PERGAMON

Vision Research 40 (2000) 3873–3894

Vision Research

www.elsevier.com/locate/visres

Perception of heading during rotation: sufficiency of dense motion parallax and reference objects

Li Li, William H. Warren Jr *

Department of Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912, USA

Received 20 January 2000; received in revised form 29 June 2000

Abstract

How do observers perceive the path of self-motion during rotation? Previous research suggests that extra-retinal information about eye movements is necessary at high rotation rates (2–5°/s), but those experiments used sparse random-dot displays. With dense texture-mapped displays, we find the path can be perceived from retinal flow alone at high simulated rotation rates if (a) dense motion parallax and (b) at least one reference object are available. We propose that the visual system determines instantaneous heading from the first-order motion parallax field, and recovers the path of self-motion by updating heading over time with respect to reference objects in the scene. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Eye movement; Heading; Motion parallax; Optic flow; Self-motion

1. Introduction

Optic flow contains information about an observer's path of self-motion that may be useful in perceiving and controlling locomotion. When travelling on a straight path, the optic flow at the moving observation point forms a radial pattern with a focus of expansion in the direction of translation (Fig. 1a). However, this pattern must be detected by a moving eye in a moving head, and such observer rotations alter the flow pattern on the retina, shifting or eliminating the focus of expansion (Fig. 1b and c). Retinal flow is thus made up of two components at any instant in time, a *translational* component of radial flow due to observer translation (T), and a *rotational* component of parallel flow due to observer rotation (R). The problem we consider is how the observer perceives the path of self-motion from such compound flow patterns. There are two interrelated issues: the *rotation problem* concerns how the visual system disentangles the two components of retinal flow to determine the instantaneous direction of translation, or heading[1]; and the *path ambiguity* concerns how the visual system distinguishes a straight path from a curved path of self-motion. Note that the concepts of heading and path must be distinguished: for straight paths they are identical, but for curved paths the instantaneous heading is tangent to the path of self-motion.

1.1. The rotation problem

The rotation problem concerns how the instantaneous direction of translation is determined during a rotation. Two general approaches have been proposed, referred to as *extra-retinal* and *retinal flow* theories. On the first approach, the rotational component of the flow pattern might be determined from extra-retinal information about eye and head rotation (Royden, Crowell, & Banks, 1994). Such information could include efferent signals sent to the extra-ocular and neck muscles,

[1] The American Heritage Dictionary (third edition) defines *heading* as, 'The course or direction in which a ship or aircraft is moving', and we use the term in this common sense to refer to the direction of observer translation. In nautical circles, 'heading' can refer to either the direction a ship is moving, also called the 'course', or the direction it is pointing; although this ambiguity is unfortunate, we retain the term because of its colloquial meaning and ease of use.

* Corresponding author. Fax: +1-401-8633980.
E-mail address: bill_warren@brown.edu (W.H. Warren Jr).

PII: S0042-6989(00)00196-6

proprioceptive information about eye and head rotation, and vestibular information about head rotation. The rotational component could then be subtracted from the retinal flow in order to recover the translational component and the instantaneous direction of heading.

On the second approach, heading might be determined from the retinal flow pattern itself. Numerous computational models in the past two decades have demonstrated that it is formally possible to recover heading from the instantaneous velocity field, also known as the *first-order flow field* (see Hildreth & Royden, 1998, for a review). One class of theories first estimates observer rotation from the motion of distant elements, which is dominated by the rotational component (Perrone, 1992). The rotation is then subtracted from the flow pattern to obtain the translational component. Binocular or perspective depth information could contribute by specifying distant elements (van den Berg & Brenner, 1994a,b). A second class of theories determines the translational component directly from the retinal flow. Such theories usually rely on some aspect of *motion parallax* between elements at different depths[2], which is invariant under rotation. For example, the point of zero parallax in a dense 3-D scene uniquely corresponds to the current heading, and heading is also specified by the surrounding parallax field. Longuet-Higgins and Prazdny (1980) showed that relative motion at depth edges, or *edge parallax*, defines a set of difference vectors that radiate from the heading point. Rieger and Lawton (1985) generalized this observation to elements within a spatial neighborhood, although heading error increases with the angular separation between elements (see also Heeger & Jepson, 1992; Thomas, Simoncelli, & Bajcsy, 1994; Lappe, Bremmer, Pekel, Thiele, & Hoffmann, 1996; Royden, 1997). Information about the depth order of a pair of elements might contribute by resolving the sign of local parallax, allowing one to determine whether one is heading to the right or left of the pair (Cutting, Springer, Braren, & Johnson, 1992; Wang & Cutting, 1999). A third class of theories is based on spatial derivatives of the flow field, which depend on locally smooth surfaces that are densely sampled (Longuet et al., 1980; Koenderink & van Doorn, 1981; Waxman & Ullman, 1985; Poggio, Verri, & Torre, 1991). The retinal flow approach thus offers a formally elegant and purely optical solution to the rotation problem.

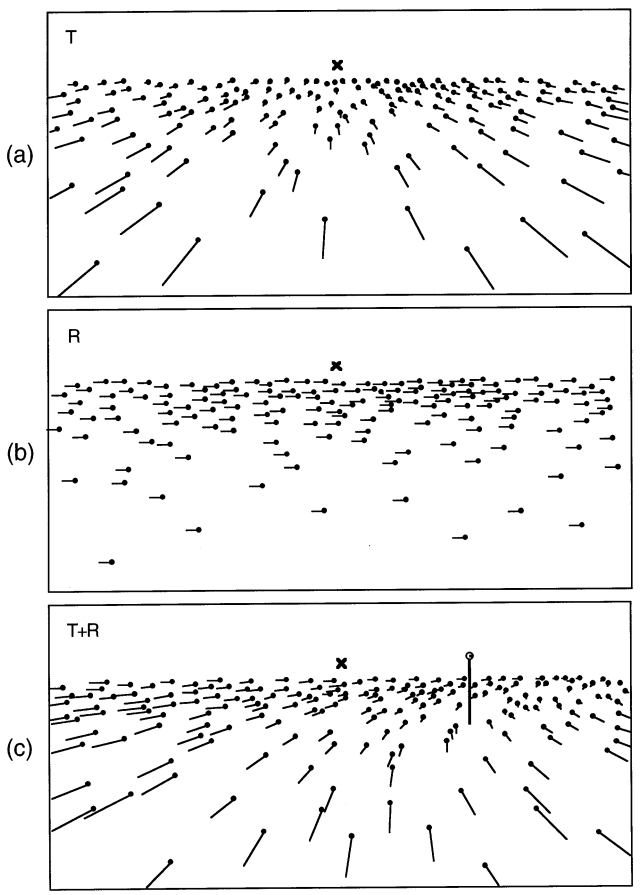

Fig. 1. Retinal velocity field for movement over a ground plane. Each vector represents the retinal velocity of an element in the scene. (a) *Translational component* produced by observer translation (T) toward the X. (b) *Rotational component* produced by eye rotation (R) to the right about a vertical axis. (c) Retinal flow field produced by the sum of translation and rotation, produced by translating toward the X while fixating the O on top of the fixation post.

Finally, it is possible that the visual system exploits both extra-retinal and retinal flow solutions to determine heading, which we call a *hybrid* approach. For instance, the visual system might compare independent extra-retinal and retinal flow estimates of heading (Banks, Ehrlich, Backus, & Crowell, 1996), combine retinal and extra-retinal estimates of rotation (Ehrlich, Beck, Crowell, Freeman, & Banks, 1998), or use an extra-retinal signal to tune selectivity to retinal flow patterns (Lappe, 1998; Beintema & van den Berg, 1998).

It is important to note that most models yield a heading direction in an oculo-centric (i.e. retinal) reference frame, not a head-centric or body-centric reference frame. As Gibson (1950) emphasized, this is sufficient for perceiving one's *object-relative heading* with respect to objects in the scene, defined as the visual angle between the instantaneous heading and a reference object. It thus provides a description of heading with respect to the world. On the other hand, it is inadequate for perceiving one's *absolute heading* in a body-centric frame, and hence for performing body-centered tasks such as pointing with an unseen hand. Extra-retinal information about eye and head position could be used to transform an oculo-centric heading estimate to head- and body-centric frames (Warren, 1998).

[2] Also known as differential motion, relative motion, or for elements on a continuous ground surface, motion perspective (Gibson, Olum, & Rosenblatt, 1955).

1.2. The path ambiguity

Perceiving self-motion is complicated by the fact that a rotational component of retinal flow may be produced not only by eye and head rotation, but also by a curved path of movement in which the body rigidly rotates to face in the direction of travel. In fact, the standard characterization of the flow pattern as an instantaneous velocity field is formally ambiguous: the same velocity field (such as Fig. 1c) can be generated by a straight path (T) plus rotation (R), or by a circular path with radius $\rho = T/R$ (assuming that the direction of gaze remains at a constant angle to the path's tangent), or by various combinations of rotation rate and path curvature. Thus, the first-order flow field is in principle insufficient to disambiguate straight and curved paths of self-motion. Indeed, observers often report seeing curved paths under conditions that simulate translation plus rotation (Royden, 1994).

The extra-retinal approach resolves this ambiguity in a straightforward manner (Banks et al., 1996). The visual system could use extra-retinal information to estimate eye and head rotation and then subtract it from the retinal flow field. If the remaining retinal flow consists of a purely translational component, the observer is on a straight path, whereas if it retains a rotational component, the observer is on a curved path (with the direction of gaze at a constant angle to the path). Determining path curvature from the remaining rotation is a further problem (Warren, Mestre, Blackwell, & Morris, 1991b). One hybrid approach (Ehrlich et al., 1998) first decomposes the retinal flow field to estimate the translation T and total rotation R, then subtracts the extra-retinal estimates of eye and head rotation from the total rotation ($R - R_e - R_h = R_p$). Any remainder R_p is attributed to a curved path of self-motion, allowing an exact estimate of its radius of curvature ($\rho = T/R_p$). Note however, that this theory presumes a retinal flow solution to the rotation problem.

Retinal flow models based on the first-order flow field generally do not attempt to resolve the path ambiguity, but broadening the description of the flow pattern provides sufficient information to do so, in principle. First, Rieger (1983) and Tsai & Huang (1984) showed that straight and curved paths of self-motion are distinguished by the second-order accelerative components of flow. However, human sensitivity to acceleration is quite poor, with Weber fractions on the order of 0.5–1.0 (Schmerler, 1976; Calderone & Kaiser, 1989), making this solution unlikely. Second, as a consequence of acceleration, the velocity fields diverge over time, such that a circular path generates a stationary (unchanging) velocity field whereas translation plus rotation generates a deforming field (Warren, Blackwell, Kurtz, Hatsopoulos, & Kalish, 1991a). However, these velocity differences are small and would be difficult to discriminate over short time intervals (see Ehrlich et al., 1998, for an analysis). Third, the trajectories of individual elements also diverge over time, yielding differences that are small but theoretically discriminable (Warren et al., 1991b; Royden, 1994; Banks et al., 1996). This would require that the visual system track local features over extended time intervals

Finally, the path ambiguity could be resolved by reference objects that are attached to the scene and can be tracked over time. In particular, one or more reference objects would allow the instantaneous heading to be updated over time with respect to constant locations in the scene. Assume that translational heading can be determined from the velocity field. If the observer is traveling on a straight path, the heading point will remain fixed in the environmental layout defined by reference objects. In contrast, if the observer is on a curved path, the heading point will shift in the environment. Another way of exploiting reference objects is based on the observation that the path of self-motion is specified by the change in perspective structure, the changing view of objects as one moves through an environment (Gibson, 1979). This can be formalized in terms of triangulation, by which the relative bearings to a set of three landmarks (or the absolute distances to two landmarks) uniquely determine the observer's position and, over time, the path of self-motion. The most elaborate proposal comes from Vishton and Cutting (1995), who suggested that the visual system uses depth information to create a 3-D mental map of the scene and recovers self-motion from the distances and displacements of landmarks.

In sum, the path ambiguity could be resolved on the basis of flow information alone if the velocity field, element trajectories, or one's heading with respect to reference objects were integrated over time.

1.3. Empirical evidence

Both the retinal flow and hybrid approaches rest on the assumption that the visual system can decompose the flow pattern into its translational and rotational components without extra-retinal information. However, after a decade of empirical research, the data on this point remain controversial (see Warren, 1998, for a critical review).

The rotation problem has been investigated using a technique known as the simulated rotation paradigm (Regan & Beverley, 1982; Rieger & Toet, 1985; Warren & Hannon, 1988). Computer displays depicting self-motion through an environment are presented on a monitor, and the observer indicates the perceived heading or path of self-motion. In the *actual rotation condition*, a radial flow pattern is presented on the screen (Fig. 1a), with a moving fixation point that induces an

active pursuit eye movement; thus, the flow pattern on the observer's retina is produced by the sum of a translation and an actual eye rotation. In the *simulated rotation condition*, the flow pattern on the screen itself simulates the sum of a translation and rotation (Fig. 1c), but the fixation point is stationary so no actual eye movement occurs[3]. The flow pattern on the retina is thus the same in both conditions, while extra-retinal information is manipulated: it specifies an eye rotation in the actual condition, consistent with the rotational component of flow, but it specifies zero rotation in the simulated condition, inconsistent with the flow. If path judgments are as accurate in the simulated condition as they are in the actual condition, this implies that heading and path can be perceived from retinal flow alone, even in the presence of conflicting extra-retinal information. On the other hand, if judgments are markedly less accurate in the simulated condition than the actual condition, this implicates a role for extra-retinal information. Subjects are typically asked to judge their heading, but together with Ehrlich et al. (1998) we are concerned that this task is not well-defined. Specifically, if a curved path is perceived, the tangential heading in the scene varies over the course of a trial and thus the heading estimate depends on when the judgment is made. Royden (1994) observed that heading errors are consistent with a perceived curved path, suggesting that subjects report the perceived path even when subjects were instructed to judge heading. In the present study, we thus explicitly ask subjects to judge their future path.

Warren and Hannon (1988, 1990) initially reported that heading judgments in the simulated rotation condition were comparable to those in the actual rotation condition, with thresholds below 1.5° for displays of a random-dot ground plane and a 3-D cloud of dots but large errors for a frontal plane. This indicated that the direction of translation could be perceived accurately with or without extra-retinal information as long as motion parallax was available. Further, in the simulated rotation condition observers reported the illusion that their eye was actually rotating, consistent with the flow pattern. However, the mean simulated rotation rate over the course of a trial was quite low, less than 1°/s.

In contrast, Royden, Banks, and Crowell (1992) and Royden et al. (1994) tested higher rotation rates (1–5°/s)[4] and obtained large constant heading errors in the direction of simulated rotation. Constant error increased as a function of simulated rotation rate, up to 15° of error at 5°/s, whereas judgments in the actual rotation condition remained accurate. These findings indicated that extra-retinal information about eye rotation may be necessary at high rotation rates, and similar results have been reported for head rotation (Crowell, Banks, Shenoy, & Andersen, 1998). As might be expected from the path ambiguity, Royden (1994) and Ehrlich et al. (1998) showed that such judgments are consistent with a curved path of self-motion, and observers frequently reported this percept. When displays consisted of mixtures of actual and simulated eye rotation, heading errors increased in proportion to the simulated component, as though the portion of rotation not ascribed to a real eye movement was attributed to a curved path (Banks et al., 1996). This line of research is thus consistent with the extra-retinal approach.

At the same time, experiments by van den Berg (1992, 1993) and van den Berg and Brenner (1994a,b) supported the retinal flow approach. Heading thresholds in the order of 2° for a random-dot ground and 4° for a cloud were reported at simulated rotation rates up to 5°/s. In addition, the presence of depth information (elevation angle or binocular disparity) seemed to make judgments more robust to noise, consistent with retinal flow theories that estimate the rotational component from distant motion. However, when Ehrlich et al. (1998) tested a stereoscopic ground plane and cloud, they obtained large path errors, up to 15° at 5°/s. By instructing subjects to judge their direction of heading based on the illusory motion of the fixation point, van den Berg (1996) found heading judgments that were more accurate than path judgments of the same displays. Taking another tack, Stone and Perrone (1997) asked subjects to judge instantaneous heading while traveling on circular paths and reported constant errors below 5° at 4°/s. However, this required extensive practice and a number of naïve observers could not perform reliably.

Using displays of stick-figure trees on a flat-shaded ground plane, Cutting (1986, 1996), Cutting et al. (1992) and Vishton & Cutting (1995), also reported data consistent with a retinal flow approach. However, this finding is tempered by their use of a *nominal heading* task, in which subjects judge whether they are heading to the left or right of the fixated object (see Warren, 1998). One experiment that used a relative heading task at high simulated rotation rates (Cutting, Vishton, Flückiger, Baumberger, & Gerndt, 1997) found significantly smaller errors with the tree display ($\sim$5° at 5°/s) than with a random-dot cloud ($\sim$12° at 5°/s). This result suggests that more complex scene structure may allow more accurate judgments from retinal flow.

In sum, the evidence that heading or path can be perceived from retinal flow at high simulated rotation

[3] The ability of observers to track the fixation point to within 1° over the course of a trial was verified in experiments using an eye tracker (Warren & Hannon, 1990).

[4] 1°/s would be produced by fixating the ground plane 10 m ahead while walking, 5°/s by fixating 4 m ahead.

rates is mixed. On the other hand, accurate judgments are reliably observed in the actual rotation condition, demonstrating that extra-retinal information about eye and head rotation contributes to the perception of self-motion. On our reading, the evidence thus currently favors an extra-retinal approach.

1.4. Motivation for the present experiments

It is important to recognize that the simulated rotation condition is actually a conflict situation, with at least two competing interpretations. On the one hand, the retinal flow over time specifies that the observer is on a straight path relative to reference objects in the scene, and the eye or head must be rotating. On the other hand, the extra-retinal information specifies that eye and head rotation are zero; since the heading point drifts across the screen, the observer must be on a curved path in the body-centric frame. Thus, there is conflicting information about whether the observer is on a straight or curved path. In previous research, extra-retinal information tends to dominate at high rotation rates, leading to the perception of a curved path. However, it remains possible that the class of displays tested thus far has not provided sufficient information for the visual system to determine the instantaneous heading or the path through the scene from retinal flow.

Nearly all of the relevant research has been carried out using minimal random-dot displays. Although random dots provide a broad-band motion signal, they have several inherent limitations. First, such displays necessarily have sparse structure. To avoid bright clustering of dots in the distance, the overall dot density must be kept low, resulting in very few dots in the foreground where velocities are highest. This greatly reduces the density of motion parallax in the display (i.e. the number of differentially moving elements within a neighborhood). Second, also to avoid clustering, random-dot displays have a limited depth range. Ground planes and dot clouds tend to extend no more than 20 m in depth, which acts to reduce the magnitude of motion parallax in the display (i.e. the velocity difference between near and far elements). This might also reduce the accuracy of a rotation estimate based on the motion of distant elements. Third, random dot displays lack properties of natural scenes such as continuous surfaces and dynamic occlusion at depth edges, which could simplify the decomposition of translation and rotation. Fourth, random dots are highly confusable, so the displays do not contain salient features or reference objects that are easily tracked over time. This could make it difficult for the visual system to resolve the path ambiguity.

In the present experiments, we examined the question of whether the path of self-motion can be perceived at high rotation rates from retinal flow alone. Our strategy was to begin by testing displays with a complex 3-D scene structure, including smooth texture-mapped surfaces, dense motion parallax, edge parallax, and salient reference objects. If accuracy in the simulated condition improved, we could then reduce the displays in an effort to identify the effective information. We find that most observers are quite accurate in the simulated condition when dense motion parallax and reference objects are present. This leads us to propose that the visual system solves the rotation problem by relying on the motion parallax field, and resolves the path ambiguity by integrating heading over time with respect to reference objects in the scene.

2. General method

2.1. Displays

Displays simulated observer translation parallel to a ground plane at a fast walking speed of 2 m/s, with a fixation point off to one side. The heading direction was varied along a horizontal axis and the observer's task was to position a probe on the perceived future path at the end of each 1.5 s trial. The fixation point was a red circle attached to an object in the scene at eye level; it only moved horizontally and thus eye rotation was about the vertical axis. In the actual rotation condition, the fixation point moved horizontally on the screen and observers tracked its motion with a pursuit eye movement. Any extra-retinal information thus corresponded to the actual eye rotation. In the simulated rotation condition, the effect of eye rotation on the flow pattern was simulated in the display, while the fixation point remained stationary on the screen. Any extra-retinal information thus corresponded to zero eye rotation, although the display simulated both translation and rotation.

With a fixation point attached to the scene, the rate of eye rotation necessarily increases as the observer travels forward, so we specified the mean rotation rate during a given display (0, ± 1, ± 3, ± 5, or ± 7°/s)[5]. The instantaneous rotation rate ($\dot{\beta}$) is jointly determined by the distance of the fixation point from the observer (D) and its visual angle from the heading direction (β),

$$\dot{\beta} = \frac{T \sin \beta}{D} \quad (1)$$

where T is the observer's translation speed (Nakayama & Loomis, 1974). The mean rotation rate is determined by the average change in the instantaneous rotation rate over a certain time period ($t_2 - t_1$).

[5] Positive values are to the right, negative values to the left of center screen.

$$\bar{\beta} = \frac{\int_{t_1}^{t_2} \dot{\beta}\, \mathrm{d}t}{t_2 - t_1} \quad (2)$$

To decouple the fixation angle β from the mean rotation rate, we used three different final β angles at each mean rotation rate and calculated the corresponding fixation point distances from Eqs. (1) and (2). Specifically, for a mean rotation rate of 0°/s the final β angle was always 0° (with initial fixation distances of 7, 10, and 13 m); for 1°/s the final angles were 5, 10, or 15° (fixation at 10, 20, or 29.6 m); for 3°/s they were 10, 15, or 20° (fixation at 6.6, 10, or 13 m); for 5°/s they were 15, 20, or 25° (fixation at 6, 7.9, or 9.7 m); and for 7°/s they were 20, 25, or 30° (fixation at 5.6, 7, or 8.2 m). Negative rotation rates (to the left) were produced by using negative final angles with the same values.

A trial was created in the following way. The nominal Z axis of the simulated environment was initially perpendicular to the center of the screen. The direction of heading was randomly chosen within $\pm 10°$ about this Z axis. Next the base of the fixated object was positioned on the ground plane at an initial β up to 19.5° (final β up to 30°) from the heading direction, so that it would produce a specified mean rotation rate for the trial. In the actual condition, the heading remained in this screen location while the fixation point moved, generating the effects of an actual pursuit eye movement. In the simulated condition, the 'camera' at the eye point was initially positioned so that the fixation point appeared randomly within $\pm 10°$ of the center of the screen at the start of a trial, and the heading within $\pm 29.5°$ from the center of the screen. This was done to limit the fixation point's initial eccentricity and keep both it and the heading point on screen throughout the trial. The 'camera' then rotated during the trial to keep the fixation point stationary on the screen, simulating the effects of a pursuit eye movement.

The displays were generated on a Silicon Graphics Crimson Reality Engine at a frame rate of 30 Hz, and were rear-projected on a large screen (112° H × 95° V) with a Barco 800 graphics projector with a 60 Hz refresh rate. Observers viewed the screen binocularly from a chin rest at a distance of 1 m, at the display's center of projection. Simulated eye height above the ground plane was 1.6 m, and the nearest point on the ground plane visible at the bottom of the screen was at a simulated distance of 1.45 m. The edges of the screen were in the periphery against a black background in a dark room, minimizing the possibility that they might provide a stationary frame of reference.

2.2. Procedure

On each trial, the first frame appeared for 1 s to allow observers to fixate the fixation point, followed by 1.5 s of motion. The motion then stopped, the last frame remained visible, and a blue probe line (9.1° tall) appeared on the ground at a distance of 10 m. The azimuthal position of the probe could be adjusted along an arc with a 10 m radius using the mouse. Observers were instructed to track the fixation point throughout the trial and, at the end of the trial, to position the probe on their perceived future path, assuming they continued to travel on their present path. The probe and the last frame remained visible until they clicked a mouse button to go to the next trial. To make sure observers understood the task and the response device, they received a set of practice trials before each condition. No feedback was provided on any trial. An experimental session typically lasted less than 1 h.

3. Experiment 1: scene structure

The purpose of the first experiment was to see whether path accuracy during simulated rotation could be improved by more complex 3-D scene structure that would, in theory, provide information supporting the recovery of the path of self-motion. We tested four display conditions of increasing complexity. First, the *random-dot ground plane* (Fig. 2a) replicated previous studies, with display parameters that were matched to the other conditions in the present experiment. It provided a minimal flow field with sparse motion parallax over a restricted depth range of 20 m, with a fixation post providing a single reference object. Second, the *textured ground plane* (Fig. 2b) provided a much denser motion pattern, particularly in the foreground, thereby defining a continuous surface and dense motion parallax. In addition, the depth range of the ground was extended to 120 m, increasing the magnitude of motion parallax in the display. Improved accuracy in this condition could thus be based on motion parallax, estimating rotation from the motion of distant elements, or spatial derivatives of the motion field. Third, the *textured ground + posts* display (Fig. 2c) added granite-textured posts on the ground plane over a depth range of 4–16 m, in such a way that they did not occlude each other during the course of a trial. This could improve accuracy by adding motion parallax at eye level, a local point of zero parallax, or multiple reference objects with perspective structure. Fourth, the *textured ground + tombstones display* (Fig. 2d) instead used wide granite-textured 'tombstones' that occluded each other, adding edge parallax between objects and perspective structure from object surfaces. Note that in all conditions, the fixated object constituted a reference object attached to the ground plane, and the ground plane itself provided static perspective information for relative depth.

If the visual system can determine the current path from retinal flow information, we would expect errors to be small in the simulated condition, despite the

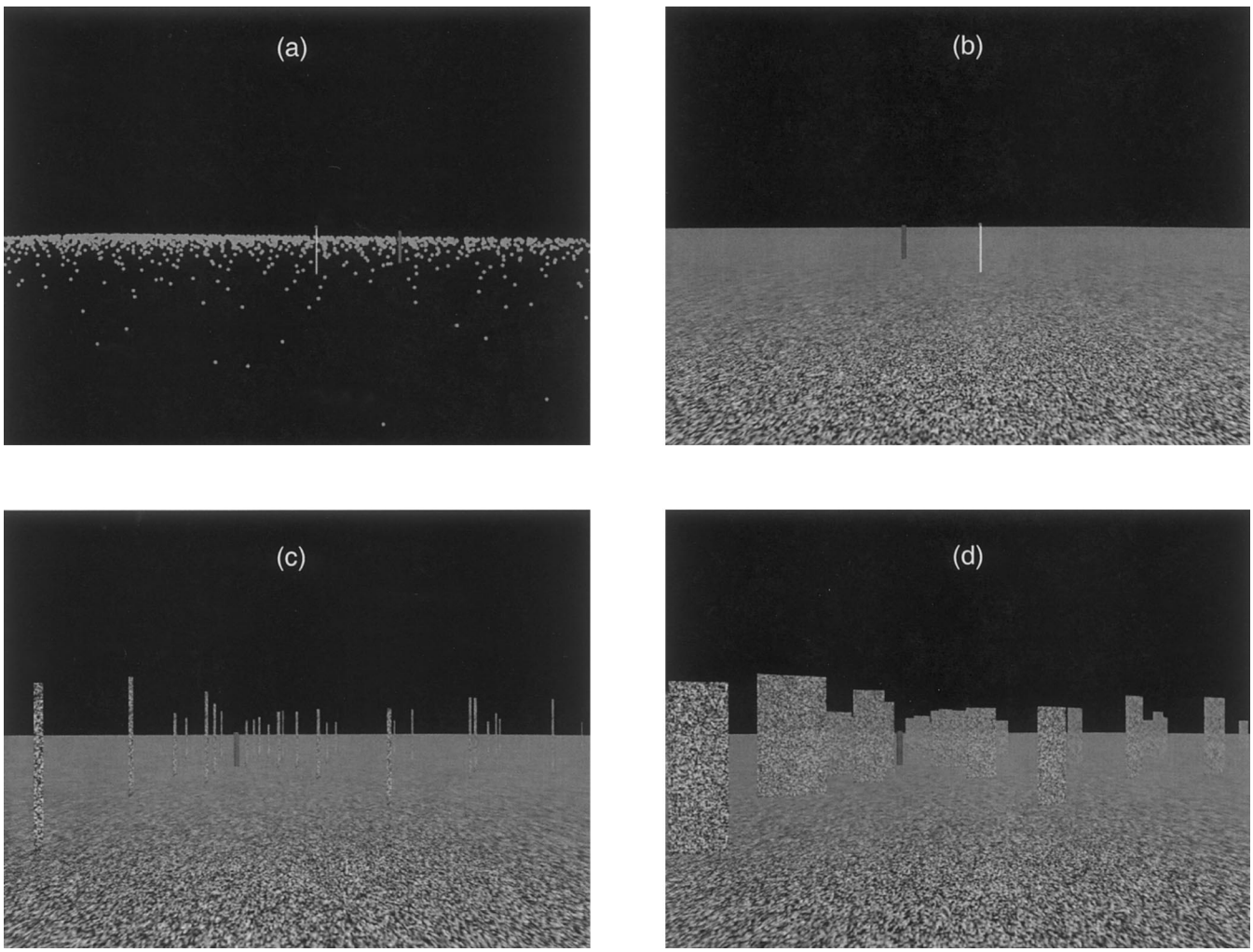

Fig. 2. Scene conditions in Experiment 1. (a) Random-dot ground. (b) Textured ground. (c) Textured ground + posts. (d) Textured ground + tombstones. White line is fixation post, dark line is response probe.

conflicting extra-retinal information for zero rotation. In contrast, if the visual system relies on extra-retinal information, we would expect observers to perceive curved paths of self-motion in the simulated condition, and hence the constant error should increase markedly with rotation rate.

3.1. Method

3.1.1. Participants

Forty-two students and staff at the University of California at Berkeley and Brown University were paid to participate, all with normal or corrected-to-normal vision. Three were removed from the sample because they could not do the task (making large errors in the actual condition), and one was removed due to experimenter error, leaving 38 in the final group. Four at each site were experienced observers who participated in every condition at that site, while the rest were participating in a heading experiment for the first time. At Berkeley, we tested the simulated random-dot condition ($N = 8$); at Brown, we tested the actual random-dot condition ($N = 12$), the textured ground condition ($N = 13$), and the ground + post and ground + tombstone conditions ($N = 13$).

3.1.2. Displays

In the simulated condition, mean rotation rates were 0, ± 1, ± 3, ± 5, and ± 7°/s; in the actual condition, we tested ± 1 and ± 5°/s as a control. These were crossed with the four display conditions: (1) *Random-dot ground*. 700 white dots were distributed on the ground plane (50 m wide × 20 m deep) with a density of 0.7 dots per m^2. The background was black. One dot was positioned in each cell (1.43 × 1 m) of a rectangular grid, randomly jittered from the center of the cell on each trial. Each dot consisted of a 2 × 2 cluster of pixels, and an anti-aliasing routine was used so that the centroid of the cluster moved smoothly over time. The fixation point appeared at the top of a white post that

rested on the ground plane. (2) *Textured ground*. The ground plane (240 m wide × 120 m deep) was texture-mapped with a green Julesz pattern and the sky was black. A 0.94 × 0.94 m texture path was composed of square grid of cells, each cell mapped with one color value randomly chosen from 256 shades of green. The image was anti-aliased with a mipmap-bilinear minification filter. The fixation point appeared at the top of a white post. (3) *Ground + Posts*. Forty five gray granite-textured posts were added on the textured ground surface. The posts were planar, varied randomly in height (2–3 m) and width (0.1–0.15 m). They were positioned in five rows spanning a depth range of 4–16 m, within a visual angle of 170° from the initial observation point. On each trial, the distance between rows varied randomly from 2 to 4 m, and the positions in each row were computed so that the posts would not occlude one another at any point during a trial. To prevent observers from using their obliquity as a cue, each post was randomly rotated out of the frontal plane by an angle of −20 to 20° about the vertical axis. The fixation point appeared on one of the textured posts. (4) *Ground + Tombstones*. Planar granite-textured tombstones were placed on the textured ground plane in the same locations as the posts. Tombstones varied randomly in height (2–3 m) and width (0.9–1.0 m) on each trial, and often occluded one another in the course of a display, creating strong edge parallax. The fixation point appeared on one of the tombstones.

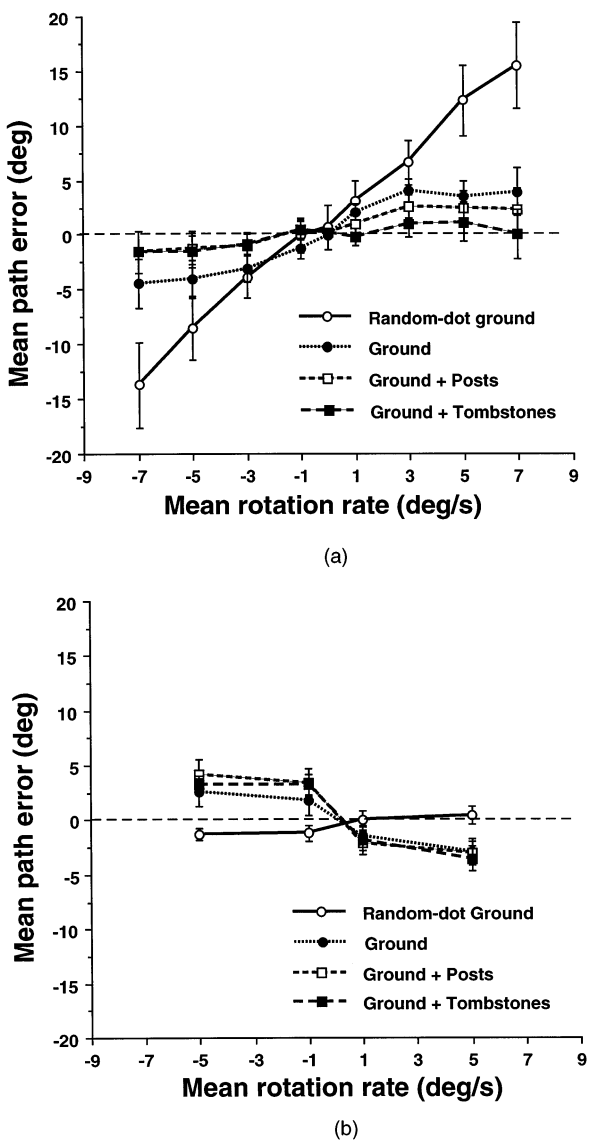

Fig. 3. Mean path error as a function of mean rotation rate during a trial, for the four scene conditions in Experiment 1. (a) Simulated rotation condition. (b) Actual rotation condition. Error bars represent between-subject S.E.

3.1.3. Procedure

For the three texture-mapped displays, each observer participated in both the actual and simulated rotation conditions in a counterbalanced order. For the random-dot displays, separate groups were run in the actual and simulated conditions. The simulated condition was tested at Berkeley using a Power Macintosh 9500/132 at a frame rate of 75 Hz, and presented with an Electrohome ECP-4000 video projector (69° H × 59° V, viewing distance 70 cm); otherwise conditions were the same. Each observer received 54 practice trials, followed by 270 test trials in the simulated condition (30 at each mean eye rotation rate) and 120 trials in the actual condition. Trials were blocked by display condition and simulated/actual condition, randomized within blocks.

3.2. Results

Mean constant path error in the simulated condition is plotted as a function of mean rotation rate in Fig. 3a. A flat function indicates that path judgments were unaffected by rotation rate, whereas a positive slope indicates that error increased in the direction of rotation, and a negative slope that error increased in the opposite direction. The data for the random-dot ground replicate Royden et al.'s (1992, 1994) finding of large errors in the direction of rotation, up to 15° at high simulated rotation rates. This is consistent with the perception of a curved path of self-motion, which subjects frequently reported.

On the other hand, path errors were much smaller with the texture-mapped displays, remaining below 4° at all rotation rates, and below 2° in the ground + tombstones condition. A multivariate regression analysis reveals that the slope in the textured ground condition (0.72) is significantly shallower than that in the random-dot ground condition (2.04), $t(415) = -7.20$, $P < 0.001$. There are also indications that adding more structure to the scene reduces errors further, for the slope in the ground + posts condition (0.34) is

significantly smaller than for the textured ground, $t(415) = -2.38$, $P < 0.02$. There is no statistical difference between the ground + posts and the ground + tombstones (0.18) conditions, $t(415) = -0.97$, *ns*. Separate analyses show that the ground + tombstones slope is not significantly different from zero, $t(165) = 1.78$, *ns*, and is thus statistically flat, although both the ground + posts, $t(165) = 3.36$, $P < 0.001$, and the textured ground, $t(165) = 6.72$, $P < 0.001$, are greater than zero.

In the actual rotation condition (Fig. 3b), the slope for the random-dot ground (0.19) is not statistically different from zero, $t(196) = 1.27$, *ns*, and it is significantly smaller than the slope in the random-dot simulated condition $t(116) = 13.49$, $P < 0.001$. This accuracy during actual eye movements confirms the contribution of extra-retinal information in determining the rotation. Curiously, for the texture-mapped displays there are mean negative slopes in the actual condition: the textured ground (-0.59), $t(165) = -3.1$, $P < 0.01$; the ground + posts (-0.80), $t(165) = -4.51$, $P < 0.001$; and ground + tombstones (-0.75), $t(165) = -4.22$, $P < 0.001$, are all statistically different from zero. This indicates that some observers were overestimating the visual angle between the direction of heading and the direction of fixation by a few degrees in the actual condition. If this were also true in the simulated condition, it would artificially reduce the path error. Note, however, that these mean negative slopes were primarily due to the performance of only two observers (Fig. 4), and have not recurred in our subsequent experiments.

Nevertheless, to control for this possible source of bias, we subtracted each observer's path error in the actual condition from his/her error in the corresponding simulated condition (at rotation rates of -5, -1, $+1$, and $+5°$/s), and recomputed the regression analysis on these mean difference scores[6]. The slope for the random-dot ground condition was still significantly greater than that for the textured ground condition, $t(180) = -2.15$, $P < 0.05$. Thus, accuracy during simulated rotation improves for the texture-mapped displays even when this possible bias is removed. There was no longer a difference in slope between the textured ground and the ground + posts, $t(180) = -0.79$, *ns*, or between the textured ground and ground + tombstones, $t(100) = 0.15$, *ns*, suggesting that more complex scene structure may not improve performance further.

Path errors for individual subjects are represented by separate curves in Fig. 4. We observed no systematic differences in performance between experienced and naïve observers. However, there are notable individual differences in the simulated condition. A number of observers have fairly flat functions, whereas others exhibit a range of slopes. Note that this is the case for both random dot and texture-mapped displays. This effect is consistent with a conflict situation that may be resolved differently by different individuals.

3.3. Discussion

First, these results confirm previous findings that extra-retinal information can contribute to the perceived path of self-motion. For random-dot displays, high rotation rates induced large path errors in the simulated condition, but accurate performance in the actual condition, when extra-retinal signals were concordant with retinal flow. Second, however, the results also demonstrate that the path of self-motion can be determined from retinal flow information alone, even at high rotation rates. In the simulated condition texture-mapped displays significantly improved accuracy in most observers, to the point that for the ground + tombstones display there is no statistical effect of rotation rate. Thus, given sufficient optical structure, the visual system can determine the path of self-motion from retinal flow to within a few degrees, even at high rotation rates and with conflicting extra-retinal information. In the balance of this article, we seek to determine the optical information upon which such a retinal flow solution is based.

The improvement is largely due to replacing a sparse random-dot ground plane (20 m in depth) with a densely textured ground plane (120 m in depth). The addition of multiple objects on the ground (posts or tombstones) yielded only small further improvements, indicating that edge parallax, dynamic occlusion, and complex perspective structure are not essential. On the other hand, at least one reference object (the fixation post) was present in all displays, suggesting that reference objects may play a role.

These results are consistent with four hypotheses regarding the rotation problem. (1) The observer may base a heading estimate on the pattern of radial flow that remains in the foreground. We test this hypothesis in Experiment 2. (2) The visual system may determine rotation from the motion of distant elements. Increasing the depth range from 20 to 120 m may allow a better estimate of rotation rate because more distant elements are less influenced by the translational component. This hypothesis will be pursued in Experiment 3. (3) The visual system may make use of spatial derivatives of the continuous flow field defined by the ground plane. This approach assumes locally smooth surfaces and will be tested in Experiment 3. (4) Finally, the visual system may rely on dense motion parallax to

[6] With the random-dot display, we subtracted the *mean* actual error from each observer's simulated error because different groups were tested in these two conditions. This is conservative because it does not remove between-subject variance from the Random-dot difference scores.

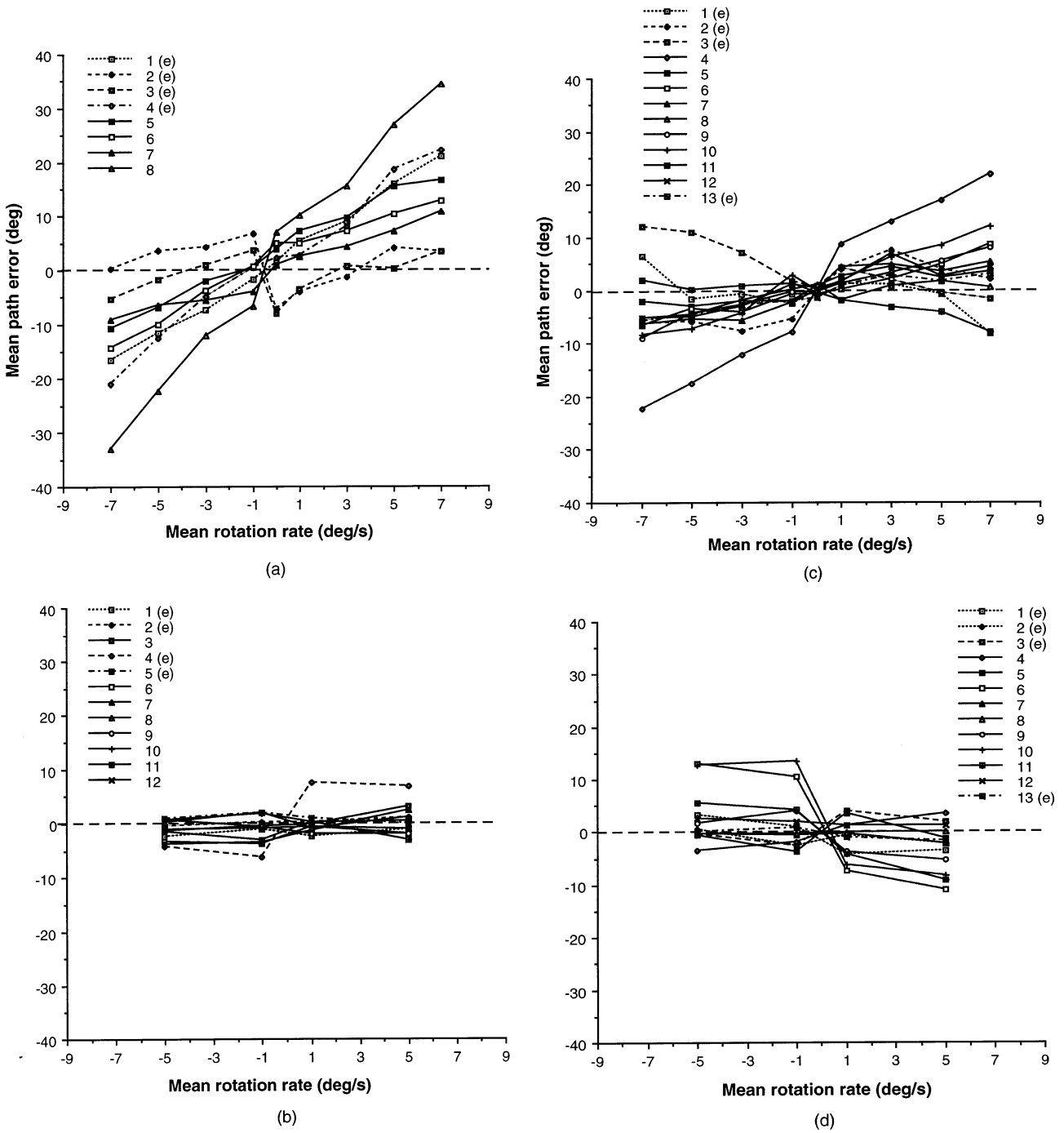

Fig. 4. Individual path error (Experiment 1). Each curve represents the data from one subject. Dotted lines are of the experienced subjects (e), and solid lines are of the naïve subjects. (a) Random-dot ground, simulated rotation. (b) Random-dot ground, actual rotation. (c) Textured ground, simulated. (d) Textured ground, actual. (e) Textured ground + posts, simulated. (f) Textured ground + posts, actual. (g) Textured ground + tombstones, simulated. (h) Textured ground + tombstones, actual.

determine the instantaneous heading. The textured ground provides a much denser parallax field, particularly in the foreground where random-dots are very sparse, and its greater depth range enhances the magnitude of motion parallax in the display. In Section 7, we describe a simulation based on Rieger and Lawton (1985) differential motion algorithm which demonstrates that use of local motion parallax alone can account for the present pattern of results.

It is important to recognize that there were marked individual differences in both the random dot and texture-mapped conditions. Some subjects showed very little influence of simulated rotation, while others had large slopes. We believe this reflects the inherent ambi-

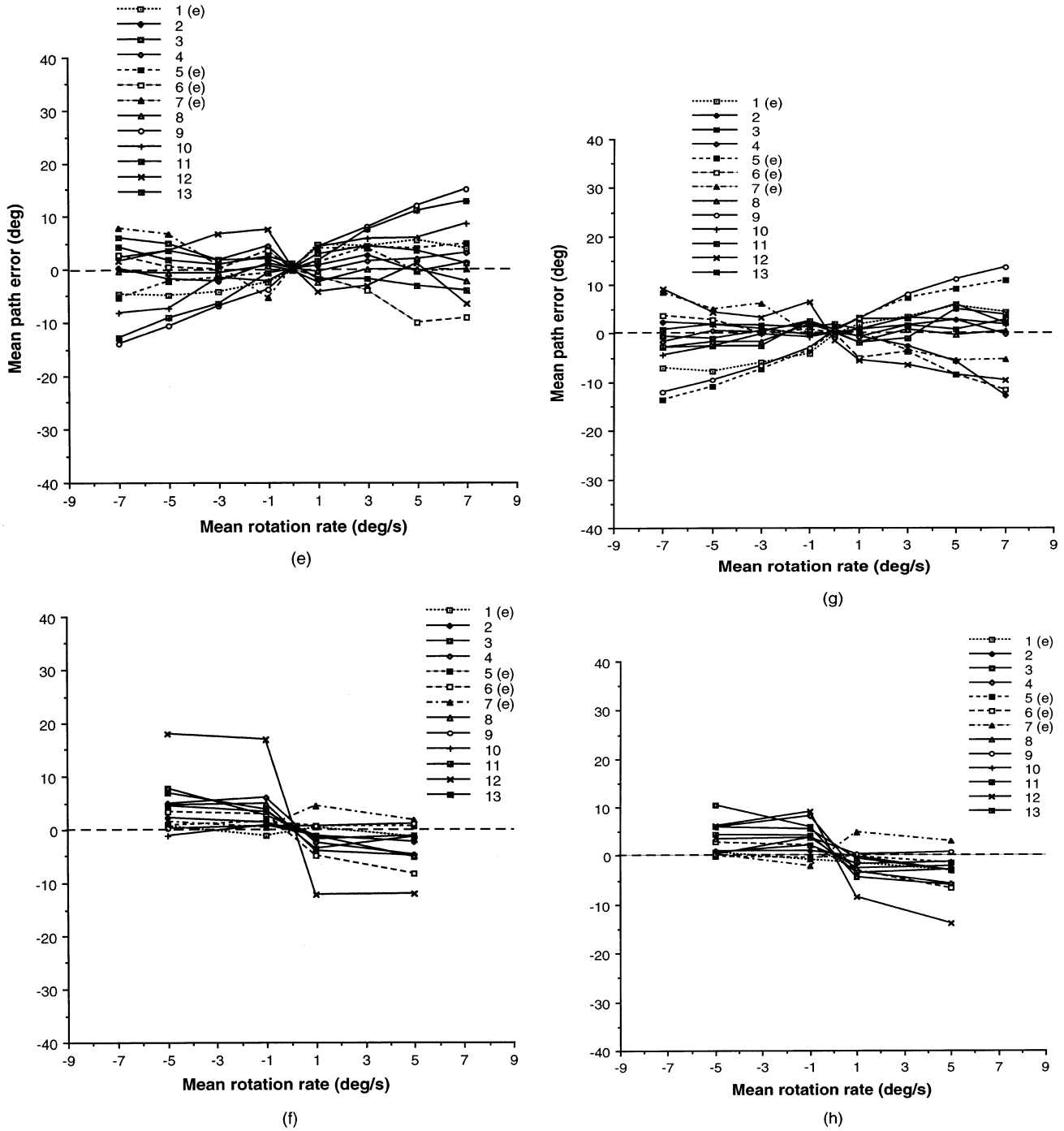

Fig. 4. (*Continued*)

guity in the instantaneous velocity field between straight and curved paths of self-motion, suggesting that the visual system relies on the first-order optic flow and not higher-order components. Individual differences might reflect differential reliance on retinal information (such as reference objects) or extra-retinal information to resolve the path ambiguity, at least under these conflict conditions.

4. Experiment 2: texture bands

During simultaneous translation and rotation, a radial flow pattern remains in the foreground. As noted earlier, the magnitude of the translational component decreases with distance from the observer (Fig. 1a), whereas the rotational component is independent of distance (Fig. 1b). Thus, the rotational component

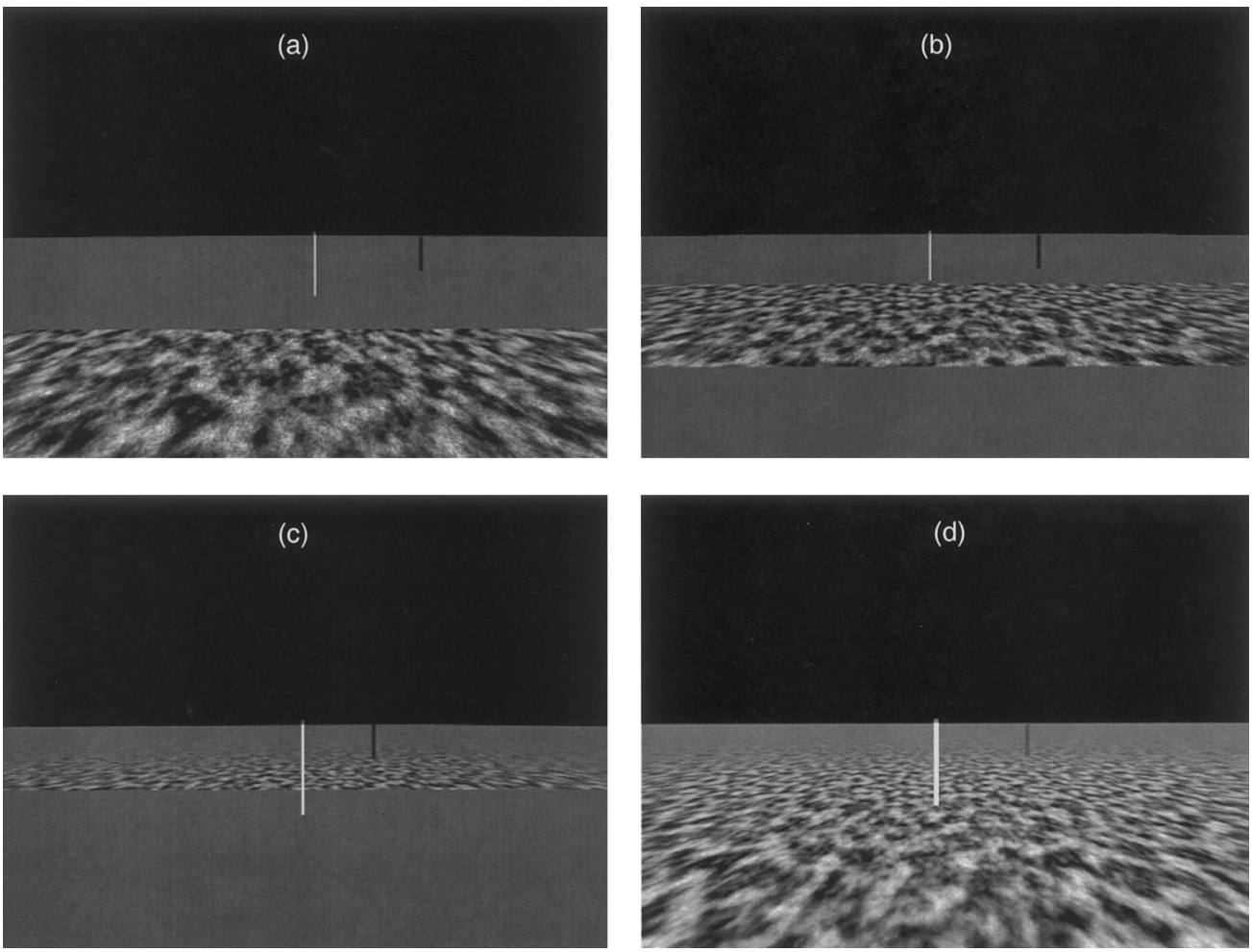

Fig. 5. Texture band displays used in Experiment 2. (a) Near band (1.4–3.5 m). (b) Middle band (2.5–6.5 m). (c) Far band (5–120 m). (d) Full ground (1.4–120 m).

dominates in far space, but the translational component dominates in near space, creating a pattern of radial foreground flow (Fig. 1c). If observers adopted a simple strategy of attending to this foreground flow and its implicit focus of expansion, path judgments could remain reasonably accurate during simulated rotation, slightly biased in the direction of rotation. With the textured ground there is a much denser motion field in near space than with the random-dot ground, so radial foreground flow might account for the improved performance.

To test this hypothesis, we presented texture in a horizontal band on the ground plane, spanning four different depth ranges (Fig. 5): the *near band*, the *middle band*, the *far band*, and the *full ground*. The rest of the ground plane was shaded a flat green. If observers rely on foreground flow, path errors in the simulated condition should be low for the full ground and near band, but increase with the middle and far bands.

4.1. Method

4.1.1. Participants

Sixteen observers were paid to participate, all students or staff at Brown. Three observers were removed from the sample, two who reported arbitrary strategies to do the task during debriefing, and one with large errors in the actual condition. This left 13 observers in the final group, four of them experienced subjects who had participated in Experiment 1 and the rest naïve. We observed no systematic differences between experienced and naïve observers.

4.1.2. Displays

Five mean rotation rates were tested in the simulated rotation condition (0, ± 3, and ± 5°/s), and two in the actual rotation condition (± 5°/s). These were crossed with four texture band conditions, which covered roughly the same visual angle. (1) In the *near band* display, the ground plane was textured in a depth range of 1.4–3.5 m, which subtended a visual angle of 22.9°

vertically. (2) The *middle band* was textured from 2.5 to 6.5 m (18.8°). (3) The *far band* was textured from 5 to 120 m (17.0°). (4) The *full ground* was completely textured from 1.4 to 120 m (45°). In this and the following experiments, we replaced the previous ground texture, which was composed of elements at only one spatial scale, with a multi-scale texture that presents visible structure over a larger depth range. The texture map was composed of a filtered noise pattern with a power spectrum of $1/f^2$ for the range of frequencies from 8 to 32 cycles per texture path. The untextured portion of the ground plane was shaded a flat green.

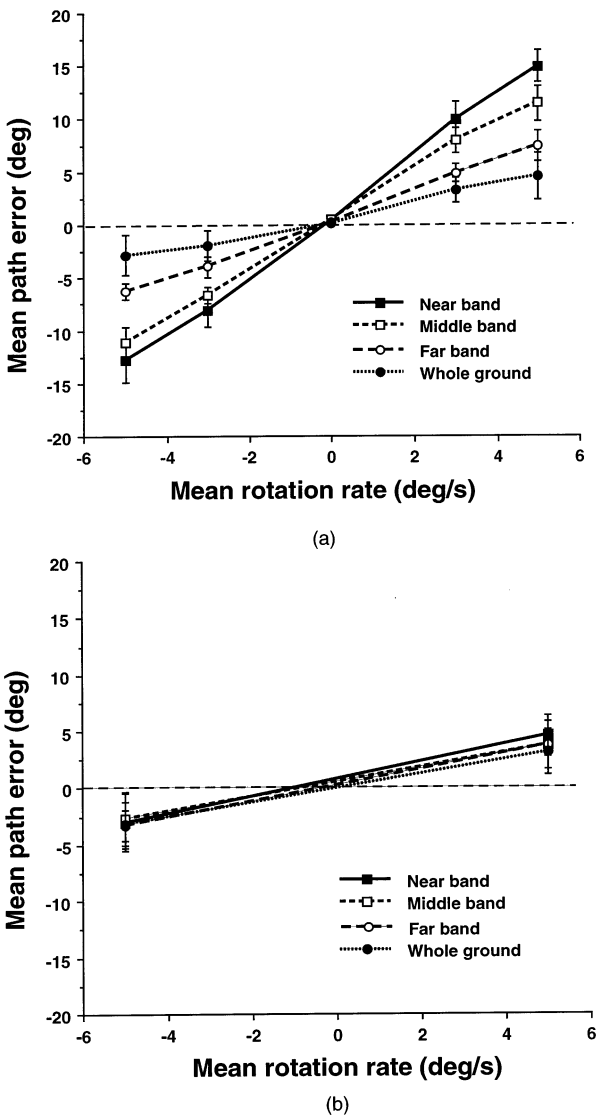

Fig. 6. Mean path error as a function of mean rotation rate for the texture band conditions (Experiment 2). (a) Simulated rotation condition. (b) Actual rotation condition. Error bars represent between-subject S.E.

4.1.3. Procedure

Each observer received ten practice trials followed by the four displays in a counterbalanced order. Trials were blocked by display type and viewing condition and randomized within blocks. One group ($N = 9$) was tested in the simulated condition, with 105 trials for each display (21 at each mean rotation rate), yielding a total of 430 trials. A control group ($N = 5$) was tested in the actual condition, with 42 trials for each display (21 at each mean rotation rate), a total of 178 trials.

4.2. Results

Mean path errors in the simulated condition appear in Fig. 6a. Directly contrary to the foreground flow hypothesis, performance is actually worst for the near band (slope, 2.84), with errors close to 15° at 5°/s, but steadily improves with the middle band (2.31), far band (1.40), and full ground (0.77). A multivariate regression analysis shows that these successive slopes are significantly different, near versus middle, $t(172) = -2.54$, $P < 0.05$; middle versus far, $t(172) = -4.37$, $P < 0.001$; and far versus full, $t(172) = -3.02$, $P < 0.01$. Performance is quite accurate for the full ground display, with path errors smaller than 4° at a simulated rotation of 5°/s. Separate analyses indicate that this slope is significantly different from zero, $t(51) = 4.51$, $P < 0.001$, but not statistically different from the corresponding actual condition, $t(51) = -0.44$, *ns*. This confirms that retinal flow is sufficient to perceive the path of self-motion under simulated rotation.

In the actual condition (Fig. 6b), there are no significant differences between the texture bands. The functions in all conditions are fairly flat, with errors less than 4° at rotations of 5°/s. This again confirms a contribution of extra-retinal information.

4.3. Discussion

The results replicate the finding of Experiment 1 that a textured ground surface with a single fixation post is sufficient for observers to judge their path quite accurately from retinal flow alone. Moreover, the data contradict the hypothesis that this improved performance is due to a simple strategy of using radial foreground flow, for path error with the near band is as large as that with the previous random-dot ground, and is reduced with the middle and far bands.

One account of this reversed pattern of results is that path judgments improve with the depth range of the texture band, consistent with the motion parallax hypothesis. However, the magnitude of parallax within a texture band actually decreased from the near to the far band. A second explanation is that rotation is estimated from the motion of distant elements, and is hence performance is increasingly accurate with the near,

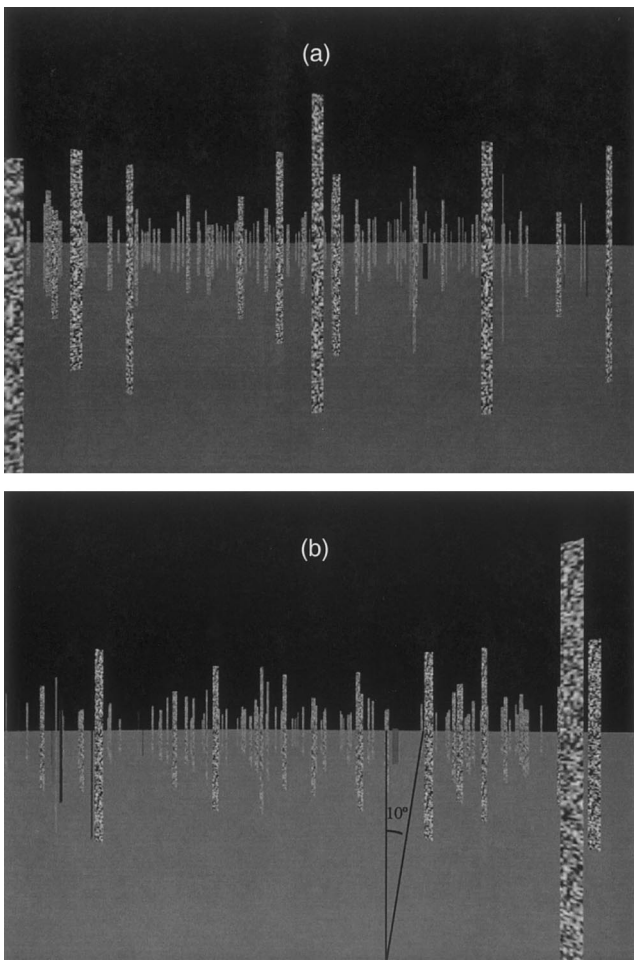

Fig. 7. Dense posts displays used in Experiment 3. (a) Dense posts. (b) Dense posts with 10° wedge of posts removed.

middle, and far bands. However, this does not account for the difference between the far band and full ground, which extend to the same depth (120 m) and should yield the same rotation estimate. A third possibility is that the fixation post never appeared in the near band, occasionally appeared in the middle band, and always appeared in the far band for some portion of the trial, thereby providing a reference object increasingly anchored to the ground texture. This hints at a reliance on reference objects for accurate judgments under simulated rotation, which we test in Experiment 4. Finally, it is possible that observers use a combination of distant elements to estimate rotation and radial foreground flow to estimate translation, which would account for the improvements with far band and the full ground, respectively. In the next experiment, we investigate motion parallax and rotation estimation directly.

5. Experiment 3: dense posts

Thus far, we have observed quite accurate path judgments during simulated rotation with a continuous textured ground plane. This result is consistent with theories based on motion parallax, rotation estimation from distant motion, or spatial derivatives of a continuous motion field. In order to test whether the perceived path is uniquely dependent on a continuous ground surface, in Experiment 3 we created a highly discontinuous 3-D scene consisting of a dense array of vertical posts, spanning a depth range of 1–18 m (Fig. 7a). This dense post display contained motion parallax, edge parallax, and multiple reference objects. It also presented static and dynamic depth information in the form of elevation angle, relative size, and dynamic occlusion. The results would have implications for the three hypotheses.

First, if judgments are based on motion parallax and/or reference objects, we would expect accurate performance under simulated rotation. Second, if the visual system relies on spatial derivatives of continuous fields, which assume locally smooth surfaces, we might expect a decline in performance. Third, to test the rotation estimation hypothesis, the depth range of the posts was restricted to 18 m, slightly less than that of the random-dot ground (20 m). If the visual system uses the motion of distant elements to estimate the rotation, we would expect large errors in the simulated condition comparable to those for the random-dot display.

In a second manipulation, we also tested the necessity of a local point of zero parallax. When moving through a 3-D scene, the relative motion between neighboring elements at different depths goes to zero along the heading axis, such that neighboring flow vectors have the same direction and magnitude. To eliminate this zero point, we removed all the posts from a horizontal wedge in the scene, subtending a visual angle of 10° (Fig. 7b). This did not noticeably alter the appearance of the display. On half the trials, the heading was randomly positioned within the wedge, so no point of zero parallax was defined. On the other half, the heading was located outside the wedge, so a zero parallax point was present. If the visual system relies on a point of zero parallax, we expect path judgments under simulated rotation to be better for the *outside-the-wedge* condition than the *in-the-wedge* condition. Conversely, if the visual system can use the surrounding motion parallax field to determine heading, we would expect little difference between conditions.

5.1. Method

5.1.1. Participants

Ten observers were paid to participate, all Brown students or staff. One observer reported using arbitrary

strategies to do the task during debriefing and was thus removed from the sample. This left nine observers in the final group, including three experienced observers who had participated in Experiments 1 and 2.

5.1.2. Displays

Displays consisted of granite-textured posts on a flat-green ground plane with a horizon at 120 m (Fig. 7). The sky was black and a red fixation point was attached to one of the posts. Each post was 0.11 m wide, varied randomly in height between 2 and 3 m, and was randomly rotated out of the frontal plane from -35 to $+35°$ about its vertical axis. There were three display conditions: (1) The *dense posts* condition contained 231 posts positioned in seven rows of 33 posts each, spanning a depth range of 1–18 m. The distance between rows varied randomly from 1.5 to 4.5 m, and the distance between posts within a row varied randomly from 0.4 to 1.2 m. (2) The *in-the-wedge* condition was the same as the dense posts, except that posts were removed from a 10° horizontal wedge (with its apex at the eye) on one side of the screen. The heading direction was randomly located within the wedge. (3) The *outside-the-wedge* condition was identical to the previous display, except that the heading direction was located outside of the wedge in a symmetrical region on the opposite side of the screen.

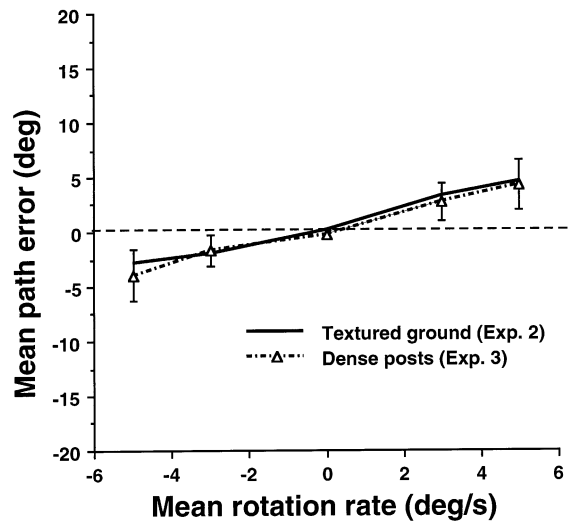

Fig. 8. Mean path error as a function of simulated rotation rate for the dense post display (Experiment 3) and the textured ground display (Experiment 2). Error bars represent between-subject S.E.

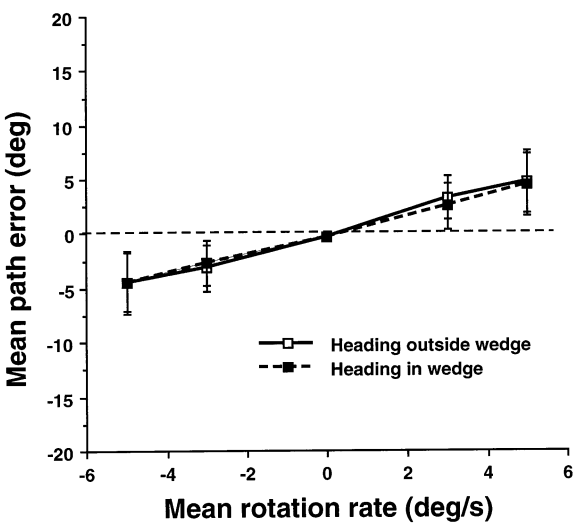

Fig. 9. Mean path error as a function of simulated rotation rate when heading is in the wedge or outside the wedge (Experiment 3). Error bars represent between-subject S.E.

5.1.3. Procedure

The simulated rotation condition was tested using the same rotation rates as in the previous experiment. Trials from the *in-the-wedge* and the *outside-the-wedge* condition were randomized together. Each observer viewed the dense post and wedge conditions in a counterbalanced order, with ten practice trials before each condition. There were 105 test trials in each display condition, for a total of 335 trials in a session.

5.2. Results

First, compare the mean path error in the dense posts condition with that from the full textured ground in Experiment 2 (Fig. 8). Errors remained low, less than 4° at 5°/s, with nearly identical slopes for the dense posts (0.80) and textured ground (0.77), $t(86) = 4.15$, *ns*. This demonstrates that a continuous textured ground plane is not necessary to perceive one's path during simulated rotation, for the same level of accuracy occurs with a discontinuous array of posts.

Next, compare mean path error in the wedge conditions (Fig. 9). Performance when heading in-the-wedge (slope, 0.88) is just as accurate as when heading outside-the-wedge (0.95), $t(129) = -0.21$, *ns*, and is comparable to the dense post condition, $t(129) = 0.46$, *ns*. This indicates that a local point of zero parallax is not needed to recover heading under simulated rotation.

5.3. Discussion

The results demonstrate that the path of self-motion can be perceived just as accurately with a discontinuous 3-D scene as with a continuous textured ground plane during simulated rotation. This implies that smooth surfaces and continuous flow fields are not necessary to decompose translation and rotation, casting doubt on theories based on spatial derivatives of the motion field. Second, we find a high level of accuracy with posts that span a depth range of only 18 m, whereas the random-dot ground with a comparable depth range (20 m) yielded large heading errors (Experiment 1, simulated

condition). This discrepancy casts doubt on the view that the visual system relies on the motion of distant elements to estimate the rotation.

Third, the findings are consistent with the use of dense motion parallax. Although there were more dots (700) than posts (231) in these displays, the texture-mapping of the posts defined many more moving elements over a larger region of the image, particularly in near space, providing denser motion parallax within smaller spatial neighborhoods. The simulation reported below confirms that these differences in motion parallax can, in principle, account for better performance with the dense posts than the random dot ground. The data from the wedge conditions further indicate that a local point of zero parallax is not essential, for path judgments are just as accurate whether or not a zero parallax point was present. This implies that the direction of heading does not depend on a local feature of the flow pattern but can be derived from the surrounding motion parallax field.

In sum, we believe the key difference between texture-mapped and random-dot displays is the density of motion parallax between sufficient numbers of near and far elements. A dense motion parallax field could allow the visual system to determine the instantaneous heading with or without concurrent rotation, potentially solving the rotation problem. This does not, however, uniquely determine the future path of the observer, so we now turn to the question of the path ambiguity.

6. Experiment 4: reference object

All displays tested thus far contained at least one reference object that could be tracked over time — the fixation post. Assuming that the instantaneous heading can be determined from retinal flow, the heading could be updated over time with respect to this constant reference point in the environment. This would allow the visual system to integrate a sequence of instantaneous headings into a path of self-motion through the scene, thereby resolving the path ambiguity. For example, if one is traveling on a straight path, the direction of translation will remain fixed with respect to the environmental layout defined by the fixation post, whereas if one is traveling on an equivalent curved path, the instantaneous direction of translation will drift closer to the post as one approaches it. With an array of objects such as the dense post display, the instantaneous heading in the image plane will remain fixed with respect to objects near a straight path, whereas it will drift with respect to all objects if one is on a curved path.

In Experiment 4, we manipulated the presence of the fixation post to determine whether a reference object is necessary for accurate path judgments during simulated rotation. The *fixation post* condition replicated our previous textured ground display. The fixation post appeared to be located in the scene at the place where its base occluded the ground surface (Gibson, 1950). During a trial, the direction of translation over the ground plane remained constant with respect to this location, consistent with a straight path of self-motion. In the matched *no post* condition, the fixation post was removed so the fixation point floated just above the horizon, as in many earlier experiments (Royden et al., 1992, 1994). Without a reference object at a fixed location in the scene, we would expect observers to be subject to the path ambiguity.

6.1. Method

Sixteen Brown students and staff were paid to participate. Four were experienced observers who had participated in the previous experiments. The *fixation post* display was the same as the full textured ground used in Experiment 2. The fixation point appeared at eye level, on top of a white fixation post (1.6 m high) positioned on the ground surface. The *no post* display was identical except that the white fixation post was removed. Only simulated rotation (0, ± 3, ± 5°/s) was tested in this experiment. The four experienced observers viewed both types of displays in a counterbalanced order, two naïve observers viewed the fixation post display and six naïve observers viewed the no post display. Viewing was monocular with the preferred eye. They received ten practice trials, followed by 105 trials in each simulated display condition in a random order.

6.2. Results

Mean path errors appear in Fig. 10. In the simulated condition, performance is reasonably accurate for the fixation post display, with a slope of 1.27 and mean errors less than 6° at 5°/s, but errors increase dramatically for the no post display, with a slope of 3.33 and errors on the order of 16° at 5°/s, a significant difference in slope, $t(76) = 4.45$, $P < 0.001$. Thus, removing the fixation post leads to large errors in the direction of simulated rotation, consistent with a perceived curved path. Although the fixation post slope is significantly greater than zero, $t(3.48) = 3.48$, $P < 0.001$, a separate analysis demonstrates that it is not statistically different from that of the ground plane with fixation post in Experiment 2 (0.77), $t(71) = 1.44$, *ns*. This replicates our previous finding that the textured ground plane, together with a reference object, is sufficient for perceiving one's path during simulated rotation.

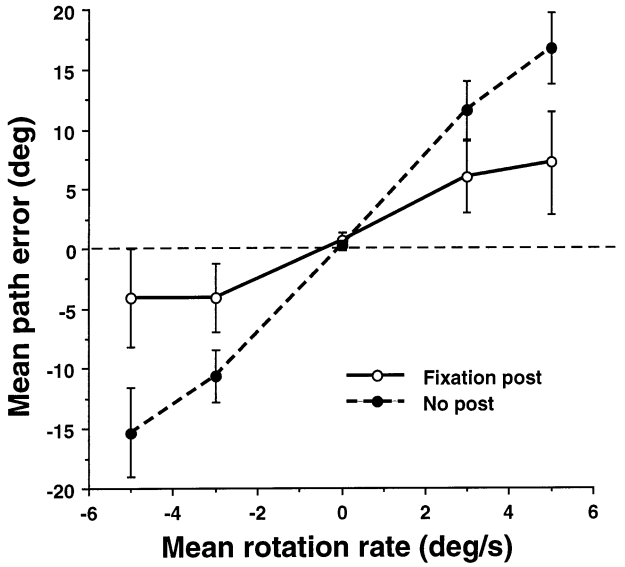

Fig. 10. Mean path error as a function of simulated rotation rate for the fixation post and no post conditions (Experiment 4). Error bars represent between-subject S.E.

6.3. Discussion

The results demonstrate that at least one reference object is necessary for good path judgments during simulated rotation. Significantly, a textured ground plane by itself — with its dense parallax and static depth information — is insufficient. This is consistent with the fact that the first-order optic flow only specifies the instantaneous heading, and strongly indicates that the visual system does not make use of higher-order temporal components. To resolve the path ambiguity, additional information is needed that allows the visual system to integrate a sequence of instantaneous headings over time and recover the path of self-motion.

The present experiments show that observers have a comparable level of accuracy with a single reference object and with multiple objects, such as the dense posts in Experiment 3. This rules out a solution based on perspective structure, for recovering the path of self-motion from triangulating the observer's position requires at least three landmarks, or two landmarks along the line of sight (Vishton & Cutting, 1995). In contrast, the results are consistent with a solution in which the instantaneous heading is continually updated with respect to a fixed reference point in the scene, allowing the observer to determine the path through that scene. Small improvements with multiple objects, as suggested in Experiment 1, could be due to the availability of reference objects near the heading point, allowing for more accurate updating.

The presence of a fixation point attached to the scene might provide other advantages. First, an attached fixation point could simplify the rotation problem by limiting the class of relevant retinal flow patterns (Perrone & Stone, 1994; Daniilidis, 1997). However, the retinal flow was identical in the fixation post and no post conditions, yet the pattern of errors was not. Further, we have also found that performance is accurate with an unattached moving fixation point as long as other reference objects are present (Li & Warren, 1999). Second, van den Berg (1996) reported that path judgments during simulated rotation are more accurate when observers attend to the illusory motion in depth of an attached fixation point than when they attend to the entire flow pattern. However, at least half of this effect is due to the distance of the point on the path that was judged (see his Figure 4): observers were reasonably accurate when indicating the point at which their path would pass the fixation point at 2.5 m, but they made large constant errors when indicating the path's endpoint at 20 m. This is what would be expected if they perceived a curved path of self-motion (Ehrlich et al., 1998). Any remaining effect could result from drawing attention to the fixation point as a reference object, such that the path is judged relative to that location in the scene. Finally, it is possible that the fixation post serves as a particularly good reference object because it is tracked by pursuit eye movements and appears at the locus of highest acuity and preferred attention. However, as just noted, we have also observed accurate performance with a moving fixation point when the reference objects are not fixated (Li & Warren, 1999).

It is interesting to note that when the fixation point is not attached to the scene in the no post condition, it appears to recede in depth toward the horizon and to move horizontally along the horizon in the direction of simulated rotation. This is another form of illusory motion of the fixation point, which appears to be a byproduct of perceiving self-motion on a curved path. It is consistent with gazing along the tangent to the curved path at a moving object. The visual system thus seems to arrive at a consistent interpretation of the scene even in the case of an erroneous curved path.

7. General discussion

We can draw several conclusions from the present experiments. First, we concur with Banks and his colleagues (Royden et al., 1992, 1994; Banks et al., 1996; Ehrlich et al., 1998) that the perceived path of self-motion can be influenced by extra-retinal signals about eye rotation. We often observed large errors in the simulated rotation condition but small errors in the actual rotation condition — with the random-dot ground in Experiment 1, the near and middle bands in Experiment 2, and the no post display in Experiment 4. This confirms that extra-retinal information accompanying

active pursuit movements allows the visual system to compensate for rotation and recover the path in body-centric space. The slightly positive slopes in the actual condition of Experiments 2 are consistent with Freeman's (1999) finding that this compensation is not complete even during real eye movements (Freeman, 1999).

Second, however, observers can also perceive the path of self-motion during fast rotations on the basis of retinal flow alone. With sufficient optical information, path judgments in the simulated condition show only a small influence of rotation rate, even up to 7°/s. We replicated this finding with the textured ground in Experiments 1, 2, and 4, and the dense posts in Experiment 3. This implies that sufficient retinal flow information tends to dominate extra-retinal information, such that the object-relative path is preferred over the absolute body-centric path. Normally, these paths are congruent, but when placed in conflict the path relative to objects in the scene tends to be perceived. Thus, both the retinal flow and extra-retinal approaches turn out to be correct: either is sufficient and neither is necessary. Precisely how retinal and extra-retinal information are combined in recovering the path of self-motion remains to be determined.

Third, we find that path judgments in the simulated condition are generally accurate when both (a) dense motion parallax and (b) reference objects are present in the display. This information is thus jointly sufficient to resolve both the rotation problem and the path ambiguity. We propose that the visual system solves the first problem by determining heading from the motion parallax field. The alternative hypothesis of radial foreground flow is inconsistent with the near band in Experiment 2. The hypothesis that rotation is estimated from distant motion cannot account for the difference between the random-dot ground (20 m) and the dense posts (18 m), which extended to similar depths (Experiments 1 and 3). Reliance on spatial derivatives of a continuous flow field is undermined by the finding that path judgments with the highly discontinuous Dense Posts were as accurate as those with a smooth textured ground. Both of these displays, on the other hand, contained dense parallax. The finding that a point of zero parallax is not required indicates that heading can be determined from the surrounding parallax. Thus, consistent with the predominant class of current theories, we suggest that the visual system recover instantaneous heading on the basis of the motion parallax field.

Further, we propose that the visual system resolves the path ambiguity by relying on reference objects. Specifically, the path of self-motion through the scene is determined by updating the instantaneous heading over time with respect to at least one reference object. The necessity of an object that can be tracked over time was demonstrated in Experiment 4, when removing the fixation post from the textured ground yielded large errors in the simulated condition, consistent with a curved path. By themselves, motion parallax and higher-order components of flow are thus insufficient to determine the path of self-motion. In principle, a reference object allows the observer to update the instantaneous heading with respect to a fixed location in the scene and thereby recover the path through the scene. In other words, we suggest that the path of self-motion is perceived relative to objects in the scene. We have recently replicated this joint dependence on dense parallax and reference objects in an active steering task (Li & Warren, 1999).

Finally, these results suggest that the visual system relies on the first-order flow field and does not make use of higher-order temporal components of flow to determine the path. This has important consequences, for the first-order flow is ambiguous and supports multiple interpretations. In the simulated condition, optical information such as reference objects can push the visual system toward a straight path interpretation, whereas extra-retinal information about eye rotation can push it toward a curved path interpretation. The individual differences observed in the simulated condition (Experiment 1) may be a consequence of differential reliance on retinal and extra-retinal information under conflict conditions. We have subsequently found that instructing observers that they are traveling on a straight or a curved path is sufficient to induce a shift between these two percepts — and observers actually believe they have viewed two different classes of displays (Li & Warren, 1998). Neutral instructions yield results similar to straight path instructions, indicating that the retinal flow solution is normally preferred.

7.1. Simulation

To assess whether the local motion parallax field could in principle account for the present pattern of data, we performed simulations based on Rieger and Lawton's (1985) differential motion algorithm, which we describe here briefly. The simulation was not intended as a biologically plausible heading model, but merely to assess the adequacy of motion parallax. The first layer was composed of differential motion units, each of which calculated the difference vectors between a central motion and the surrounding motions within a local neighborhood. The second layer was a heading map composed of expansion units, each of which acted like a template for a radial flow pattern centered in the unit's receptive field. The goal of the model was to estimate the instantaneous heading from an array of local motion parallax measurements, and thus it did not address the path ambiguity.

The flow field for translation and rotation with respect to the objects in the scene was computed in a

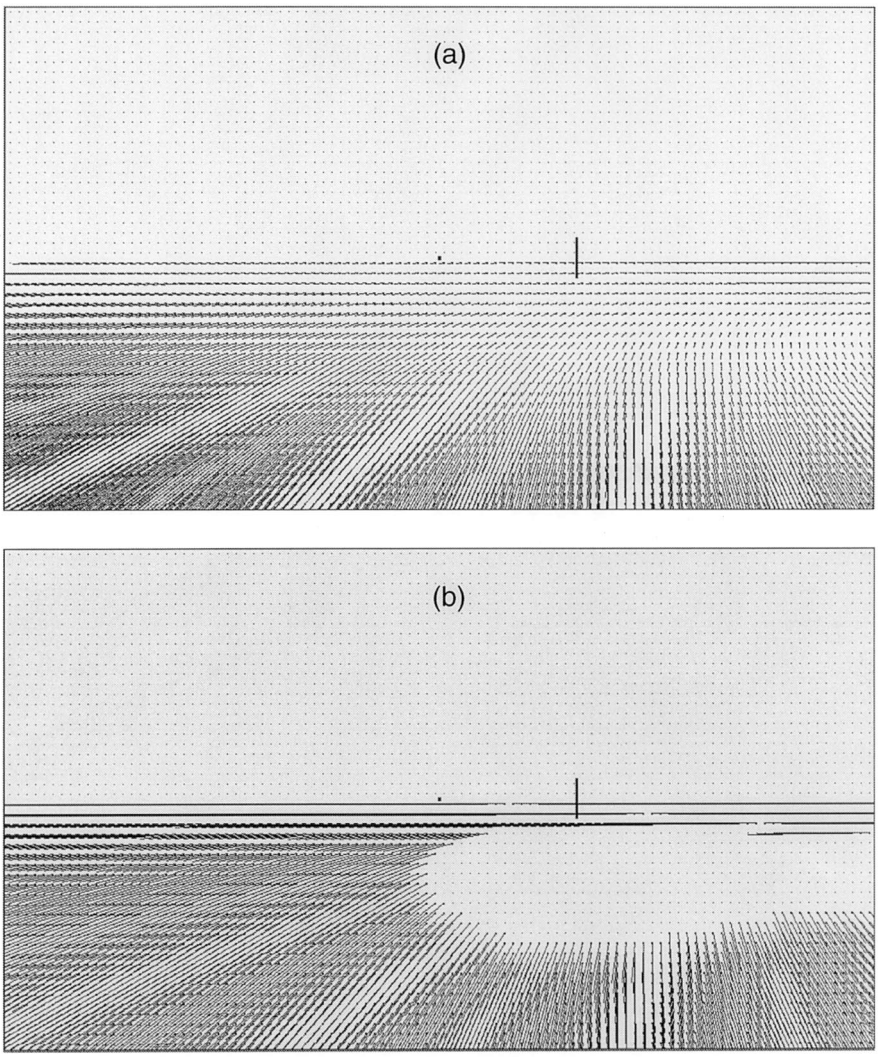

Fig. 11. Velocity fields from the differential motion model. (a) Sample input velocity field generated by translation over a textured ground plane toward the vertical line (2 m/s), plus simulated rightward eye rotation about a vertical axis (5°/s). (b) Differential motion field computed by the first layer. Note that vector lengths are normalized and only represent direction.

400×200 unit grid in the image plane (Fig. 11a). One velocity vector was measured in each cell, due to motion from the ground plane or an object. The grid was divided into 40×20 contiguous neighborhoods (each containing 10×10 cells), corresponding to the receptive fields of units in the first layer. To compute the differential motion in each neighborhood, the mean velocity vector computed for the four cells in the center (μ_c) was subtracted from each of the surrounding velocity vectors (μ_i) in the neighborhood to obtain a set of difference vectors ($\Delta\mu_{ci}$). Then the average orientation of these difference vectors was determined by fitting a line (l_c) through the center of the neighborhood that maximized the sum of projections of the difference vectors ($\Delta\mu_{ci}$) onto the line. To reduce noise in this estimate, an anisotropic criterion was adopted so that the ratio between the small and the large eigenvalue of the fitted line did not exceed a constant ($\lambda_{min}/\lambda_{max} \leq 0.01$)[7]. This insured that only neighborhoods with a dominant orientation (i.e. a distribution of difference vectors that was strongly anisotropic) were passed to the output layer. The result of this procedure was a set of differential motion vectors (d_c) with an approximately radial pattern, roughly aligned with the translational component of the flow field (Fig. 11b).

To determine the instantaneous heading, these local differential motions were pooled by completely connecting every differential motion unit in the first layer to each expansion unit in the second layer. Because heading only varied along 1-D (azimuth) in the experiments, the second layer consisted of a horizontal row of

[7] Rieger and Lawton (1985) showed that a criterion of 0.01 was reliable for small neighborhoods with 10×10 velocity vectors.

201 expansion units spanning 100° of visual angle, their centers evenly spaced 0.5° apart, corresponding to set of output headings. Each expansion unit had a preferred pattern of vector directions ($\theta_h(x)$) associated with its heading, and it was insensitive to vector magnitudes. The response $R(h)$ of the unit was thus a measure of the overall consistency of the input direction field with the unit's preferred direction field:

$$R(h) = \int \cos(\theta_h(x) - \theta_i(x)) \quad (3)$$

where $\theta_h(x) - \theta_i(x)$ is the local angular difference between the differential motion vector and the preferred direction at that location. The maximum of this function over the second layer was taken to be an estimate of the instantaneous heading.

Simulations were conducted on the five basic displays from Experiments 1 and 3: random-dot ground (Fig. 2a); textured ground (Fig. 2b); ground + posts (Fig. 2c); ground + tombstones (Fig. 2d); and dense posts (Fig. 7a). A complete set of trials was simulated for each display, with fixation distances and rotation rates matched to the experimental conditions. For the random-dot display, the input to the model was a sparse velocity field consisting of a vector at each dot location, and a different flow field was tested on each trial. To approximate the densely textured displays, 400×200 vectors were sampled, creating a dense flow field on each trial. Because the flow pattern changed over the course of a 1.5 s experimental trial, the first and last velocity field in each trial were tested and the two heading estimates averaged.

Mean heading errors for the simulation are plotted in Fig. 12. The results capture the qualitative pattern of the human data (Fig. 3a), with large errors at higher rotation rates for the sparse random-dot display, but small errors for the dense textured displays. The slight asymmetry in the random-dot curve is due to the random sampling of input vector fields. Interestingly, there are no differences among the textured displays, indicating that sufficient motion parallax is present with both the ground and the dense posts to achieve the same level of accuracy. The simulation simply demonstrates that the observed difference between random-dot and textured displays can be accounted for by the density of motion parallax, thereby providing converging evidence for the dense parallax hypothesis.

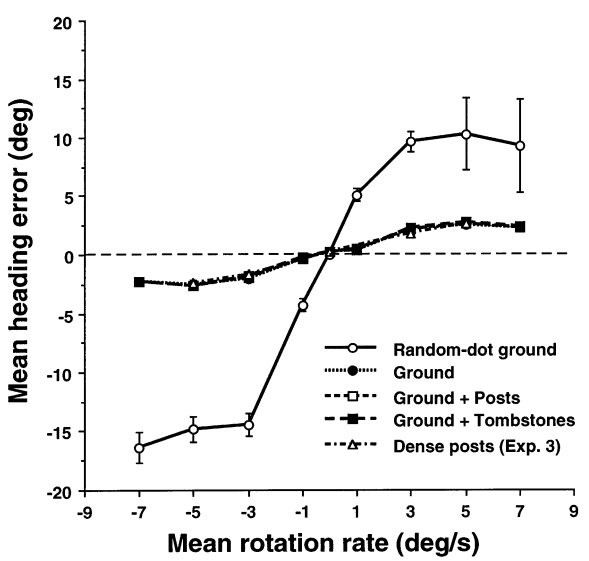

Fig. 12. Model data. Mean heading error as a function of mean rotation rate for the five displays.

Let us clarify the relation between our view and Cutting's 'differential motion parallax' theory (Cutting, 1996; Cutting et al., 1992; Wang & Cutting, 1999). Although they share an emphasis on motion parallax (as do most current theories) and naturalistic displays, the proposals are quite different in their specifics. First, Cutting and colleagues are committed to local motion parallax between a few objects, rather than the motion parallax field. We do not doubt that observers will do the best they can with displays of only a few objects (e.g. Wang & Cutting), but it is likely that the visual system makes use of dense motion parallax when it is available, as we have found here. Second, Cutting and colleagues promote particular variables of relative motion with respect to a fixated object — most recently, converging, decelerating diverging, and accelerating diverging motion (Wang & Cutting). Two of these are second-order variables, to which the visual system has poor sensitivity, one is a probabilistic cue (accelerating divergence), and all depend on auxiliary depth information. We do not think this is a promising general basis for the perception of heading, in contrast to the pattern of motion parallax that is invariant under rotation. Further, a fixated object appears not to be necessary, for we have recently found accurate simulated performance with a moving fixation point as long as other reference objects are present (Li & Warren, 1999). Third, whereas Cutting and colleagues attribute accurate heading judgments to 'naturalistic displays' (Vishton & Cutting, 1995; Cutting et al., 1997), we specifically suggest that textured surfaces provide dense motion parallax and that reference objects provide fixed locations in the environment that allow heading to be updated.

In sum, we propose that two sorts of optical information are required to determine the path of self-motion under rotation: dense motion parallax is used to solve the rotation problem, and reference objects are used to resolve the path ambiguity. This may explain why there are few data firmly supporting the retinal flow approach, for standard random-dot displays do not typically contain both types of information. As it turns out, retinal and extra-retinal solutions are each sufficient for perceiving the path of self-motion during eye rotations. The former yields the path relative to objects in the scene, whereas the latter yields the absolute path in a body-cen-

tric frame. Under normal circumstances these paths are congruent and retinal and extra-retinal solutions are redundant, but when placed in conflict retinal flow tends to dominate.

Acknowledgements

Part of this research was carried out while the second author was on sabbatical at the University of California, Berkeley. We would like to thank the Department of Optometry for its hospitality, and Marty Banks and Sheryl Ehrlich for ongoing interactions. Thanks also to Andy Forsberg, Seung Hong, and Lindsay Mann for their assistance. This research was supported by grants from the National Institutes of Health, EY10923 and K02 MH01353.

References

Banks, M. S., Ehrlich, S. M., Backus, B. T., & Crowell, J. A. (1996). Estimating heading during real and simulated eye movements. *Vision Research*, *36*, 431–443.

Beintema, J. A., & van den Berg, A. V. (1998). Heading detection using motion templates and eye velocity gain fields. *Vision Research*, *38*, 2155–2179.

Calderone, J. B., & Kaiser, M. K. (1989). Visual acceleration detection: effect of sign and motion orientation. *Perception and Psychophysics*, *45*, 391–394.

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.

Cutting, J. E. (1986). *Perception with an eye to motion*. Cambridge, MA: MIT Press.

Cutting, J. E. (1996). Wayfinding from multiple sources of local information in retinal flow. *Journal of Experimental Psychology*: *Human Perception and Performance*, *22*, 1299–1313.

Cutting, J. E., Springer, K., Braren, P. A., & Johnson, S. H. (1992). Wayfinding on foot from information in retinal, not optical, flow. *Journal of Experimental Psychology*: *General*, *121*, 41–72.

Cutting, J. E., Vishton, P. M., Flückiger, M., Baumberger, B., & Gerndt, J. D. (1997). Heading and path information from retinal flow in naturalistic environments. *Perception and Psychophysics*, *59*, 426–441.

Daniilidis, K. (1997). Fixation simplifies 3-D motion estimation. *Computer Vision and Image Understanding*, *68*, 158–169.

Ehrlich, S. M., Beck, D. M., Crowell, J. A., Freeman, T. C. A., & Banks, M. S. (1998). Depth information and perceived self-motion during simulated gaze rotations. *Vision Research*, *38*, 3129–3145.

Freeman, T. C. A. (1999). Path perception and Filehne illusion compared: model and data. *Vision Research*, *39*, 2659–2667.

Gibson, J. J. (1950). *Perception of the visual world*. Boston: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gibson, J. J., Olum, P., & Rosenblatt, F. (1955). Parallax and perspective during aircraft landings. *American Journal of Psychology*, *68*, 372–385.

Heeger, D. J., & Jepson, A. D. (1992). Subspace methods for recovering rigid motion I: algorithm and implementation. *International Journal of Computer Vision*, *7*, 95–117.

Hildreth, E. C., & Royden, C. S. (1998). Computing observer motion from optical flow. In T. Watanabe, *High-level motion processing* (pp. 269–293). Cambridge, MA: MIT Press.

Koenderink, J. J., & van Doorn, A. J. (1981). Exterospecific component of the motion parallax field. *Journal of the Optical Society of America*, *71*, 953–957.

Lappe, M. (1998). A model of the combination of optic flow and extraretinal eye movement signals in primate extrastriate visual cortex: neural model of self-motion from optic flow and extraretinal cues. *Neural Networks*, *11*, 397–414.

Lappe, M., Bremmer, F., Pekel, M., Thiele, A., & Hoffmann, K.-P. (1996). Optic flow processing in monkey STS: a theoretical and experimental approach. *Journal of Neuroscience*, *16*, 6265–6285.

Li, L., & Warren, W. H. (1998). Ambiguity of the flow field during simulated rotation: influence of instructions and fixation. *Investigative Ophthalmology and Visual Science*, *39*, S1085.

Li, L., & Warren, W. H. (1999). Active control of steering during translation and rotation. *Investigative Ophthalmology and Visual Science*, *40*, S799.

Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London*, *B*, *208*, 385–397.

Nakayama, K., & Loomis, J. M. (1974). Optical velocity patterns, velocity sensitive neurons, and space perception: a hypothesis. *Perception*, *3*, 63–80.

Perrone, J. A. (1992). Model for the computation of self-motion in biological systems. *Journal of the Optical Society of America A*, *9*, 177–194.

Perrone, J. A., & Stone, L. S. (1994). A model of self-motion estimation within primate extrastriate visual cortex. *Vision Research*, *34*, 2917–2938.

Poggio, T., Verri, A., & Torre, V. (1991). *Green theorems and qualitative properties of the optical flow (AI Memo 1289)*. MIT.

Regan, D., & Beverley, K. (1982). How do we avoid confounding the direction we are looking and the direction we are moving? *Science*, *215*, 194–196.

Rieger, J. H. (1983). Information in optical flows induced by curved paths of observation. *Journal of the Optical Society of America*, *73*, 339–344.

Rieger, J. H., & Lawton, D. T. (1985). Processing differential image motion. *Journal of the Optical Society of America A*, *2*, 354–360.

Rieger, J. H., & Toet, L. (1985). Human visual navigation in the presence of 3-D rotations. *Biological Cybernetics*, *52*, 377–381.

Royden, C. S. (1994). Analysis of misperceived observer motion during simulated eye rotations. *Vision Research*, *34*, 3215–3222.

Royden, C. S. (1997). Mathematical analysis of motion-opponent mechanisms used in the determination of heading and depth. *Journal of the Optical Society of America A*, *14*, 2128–2143.

Royden, C. S., Banks, M. S., & Crowell, J. A. (1992). The perception of heading during eye movements. *Nature*, *360*, 583–585.

Royden, C. S., Crowell, J. A., & Banks, M. S. (1994). Estimating heading during eye movements. *Vision Research*, *34*, 3197–3214.

Schmerler, J. (1976). The visual perception of accelerated motion. *Perception*, *5*, 167–185.

Stone, L. S., & Perrone, J. A. (1997). Human heading estimation during visually simulated curvilinear motion. *Vision Research*, *37*, 573–590.

Thomas, I., Simoncelli, E., & Bajcsy, R. (1994). *Linear structure from motion (GRASP Lab Tech Report MS-CIS-94-61)*. Philadelphia, PA: University of Pennsylvania.

Tsai, R.Y., & Huang, T.S. (1984). Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-6*, 13–27.

van den Berg, A. V. (1992). Robustness of perception of heading from optic flow. *Vision Research*, *32*, 1285–1296.

van den Berg, A. V. (1993). Perception of heading. *Nature*, *365*, 497–498.

van den Berg, A. V. (1996). Judgements of heading. *Vision Research*, *36*, 2337–2350.

van den Berg, A. V., & Brenner, E. (1994a). Humans combine the optic flow with static depth cues for robust perception of heading. *Vision Research*, *34*, 2153.

van den Berg, A. V., & Brenner, E. (1994b). Why two eyes are better than one for judgments of heading. *Nature*, *371*, 700–702.

Vishton, P. M., & Cutting, J. E. (1995). Wayfinding, displacements, and mental maps: velocity fields are not typically used to determine one's heading. *Journal of Experimental Psychology*: *Human Perception and Performance*, *21*, 978–995.

Wang, R. F., & Cutting, J. E. (1999). Where we go with a little good information. *Psychological Science*, *10*, 71–75.

Warren, W. H. (1998). The state of flow. In T. Watanabe, *High-level motion processing* (pp. 315–358). Cambridge: MIT Press.

Warren, W. H., & Hannon, D. J. (1988). Direction of self-motion is perceived from optical flow. *Nature*, *336*(6195), 162–163.

Warren, W. H., & Hannon, D. J. (1990). Eye movements and optical flow. *Journal of the Optical Society of America A*, *7*(1), 160–169.

Warren, W. H., Blackwell, A. W., Kurtz, K. J., Hatsopoulos, N. G., & Kalish, M. L. (1991a). On the sufficiency of the velocity field for perception of heading. *Biological Cybernetics*, *65*, 311–320.

Warren, W. H., Mestre, D. R., Blackwell, A. W., & Morris, M. W. (1991b). Perception of circular heading from optical flow. *Journal of Experimental Psychology*: *Human Perception and Performance*, *17*, 28–43.

Waxman, A. M., & Ullman, S. (1985). Surface structure and 3-D motion from image flow: a kinematic analysis. *International Journal of Robotics Research*, *4*, 72–94.