rigid motion (Braunstein, Hoffman, & Pollick, 1990). Taken that the stimuli used sufficiently specify the position and velocity information for the human visual system, Todd and Bressan (1990) conclude that performance of metric tasks is limited by the inability to process the information added by the third projection: optic acceleration. Such conclusions are supported by findings that the human visual system performs badly in acceleration detection tasks (Snippe & Werkhoven, 1993; Snowden & Braddick, 1991; Werkhoven, Snippe, & Toet, 1992).

The inability to judge metric properties in structure-from-motion tasks suggests that human subjects work

The present research was undertaken at the Buys Ballot Laboratory of the Utrecht Biophysics research institute, Utrecht University. Correspondence concerning this article should be addressed to P. Werkhoven, who is now with the TNO Human Factors Research Institute, Kampweg 5, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands (e-mail: wekh@tm.tno.nl).

with weaker (e.g., affine) representations of environmental objects. Affine representations may still contain sufficient information for successful performance of a variety of tasks in a 3-D environment based on affine invariant properties (e.g., alignment tasks). Moreover, a priori partial knowledge about metric object properties may settle the calibration in such a way that a metric representation is obtained.

Our aim in this paper is to study the ability of human subjects to extract affine properties in structure-from-motion tasks. More specifically, we studied subjects' ability to discriminate the distance ratio of two points with respect to a plane that was slanted with respect to the viewing direction. We call this *relief*. To extract relief, the extraction of acceleration is not needed and two-frame motion sequences (two projections) are theoretically sufficient.

We sought to answer the following specific questions: (1) How accurately is relief information extracted in rich (over-specified) conditions: a densely sampled sequence of projections of the dynamic scene? (2) Can the visual system extract relief information in the minimal condition of "two-frame" motion sequences? And (3) Do the characteristics of "low-level" velocity extraction determine those of "high-level" relief extraction?

METHOD

Subjects viewed the projection of five points in 3-D space (see Figure 1). Three points p_i ($i = 1 \dots 3$; see the filled circles in Figure 1) were the projections of points on a plane P . Two points q_1 and q_2 (see the open circles in Figure 1) were projections of points out of plane P at 3-D distances d_1 and d_2 from the plane. This configuration oscillated rigidly around the vertical axis. Subjects in-

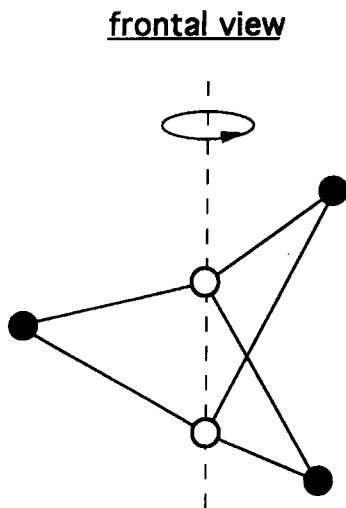


Figure 1. A sketch of the frontal view of the configuration of points. The filled circles are the projections of the points that specify plane P . The open circles are projections of the points that are out of plane P . Points are connected with solid lines as sketched. The configuration oscillates rigidly around the vertical axis (dashed line, not shown during the experiment). This frontal view is the *average attitude* of the configuration during viewing.

dicated which of the two points Q_1 and Q_2 stood out farther from plane P . Details are presented next.

Apparatus

The stimuli were generated on a Macintosh IIfx computer with a GS/C video board driving a 71-Hz Radius TPD/19 monitor. The screen dimensions were 35.6×26.9 cm, with $1,152 \times 882$ pixels. The stimuli were displayed at high contrast.

Stimulus

The stimulus was a projection of five points of an object that rotated rigidly in 3-D space about a vertical axis. The background of the projection was dark. Three points P_1 , P_2 , and P_3 were shown as light filled circles and specified a plane P . Two points Q_1 and Q_2 (the probes) were shown as light open circles and had 3-D distances d_1 and d_2 from the plane P . Each probe Q_i was connected with all points P_i by light lines, except in one experiment in which we studied the influence of the type of graphical presentation of the configuration.

The projections q_1 and q_2 of probes Q_1 and Q_2 were near the center of the projected triangle specified by the projections p_i of P_i (see Figure 1). The exact position of these projected points was determined by stochastic variables. We will first describe the "generating" configuration C_0 of five points in arbitrary dimensionless units. We take the 2-D projection plane to be the x, y plane and the viewing distance to be the z -axis. The coordinates (x, y, z) of points Q_i in C_0 are: $Q_1 = (0, y_1, d_1)$ and $Q_2 = (0, y_2, R d_1)$, with $d_1 = [0.2 \dots 0.4]$, $y_1 = [-0.3 \dots -0.1]$, $y_2 = [0.1 \dots 0.3]$, and R the depth ratio of probes Q_2 and Q_1 relative to the plane. The notation $[a \dots b]$ indicates a stochastic variable with a uniform distribution between values a and b .

The points P_i of C_0 are randomly positioned at equal distances on a circle in the frontoparallel plane with its center at the origin and a diameter of 2 followed by a linear transformation of the form

$$D = \begin{pmatrix} 1+\epsilon & \epsilon & 0 \\ \epsilon & 1+\epsilon & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (1)$$

which varies the shape of the triangle ($\epsilon = [-0.2 \dots 0.2]$) specified by P_i .

We simply get the simulated configuration C by (1) rotating all points about the x -axis by an angle equal to the desired slant of the plane P , followed by (2) a translation along the z -axis with a magnitude d_0 , where $d_0 = -[0.5 \dots 1.5] d_1 (1+R)/2$, and by (3) scaling (multiplying all coordinates by a factor of $S = 160$ pixels).

Rotation (1) determines the slant of the configuration and thus the slant of plane P . Note that the tilt direction of plane P is always downward (see Figure 2A). Depth translation (2) causes the averaged projected velocities of the points Q_i to be zero, so that the use of simple velocity cues in specific conditions is discouraged. The scaling (3) determines the size of the desired configuration.

Configuration C oscillated about a rotation axis Ω . In image I_n , the *angle of rotation* was $[2(n-1)/(N-1)-1]\rho$, where $2N$ is the number of frames in a single cycle of the oscillation, n is the frame index ranging from 1 to N , and ρ is the oscillation amplitude. The images were shown consecutively in time. The number of Vsync pulses (time segments of 14.1 msec) that pass while one frame is presented (stimulus onset asynchrony, or SOA) determines the speed of rotation. The number of cycles of oscillations determines the total presentation time of the stimulus.

In *smooth* motion conditions, the oscillation was densely sampled: two cycles were shown with $N = 16$, SOA = 42 msec, and $\rho = 45^\circ$. In *two-frame* motion conditions, we showed four cycles of oscillation with $N = 2$. The SOA and values of ρ were chosen so that performance would be optimal. These values are given in the Results sections.

The slant of a plane P is usually defined as the angle between the normal of plane P and the viewing direction. Following this definition, the slant of P varies slightly when the plane rotates forth and back around the vertical axis. As may be clear from the description above, the term *slant* as used in this paper is the slant value of P at the time when the angle of rotation is zero degrees.

Subjects viewed the stimulus from a distance of 100 cm, monocularly with natural pupils in a room with dimmed light. At this viewing distance, the size of a pixel was 1.06×1.05 arcmin. Consequently, the projected points p_i had an average eccentricity of 2.8° (for a frontoparallel plane P). With a projected diameter of points p_i and q_i of 6 pixels (or 6.3 arcmin), all points were clearly resolvable.

Procedure

Subjects were instructed to fixate at the center of the configuration of points. A fixation point was presented between trials but not during the rotation of the configuration, to avoid perceptual interactions with the points of the stimulus.

The task of the subject was to indicate which of the points Q_i (open circles) had the largest distance to the plane specified by the points P_i (filled disks). A "down" answer indicates that Q_2 (always the lower dot in Figure 2) was perceptually farther away from plane P than Q_1 (the upper dot).

For each experimental condition, we measured the probability of a "down" answer as a function of the ratio R of the distances of Q_2 and Q_1 to the plane. An adaptive psychometric procedure was used

to estimate two parameters that characterize the psychometric function—that is, the probability of a "down" answer as a function of the ratio R . These two parameters are the *point of subjective equality* μ , and *threshold* σ . The point of subjective equality μ (sometimes shortened to PSE) is defined as the objective ratio R for which the distance of point Q_2 from plane P is subjectively equal to that of Q_1 . At this point of subjective equality μ , the probability of "down" answers is 50%. The *threshold* σ (also called *discrimination threshold*) is defined as the increment relative to the point of subjective equality μ for which the percentage of "down" answers is 84%.

The psychometric function is modeled by an error function Erf: the probability of "down" answers $P(R)$ as a function of ratio $R = d_2/d_1$ is

$$P(R) = \text{Erf} \left(\frac{R - \mu}{\sigma} \right). \quad (2)$$

For each experimental condition, we presented a set of 70 stimuli to determine the point of subjective equality μ and threshold σ . After each trial in a *set*, maximum likelihood estimates of μ and σ were calculated. A next trial was placed at the current most probable estimate of $\mu + \sigma$ or $\mu - \sigma$ (with equal probability). The values of μ and σ for each condition presented in this paper are the average of the results for six such sets. The error bars shown in the graphs are the standard deviation of the distribution of the averaged six values.

The trials of sets for different experimental conditions were mixed in several sessions.

Subjects

Three subjects participated in the experiments: both authors (P.W. and H.V.) and S.P. P.W. and S.P. were emmetropic. H.V. was myopic, but spectacle corrections were used throughout. All subjects were experienced in psychophysical experiments. P.W. and H.V. were aware of the objectives of research; Subject S.P. was not.

EXPERIMENT 1 Smooth Motion

In this experiment, we measured the ability to extract relief information (the distance ratio of the two points Q_1 and Q_2) relative to plane P specified by points P_i for slanted configurations. In smooth motion conditions, the configuration oscillated smoothly ($N = 16$, SOA = 42 msec) around the vertical axis.

Before measuring the dependence of performance on the slant of the configuration, we studied the influence of the type of graphical presentation on the point of subjective equality μ (or PSE) and the threshold σ . We did not want our results to depend on specific graphical properties of our stimulus.

Dependence on Graphical Presentation

There are reasons to believe that the type of graphical presentation of a moving object can affect its perceived motion (Kaiser & Calderone, 1991; Werkhoven & Koenderink, 1993) and structure (Todd, Akerstrom, Reichel, & Hayes, 1988). In fact, Todd et al. (1988) showed that moving configurations of connected lines appear to work better than configurations of isolated points. Therefore, we have tried connecting lines and shaded planes to possibly enhance the percept of a coherently moving 3-D configuration of dots. The normal components of the ve-

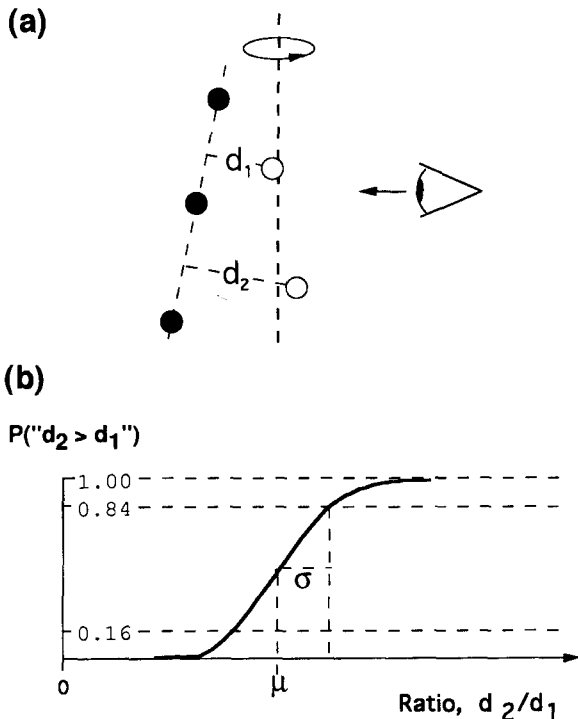


Figure 2. (a) A sketch of the configuration of points. The filled circles are the projections of the points P_i that specify plane P . The distances of Q_1 and Q_2 from plane P are d_1 and d_2 , respectively. In this side-view projection, the slant of plane P is defined as the angle between P (the dashed line through points P_i) and the axis of rotation (the vertical dashed line). The tilt direction of plane P is always downwards. (b) A model of the probability $P(d_2 > d_1)$ that a subject indicates that Q_2 was further from plane P than Q_1 as a function of the distance ratio d_2/d_1 . This psychometric curve is a cumulative normal distribution parameterized by its PSE μ and its threshold (reciprocal steepness) σ .

locities of points along the contours made visible by these cues depend uniquely on the velocities of points P_i and Q_i . Thus, in principle, these added cues do *not* add information about the structure of the configuration: structural properties are fully specified by the projected velocities of points P_i and Q_i *only*. In fact, Rubin, Solomon, and Hochstein (in press) have shown (both theoretically and psychophysically) that replacing all connecting line segments by infinitely long lines makes structure-from-motion tasks generally impossible. However, additional cues may serve spatial integration or segmentation processes that bias the extraction of structural properties such as relief.

Here, we measure the PSE μ and the threshold σ for the extraction of the relative distances of the probes Q_1 and Q_2 from plane P for different types of graphical presentations as shown in Figure 3.

The projected probes q_i and projected points p_i of plane P always moved as described in the Method section. The types of presentation differed only in the way

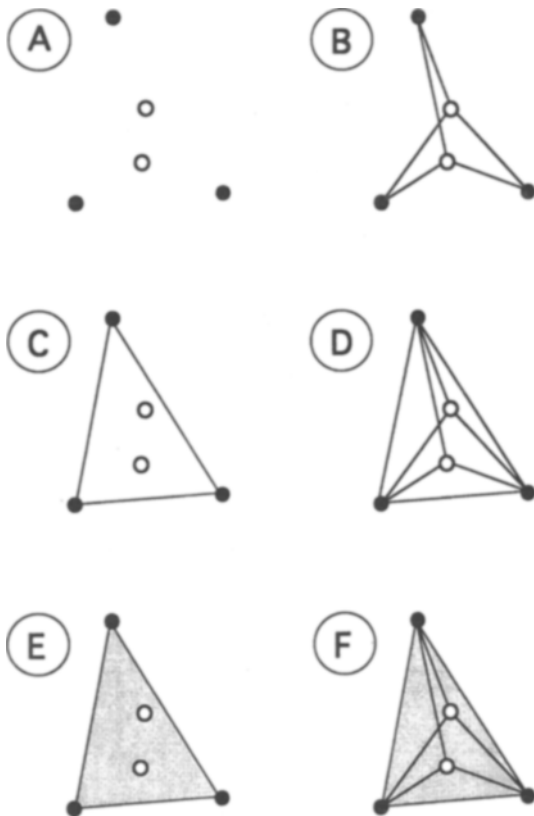


Figure 3. Types of graphical presentation of the configuration of points. The actual presentation is in reversed contrast. The following description is in terms of lightness as seen on our displays. Type A: points P_i are shown as light filled circles on a dark background. Points Q_i are shown as light open circles. Type C: as type A, but now points P_i are interconnected with light lines. Type E: as type C, but the area enclosed by points P_i has a luminance value which is the average of the luminance value of the light circles and the background. Types B, D, and F (right column) are similar to types A, C, and E, respectively, but in addition, each point Q_i is connected with all points P_i by light lines.

that plane P was drawn and the way that the probes q_i were connected to points p_i . Note, however, that all presentations have in common that only five identifiable points (the corners of the plane and the probes) have unique projected velocities. The slant of plane P was random across trials and was uniformly distributed between 0° and 45° . Trials with different types of graphical presentations were mixed in one session.

Results. Figure 4 shows the results (PSE μ and threshold σ) for different types of presentation and different subjects.

First, performance is generally characterized by a PSE larger than one (except in the case of Subject H.V., type F). Second, the PSE values (Figure 4A) are lower (averaged across subjects) for types where points Q_i are connected with points P_i (compare the results of type B with A, those of D with C, and those of F with E). Introspectively, the connecting lines of types B, D, and F enhance the percept of coherency; that is, points P_i and Q_i more strongly appear to belong to the same 3-D object. Third, the PSE μ generally decreases with the number of cues (contour, shading) that suggest that the points P_i belong to the same 3-D object (e.g., compare the results for set A,B with those for set E,F).

Although PSE values are affected by adding connecting lines between points P_i and filling the enclosed area, the averaged thresholds (Figure 4B) are not.

Discussion. A configuration of type B emphasizes the role of the local velocities of points P_i (the connecting lines between them are omitted), while appearing more coherent than type A (because points Q_i are connected with points P_i). We used this type of graphical presentations for the next experiments, in which we examined the dependence of μ and σ on the slant of plane P .

Dependence on Slant

To measure the point of subjective equality μ and threshold σ as a function of the slant of plane P , we had mixed trials with different slant values in one session. The type of presentation was always type B of Figure 3.

Results. In Figure 5, we present the PSE μ (Figure 5A) and the ratio threshold σ (Figure 5B) as a function of the slant of the configuration.

For a frontoparallel relief (slant 0°), human subjects are quite accurate in discriminating the distances of two points relative to a plane: $\sigma = 0.11$ (Figure 5B, averaged across subjects). Furthermore, points at objectively equal distances to the plane are indeed perceived as having equal distances to the plane: the PSE $\mu = 1.00$ (Figure 5A, averaged across subjects). The zero bias in this frontoparallel condition shows that subjects had no preference to see the upper or lower point farther from plane P .

The accuracy in extracting relief becomes progressively worse when the slant of plane P is increased from 0° to 45° . At a slant of 45° , the averaged threshold value σ is 0.25. Furthermore, the averaged PSE μ is 1.27.

When the slant of the configuration is increased farther, to 80° , both the PSE μ and the threshold ratio σ decrease again.

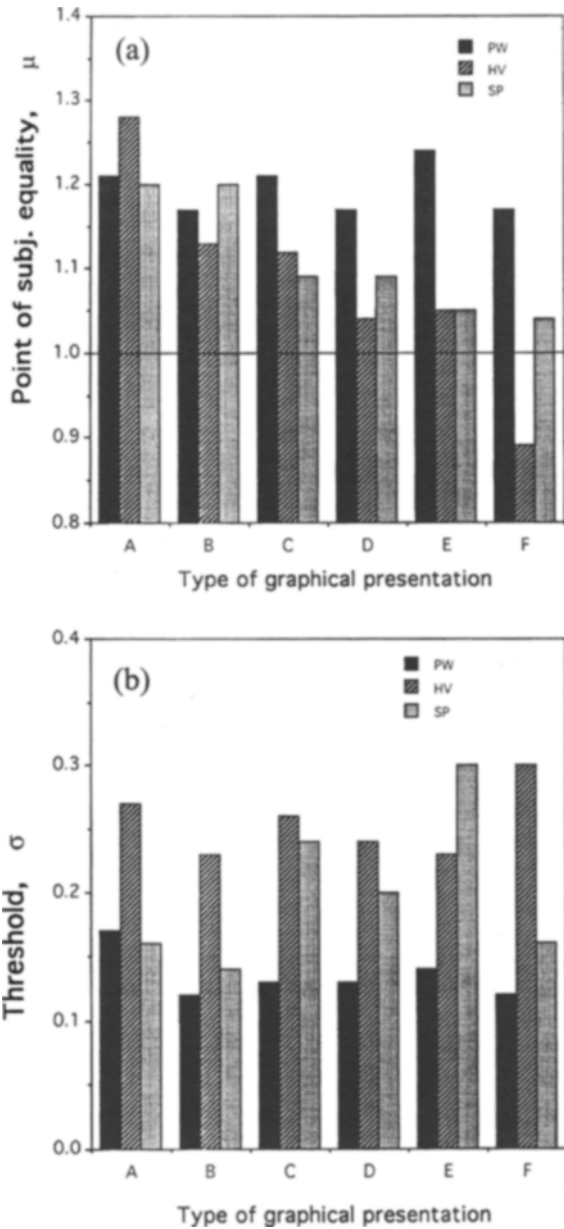


Figure 4. Relief-from-motion results for different graphical presentations of the configuration. The slant of plane P is random across trials and is uniformly distributed between 0° and 45° . Types of presentation A, B, C, D, E, and F correspond to those shown in Figure 3. For each type, the PSE μ and the threshold σ are given in (a) and (b), respectively, for different subjects.

At a slant of 90° , the viewing direction is tangential to the plane and the distance-ratio discrimination task becomes a 2-D task: discriminating the difference in projected positions of the two points Q_1 and Q_2 in a direction orthogonal to their velocity direction. Although the points oscillate in the projection, such position difference is easily determined visually. At 90° slant, the average distance of the projections q_i of the points Q_i relative to the projection of the plane (which is a line now)

is 48 pixel units. A difference of 1 pixel yields a smallest ratio increment of 0.02. A control experiment has shown that this smallest ratio increment can still be discriminated with an accuracy of 100% correct. This upper limit of 0.02 for the thresholds at 90° slant is also shown in Figure 5A. The PSE becomes 1, for this condition.

Interpretation of the point of subjective equality. Any computation that comes up with a 3-D representation of the configuration that is a linear deformation of the real stimulus configuration leaves distance ratios along the

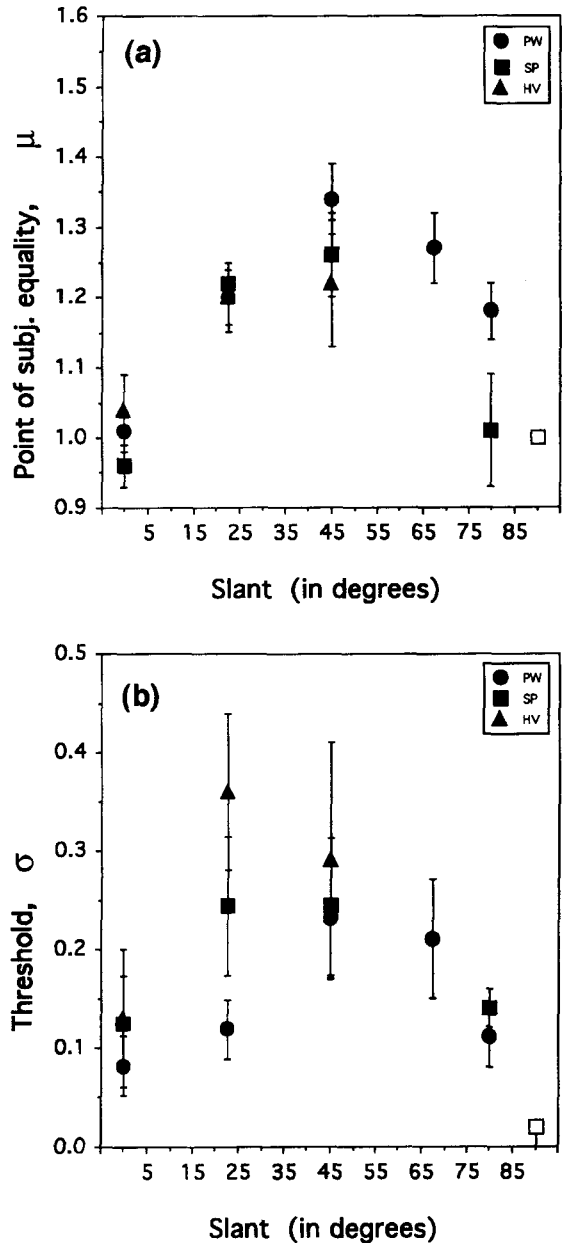


Figure 5. Relief extraction in smooth motion conditions ($N = 16$, $SOA = 42$ msec). Point of subjective equality μ (a) and ratio threshold σ (b) as a function of the slant of the configuration. Results are shown for 3 subjects. The error bars indicate the standard deviation in the measurements.

same direction invariant. Hence, μ is still expected to equal 1 for such computations. A consequence of this observation is that, when $\mu \neq 1$, the visual system necessarily builds up a representation that is a *nonlinear* deformation of the real 3-D stimulus configuration. Such transformations include partial deformations. One hypothesis is that the slant of plane P could be underestimated, whereas the 3-D positions of points Q_1 and Q_2 are correctly extracted. According to this hypothesis, the measured μ values can be directly translated into an underestimation of the slant of plane P by an angle φ , given the average configuration parameters described in the Method section: $\varphi = \arctan[0.75(\mu - 1)]$. For example, for a PSE of $\mu = 1.4$, this underestimation φ becomes 16.7° . Largest underestimations occur for medium slant values (45°). The question, however, *why* the slant of plane P is underestimated still remains unanswered. Furthermore, one may raise the question of whether a nonlinearly deformed representation of the configuration can appear rigid over time (as was reported by our subjects): The characteristics of the nonlinear deformation may well change over time due to the changing positions and velocities of points. Subjects' reports of rigid motion under these conditions suggest a remarkable insensitivity to changing structure or nonrigidity. Similar conclusions were reached by Norman and Todd (1993).

We will address similar questions in the section Model and Simulation.

EXPERIMENT 2 "Two-Frame" Motion

In principle, relative distances from a plane (relief) are specified by the instantaneous projected velocities of the configuration as contained in a sequence of two projections or two frames. It is an open question, however, how well human subjects perform in such a minimal condition of two frames.

As in our smooth motion conditions, we presented a few cycles of oscillating motion. However, in the two-frame motion condition, one cycle contained only two projections or frames. Thus, the term *two-frame* refers to the number of *independent* projections contained in the motion sequence.

In pilot experiments, we found that human performance in our relief extraction task was strongly influenced by two parameters: the amplitude of oscillation ρ and the SOA value that determines the frequency of oscillation.

Before examining the influence of slant on relief extraction, we carried out an experiment to determine the motion parameters ρ and SOA that would yield optimal performance for relief extraction.

Optimal Frequency and Amplitude of Oscillation

Results. In Figure 6, we present the thresholds for discriminating the distances of two points from a frontoparallel plane P . The similarity in performance between subjects allowed us to average across Subjects P.W. and S.P. Averaged points of subjective equality were close to

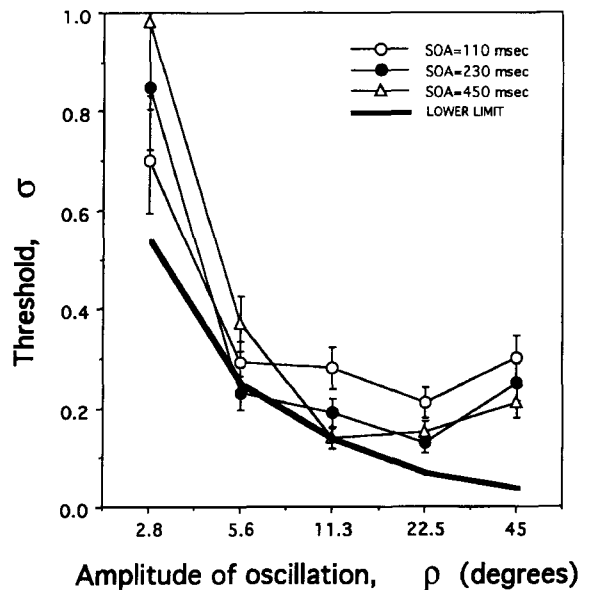


Figure 6. Relief-from-motion thresholds for different stimulus onset asynchrony values as a function of the amplitude of oscillation ρ . Here, the plane was frontoparallel. The thick line shows the theoretical lower limit for the thresholds: a perfect relief from motion computation based on the pixel coordinates of the projected points (simulated thresholds are the result of the discretization of positions on a raster display). The thresholds σ are averaged across Subjects P.W. and S.P.

one, as was expected for frontoparallel conditions; they are not presented.

The thresholds σ are presented as a function of the amplitude of rotation ρ for different SOA values (110, 230, and 450 msec). The thick line in the graph represents the theoretical lower limit of the measured thresholds. These lower limits are the result of an ideal relief-from-motion computation (see the Model and Simulation section) based on the projected positions of the points in both frames. Inherent to the use of raster displays with a finite resolution, these projected positions are rounded to integer pixel units. We call this (spatial) *discretization noise*. Discretization noise acts as an external noise source and determines the limits of the performance of a human observer.

For small (2.8°) and large (larger than 5.6°) amplitudes of oscillation, thresholds for human observers exceed the lower limit indicating that performance is limited by noise sources in our visual system and not by discretization noise. When ρ is 45° , subjects reported having problems with the correspondence between points over time. Amplitudes of 11.3° and 22.5° elicit a compelling impression of a rigidly rotating configuration, and thresholds for this range are lowest. For amplitudes larger than 5.6° , a small SOA value of 110 msec systematically yields higher thresholds than a large SOA value of 450 msec. For the Subjects P.W. and S.P., performance is optimal ($\sigma = 0.13 \dots 0.19$) for medium angles of rotation ($\rho = 11.3^\circ$ or 22.5°) and for slow oscillations ($\text{SOA} \geq 230$ msec).

We should be careful with this conclusion, however, because the thresholds found for an amplitude of 5.6° ap-

proximately equal the lower limit due to discretization noise ($\sigma = 0.25$). Hence, performance might have been better (and perhaps best of all) if we had used a higher display resolution. Therefore, we scaled both the stimulus and the viewing distance by a factor of two (an effective increase of the display resolution by a factor of two) and again measured thresholds for an amplitude of rotation of 5.6° and an SOA of 230 msec. At this viewing distance of 200 cm, the size of a pixel was 0.53 arcmin. The theoretical lower limit for this *higher resolution condition* becomes 0.12. Psychophysically, we measured similar thresholds ($\sigma = 0.20$) as for the *lower resolution condition*, indicating that performance is not limited by the discretization noise, but by noise sources in our visual system. Also, the threshold for an amplitude of 22.5° under this higher resolution condition are invariant. Therefore, we conclude that a medium amplitude of 11.3° or 22.5° yields optimal performance.

We also tested Subject H.V. Although Subject H.V. reported to perceive rigidly moving configurations in smooth motion conditions (and performed similarly to Subjects P.W. and S.P. for those conditions), he reported nonrigid motion for more than 50% of the trials for all temporal parameters and oscillation amplitudes examined. Owing to the nonrigidity of the perceived structure, our structure-from-motion task was not well defined for him and H.V. was excluded from the two-frame motion experiments.

Discussion. Todd et al. (1988) thoroughly examined which display parameters can reliably elicit a perceptually compelling impression of rigid rotation in 3-D space (rigidity rating). Although such a compelling impression of rigidity does not predict performance in our affine-structure-from-motion task directly, it is of interest to compare their optimal motion parameters with those presented in Figure 6. Given a sequence containing two distinct frames, they reported optimal rigidity rating for a zero interstimulus interval value (as used in our experiments), for large SOA values (larger than 200 msec). Their finding that large SOA values yield optimal rigidity rating is consistent with our finding that relief extraction is better for values larger than 200 msec than for the smallest value of 110 msec.

An *amplitude* of rotation of ρ degrees in our experiment corresponds to a *total* rotation of 2ρ degrees in Todd et al.'s (1988) experiments. Todd et al. examined ρ values of 3° , 9° , and 15° and found only a small effect of the amplitude of rotation: rigidity rating was only slightly better for a 3° amplitude than for a 15° amplitude. In contrast, we have found that a small ρ value of 2.8° yields very poor performance for all SOA values, which is worse than had been expected on the basis of the simulations mentioned above. This discrepancy is probably due to the fact that the 3-D positions of the points in our configuration were constrained differently. Here, the average attitude of plane P was frontoparallel and the points Q_i were in front of the P within a certain range around the axis of rotation, whereas Todd et al. have randomly positioned the points in 3-D space at a distance of be-

tween 80 and 120 pixels from the origin at the center of the display screen.

For Subjects P.W. and S.P., we chose $\rho = 22.5^\circ$ and SOA = 230 msec as the motion parameters to determine the influence of the slant of plane P on the extraction of relief as described in the following section. These parameters yield the most compelling percept of a rigidly rotating configuration, and thresholds are low.

Dependence on Slant

The procedures and setup for the two-frame condition described here were identical to those described for the smooth motion condition, except for the motion parameters. Figure 7 shows the PSE μ and threshold σ as a function of the slant of plane P for a two-frame oscillating motion condition.

Comparison of Figures 5 and 7 shows that the performance in two-frame motion and smooth motion conditions is qualitatively very similar for Subjects P.W. and S.P. The PSE deviates most strongly from one for medium slant values (45°) and is almost negligible for frontoparallel conditions. However, quantitatively, both the bias μ and the thresholds σ are significantly higher for two-frame motion than for smooth motion for all slant values tested. For example, when the plane is slanted 45° , the PSE is about 1.5 and the threshold about 0.45! The *qualitative* similarity suggests that a similar relief-from-motion computation is performed by the visual system for both smooth motion and two-frame motion conditions. The *quantitative* difference would then mean that the velocity estimates that are input for this computation are less accurate for two-frame motion, even when optimized for the amplitude and frequency of oscillation.

MODEL AND SIMULATIONS

Here we describe a simulated subject performing a particular relief-from-motion computation based on the positions and velocities of projected points. That is, the input to the relief-from-motion computation are the projected positions \mathbf{p}_i and velocities $\mathbf{p}'_i - \mathbf{p}_i$ of the three points \mathbf{P}_i that specify the plane P , and the projected positions \mathbf{q}_i and velocities $\mathbf{q}'_i - \mathbf{q}_i$ of the two points \mathbf{Q}_i that have to be discriminated with respect to their relative distances d_1 and d_2 from plane P . The output of the computation is the distance ratio d_2/d_1 .

In order to let our simulated subject be independent of the specific implementation of this relief-from-motion computation, we assume that this computation is infinitely precise. Errors in the extraction of relief are assumed to be caused by upfront measurement errors. That is, we assume that the velocity estimates that enter the computation of relief are imprecise.

Relief-From-Motion Computation

The scene. We describe the scene in an orthogonal coordinate system with the x -axis in the horizontal and the y -axis in the vertical direction. We take the z -axis in the direction away from the subject. The subject views the pro-

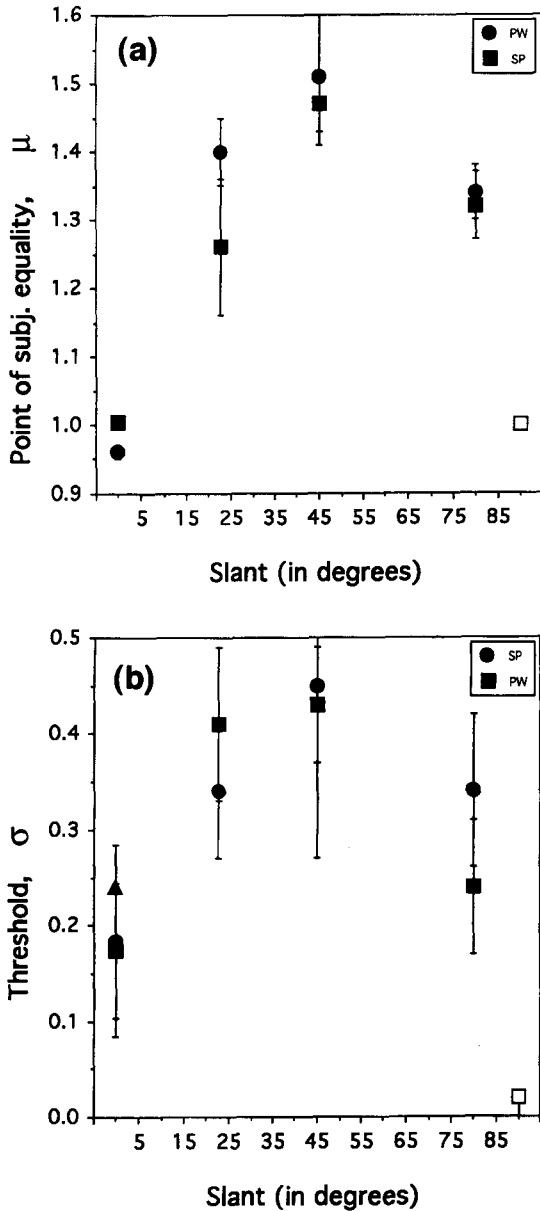


Figure 7. Relief extraction in two-frame motion conditions (two frames/cycle, SOA = 230 msec). Point of subjective equality μ (a) and ratio threshold σ (b) as a function of the slant of the configuration. Results are shown for 2 subjects (P.W. and S.P.). The error bars indicate the measurement error.

jection of an object in space onto the frontoparallel (xy) plane. We assume that a parallel projection approximates the perspective projection for this scene, that is, the projection of a point $\mathbf{P} = (x, y, z)$ onto the frontoparallel plane is a point \mathbf{P} with coordinates (x, y) .

The object. The object consists of three points \mathbf{P}_i ($i = 1 \dots 3$) in a plane P with depth gradient (g_1, g_2) , and points \mathbf{Q}_i that are noncoplanar with P . That is, \mathbf{Q}_i can be written as a coplanar point $(x, y, g_1x + g_2y)$ plus a vector $\gamma_i(g_1, g_2, -1)$, which is orthogonal to P . The values γ_i

specify the relief—that is, the relative distances of points \mathbf{Q}_i from plane P .

The object transformation. For the purpose of extracting relief, it is *not* necessary to assume a rigid transformation of the object relative to the subject. In fact, we can allow an arbitrary *linear* transformation A in space:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}. \quad (3)$$

Note that A includes arbitrary rigid rotations, but also deformations. We have left out arbitrary translations, since, with parallel projections, they merely shift the projection in the frontoparallel plane.

Extracting relief. We will extract the relative depth values γ_i of points \mathbf{Q}_i with respect to plane P . Let \mathbf{p}_i ($i = 1 \dots 3$) be the projections of \mathbf{P}_i in image I_1 and \mathbf{p}'_i those in a second image I_2 . It is obvious that the transformation that maps \mathbf{p}_i in I_1 to \mathbf{p}'_i in I_2 is linear in x and y , because both the transformation A and the projection are linear in x and y . More specifically, we have

$$\mathbf{p}'_i = M\mathbf{p}_i, \quad (4)$$

where

$$M = \begin{pmatrix} a_{11} + a_{13}g_1 & a_{12} + a_{13}g_2 \\ a_{21} + a_{23}g_1 & a_{22} + a_{23}g_2 \end{pmatrix}. \quad (5)$$

For the projections of *noncoplanar* points \mathbf{Q}_i , we find

$$\mathbf{q}'_i = M\mathbf{q}_i + \gamma_i\mathbf{z}, \quad (6)$$

where

$$\mathbf{z} = -(1 + g_1^2 + g_2^2) \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix}. \quad (7)$$

In other words, the projections \mathbf{p}_i of coplanar points \mathbf{P}_i travel conform to an affine velocity field specified by M , whereas the projections \mathbf{q}_i of *noncoplanar* points \mathbf{Q}_i deviate from this affine flow by a vector $\gamma_i\mathbf{z}$. The vector \mathbf{z} depends on the object transformation parameters a_{13} and a_{23} and on the depth gradient of the plane P and is thus the same for all points \mathbf{Q}_i . (For infinitesimal rotations a_{13} and a_{23} are the frontoparallel components of the axis of rotation!)

The affine flow (or the matrix M) is fully specified by the observed displacement vectors $\mathbf{p}'_i - \mathbf{p}_i$. Knowing M , the $\gamma_i\mathbf{z}$ are simply $\mathbf{q}'_i - M\mathbf{q}_i$.

In conclusion, we can find the depth ratio γ_i/γ_j by simply measuring the length ratio $|\gamma_i\mathbf{z}|/|\gamma_j\mathbf{z}|$ of the deviation vectors. The γ_i specify the object's relief, although no metric information is available: absolute values of d_1 and d_2 cannot be obtained from two frames.

This relief-from-motion computation can be implemented through a variety of steps, any of which can introduce errors in the final estimation of γ_2/γ_1 owing to imprecise calculations. Because the specific steps in the implementation of this computation cannot be discrimi-

nated psychophysically on the basis of the results of our experiments, we have chosen all steps to be infinitely precise. Errors in the output of this computation are assumed to be caused by imprecise velocity estimates $\mathbf{p}'_i - \mathbf{p}_i$ and $\mathbf{q}'_i - \mathbf{q}_i$ as discussed next.

Errors in Human Velocity Extraction

We discuss two types of measurement errors that are reported in the literature for human velocity estimation: a bias (or systematic error) in velocity estimates and a scatter (or variance) in velocity estimates.

A bias in velocity estimates has been reported by Johnston and Wright (1986): when human subjects are presented with a moving target (at velocity \mathbf{v}) at an eccentricity E in the periphery, the average velocity estimated by the subjects ($\bar{\mathbf{v}}_e$) is a factor lower than when the moving target is presented in the fovea. The dependence on eccentricity is roughly described by

$$\bar{\mathbf{v}}_e = \frac{E_0}{E + E_0} \mathbf{v}, \quad (8)$$

where E_0 is the characteristic eccentricity.

A scatter in velocity estimates has been reported, for example, by McKee (1981), Orban, van Calenbergh, de Bruyn, and Maes (1985) for uniform motion, and by Werkhoven and Koenderink (1991) for rotary motion. We have modeled the variance in velocity estimates by a stochastic process and the estimate \mathbf{v}_e of velocity \mathbf{v} is

$$\mathbf{v}_e = [1 + \varepsilon] \bar{\mathbf{v}}_e, \quad (9)$$

where ε is a stochastic variable with a Gaussian distribution $P(\varepsilon)$:

$$P(\varepsilon) = \frac{1}{\sqrt{2\pi\omega^2}} \exp\left[-\frac{\varepsilon^2}{2\omega^2}\right]. \quad (10)$$

The parameter ω , called the Weber fraction, specifies the level of the multiplicative velocity noise.

Together, the estimated velocity \mathbf{v}_e of a target moving at a velocity \mathbf{v} at eccentricity E is modeled by

$$\mathbf{v}_e = [1 + \varepsilon] \frac{E_0}{E + E_0} \mathbf{v}. \quad (11)$$

Simulated Subject in Our Relief-From-Motion Experiment

In our experiments, subjects indicated which of two points \mathbf{Q}_1 and \mathbf{Q}_2 is further away from a plane P specified by three points \mathbf{P}_i . Here, we compare the performance of human subjects for this task with the performance of a simulated subject as described above: a relief-from-motion computation based on imprecise local velocity estimates.

Specifics. The results for a simulated subject are collected following identical procedures as for human subjects. Instead of presenting the generated stimuli to human subjects, we fed the projected coordinates of the stimuli into our relief-from-motion computation (after corrupting the displacements following Equation 11). More pre-

cisely, the input to our relief-from-motion computation are the *exact* positions¹ of all projected points \mathbf{p}_i and \mathbf{q}_i in frame 1 and the *imprecise* estimates \mathbf{v}_e of displacements $\mathbf{p}'_i - \mathbf{p}_i$ and $\mathbf{q}'_i - \mathbf{q}_i$ of these points between frame $n = 1$ and frame $n = N$ (those two projections for which the angles of rotation were $-\rho$ and ρ , respectively). Control simulations have shown no dependence on which two frames are taken as input. The answers given by our simulated subject were input to the adaptive method to determine the psychometric functions (as for human subjects). The PSE μ and the threshold ratio σ for a simulated subject were determined in six sets of 70 trials (as for human subjects).

Simulated point of subjective equality, μ . Results of a simulated subject characterised by an eccentricity dependent scaling of velocity estimates are shown in Figure 8A for different characteristic eccentricities E_0 .

The PSE μ for relief discrimination shows a maximum at 45° slant (for the discrete slant values simulated here), similar to the results found for human subjects. The result for $E_0 = 160$ pixels (which corresponds to a characteristic eccentricity of approximately 3.5°) yields μ ratios as a function of slant that are a good approximation (by eye) of the average performance of our subjects.

Simulated threshold, σ . Here, we introduce a Weber law for the accuracy of velocity estimates, that is, the width of the distribution of noisy estimates of velocity \mathbf{v} increases linearly with \mathbf{v} . The noise is parameterized by the Weber fraction ω (see Equation 9).

Computed threshold ratios σ of the simulated subject are presented in Figure 8B as a function of slant. We have plotted curves for different Weber fractions. The characteristic eccentricity E_0 was held constant and was 3.5° as determined above.

First about the shape of these curves. As for human subjects, the accuracy of our simulated subject decreases (thresholds increase) when the slant of the configuration increases from 0° to roughly 40°–50°. Increasing the slant even further has the effect of increasing the accuracy again (thresholds go down). At 90° slant, thresholds become zero. Second, at a given slant of the configuration, relief discrimination thresholds increase approximately linearly with the Weber fraction for velocity discrimination.

For the average subject, it seems that relief discrimination thresholds (see Figure 5) are reasonably described (within the measurement error) by a Weber fraction of 30% ($\sigma = 0.3$), as can be seen in Figure 8B.

Model summary and discussion. For our stimuli, we found that relief extraction based on motion information is much more accurate for frontoparallel relief than for configurations slanted 45°. Qualitatively, the performance of human subjects is predicted well by the simple relief computation described above, based on noisy velocity estimates. A Weber fraction $\omega = 30\%$ yields a behavior that comes close to the behavior of human subjects for the smooth motion condition. Such a Weber fraction is high in comparison with the Weber fraction (5%–8%) found for simple foveal velocity discrimination tasks (McKee, 1981). However, in our tasks, subjects have to extract ve-

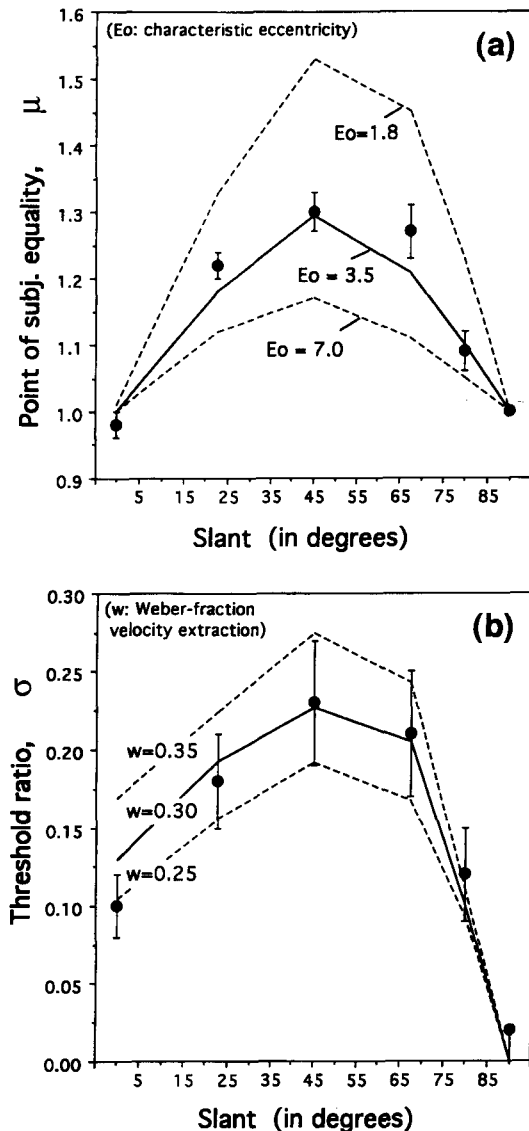


Figure 8. Simulated subjects. (a) Simulated μ values for a subject that underestimates velocity as a function of eccentricity (systematic measurement errors). Different curves are for different characteristic eccentricities E_0 . The solid curve shows results when $E_0 = 3.5^\circ$ (160 pixels). Found μ values are independent of the scatter parameter w . (b) Threshold ratio σ as a function of slant. Here the characteristic eccentricity $E_0 = 3.5^\circ$. Different curves are plotted for different Weber fractions ω . To facilitate the comparison of simulated and human subjects, we have added the average performance of human subjects for the smooth motion condition.

locity estimates at five random positions in the visual field. Furthermore, the positions in the visual field of three points are not foveal, but are positioned approximately 3° – 4° in the periphery. Velocity discrimination tasks can become quite inaccurate ($>30\%$) for such conditions (see Werkhoven & Koenderink, 1991).

There are some reasons to argue for an eccentricity-dependent scaling of velocity as one of the few candidates to explain our data. First, when the distribution of veloc-

ity estimates v_e of a given velocity v is symmetric around v , the average extracted distance ratio μ will be the correct ratio in first approximation; that is, the predicted PSE will be 1. The width of the distribution for velocity estimates merely determines the threshold σ for relief extraction, but leaves the PSE μ invariant. So, to explain that the mean estimated ratio μ deviates from the correct ratio, we need to assume that the mean of the distribution of velocity estimates v_e of v deviates from the true v . Furthermore, we know that any linear transformation of the projected velocities (for example $v_e = \alpha v$) leaves affine invariant properties such as the ratio γ_2/γ_1 unchanged.

Hence, a linear scaling of velocity (under-/overestimation) independent of the spatial position at which this velocity is measured would leave the distance ratio in our experiment invariant (this would simply correspond to an affine transformation between two consecutive projections). To explain our results, we have to consider *non-linear* transformations of the velocity field. An example of such a transformation is eccentricity-dependent scaling of velocity (see Equation 8) that characterizes human motion perception. Eccentricity-dependent scaling indeed explains our data when we assume that the characteristic eccentricity is approximately 3.5° . Such characteristic eccentricity is consistent with the findings of Johnston and Wright (1986).

GENERAL DISCUSSION

We presented the dynamic projection of a configuration moving in space that contained the minimum information required to extract relief: the relative distance of two points from a plane. Human subjects were able to extract relief with an accuracy that depends strongly on the slant of the configuration. Qualitatively, this dependence is very similar for both two-frame motion and smooth motion conditions (worse for medium slant values than for frontoparallel and fully slanted conditions). Thresholds for two-frame motion, however, were systematically higher than for smooth motion.

Relief Discrimination Versus Relief Detection

Although we must conclude that relative distances from a plane are poorly *discriminated*, human subjects are quite sensitive with respect to *detecting* deviations from planarity. Todd and Bressan (1990) reported that subjects could correctly discriminate planar and nonplanar configurations of moving line segments in 84% of the presentations when the dihedral angle of two connected facets was only 4° (showing only a weak effect of the number of frames). That is, the *nonplanarity* of four points was accurately *detected* in their experiments even when a minimum motion sequence of two frames was presented. In our experiments, the points Q_i were clearly out of the plane P and thus above detection threshold, but discriminating their distances from plane P was difficult.

Planarity detection has also been studied in a different context. Wagemans, van Gool, de Troy, Foster, and Wood (1993) have studied the ability to determine "affine shape

equivalence." They presented subjects with two (spatially) separated static configurations, each containing four points. The subjects indicated whether one configuration of four points was the 2-D affine transform of the other. Such 2-D affine equivalence of the projections could be interpreted as the planarity of two facets in 3-D space when the two configurations had been presented in a motion sequence. Wagemans et al. found that subjects performed remarkably well.

Together with Todd and Bressan's (1990) experiments and the experiments of Wagemans et al. (1993), our experiments seem to indicate that the angle between two facets in 3-D space is accurately coded for small angles (approximating planarity), but very coarsely for larger angles. Therefore, we conclude that the visual system is particularly tuned to planarity's being an important structural property.

What Limits Relief Extraction in "Smooth" Motion Conditions?

We have shown that human subjects make errors in our relief-from-motion task that are consistent with the characteristics of human velocity discrimination (Werkhoven & Koenderink, 1991) and matching (Johnston & Wright, 1986) as reported for local targets.

We should note that our findings are probably open to alternative explanations in terms of the relief from computation that follows velocity extraction. The adoption of ad hoc errors in the combination of correct local velocity estimates may no doubt yield the slant-dependent performance revealed by our experiments.

To distinguish errors in the local velocity estimation from errors in the combination of local velocity estimates, we tested a specific prediction based on our eccentricity-dependent scaling model. We carried out a control experiment and measured performance under smooth motion conditions at double the viewing distance (200 cm). Otherwise, the experimental conditions were identical to those described above for smooth motion conditions. The average bias measured experimentally for this condition is 0.18 (with a measurement error of 0.03). The predicted points of subjective equality depend on the ratio of the extent of the stimulus and the characteristic eccentricity. Doubling the viewing distance to 200 cm reduces this relative extent by a factor of two. In our simulations, this is identical to increasing the characteristic eccentricity by a factor of two and keeping the viewing distance constant at 100 cm; the predictions for this condition are given by the lower dashed curve in Figure 8 ($E_0 = 7^\circ$). The maximum bias predicted is found at a slant of 45° and is 1.17. Although more extensive experiments are certainly useful, the good agreement between the experimental and predicted values supports the idea that eccentricity-dependent scaling can explain the data.

What Is The Additional Problem With "Two-Frame" Motion Conditions?

Our finding that human subjects perform systematically better in smooth motion conditions than in two-frame

conditions even when those conditions are optimized for the frequency and amplitude of oscillation suggests that our visual system explores the information contained in more than two consecutive frames. With some speculation, we will discuss three distinct stages at which the information contained in more than two frames may enter the structure-from-motion computation.

First, temporal integration may improve the accuracy of velocity estimates that are the input to a two-frame *affine* structure-from-motion computation. Consider a three-frame motion sequence, for example: the velocities extracted from the 1,2 transition (the velocities specified by frame 1 and 2) can be averaged with those extracted from the 2,3 transition. Reduced variance in the averaged velocity estimates will yield a better estimate of the computed distance ratio d_2/d_1 . We call this a temporal integration at the *input* stage of an affine reconstruction algorithm.

Second, a temporal integration of the d_2/d_1 estimate computed from the 1,2 transition and the one computed from the 2,3 transition will reduce the variance of the final average d_2/d_1 estimate. This is a temporal integration at the *output* stage of an affine reconstruction algorithm. One can show that for small variances in the extraction of velocity, integration at the output and input stages yields equivalent reductions in the variance of the final distance-ratio estimate and cannot be distinguished psychophysically on the basis of our data.

Third, temporal differentiation may give an estimate of acceleration such that a metric reconstruction algorithm becomes possible. However, temporal differentiation of the velocity of local targets is not accurate (Snippe & Werkhoven, 1993; Werkhoven et al., 1992), and it is not clear why a metric representation would yield better estimates of *affine* properties than would affine representations.

In short, in two-frame motion conditions, the visual system loses the option to reduce the variance in velocity estimates and/or distance-ratio estimates and loses the option to construct a metric representation of the configuration because the acceleration is no longer specified.

Necessity of Cue Integration.

It has previously been shown that metric structure based solely on motion information is extracted poorly (Todd & Bressan, 1990), probably because of the inability to accurately extract visual acceleration. Here we have shown that the extraction of affine properties of the motion stimuli studied is quite poor. Affine representations based on projected motion information may suffer strongly from scatter and from biases in the extraction of local velocity estimates.

These and other results emphasize the necessity to integrate other available cues (stereo, texture, shading, etc.) with motion information to construct accurate representations when needed. However, early preliminary evidence on cue integration would seem to indicate that combining multiple sources of information about 3-D form does not necessarily lead to accurate representations of metric structure (Norman, Todd, & Phillips, 1994).

A priori knowledge about sizes, angles, or motion parameters may also calibrate affine representations such that metric properties become available.

REFERENCES

- BRAUNSTEIN, M. L., HOFFMAN, D. D., & POLLICK, F. E. (1990). Discriminating rigid from nonrigid motion: Minimum points and views. *Perception & Psychophysics*, **47**, 205-214.
- JOHNSTON, A., & WRIGHT, M. J. (1986). Matching velocity in central and peripheral vision. *Vision Research*, **26**, 1099-1109.
- KAISER, M. K., & CALDERONE, J. B. (1991). Factors influencing perceived angular velocity. *Perception & Psychophysics*, **50**, 428-434.
- KANATANI, K. (1993). Unbiased estimation and statistical analysis of 3-D rigid motion from two views. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **15**, 37-50.
- KOENDERINK, J. J., & VAN DOORN, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A*, **8**, 377-385.
- MCKEE, S. P. (1981). A local mechanism for differential velocity detection. *Vision Research*, **21**, 491-500.
- 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., & PHILLIPS, F. (1994). The perception of 3-D structure from multiple sources of optical information. *Investigative Ophthalmology & Visual Science*, **35**, 1742.
- ORBAN, G. A., VAN CALENBERGH, F., DE BRUYN, B., & MAES, H. (1985). Velocity discrimination in central and peripheral visual field. *Journal of the Optical Society of America A*, **2**, 1836-1847.
- RUBIN, N., SOLOMON, S., & HOCHSTEIN, S. (1995). Restricted ability to recover 3D global motion from 1D local signals: Theoretical observations. *Vision Research*, **35**, 569-570.
- SNIPPE, H. P., & WERKHOVEN, P. (1993). Pulse modulation detection in human motion vision. *Vision Research*, **33**, 647-656.
- SNOWDEN, R. J., & BRADDICK, O. J. (1991). The temporal integration and resolution of velocity signals. *Vision Research*, **31**, 907-914.
- TODD, J. T., AKERSTROM, R. A., REICHEL, F. D., & HAYES, W. (1988). Apparent rotation in three-dimensional space: Effects of temporal, spatial, and structural factors. *Perception & Psychophysics*, **43**, 179-188.
- 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.
- TODD, J. T., NORMAN, J. F., PEROTTI, V. J., & TITTLE, J. T. (1993). The discrimination of 3D length from motion and stereopsis. *Investigative Ophthalmology & Visual Science*, **34**, 1131.
- ULLMAN, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.
- WAGEMANS, J., VAN GOOL, L., DE TROY, A., FOSTER, D. H., & WOOD, J. R. (1993). Minimal information to determine affine shape equivalence. *Perception*, **22**, 107.
- WERKHOVEN, P., & KOENDERINK, J. J. (1991). Visual processing of rotary motion. *Perception & Psychophysics*, **49**, 73-82.
- WERKHOVEN, P., & KOENDERINK, J. J. (1993). Visual size invariance does not apply to geometric angle and vorticity. *Perception*, **22**, 177-184.
- WERKHOVEN, P., SNIPPE, H. P., & TOET, A. (1992). Visual processing of optic acceleration. *Vision Research*, **32**, 2313-2329.

NOTE

1. By *exact positions*, we mean the positions in screen coordinates (integer pixel units). Thus, discretization noise is taken into account. We assume that these discrete positions are measured with high precision.

(Manuscript received February 10, 1994;
revision accepted for publication January 1, 1995.)