

Perception of plane orientation from self-generated and passively observed optic flow

Jeroen J. A. van Boxtel

Laboratoire de Physiologie de la Perception et de l'Action
CNRS, Collège de France, Paris, France

Mark Wexler

Laboratoire de Physiologie de la Perception et de l'Action
CNRS, Collège de France, Paris, France



Jacques Droulez

Laboratoire de Physiologie de la Perception et de l'Action
CNRS, Collège de France, Paris, France

We investigated the perception of three-dimensional plane orientation—focusing on the perception of tilt—from optic flow generated by the observer's active movement around a simulated stationary object, and compared the performance to that of an immobile observer receiving a replay of the same optic flow. We found that perception of plane orientation is more precise in the active than in the immobile case. In particular, in the case of the immobile observer, the presence of shear in optic flow drastically diminishes the precision of tilt perception, whereas in the active observer, this decrease in performance is greatly reduced. The difference between active and immobile observers appears to be due to random rather than systematic errors. Furthermore, perceived slant is better correlated with simulated slant in the active observer. We conclude with a discussion of various theoretical explanations for our results.

Keywords: active vision, 3D shape perception, structure from motion, optic flow, surface perception, sensorimotor integration

Introduction

While moving through a three-dimensional (3D) environment, observers extract important visual characteristics of that world from the two-dimensional (2D) image flow on their retina. From this optic flow, they can sometimes reconstruct the original 3D layout with a fairly high level of accuracy, given its severely underdefined nature. Under certain conditions, immobile observers can also extract 3D structure and motion of objects from optic flow alone. The reconstruction of structure from motion (SfM) has for a long time attracted much attention (Von Helmholtz, 1867; Wallach & O'Connell, 1953; Rogers & Graham, 1979). And although it is well established that object shape can be recovered in SfM tasks, it is not fully known under what conditions and to what degree of accuracy.

It has generally been assumed that, at least as far as the perception of 3D shape is concerned, SfM depends only on retinal input. Therefore, according to this assumption, an active observer moving around a stationary 3D object should perceive the shape of that object in the same way as an immobile observer receiving the same optic flow, which would be generated by an equal-and-opposite rigid motion of the object (e.g., Wallach, Stanton, & Becker, 1974). In fact, the hypothesis that SfM depends solely on optic flow, coupled with the ecological prevalence of observer motion through a largely stationary environment, was used by

Wallach et al. (1974) as a justification for the rigidity assumption for immobile observers.

However, a number of studies have compared the SfM performance of actively moving observers with that of immobile observers receiving more-or-less accurate replays of the same optic flow, and have found differences between the two conditions. From these results, we can conclude that the purely retinal theory of SfM cannot be the whole story. These studies fall into two groups.

In the first group of studies (Dijkstra, Cornilleau-Pérès, Gielen, & Droulez, 1995; Rogers & Rogers, 1992; Wexler, Lamouret, & Droulez, 2001a; Wexler, Panerai, Lamouret, & Droulez, 2001b), subjects were presented with stimuli that admitted a small number of different solutions. The frequencies with which the solutions from this discrete set were perceived were different in active and immobile conditions. Subjects in the experiments of Wexler et al. (2001a), for example, could perceive 3D surfaces based on either perspective or motion cues, which were in conflict. It was found that in the active condition subjects made use of motion cues more often than they did while immobile, despite receiving the same retinal stimulation in the two conditions. The results from this group of studies can be summarized by the stationarity assumption, namely a bias toward perceiving objects whose 3D motion is minimal in an allocentric or earth-fixed reference frame. The stationarity assumption will be discussed further below.

In the second group of studies (Ono & Steinbach, 1990; Panerai, Cornilleau-Pérès, & Droulez, 2002; Peh, Panerai, Droulez, Cornilleau-Pérès, & Cheong, 2002; Rogers & Graham, 1979), differences in SfM performance were found in active and immobile observers, but in tasks that involved the perception of absolute length (either distance or depth). Immobile observers could not in principle perform such tasks with metric accuracy, as no absolute length scale is available. The performance differences between active and immobile conditions that have been found in these studies thus also provide evidence for extra-retinal contributions to depth perception.

Additionally, a number of studies have shown the effects of other extra-retinal information on the perception of self-motion, for example, heading (Crowell, Banks, Shenoy, & Andersen, 1998; Royden, Banks, & Crowell, 1992) or slant (Freeman & Fowler, 2000). Furthermore, haptic extra-retinal information has recently been shown to affect the visual perception of depth (Ernst, Banks, & Bühlhoff, 2000).

The goal of this work was to compare the precision of active and immobile observers in an SfM task that (1) can be done in both self-motion conditions (and in this way differs from the second group of studies cited above) and (2) where any active-immobile differences would not be due to different choices from among a discrete set of solutions (as in the first group of studies cited). In order to satisfy these two goals, we used a task in which the subject has to indicate the 3D inclination of a planar surface. The subject perceives this 3D information from optic flow that is either generated by active head motion around a stationary, virtual object (the ACT condition) or while remaining immobile but experiencing a replay of the same optic flow (the IMMOB condition).

Self-Motion and the Stationarity Assumption

The advantage of incorporating information from extra-retinal sources is evident because 3D structure and motion are mixed in a nonlinear way in optic flow. Confronted with a SfM task, an immobile observer has to simultaneously extract both structure and motion, and, therefore, must solve a complex, nonlinear problem (see “[Linearization of SfM by SelfMotion Information](#)”). A moving observer, on the other hand, has additional extra-retinal information about motion, such as a copy of the motor command (in the case of voluntary motion) as well as proprioceptive information. This additional motion information can transform the SfM task into a linear problem, provided that relative motion between observer and object is due entirely to the observer’s self-motion (i.e., that the perceived object is stationary in an earth-fixed or allocentric reference frame). Because much of the optic flow that we experience is due to self-motion in a

stationary environment, under many or even most circumstances, this last provision is met.

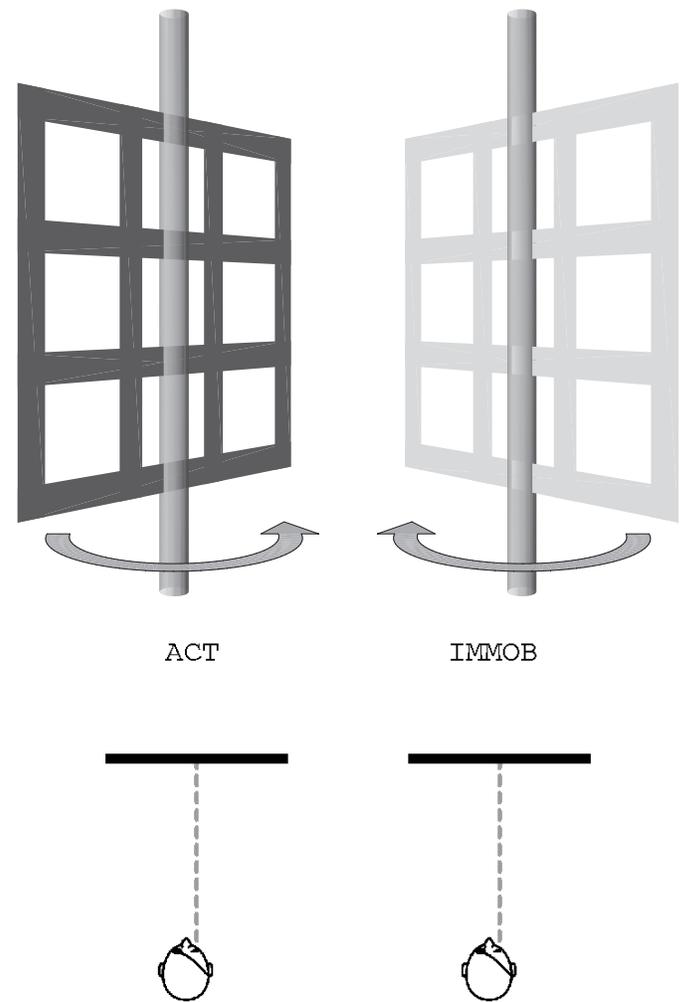


Figure 1. The optic flow generated by the moving plane in our experiment (left) is approximately the same as the optic flow generated by a plane with its tilt rotated by 180° and reversed angular velocity (right). In the limit of small stimuli, the difference between the optic flow generated by the two planes disappears, and they are equally likely to be seen by an immobile observer. The animation schematically shows the motion, in an allocentric reference frame, of the simulated (black) and tilt-reversed (red) planes for active and immobile observers. Small in-depth translations in IMMOB are not depicted in the animation, but were present in the stimuli.

Indeed, it has recently been shown that, in the perception of 3D shape, the visual system does make use of the heuristic assumption that objects are stationary in an allocentric frame – the stationarity assumption (Wexler et al., 2001a) – and that extra-retinal information about self-motion is used in this process (Wexler et al., 2001b). One way in which we can see this is the reduction of tilt reversals in active observers compared to immobile observers. A tilt reversal arises

from the following symmetry: simultaneously adding 180° to the tilt of a rotating plane and reversing its direction of rotation result in approximately the same optic flow (see Figure 1). Therefore, an immobile observer who has no a priori knowledge of the direction of rotation is equally likely to perceive the simulated plane and its reversal (whose tilts differ by 180°). If the same optic flow is generated by an active observer moving about a stationary plane, however, the simulated and the reversed planes have very different motion in an allocentric frame: the simulated plane is stationary (by construction), whereas the reversed plane rotates with twice the speed of the observer. It is known that active observers perceive the reversed plane much less frequently than do immobile observers (Dijkstra et al., 1995; Rogers & Rogers, 1992), and it has recently been demonstrated that difference is due to the visual system's use of the stationarity assumption (Wexler et al., 2001a).

Surface Perception From Optic Flow

Although many 3D objects have been used in experiments that probe SfM performance, one of the simplest is a plane. The orientation of a plane relative to the eye can be fully described by two variables, tilt and slant. Tilt is the orientation of the plane's normal projected onto the fronto-parallel plane (in our convention, the direction of zero angle is to the subject's right, with values increasing counterclockwise; see Figure 2). Slant is the angle, in 3D, between the normal and the direction perpendicular to the fronto-parallel.

For a small rotating planar object, slant and angular speed cannot be extracted independently from optic flow, which depends only on the product of its tangent and the angular speed (Domini & Caudek, 1999; Hoffman, 1982). (Strictly speaking, the above ambiguities in tilt and slant hold only in first-order optic flow, i.e., in the limit of small objects or parallel projection, but they are approximately true for objects that subtend 5° – 10° as in our experiments. The perspective [polar] projection approaches parallel projection in the limit of small objects.) Because slant is poorly defined in passive vision (and wholly ambiguous from first-order optic flow), an observer's judgment of this variable is not suitable as a measure to compare active and passive vision. Tilt, on the other hand, is theoretically well defined up to the 180° ambiguity (Longuet-Higgins & Prazdny, 1980; Ullman, 1979), even in immobile conditions, and serves this goal perfectly.

Whereas slant has received a great deal of attention in experimental studies (e.g., Braunstein, 1968; Domini & Caudek, 1999; Meese, Harris, & Freeman, 1995; Rogers & Graham, 1979), tilt has received relatively little. This is partly because tilt is theoretically always well defined¹ and partly because, in most experimental studies, tilt is perceived without large errors, either random or

systematic; this seems to be true in SfM tasks (Domini & Caudek, 1999; Norman, Todd, & Phillips, 1995; Stevens, 1983; Todd & Perotti, 1999), as well as in experiments where depth could be perceived from other depth cues, such as texture and shading (Norman et al., 1995).

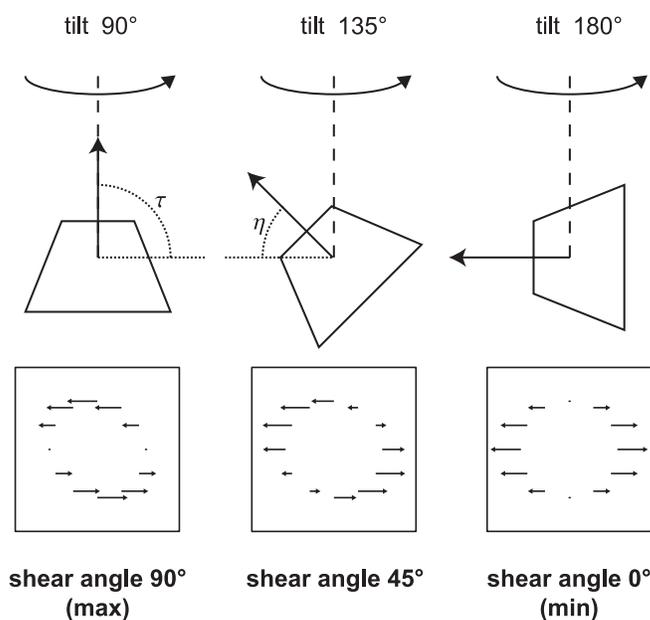


Figure 2. Definitions of shear angle and tilt. Upper panels. Three different orientations of a plane are depicted, with different values of tilt. In this example, the axis of rotation is vertical (which, in our experiment would be the case in the HORIZ motion condition). The left panel depicts how the tilt (τ) is defined, namely the orientation of the plane's normal (indicated by the arrow attached to the plane) projected onto the fronto-parallel plane. The middle panel depicts the definition of the shear angle, η (90° minus the smallest difference between the axis of rotation and the tilt). Lower panels. The approximate optic flow associated with the conditions drawn in the upper panels.

Cornilleau-Pérès, Wexler, Droulez, Marin, Miège, & Bourdoncle (2002) found that errors in tilt perception depend on the angle between the plane normal and the direction of dot motion. We will call this angle the shear angle (see Figure 2) because as it approaches 90° , the shear component of the optic flow increases as well.² More precisely, for rotations about axes in the frontal plane (the only 3D motion we study here), optic flow is approximately perpendicular to the axis of rotation (which is defined modulo 180° , of course); if β is the angle of the rotation axis in the frontal plane, and τ the tilt, the shear angle η is defined as

$$\eta = 90^\circ - \min[|\tau - \beta|, |\tau - (\beta + 180^\circ)|] \quad (1)$$

with angular differences taken the short way around the circle. As defined in Equation 1, shear angle (to which we will sometimes refer simply as “shear”) runs from 0° (no shear in optic flow) to 90° (maximal shear). Figure 2

provides a graphical example. Cornilleau-Pérès et al. (2002) found that when the shear angle increased, the perception of tilt badly deteriorated, which is surprising, given that tilt was believed to be easily and correctly found in SfM tasks.

No research has addressed the question of whether active vision increases the perceptual precision or accuracy because the earlier studies concerned with tilt perception (Cornilleau-Pérès et al., 2002; Domini & Caudek, 1999; Norman et al., 1995; Stevens, 1983; Todd & Perotti, 1999) used objects presented only to immobile observers. Such results cannot automatically be extrapolated to active vision (Wexler et al., 2001b). Our work is the first to study perceptual precision of tilt perception in active vision, and to compare that performance to that in passive vision.

Methods

Stimulus

In the reference frame used to describe the experiment, the xy -plane is co-planar with the monitor screen, with the x -axis pointing to the subject's right, the y -axis upward, the z -axis toward the subject, and the origin at the center of the monitor. Lengths will be expressed in centimeters.

The stimulus was a planar patch, inclined in depth. The patch was filled with a random-dot texture, with the dot distribution chosen to be uniform (on the average) in the projected image. This was done to remove texture cues to any particular tilt as much as possible (see Figure 3). The only depth cue from texture in our stimuli is the spatial distribution of texel positions, which is nearly uniform (it is not exactly uniform because of motion), but which is of minor importance compared to other texture gradients (Cutting & Millard, 1984). The 200 dots were chosen so that their projections fell inside a circle of radius 5.25 cm in the image; therefore, the texture on the stimulus plane was an ellipse with a nonuniform distribution of dots. (At the approximate mean observation distance of 60 cm [see below], this resulted in a radial angular stimulus size of 5° .) During each moment (more precisely, display monitor frame) that the stimulus was visible, the texture elements were projected onto the screen using a perspective projection from the subject's current eye position, and drawn as white pixels. The subject's position was measured by a head tracker with almost no latency (see below) that was sampled at the same rate as the monitor refresh and stimulus update rate, 96 Hz. The center of the stimulus lay in the xy -plane. The stimulus was centered at the point directly opposite the subject's eye at the beginning of each trial [i.e., if the subject's eye was at point (x,y,z) , the center of the stimulus was at $(x,y,0)$].

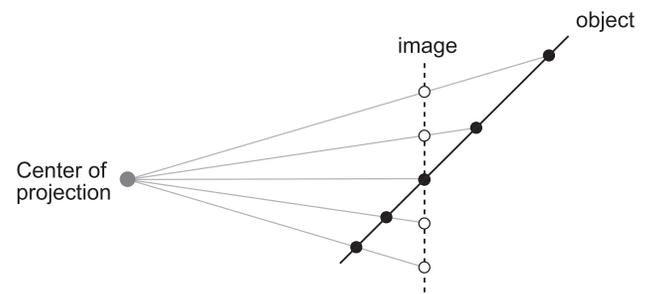


Figure 3. A schematic diagram of texture in our stimuli. The goal is to remove texture (i.e., nonmotion) cues to 3D structure as much as possible (Julesz, 1964; Rogers & Graham, 1979). We start with a uniform distribution of dots in the image plane (the white circles). These are then projected onto the inclined stimulus plane (the black circles). The distribution of the black circles is therefore nonuniform in the stimulus plane. With only small movements of the object or observer, the distribution of texture elements in the image remains nearly uniform.

The plane's normal was $(\sin\sigma \cos\tau, \sin\sigma \sin\tau, \cos\sigma)$, where σ is the slant and τ is the tilt. Tilt varied from 0° to 345° in increments of 15° . Slant was 30° , 45° , or 60° . A red fixation mark (a circle of size 0.05 cm) was visible in the center of the stimulus during the entire duration of the trial. Other than the stimulus, nothing was visible, including the borders of the display monitor.

Procedure

The experiment was performed in monocular viewing conditions using the subject's dominant eye, with the other eye covered. We will refer to the position of the center of the dominant eye (as measured by the head tracker, see below) simply as the "subject's position." For a trial to begin, the subject's position had to be inside a cube with sides of length 10 cm, centered on the point (0, 0, and 60), so that the stimulus stayed within the monitor screen (our reference frame is defined at the beginning of the "Stimulus" section).

In the beginning of every ACT trial, the subject was verbally cued to begin moving his head in one of four directions: right, up, left, or down (from the subject's point of view). Initial motion cycled on every subsequent trial through these four directions. The DIRECTION variable grouped trials by direction of motion: left and right trials will be called HORIZ and up and down trials VERT. Note that, in terms of relative rotation between the subject and the plane, HORIZ trials resulted in a vertical axis of rotation, whereas VERT trials resulted in a horizontal axis. We used both horizontal and vertical motion to decorrelate shear and tilt.

Motion continued until displacement along the required direction reached 3 cm, at which point a beep was heard. This was the signal to reverse the direction of motion, which occurred somewhat after the 3 cm, of course. When, after reversing direction, the subject's

position reached -3 cm along the direction of motion, another beep was sounded, and so on. In this way, we produced oscillatory motion in a given direction. The subject performed this oscillatory motion for 2.5 cycles in each trial. During the first half-cycle, only the fixation mark was visible, while the stimulus appeared during the next 2 cycles. Following the 2.5 cycles, the stimulus disappeared, and was replaced by a response probe. This was the subject's signal to stop moving the head.

To control variability in motion trajectories, we implemented some additional restrictions on the subject's motion. The amplitude was controlled by aborting the trial if displacement along the required motion direction (x - or y -axis in HORIZ and VERT trials, respectively) exceeded 6 cm from the central point. To make sure that motion was primarily in the required direction, at the end of each trial, we calculated the RMS of the subject's motion in that direction, normalized by the RMS of the motion in the two perpendicular directions; if this ratio exceeded 0.5, the trial was restarted. Furthermore, a trial was restarted when the duration of the visible stimulus was less than 2 s or greater than 5 s.

In the second condition, IMMOB, subjects moved very little, but regardless of their head movements, they experienced the exact same optic flow as in corresponding ACT trials. Subjects were instructed to remain still for the duration of the trial, an instruction that was followed to a great extent (see "Analysis of Movement Trajectories" section in the "Results"). Nevertheless, any motion that the subject did produce in IMMOB was taken into account to exactly reproduce the optic flow from the corresponding ACT trial, as described in Appendix A. Therefore, unlike the ACT condition, subjects' head movements did not result in motion parallax in IMMOB. Furthermore, all other cues from the ACT trial (the verbal cue indicating the previous initial direction of motion and the beeps to control the subject's movement) were replayed during the IMMOB trial.

Following the presentation of the stimulus in each trial, subjects indicated the perceived orientation (that is, perceived slant and tilt) of the plane by adjusting a visual probe with a joystick. Because in IMMOB the object moved in an allocentric frame, the subjects were told to respond with the average tilt and slant they had seen. (Previous results and post hoc analysis revealed that this averaging was carried out accurately; see "Change of Surface Normal During a Trial" in "Discussion.")

We used the projection of an inclined square grid as a probe because it was previously found that the performance in the perception of such objects is independent of shear (Cornilleau-Pérès et al., 2002). The square grid was subdivided into 6×6 squares (each 1.75 cm wide; total size 10° if the probe had zero slant and the subject was 60 cm from the screen). The orientation of the grid texture was random on each trial, in order to remove any reliable 2D cues; probe texture orientation in IMMOB trials was the same as that in

corresponding ACT trials. The joystick's base was approximately horizontal. The probe tilt was equal to the joystick azimuth (i.e., the direction that the joystick shaft was inclined, as seen from the top, with tilt 90° corresponding approximately to the direction away from the subject), whereas the slant of the probe was proportional to the angle between the joystick shaft and the vertical. The probe had a maximum slant of 80° .

We used a factorial design in which each subject performed 576 trials: 2 SELF MOTION (ACT and IMMOB) conditions, 3 slant values, 24 tilt values, and 4 directions of initial motion. The experiment was performed in 3 ACT and 3 IMMOB blocks, with each ACT block followed by the corresponding IMMOB block. The order of ACT trials within each block was random. The IMMOB blocks reproduced each trial in the same order from the previous ACT block. Before the experiment started, subjects were given 2 practice blocks, one active and one immobile.

Apparatus

The translational eye displacements of the subject were measured by a mechanical head tracker (Panerai, Hanneton, Droulez, & Cornilleau-Pérès, 1999), which has submillimeter within-trial precision. Sampling of the head tracker was at the same frequency as the display monitor, 96 Hz. The latency exhibited by the tracker was lower than the sample interval. A Pentium II 400 MHz computer both sampled the tracker (using a National Instruments PCI-6602 card) and controlled the stimulus display (Sony GDM-F500 CRT monitor with 1600×1200 pixels on a 40.2×29.6 cm screen, driven by a Matrox G400 video card). The resolution was about 1.4 arcmin/pixel at a distance of 60 cm. A Microsoft Sidewinder Precision Pro digital joystick was used to direct the probe. Subjects viewed the stimulus monocularly with their dominant eye, the other being covered by an opaque patch. The experiment was performed in a dark room. To prevent anything other than the stimulus from being seen, the observers wore a pair of ski goggles.

Subjects

Five subjects participated (2 men and 3 women). They were all between 20 and 25 years old and were naïve to the experimental purpose. All had normal or corrected-to-normal vision.

Results

It was found that trials in which the initial motion was to the left gave the same results as trials in which the initial direction was to the right, and likewise for up and down initial motion. Nor were any differences found between the first, second, and third sessions. We therefore collapsed all data across these variables.

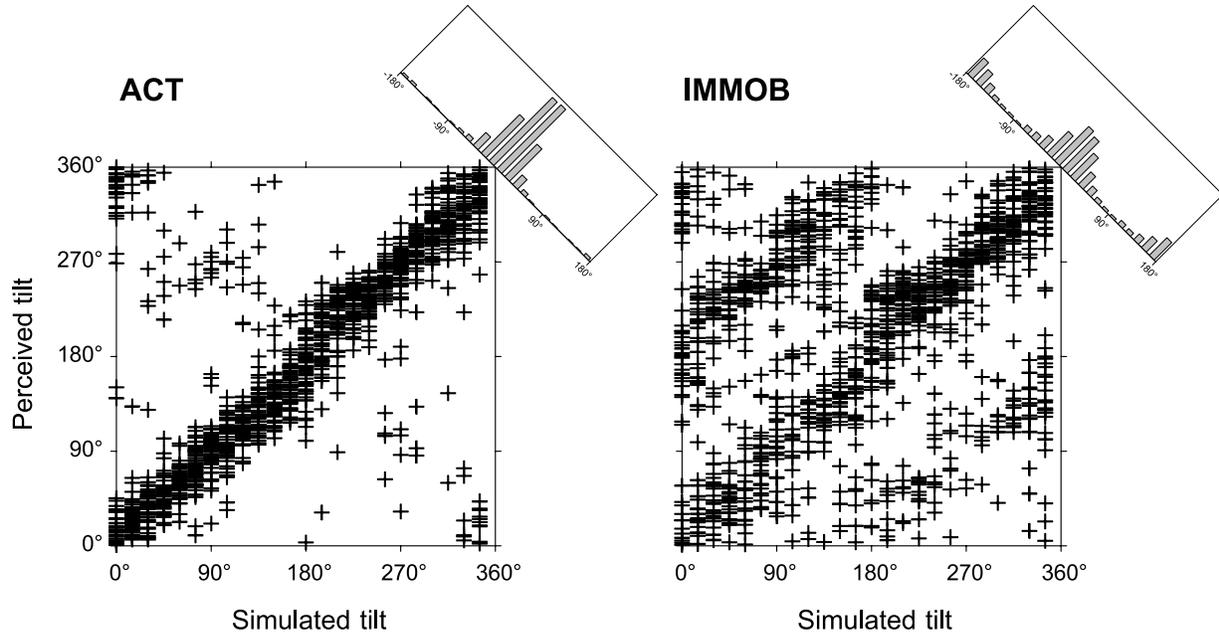


Figure 4. Perceived versus simulated tilt for all subjects in the ACT and IMMOB conditions. The histograms show the distributions of the differences Δ_τ between response and simulated tilt.

Tilt Errors

In Figure 4, we plot the perceived (τ_p) versus simulated (τ_s) tilt. We define the quantity

$$\Delta_\tau = \tau_p - \tau_s, \tag{2}$$

with angular differences always taken in the shortest way around the circle, and therefore ranging from -180° to 180° . The histograms in Figure 4 show the distributions of Δ_τ . The figure shows that, especially in the immobile (IMMOB) condition, in many trials Δ_τ was close to $\pm 180^\circ$. This corresponds to the perception of a reversal (see “Introduction” and Figure 1). Because the optic flow is almost ambiguous – there are really two simulated tilts differing by 180° – we also define a second tilt error, with respect to the reversed tilt:

$$\Delta'_\tau = \tau_p - (\tau_s + 180^\circ), \tag{3}$$

(with angular differences as above). Using these two quantities, we introduce an absolute-value tilt error measure, E_τ , which is the absolute value of the angular difference between the response and either the regular or the reversed simulated tilt, whichever is closer:

$$E_\tau = \min(|\Delta_\tau|, |\Delta'_\tau|). \tag{4}$$

As defined, E_τ ranges from 0° to 90° . In the remainder of this article, we will refer to E_τ as the tilt error.

Figure 5 shows the dependence of the tilt error on the shear angle.

There is a clear difference between the ACT and IMMOB conditions: in ACT the mean tilt error is 17.3° , whereas in IMMOB it is 24.1° . This difference is

significant ($p < .01$). A higher tilt error in IMMOB than in ACT was observed in all subjects. However, the average error is not fully informative, as there is a large effect of shear angle.

A SELF MOTION (ACT, IMMOB) \times DIRECTION (VERT, HORIZ) \times slant \times shear angle ANOVA with tilt error as a dependent variable showed that shear had a significant effect on the precision of tilt perception ($p < 10^{-4}$). Further analysis showed that in both IMMOB and ACT, tilt error increased significantly with increasing shear (both $p < 10^{-3}$). However, the tilt errors increased differently in the two conditions. The ANOVA showed a significant SELF MOTION \times shear angle interaction ($p < 10^{-4}$): the tilt error rose faster in the IMMOB than in the ACT condition. The magnitude of this effect can be demonstrated by a linear regression: the mean slope of the tilt error versus shear is 0.224 in IMMOB but only 0.067 in ACT. Moreover, this slope is lower in ACT than in IMMOB in all subjects. As far as the effect of slant was concerned, the ANOVA revealed that tilt errors decreased with increasing slant ($p < 10^{-3}$). Finally, the ANOVA showed a significant influence of DIRECTION, which we will return to in section “Anisotropy With Respect to Movement Direction” below.

Slant Errors

Errors in slant perception were analyzed using a SELF MOTION \times DIRECTION \times slant \times shear angle ANOVA with slant response as a dependent variable. This revealed a dependence on the simulated slant ($p < 10^{-4}$), although this dependence is rather weak (the slope of the linear regression is 0.21); however, all subjects showed a significant positive correlation between simulated and

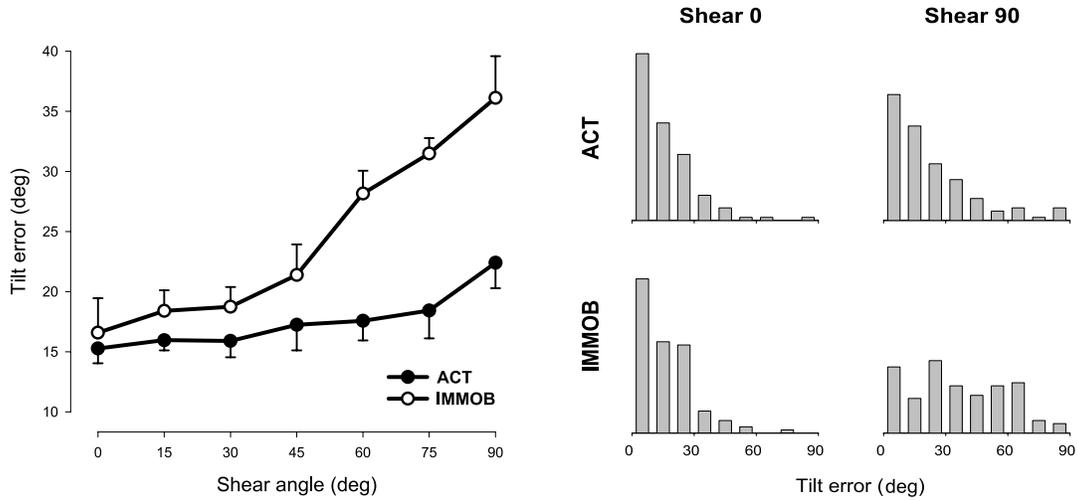


Figure 5. The influence of the shear angle on tilt error E_{τ} . In IMMOB, a clear increase of tilt error is observed with increasing values of the shear angle. The increase is present but significantly smaller in ACT. In the right panel, four histograms show the tilt error distribution in both ACT and IMMOB, for shear angle 0° and 90° . For shear 0° , tilt errors in both ACT and IMMOB are small. For shear 90° , errors have increased, especially for IMMOB, where performance is near chance level (flat). Error bars represent between-subject SEs.

response slant in both SELFOTION conditions (all $p < 10^{-2}$). However, slant response was better correlated with simulated slant in ACT than IMMOB (mean slope 0.25 in ACT, 0.16 in IMMOB). Indeed, there was a significant SELFOTION-slant interaction ($p < .05$).

The shear angle, apart from its effect on tilt perception, also influences the ability of subjects to estimate slant: absolute slant error (i.e., the absolute difference between the response slant and the simulated slant) increases with increasing shear ($p < .05$).

Reversals

We neglected tilt reversals in the preceding analyses. In this section, we specifically look at the rate of reversals and at the effect that reversals have on tilt and slant errors.

As stated above, we define tilt reversals as those trials in which the response tilt differed from simulated tilt by more than 90° . In IMMOB, reversals occurred on 35.3% of all trials; in ACT they occurred on only 4.4% (the difference in rates was significant: $p < 10^{-4}$, z test for independent proportions). The rate of reversals in IMMOB is significantly lower than 50% ($p < 10^{-4}$) – a 50% reversal rate would have been expected if subjects had ignored second-order information in optic flow (Wexler et al., 2001a)

Figure 6 shows that when reversals occurred both tilt and slant errors increased significantly in ACT ($p < .05$), but not in IMMOB. Nevertheless, even when errors were greater in reversal trials, the responses were not random: tilt responses in reversal trials were centered around reversed tilt and almost absent in the region which could be interpreted as large deviations in the percept of the simulated tilt—see, for example, the histograms in Figure

4. Absolute slant errors in IMMOB did not increase when reversals occurred, but a significant increase was seen in ACT ($p < .05$).

Systematic Errors in Perception of Tilt

We do not yet know whether the differences that we have found between the active and immobile conditions are due to random or systematic errors. In this section, we demonstrate that there indeed were systematic errors that amounted to directional biases in tilt perception, but that they most likely did not differ in the active and immobile conditions.³

Up to now, we have used the absolute-value tilt error, E_{τ} , which confounds random with systematic errors in tilt. Here, we wish to examine systematic errors in tilt corrected for reversals, and therefore we define a new, signed error measure:

$$S_{\tau} = \begin{cases} \Delta_{\tau} & \text{if } |\Delta_{\tau}| < |\Delta'_{\tau}| \\ \Delta'_{\tau} & \text{if } |\Delta'_{\tau}| < |\Delta_{\tau}| \end{cases} \quad (5)$$

with Δ_{τ} and Δ'_{τ} defined in Equations 2 and 3. S_{τ} is a signed tilt error that corrects for possible tilt reversals (i.e., the error is with respect to either the regular or the reversed simulated tilt, whichever is closer); it therefore ranges from -90° (clockwise errors) to $+90^{\circ}$ (counterclockwise errors). Averaging S_{τ} for a given value of simulated tilt permits us to study any systematic bias that is present at that point, independently of reversals.

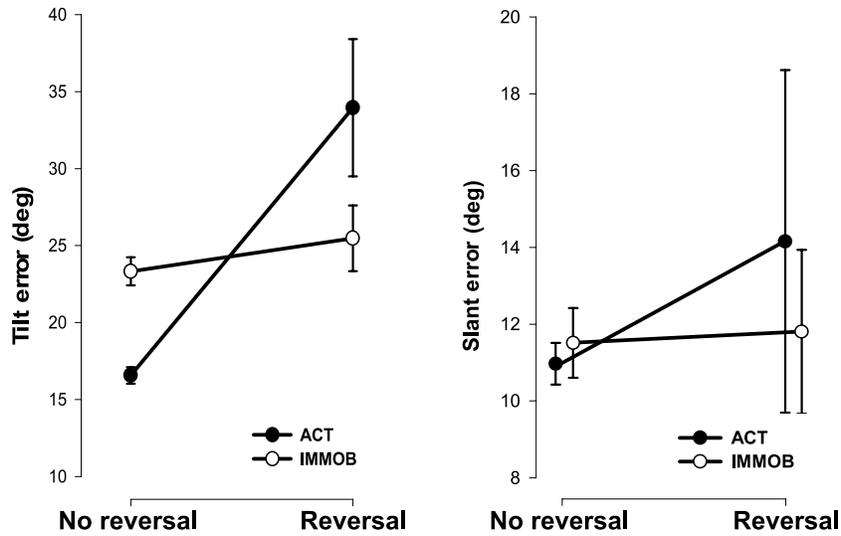


Figure 6. Reversals were found to have a significant effect on both tilt and slant perception. Left panel. In trials without reversals, the tilt error was significantly smaller in ACT than in IMMOB. In reversal trials, the errors in the IMMOB condition changed little, but the errors in ACT increased significantly and even became significantly greater than in IMMOB. Right panel. Slant errors were low in nonreversal trials for both ACT and IMMOB. For reversal trials, the slant errors in IMMOB stayed at this level, but in the ACT condition, the errors increased significantly, although there was a wide scatter between subjects. Error bars represent between-subject SEs.

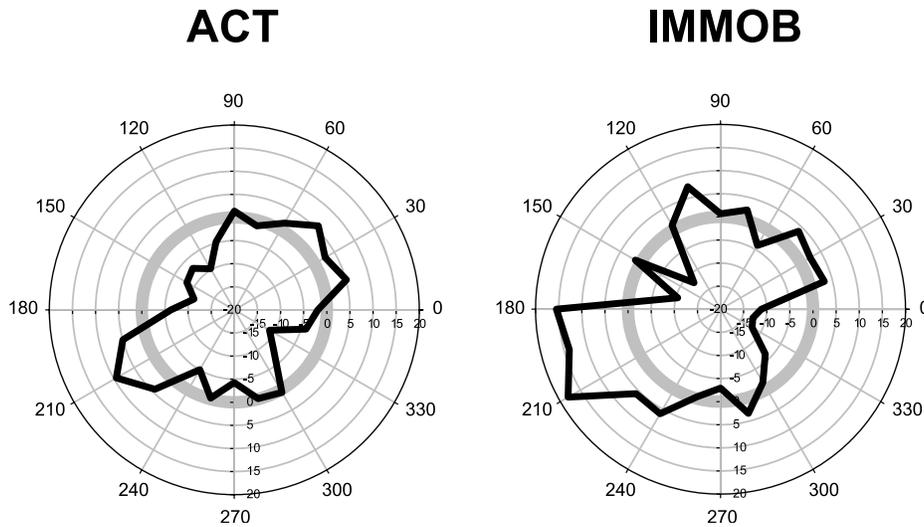


Figure 7. The dependence of systematic tilt error S_t on simulated tilt. The thick gray line represents zero systematic bias, with anti-clockwise biases positive. A bimodal Rayleigh test showed that mean bias was toward 84.9° (and 264.9°) in ACT and 88.3° (and 268.3°) in IMMOB (i.e., roughly horizontal surfaces).

Indeed, systematic errors were present in our data, as can be seen from Figure 7. Given the qualitatively bimodal trend in the data (Figure 7), we carried out Rayleigh tests (Batschelet, 1981) for bimodal distributions on the overall data and for subjects individually. The overall data were significantly bimodal in both ACT and IMMOB ($p < 10^{-4}$, Bonferroni corrected, for both), with mean bias tilt of 85° in ACT and 88° in IMMOB. In individual subject data, mean tilt biases were 91° , 84° , 85° , 82° , and 176° in ACT, and 99° , 85° , 80° , 100° , and 178° in IMMOB (keeping

the order of the subjects the same). All 10 tests were significant at $p < 10^{-4}$ and remained so when Bonferroni corrected.

However, we find no evidence for any differences in the tilt anisotropies in the ACT and IMMOB conditions. We carried out a t test for the mean values of S_t on individual subject data at the 24 values of simulated tilt. None of the tests reached the Bonferroni-corrected threshold for significance at $p = .05$.

Anisotropy With Respect to Movement Direction

The absolute tilt error E_τ was significantly different in the two DIRECTION conditions, being greater in the VERT (26.8°) than in the HORIZ condition (22.5° , $p < .05$). Further analysis showed that the two curves in IMMOB were not significantly different, but the ones in ACT were. Quantitatively, however, the difference between the two curves of the IMMOB conditions and the difference between the two ACT curves was very similar.

Analysis of Movement Trajectories

Because in the IMMOB condition, the subject's head was not held immobile by, say, a bitebar, the subject certainly performed some movements. Although the optic flow was identical in ACT and IMMOB trials regardless of any motion in IMMOB (see Appendix A), any head movement in IMMOB would thus provide no additional 3D information, but could have been a source of noise in the perceptual task. To compare movement in ACT and IMMOB conditions, for each trial we calculated the total 3D pathlength by summing eye displacements during the part of the trial in which the stimulus was visible. In ACT the average pathlength was 34.6 cm, whereas in IMMOB it was 3.6 cm. Therefore subjects followed, to a great extent, our instruction to remain still in the IMMOB condition.

We analyzed the movement trajectories of the ACT condition and studied kinematic quantities, such as the maximum amplitude, the velocity, and acceleration along all three axes. We divided the range of values each quantity subtended into several equally sized bins and checked whether the data (tilt error, signed tilt error, and reversals) showed a dependence on the quantity in question. No such dependence was found.

Next we compared the VERT versus the HORIZ condition to investigate the origin of the anisotropy with respect to movement. VERT trials showed larger amplitudes in the movement along the z -axis than HORIZ trials. The displacement along the x -axis during up/down movement was also greater than the displacement along the y -axis during left/right movement. We homogenized the trajectories post hoc by only considering trials whose movement amplitudes fell within a certain range. After this homogenization, the anisotropy with respect to movement was, however, still present.

Discussion

Active Vision Is More Precise Than Passive Vision

We examined the perceptual precision in a SfM task using two dependent variables: tilt and slant. To compare active and immobile conditions, it is necessary for the

dependent variable to be well defined and recoverable in both SELF-MOTION conditions. The precision of tilt perception does satisfy this criterion. However, as with other variables used in earlier research (e.g., depth and absolute distance), the precision of slant perception is not very useful, because it is poorly defined in the IMMOB condition.

We find that tilt perception in ACT is more precise than in IMMOB, and thus demonstrates, for the first time, that active vision increases the precision of surface perception compared to passive vision. For optic flow with minimal shear, tilt precision is about equal in active and immobile conditions; however, as shear increases, precision falls off rapidly in IMMOB, while remaining almost constant in ACT (see Figure 5). We find that the difference between active and immobile conditions is most likely due to random errors being greater in the immobile condition: although systematic errors are present, they appear to be the same in the two conditions.⁴

In Search for an Explanation for the Shear Effects

Given its clear importance for surface perception, the shear variable has been studied very little, and its effect is not understood theoretically. In this sub-section, we explore several different ways to account for the effect of shear. However, we warn the reader from the outset that none of our models gives, at present, a satisfactory account. For readers wishing to skip the details, a summary of our arguments is given in at the end of this section.

Change of Surface Normal During a Trial

Although the motion of the simulated surfaces in our experiment relative to the subject was the same in the ACT and IMMOB conditions, their motion in an earth-fixed or allocentric reference frame was different. Namely, the object rotated in IMMOB, while remaining motionless in ACT. During the object's rotation in IMMOB, its normal – and therefore its slant and tilt, which we define here in an allocentric frame – changed. Could this allocentrically defined change account for the different effects of shear in the two conditions?

Using standard vector algebra, we can show that for small surface rotations in the IMMOB condition (first order in rotation angle, α , a reasonable assumption for experiment where the average maximum angle was about 4°), the change in the surface normal $\hat{\mathbf{n}} = (\sin\sigma \cos\tau, \sin\sigma \sin\tau, \cos\sigma)$ is given by $(\hat{\mathbf{A}} \times \hat{\mathbf{n}})\alpha$ plus terms higher order in α , where $\hat{\mathbf{A}} = (-\sin(\tau - \eta), \cos(\tau - \eta), 0)$ is the axis of rotation and η the shear angle. The corresponding changes in slant and tilt are therefore

$$\Delta\sigma \approx (\cos\eta)\alpha \quad (6)$$

$$\Delta\tau \approx \frac{\sin\eta}{\tan\sigma} \alpha. \quad (7)$$

These equations together with the change-in-surface-normal hypothesis make several quantitative predictions, which are contradicted by our IMMOB results. First, the errors predicted in equations (6) and (7) are much too small: on the order of 4° for slant and 7° for tilt, compared to our experimental results in the IMMOB condition of 12° and 40° , respectively. Second, Equation 7 predicts a milder dependence of tilt error on shear for higher slant in IMMOB; instead, averaging tilt error for $\eta=0^\circ$ and 15° and subtracting from the average for $\eta=75^\circ$ and 90° , we found tilt error differences of 14.6° , 18.7° , and 15.6° for slant 30° , 45° , and 60° , respectively. Third, the $\sin\eta$ term in Equation 7 would predict that the slope of the tilt error versus shear curve approach zero as shear approaches 90° , which is not observed in our data. Fourth, Equation 6 would predict that slant errors decrease with shear, but they actually increase significantly.

A different analysis confirms the above result. Despite the controls put on subject motion, there were variations in the total amount of motion in the ACT condition, and therefore in the object motion in the corresponding IMMOB trials (see “Analysis of Movement Trajectories” in “Results”). The change-in-surface-normal hypothesis predicts that absolute tilt error should be positively correlated with total motion, but only in the IMMOB condition (where the object moves) and only for high values of the shear angle (where object movement results in tilt variation: see Equation 7). This prediction is spectacularly contradicted by the data, where we find small positive correlations between total motion for low shear (mean values of 0.36 and 0.07 in ACT and IMMOB, respectively, for shear angle 0°) that decrease as shear angle increases (-0.27 in both conditions for shear angle 90°).

Finally, the change-in-surface-normal hypothesis is contradicted by other recent results. This hypothesis makes no reference to specific depth cues, and should equally apply to, say, the perception of oscillating grids. It does not: Cornilleau-Pérès et al. (2002) show a strong effect of shear on tilt perception in the case of SfM, but no effect at all when grids (texture cues) undergo the same motion.

Thus, the change in the allocentric normal in IMMOB cannot account for our results.

The Stationarity Assumption

It has recently been shown in our laboratory (Wexler et al. 2001a; Wexler et al., 2001b) that a new hypothesis is needed to account for structure-from-motion performance in moving observers: the visual system makes the stationarity assumption; that is, it prefers SfM solutions that minimize motion in an allocentric or earth-fixed reference frame. The stationarity assumption has

obvious computational and ecological advantages. Similar motion-minimization criteria have classically been invoked to account for the perception of 2D movement (Wertheimer, 1912; Wallach & O’Connell, 1953; see Weiss, Simoncelli, & Adelson [2002] for recent work).

Due to the stationarity assumption, the rate of reversals should be much smaller in ACT than in IMMOB: the simulated plane in ACT is stationary and the reversed plane is not, whereas the simulated and reversed planes are equally non-stationary in IMMOB (Wexler et al., 2001a). We have indeed found this to be the case (see Figure 4). (The fact that reversals occur in less than 50% of IMMOB trials indicates that the visual system takes into account second-order optic flow.)

The stationarity assumption would predict that solutions that really are stationary (as always, we mean stationary in an allocentric frame) will be perceived more precisely, because the initial 3D motion estimate will need much less refinement. This prediction is, in fact, borne out by some of our results. When there is no tilt reversal, the solution in ACT is stationary, whereas the solution in IMMOB rotates at the same speed, ω , as the subject did in the corresponding ACT trial (see animation).⁵ Accordingly, we found that, in trials without reversals, tilt errors are smaller in ACT (16.6°) than in IMMOB (23.3°): see Figure 6. When tilt reverses, in IMMOB the reversed solution rotates in the opposite direction but with the same speed, $-\omega$. In ACT, on the other hand, the reversed solution rotates at 2ω (see animation). Accordingly, we find IMMOB tilt errors about the same in reversed trials (25.5°), whereas in ACT they are about twice as high (34.0°) as in unreversed trials. This could mean that observers do not prefer to see only an allocentrically stationary object, but that the computation of its tilt is performed in an allocentric reference frame, contrary to using only retinal data that are egocentric.

On the other hand, the stationarity assumption runs into problems in predicting the effects of shear. At first, all seems well: we take a circle in space with an arbitrary slant and tilt and rotate it by angle α about an arbitrary frontal axis. Let $R_0(\rho, \theta)$ and $R(\rho, \theta)$ be the initial and final positions in 3D space of a point on the circle with 2D polar coordinates (ρ, θ) . When we average the square length of 3D displacements generated by this rotation, we find the following expression

$$\frac{1}{\pi r^2} \int_0^r \rho d\rho \int_0^{2\pi} d\theta |R(\rho, \theta) - R_0(\rho, \theta)|^2 = \frac{r^2}{4} [5 - \cos^2 \sigma - \sin^2 \sigma (2 \cos 2\eta - 1)] \sin^2 \alpha / 2 \quad (8)$$

Equation 8 shows that nonstationarity rises with the shear, η , which would seem to be in agreement with our tilt error results in IMMOB (see Figure 5). However, our virtual objects were not circles in space but in the image

plane, and were then projected onto the simulated surface; therefore, in space, these objects were ellipses. When we perform the above calculation for these elliptical objects (in parallel projection), we find the following mean square displacement:

$$\frac{r^2}{\cos^2 \sigma} \sin^2 \frac{\alpha}{2}, \quad (9)$$

which is independent of shear. (In perspective projection, the first correction to Equation 9 is in second order, which can be safely ignored for our small stimuli.)

Therefore, the stationarity assumption seems to be in agreement with some general features of our data, but not with the dependence of tilt perception on shear.

Note that at first glance, the stationarity assumption resembles the change-of-surface-normal explanation, but the two should not be confused. The stationarity assumption has to do with how plane orientation is extracted from optic flow in moving and immobile observers; the change-in-the-normal hypothesis has to do with how perceived plane orientations are combined.

Optic Flow and Shear

No argument based solely on retinal information could account for the differences in tilt errors in the ACT and IMMOB, as optic flow was held constant across these conditions. Nevertheless, could an optic flow-based model account for the effect of shear, an effect that is significant in both active and immobile conditions?

To study this question we used the well-known model of Longuet-Higgins and Prazdny (1980), which assumes a plane undergoing rigid motion and yields 3D structure and motion from first and second spatial derivatives of optic flow at one point. The details of our calculation are given in Appendix B. Briefly, we assumed that errors are caused by noise in estimating the optic flow derivatives, and that errors in first derivatives are negligible compared to those in second derivatives. We found that, instead of an increase in tilt error with increasing shear angle as in our data, this model predicts a decrease of tilt error. Therefore, at least one well-known SfM model, perturbed in a reasonable way, does not account for the effect of shear on tilt errors, even in the case of the immobile observer.

Linearization of SfM by Self-Motion Information

On a functional level, the problem of simultaneously solving for 3D depth and motion of a moving plane from optic flow is a nonlinear one. Indeed, if we take the origin to be at the eye, \mathbf{R} and \mathbf{T} the rotation and translation of the plane, (x,y) a point on the retina, and Z the z -coordinate of the plane in that direction, we have the following optic flow:

$$u_x(x,y) = \frac{1}{Z} (T_x - xT_z) + R_y x^2 - R_x xy - R_z y + R_y \quad (10)$$

$$u_y(x,y) = \frac{1}{Z} (T_y - yT_z) + R_x y^2 - R_y xy - R_z x + R_x \quad (11)$$

If the flow $u_{x,y}$ is known and the goal is to solve for 3D structure (Z) and motion (\mathbf{R} , \mathbf{T}), Equations 10 and 11 are a complex nonlinear system, due to the \mathbf{T}/Z terms.

However, if the motion is known – for instance, if the object is assumed to be stationary in an allocentric frame and self-motion information about \mathbf{R} and \mathbf{T} is integrated into the process – the reduced SfM problem of solving Equations 10 and 11 for Z becomes linear and therefore simple. In our experiments, any self-motion information would have to be extra-retinal, as optic flow was the same in the ACT and IMMOB conditions. We hypothesize that quantitative, extra-retinal information about self-motion is integrated into the SfM process.

The strong dependence of tilt errors on shear in immobile observers, with performance that approaches chance level for large shear, can be taken as a clue to the complexities of solving a nonlinear problem. According to our hypothesis, the problem is linearized in the active condition, where indeed we find a much-reduced dependence of tilt errors on shear. Furthermore, the sharp rise of tilt errors in reversal trials in the active conditions hints that the visual system assumes that relative motion is due to self-motion; that is, that the object is stationary.

Summary of Theoretical Arguments

In this sub-section, we have shown that the allocentric change in the simulated surface normal in the immobile condition, and its constancy in the active condition, cannot account for the active-immobile differences in tilt error nor for the different effects of shear on this error in the two self-motion conditions. The stationarity assumption by itself does not seem to be able to account for the effects of shear, either. Furthermore, while the active-immobile differences rule out any explanation based solely on optic flow (because optic flow is the same in the two conditions), it might be hoped that traditional models could account for the effect of shear in the immobile condition; however, a perturbed version of the Longuet-Higgins model does not predict the effects of shear in the immobile (or in the active) condition, either.

Finally, although we cannot offer any direct proof, it seems reasonable to assume that the dramatic effect of shear on the perception of surface tilt from optic flow in the immobile condition is due to the difficulties attendant upon solving what is intrinsically a nonlinear problem, or, on the other hand, constitutes evidence that the visual system does not solve this problem with anything like complete generality. The almost complete disappearance of the deterioration of tilt response in the active condition could be taken as a sign of the linearization of the problem from self-motion information and the assumption of object stationarity.

Perception and Influence of Slant

Unsurprisingly, slant responses were higher with higher simulated slant in the stimulus. The responses are far from perfect, however: slant was overestimated for small simulated slant and underestimated for high simulated slant. There was nevertheless a difference between the ACT and IMMOB conditions: the slope of the linear regression between perceived and simulated slant was closer to 1 in the ACT condition.

There was a nonzero correlation between perceived and simulated slant even in the IMMOB condition, where slant was poorly defined (see “[Introduction](#)”). Three explanations come to mind. First, the average speed of optic flow could have been used as a heuristic measure of slant, because, for a given relative motion (either the observer’s or the object’s), the more slanted a plane is, the higher the speeds in optic flow (Todd & Perotti, 1999). Alternatively, second-order information in optic flow could have disambiguated the slant, which is ambiguous in first order. Or, possibly, the visual system uses the amount of deformation flow present in the optic flow as an indicator of the amount of slant (Domini & Caudek, 1999); however, this may not be distinct from the first explanation above. With our experimental design, it not possible to determine which of these mechanisms is at work. We do find, however, that slant perception is better correlated with the simulated slant in ACT than in IMMOB, indicating that, as with tilt perception, active vision increases the accuracy of slant perception.

Simulated slant had a marked influence on the correctness of tilt perception. With an increased slant, subjects generally had smaller errors in tilt estimation, indicating that they were better able to recover the orientation of the stimulus. This is not surprising because tilt is better defined for higher values of slant. For example, the plane normal is mis-estimated by δ in a random direction, the resulting tilt error is of order $\delta \sin\sigma$, where σ is the slant.

Anisotropy With Respect to Movement Direction

The finding that the VERT condition generally had greater tilt errors than HORIZ, combined with some subjects’ indications that up/down movements were harder to perform, raised the possibility that the complexity of the motion task in the VERT condition caused greater errors. There are two ways that this could have happened. First, if the movement was hard to perform, motion trajectories could differ from HORIZ and therefore the visual stimulus would differ. However, we analyzed movement trajectories and found that their differences were not related to differences in the responses. Second, subjects could have been preoccupied with the motor task in the VERT condition, which could have interfered with the main task, namely indicating the

orientation of the stimulus. It seems unlikely that this could be the cause of the VERT-HORIZ effect, as the difference is also present in the IMMOB condition, in which there was no motor task. Although the difference in tilt errors between the two DIRECTION conditions in IMMOB was not significant, the differences were quantitatively similar to the ACT condition. This suggests that the mechanism causing these differences has nothing to do with the movement of the subject or object, but instead is more likely the result of the visual hardware or a cognitive bias. However, because the DIRECTION effect is present but not significant in the immobile condition, we cannot exclude that the effect was due to interference with the motor task.

Summary

This report presents results on the perception of surface orientation during active observer motion. We compare this performance to that in the same subjects experiencing the same retinal stimulation, but while remaining still. The main result is that the error in tilt perception is significantly reduced in the active condition, compared to the immobile condition. Furthermore, perceived slant is better correlated with simulated slant in the active condition. Because the retinal stimulus is the same in the active and immobile conditions, these results demonstrate the contribution of extra-retinal information concerning self-motion to the perception of 3D structure.

With increasing presence of shear in optic flow, the precision of tilt perception decreased: the shear effect. However, the shear effect was severely attenuated in the active relative to the immobile condition. These findings cannot be explained by models based solely on the rigidity assumption, the stationarity assumption, or optic flow differences. We speculate that it could be due to the linearization of the structure-from-motion problem by extra-retinal self-motion information, coupled with the assumption of object stationarity.

Previous comparisons between 3D visual perception in active and immobile observers have involved tasks that are not possible for the latter (such as distance perception) or differences between active and immobile observers due to differing frequencies of choice between two discrete solutions. To our knowledge, this is the first psychophysical result demonstrating a task which can be performed by both active and immobile observers, but which is performed with higher precision in active vision.

Appendix A: Replay of the Active Trial

Here we summarize the algorithm used to generate the same optic flow in the immobile condition as in the active condition. The active observer moves both the head and eye (the other eye being covered) in order to fixate

the allocentrically stationary fixation point: in other words, the eye undergoes a simultaneous translation and rotation. We will use two reference frames: the allocentrically stationary world frame (whose origin is the fixation point, and which is used by default) and the eye frame (whose origin is the center of the eye, and which rotates and translates with eye). When we say that an active and an immobile trial have the same optic flow, we mean that at every moment (in practice, for every monitor frame) during the two trials, the stimulus is the same in the eye frames of the active and the immobile observers.

Note that although we use the term “immobile” throughout, we do not wish to suggest nor do we assume that subjects in the immobile condition remained strictly stationary. What we mean is that while in the active condition, the subject’s movement resulted in motion parallax as if a stationary object were viewed from different vantage points. In the immobile condition, on the other hand, the subject’s movements caused no parallax whatsoever; instead, the subject experienced the same optic flow and parallax as in a corresponding active trial.

In deriving the exact form of the eye’s rotation, we used Listing’s law, namely that if the eye is at position $\mathbf{p}_0 = (0, 0, z_0)$ and fixates the origin and then moves to point $\mathbf{p} = \rho(\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$ while still fixating the origin, it rotates about an axis parallel to $\mathbf{p}_0 \times \mathbf{p}$, although any other rotation would have done equally well (see below). In this law, rotations about the line of sight are neglected. The corresponding rotation matrix is

$$L(\mathbf{p}) = \begin{pmatrix} \cos \theta \cos^2 \phi + \sin^2 \phi & (\cos \theta - 1) \cos \phi \sin \phi & \sin \theta \cos \phi \\ (\cos \theta - 1) \cos \phi \sin \phi & \cos \theta \sin^2 \phi + \cos^2 \phi & \sin \theta \sin \phi \\ -\sin \theta \cos \phi & -\sin \theta \sin \phi & \cos \theta \end{pmatrix} \quad (12)$$

Because L is orthogonal, $L^{-1} = L^T$.

Now, consider that at a given moment during an active trial, the eye is at point \mathbf{p} , having rotated according to Listing’s law from point \mathbf{p}_0 . If \mathbf{r} is a point on the virtual stimulus (in the world frame), where is it in the eye’s frame (\mathbf{r}_e)? Because the eye’s frame is parallel to the world frame when the eye is on the z -axis (i.e., the rotation matrix between them is the identity), we have $\mathbf{r}_e = L(\mathbf{p})(\mathbf{r} - \mathbf{p})$.

At the same moment in the IMMOB trial, the eye is at \mathbf{P} and fixates the origin. Define point \mathbf{R} so that it is at the same position in the eye frame in the immobile trial as point \mathbf{r} was in the active trial. In other words, the changes in the active condition must be the same as in the IMMOB condition, in the eye’s frame:

$$\mathbf{r}_e = L(\mathbf{p})(\mathbf{r} - \mathbf{p}) = L(\mathbf{P})(\mathbf{R} - \mathbf{P}) \quad (13)$$

Equation 13 guarantees the same optic flow in immobile and active trials. Ideally, the observer in the immobile condition should not move, in which case $L(\mathbf{P})$ would be the identity, but as we wanted to correct for any spurious motion in the immobile condition, we need a rotation

matrix here, too. Equation 13 can be easily solved for \mathbf{R} , giving

$$\mathbf{R} = \mathbf{P} + L^T(\mathbf{P})L(\mathbf{p})(\mathbf{r} - \mathbf{p}) \quad (14)$$

In deriving Equation 14, we have made two assumptions, which may be false. First, Listing’s law is not exact, because the eye does rotate about the line of sight. As a consequence, in the active condition, the image could slightly rotate on the retina about the line of sight, while such rotations are removed in the immobile condition. However, such rotations do not contain 3D information and are uninformative for our subjects. The second possible problem would be if subjects did not fixate the fixation point perfectly. This is not accounted for by the rotation matrix, but any discrepancies would result only in rotations about the center of the eye (i.e., wholesale shifts of the retinal image), which again are uninformative about stimulus structure.

Appendix B: Tilt Error in the Perturbed Longuet-Higgins Model

Following the Longuet-Higgins algorithm for a moving plane (Longuet-Higgins, 1984), optic flow can be fitted by second-order polynomials in retinal coordinates:

$$u_x(x, y) = a_{0,1} + a_{1,1}x + a_{1,2}y + a_{2,1}x^2 + a_{2,2}xy \quad (15)$$

$$u_y(x, y) = a_{0,2} + a_{1,3}x + a_{1,4}y + a_{2,1}xy + a_{2,2}y^2 \quad (16)$$

where the coefficients $a_{i,j}$ depend on 3D structure and motion (see Equations 10 and 11). One way of solving the SfM problem, suggested by Longuet-Higgins (1984), is by forming the matrix

$$H = \begin{pmatrix} 2a_{1,1} & a_{1,3} + a_{1,2} & a_{0,1} - a_{2,1} \\ a_{1,3} + a_{1,2} & 2a_{1,4} & a_{0,2} - a_{2,2} \\ a_{0,1} - a_{2,1} & a_{0,2} - a_{2,2} & 0 \end{pmatrix} \quad (17)$$

and solving its eigenvalue problem, $Hv_i = \lambda_i v_i$. If we order the eigenvectors so that $\lambda_1 \leq \lambda_2 \leq \lambda_3$ and normalize them so that $\|v_1\|^2 = (\lambda_3 - \lambda_1)(\lambda_2 - \lambda_1)$ and $\|v_3\|^2 = (\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)$, we have a simple expression for the plane’s normal: $\hat{\mathbf{n}} = v_1 + v_3$ (Longuet-Higgins, 1984).

We can model errors in the above framework by assuming noise in the measurement of the optic flow derivatives in Equations 15 and 16. Because any noise in determining second derivatives $a_{2,i}$ will be much greater than for first derivatives, we can perturb the noise-free matrix (Equation 17) by

$$H' = - \begin{pmatrix} 0 & 0 & a_{2,1} \\ 0 & 0 & a_{2,2} \\ a_{2,1} & a_{2,2} & 0 \end{pmatrix} \varepsilon \quad (18)$$

and expand to first order in ε . Using standard perturbation-theory techniques (e.g., see Courant & Hilbert, 1953), we find the estimated normal \hat{n}_e from the perturbed eigenvectors, and from it the estimated tilt $\tau_e = \arctan n_y / n_x$, obtaining

$$\tau_e = \tau + \left(\frac{\cos^2 \sigma}{1 - \sin^2 \sigma \cos^2 \eta} \right) a_{2,2} \varepsilon + O(\varepsilon^2) \quad (19)$$

where τ is the exact tilt, obtained from Equation 17, and η is the shear angle. Terms proportional to $a_{2,1}$ disappear in the first-order correction to τ_e . Thus, the first-order tilt error “gain” (i.e., the sensitivity of the tilt estimate to noise) is the term in parentheses in Equation 19. The first-order tilt error gain as a function of shear is shown in Figure 8 for slant $\sigma = 30^\circ$, 45° , and 60° .

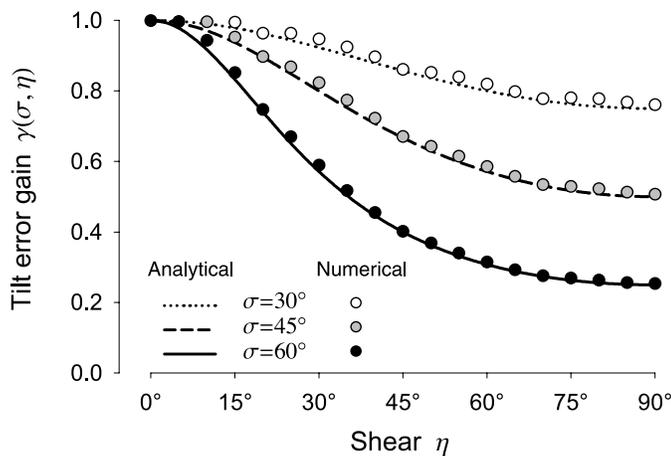


Figure 8. The tilt error gain as a function of shear, for different values of slant. The analytical results are shown as curves, the numerical results as circles.

In order to verify the perturbation-theory result (Equation 19), we also performed a Monte Carlo simulation. In each iteration, the noise parameters $a_{2,i}$ were drawn randomly from a Gaussian distribution with a SD of 0.03 centered around zero. The eigenvalues and eigenvectors of the resulting (perturbed) matrix were calculated, and the tilt error gain was calculated as above. Ten thousand iterations were performed for each slant-shear combination. The average tilt error gain, plotted as points in Figure 8, closely agrees with the analytic result (Equation 19).

Footnotes

- Slant is in principle well defined in perspective projections. However, which decreasing object size perspective projections approach parallel projections in which slant is ambiguous (see preceding paragraph).
- By “shear” (a term that is used in somewhat different ways in the literature), we mean the extent to which optic flow is perpendicular to its gradient. Cornilleau-

Pérès et al. (2002) used the “winding angle” for what we call “shear angle.”

- We also searched for but found no evidence of any oblique effect such as that found in Oomes & Dijkstra (2002).
- The fact that we found systematic errors in tilt perception, whereas other studies have not (Domini & Caudek, 1999; Norman et al., 1995; Stevens, 1983; Todd & Perotti, 1999), may be due to the way we analyzed the data. Most of the above-mentioned studies calculate regression coefficients between perceived and simulated tilts over the entire range of tilt values (i.e., 360°). Consequently, it is hardly surprising that the slope of the regression line one finds is always near unity (remember that tilt is a circularly periodic variable, in contrast to the way slant is normally defined). Such an analysis passes over the possibility of systematic errors with a period of less than 360° , nor is it a very strong indicator of random errors (errors in precision) because these errors are not necessarily uniform over the entire range of tilt. This is precisely what we have found.
- This is only true, of course, if the subject estimates relative motion to the object, and self-motion in an allocentric frame, accurately. There is converging evidence that self-motion is underestimated, but only by about 30%-40% in actively moving subjects (Wexler, in press). Therefore, the above argument still holds.

References

- Batschelet, E. (1981). *Circular Statistics in Biology*. London: Academic Press.
- Braunstein, M. L. (1968). Motion and texture as sources of slant information. *Journal of Experimental Psychology*, 78, 247-253. [PubMed]
- Cornilleau-Pérès, V., Wexler, M., Droulez, J., Marin, E., Miège, C., & Bourdoncle, B. (2002). Visual perception of planar orientation: Dominance of static depth cues over motion cues. *Vision Research*, 42, 1403-1412. [PubMed]
- Courant, R., & Hilbert, D. (1953). *Methods of Mathematical Physics*. New York: Interscience.
- Crowell, J. A., Banks, M. S., Shenoy, K. V., & Andersen, R. A. (1998). Visual self-motion perception during head turns. *Nature Neuroscience*, 1, 732-737. [PubMed]
- Cutting, J. E., & Millard, R. T. (1984). Three gradients and the perception of at and curved surfaces. *Journal of Experimental Psychology: General*, 113, 198-216. [PubMed]

- Dijkstra, T. M., Cornilleau-Pérès, V., Gielen, C. C., & Droulez, J. (1995). Perception of three-dimensional shape from ego- and object-motion: Comparison between small- and large-field stimuli. *Vision Research*, *35*, 453–462. [PubMed]
- Domini, F., & Caudek, C. (1999). Perceiving surface slant from deformation of optic flow. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 426–444. [PubMed]
- Ernst, M. O., Banks, M. S., & Bühlhoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience*, *3*, 69–73, 2000. [PubMed]
- Freeman, T. C. A., & Fowler, T. A. (2000). Unequal retinal and extra-retinal motion signals produce different perceived slants of moving surfaces. *Vision Research*, *40*, 1857–1868. [PubMed]
- Hoffman, D. D. (1982). Inferring local surface orientation from motion fields. *Journal of the Optical Society of America*, *72*, 888–892. [PubMed]
- Julesz, B. (1964). Binocular depth perception without familiarity cues. *Science*, *145*, 356–362.
- Longuet-Higgins, H. C. (1984). The visual ambiguity of a moving plane. *Proceedings of the Royal Society of London (B, Biological Sciences)*, *223*, 165–175. [PubMed]
- Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London (B, Biological Sciences)*, *208*, 385–397. [PubMed]
- Meese, T. S., Harris, M. G., & Freeman, T. C. M. (1995). Speed gradients and the perception of surface slant: Analysis is two-dimensional not one-dimensional. *Vision Research*, *35*, 2879–2888. [PubMed]
- Norman, J. F., Todd, J. T., & Phillips, F. (1995). The perception of surface orientation from multiple sources of optical information. *Perception and Psychophysics*, *57*, 629–636 [PubMed]
- Ono, H., & Steinbach, M. J. (1990). Monocular stereopsis with and without head movement. *Perception and Psychophysics*, *48*, 179–187. [PubMed]
- Oomes, A. H. J., & Dijkstra, T. M. H. (2002). Object pose: Perceiving 3-D shape as sticks and slabs. *Perception and Psychophysics*, *64*, 507–520. [PubMed]
- Panerai, F., Cornilleau-Pérès, V., & Droulez, J. (2002). Contribution of extraretinal signals to the scaling of object distance during self-motion. *Perception and Psychophysics*, *64*, 717–731. [PubMed]
- Panerai, F., Hanneton, S., Droulez, J., & Cornilleau-Pérès, V. (1999). A 6-dof device to measure head movements in active vision experiments: Geometric modeling and metric accuracy. *Journal of Neuroscience Methods*, *90*, 97–106. [PubMed]
- Peh, C.-H., Panerai, F., Droulez, J., Cornilleau-Pérès, V., & Cheong, L.-F. (2002). Absolute distance perception during in-depth head movement: Calibrating optic flow with extra-retinal information. *Vision Research*, *42*, 1991–2003.
- Rogers, B., & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception*, *8*, 125–134. [PubMed]
- Rogers, S., & Rogers, B. J. (1992). Visual and nonvisual information disambiguate surfaces specified by motion parallax. *Perception and Psychophysics*, *52*, 446–452. [PubMed]
- Royden, C. S., Banks, M. S., & Crowell, J. A. (1992). The perception of heading during eye movements. *Nature*, *360*, 583–585. [PubMed]
- Stevens, K. A. (1983). Surface tilt (the direction of slant): A neglected psychophysical variable. *Perception and Psychophysics*, *33*, 241–250. [PubMed]
- Todd, J. T., & Perotti, V. J. (1999). The visual perception of surface orientation from optical motion. *Perception and Psychophysics*, *61*, 1577–1589. [PubMed]
- Ullman, S. (1979). *The interpretation of Visual Motion*. Cambridge: MIT Press.
- von Helmholtz, H. (1867). *Handbuch der Physiologischen Optik*. Hamburg: Voss.
- Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, *45*, 205–217.
- Wallach, H., Stanton, L., & Becker, D. (1974). The compensation for movement-produced changes in object orientation. *Perception and Psychophysics*, *15*, 339–343.
- Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*, 598–604. [PubMed]
- Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie*, *61*, 161–265.
- Wexler, M. (in press). Allocentric perception of space and voluntary head movement. *Psychological Science*.
- Wexler, M., Lamouret, I., & Droulez, J. (2001a). The stationarity hypothesis: An allocentric criterion in visual perception. *Vision Research*, *41*, 3023–3037. [PubMed]
- Wexler, M., Panerai, F., Lamouret, I., & Droulez, J. (2001b). Self-motion and the perception of stationary objects. *Nature*, *409*, 85–88. [PubMed]