Eye movements and optical flow

William H. Warren, Jr., and Danial J. Hannon

Department of Cognitive and Linguistic Sciences, Brown University, Providence, Rhode Island 02912

Received January 17, 1989; accepted August 15, 1989

Translation of an observer through a static environment generates a pattern of optical flow that specifies the direction of self-motion, but the retinal flow pattern is confounded by pursuit eye movements. How does the visual system decompose the translational and rotational components of flow to determine heading? It is shown that observers can perceive their direction of self-motion during stationary fixations and pursuit eye movements and with displays that simulate the optical effects of eye movements. Results indicate that the visual system can perform the decomposition with both continuous and discontinuous fields on the basis of flow-field information alone but requires a three-dimensional environmental structure to do so. The findings are inconsistent with general computational models and theories based on the maximum of divergence, oculomotor signals, or multiple fixations but are consistent with the theory of reliance on differential motion produced by environmental variation in depth.

INTRODUCTION

As was first noted by Gibson,[1] translation of an observer through a stationary environment generates a radial pattern of optical flow at the eye, in which the focus of outflow specifies the observer's direction of self-motion, or heading (Fig. 1). In recent experiments, Warren *et al.*[2] found that observers can perceive their heading from such radial flow patterns with an accuracy of the order of 1° of visual angle and that they rely on the global structure of the flow field. However, this straightforward relationship holds only for linear translation. The addition of observer rotation, which commonly occurs with pursuit eye movements during locomotion, complicates the retinal flow pattern significantly. This raises the question of how the direction of self-motion can be determined during observer translation and rotation.

The Problem

We use the term optical flow in Gibson's original sense to refer to a temporal change in the structure of the optic array, the pattern of light intensities in different visual directions about a point of observation before the introduction of an eye.[3] Retinal flow refers to change in the retinal image, which is influenced by eye movements. Both optical flow and retinal flow are represented typically as instantaneous two-dimensional velocity fields such as that shown in Fig. 1, in which each vector corresponds to the optical velocity of an environmental element. Although eye movements affect only the retinal flow pattern and not the optic array,[4] the eye-movement problem must ultimately be addressed by a theory of how the optic array is registered by a moving eye.

Consider the basic contributions to the flow field. First, observer translation ($\mathbf{T} = T_x, T_y, T_z$) produces a radial pattern of both optical flow and retinal flow called the translational component of the flow field (Fig. 1). The vectors' directions are determined completely by the observer's heading, whereas their magnitudes depend on the distances to environmental elements. Second, rotation of the observer about the approximate nodal point of the eye ($\mathbf{R} = R_x, R_y, R_z$) produces a solenoidal pattern of retinal flow called the rotational component of the flow field. It is equivalent to a rigid rotation of the world about the observer, with rotary flow about the poles of rotation and approximately translational flow in regions near the equator, such as in central vision during an eye movement (Fig. 2). In this case, both the directions and the magnitudes of flow vectors are determined completely by observer rotation, independent of element distances. Third, curvilinear movement of the observer ($\mathbf{C} = C_x, C_y, C_z$) about a center of rotation (x_0, y_0, z_0) external to the eye produces a hyperbolic flow pattern in the image plane. Although any arbitrary movement can be described instantaneously as the sum of a translation and a rotation, curvilinear movement has a distinct meaning for locomotion and is discussed below. Finally, the combination of translation and rotation, such as a pursuit eye movement to track a point on the passing ground surface, yields the vector sum of the first two fields, annihilating the focus of outflow at the heading point and creating a new singularity at the fixation point (Fig. 3). If observers relied simply on the singularity in the field to determine heading, they would perceive themselves as heading toward the fixation point. How might they distinguish the direction in which they are heading from the direction in which they are looking?

Theories

Gibson[1,5] himself recognized that a pursuit eye movement adds a constant to the retinal flow field and suggested that the visual system could disregard the constant, effectively decomposing the field into its translational and rotational components. This turns out to be a nontrivial problem, and formal solutions were derived only recently.

Discrete Models

Discrete models[6–13] compute the six parameters of observer translation and rotation from the optical motions of some minimum number of points, often relying on iterative methods to solve a set of nonlinear equations. Four to seven points in two successive images have been found to be sufficient, and the models are general in that they do not depend on surface layout. This proves that the decomposition is

0740-3232/90/010160-10$02.00

Fig. 1. Instantaneous velocity field produced by pure observer translation parallel to the ground plane. Vertical line indicates heading; vectors (line segments) indicate optical motions of environmental elements (corresponding dots).

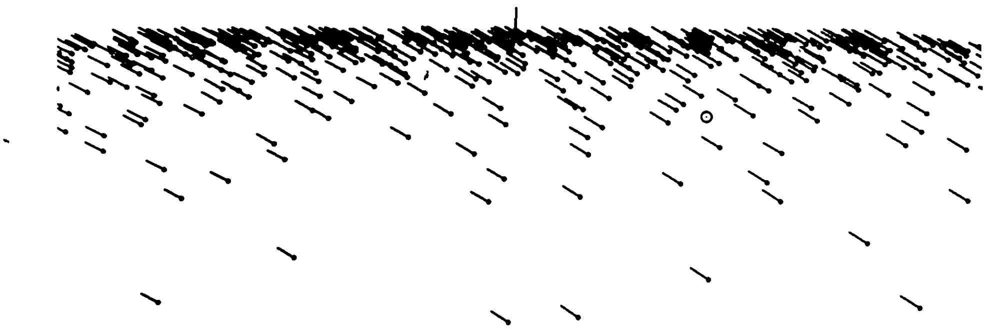

Fig. 2. Velocity field produced by pure eye rotation down and to the right. Note that this system yields approximately parallel flow in the central visual field.

Fig. 3. Velocity field produced by combined translation and rotation, resulting from translating toward the vertical line while fixating the circle on the passing ground surface.

mathematically possible, but because the method relies on precise measurements of just a few points it is highly vulnerable to noise. In contrast, biological vision must be robust; we have evidence that human observers can tolerate large amounts of directional and speed noise in local dot motion when judging translational and curvilinear heading.[14]

Differential Models

Differential models use differential invariants of the flow field to decompose observer translation and rotation. However, because they are based on spatial derivatives of the velocity field, they require locally continuous (or interpolatable) fields, implying a smoothness constraint for environmental surfaces. Thus they would fail with discontinuous or sparse flow fields. They also depend on accurate measurements of flow and thus are sensitive to noise. Several analyses have been proposed.

Maximum of Divergence. Koenderink and van Doorn[15,16] showed that any locally continuous field can be described as a sum of the divergence (rate of expansion), the curl (rate of rotation), the deformation (degree of shear), and translation. They demonstrated[17] that with movement relative to a plane the local maximum and minimum of the divergence are invariant under rotation and thus could be used to determine the direction of translation independent of eye movements. This approach has several limitations. First, unlike Gibson's focus of outflow, the maximum of the divergence does not correspond to the heading but lies in a direction that bisects the direction of heading and the surface normal, and thus it depends on surface slant. With movement parallel to a plane, the maximum ahead of the observer and the minimum behind the observer together specify the axis of translation, but they are rarely both visible. Second, in nonplanar environments the number and locations of local maxima and minima depend on environmental structure. Third, the information is undefined with discontinuous or sparse fields.

Flow Gradient. Nakayama[18] formalized Gibson's intuition directly, noting that the gradient of the velocity field is unaffected by the addition of a constant [Grad V = Grad ($V + C$)]. Thus the effects of eye movements could be disregarded by detecting the gradient of the flow field.

General Models. Longuet-Higgins and Prazdny[19] and Waxman and Ullman[20] proposed models based on related differential invariants. They are general in that they do not depend on surface layout, but they also require continuous fields.

Least-Squares Models

Least-squares models[21–23] use iterative methods to search the parameter space of possible observer movements and surface distances by minimizing the difference between the observed optical flow and that calculated from possible parameter values. Because they sample many points over a large region of the velocity field, such models are more resistant to noise, and they are also general with respect to surface layout. However, they also assume a smoothness constraint, require many elements over a wide field of view, and are biologically questionable.

Dynamical Models

A dynamical model proposed by Verri *et al.*[24] treats the flow field as the phase portrait of a dynamical system. Observer-movement parameters can then be determined from the time evolution of the local structure of flow in the neighborhood of singular points in the field. Because these qualitative properties are structurally stable, the model tolerates flow-field noise. However, it also assumes a smoothness constraint and requires dense flow about singular points and thus would fail with discontinuous or sparse flow fields.

Motion Parallax

Several hypotheses depend in different ways on motion-parallax information, or relative optical motion among elements at different depths. Motion parallax is relevant because flow velocities resulting from observer translation are influenced by element distance, whereas those resulting from observer rotation are not. The advantage to this approach is that it makes no smoothness assumptions and does not require continuous fields. In contrast to the general models above, however, it requires three-dimensional (3D) environmental structure and fails in the absence of depth differences, such as in the special case of approach to a plane.

Edge Parallax. Longuet-Higgins and Prazdny[19] first demonstrated that heading is specified by motion parallax between two elements at different depths in the same visual direction, which occurs commonly at depth edges. Because observer rotation affects two such overlapping elements equally, they have the same optical velocity resulting from rotation, and any relative motion between them results solely from observer translation. Their difference vector radiates from the direction of heading, and thus several overlapping pairs of elements yield a radial pattern of difference vectors that specifies the direction of self-motion. This requires depth edges in the field of view.

Differential Motion. Rieger and Lawton[25] generalized the notion of edge parallax to include relative motion between nonoverlapping elements within a neighborhood. The difference vector characterizing relative motion between two neighboring elements radiates approximately from the direction of heading, although individual vectors may deviate slightly depending on the angular separation of the elements. Thus the difference vectors radiate globally from the heading point, and differential motion goes to zero in the direction of heading. This model requires sufficient variation in depth within local neighborhoods to produce detectable differential motion, and it would fail in the case of an approach to a plane.

Differential-Motion Parallax. Cutting[26] proposed that the observer could use what he called differential motion parallax to guide a sequence of fixations to the focus of outflow. With sufficient numbers of elements nearer and farther than the fixation point, the highest retinal velocity across the line of fixation is opposite the direction of self-motion and thus indicates whether the heading is to the left or to the right of the fixation point. This hypothesis requires multiple fixations as well as sufficient numbers of elements at different depths along potential lines of fixation, and thus it would fail with sparse fields or in the case of an approach to a plane.

Flow Asymmetry. Several authors[27,28] suggested that the observer could use related asymmetries in the retinal flow pattern to guide a sequence of fixations to the heading point, where the flow pattern is symmetrical. Such asymmetries depend on environmental structure and require variation in depth.

Extra-Flow-Field Information

The above-described theories attempt to decompose observer translation and rotation on the basis of information contained in the flow pattern itself. Alternatively, the observer might rely on extra-flow-field information.

Oculomotor Signal. A popular hypothesis is that an oculomotor signal, such as an efference copy, is used to determine the amount of eye rotation and to compensate for the effects of eye movements.[29] It is assumed that there is a nonvisual contribution to the visual perception of optical flow, and active eye movements are required.

Ocular Drift. It has been suggested that muscle proprioception about ocular drift resulting from tracking of a moving element could be used to guide a sequence of fixations to a stationary element at the focus of outflow.[30] This hypothesis requires multiple fixations and an element at or near the focus and thus would be inaccurate in sparse fields.

Frame of Reference. Gibson[31] suggested that the visible orbit of the eye provides a visual frame of reference that yields information about eye movements relative to the head. The radial flow pattern is defined within this rotation-invariant frame of reference, independent of eye movements. Artificial frames of reference may also be provided by the edges of a car or airplane windshield or display screen, but these are not normally available during legged locomotion.

Research

Consistent with theories based on motion parallax, several studies suggest that depth differences are necessary for observers to determine heading during rotation, but the results are difficult to interpret. Regan and Beverley[32] simulated eye rotation during an approach to a surface by presenting a sine-wave grating that translated as it expanded. When the display simulated an approach to a planar surface, observers could not determine whether the focus of outflow was to the left or to the right of a target, but when it simulated a convex surface they could locate the maximum of divergence. Although Regan and Beverley interpreted this as support for the divergence maximum hypothesis, observers could have simply located the fattest bar at the end of each trial, a cue reinforced by feedback. These results have been cited widely as evidence of the ineffectiveness of optical flow, but we show here that an approach toward a plane is a special case.

Cutting[26] simulated eye rotation during an approach toward 12 vertical lines and asked subjects to judge whether the display represented a view to the left of heading, to the right of heading, or straight in the direction of heading. When the lines lay in a single plane, observers responded at the chance level, but, when parallax was added by placing the lines in three planes that were parallel in depth, performance improved, with thresholds (66% correct) at final angles of 10°–1.25° between the direction of heading and the simulated direction of gaze, depending on the amount of parallax. Although the results were interpreted as support for the differential-motion-parallax theory, fixation was not controlled (permitting anomalous retinal flow patterns), judgments were made relative to the direction of gaze rather than an environmental reference (and were thus relevant only to a multiple-fixation strategy), and because the simulated direction of gaze was always at the center of the screen observers could have distinguished the three cases by the asymmetry of flow in the display, reinforced by feedback.

Rieger and Toet[33] simulated observer rotation during approach to two random-dot planes at different depths, and observers judged whether they were heading in one of four directions relative to a fixation point at the center of the screen. Once again, performance was poor with no depth differences, but, when the two planes were separated in depth, observers were 85% correct with a final angle of 5° between the direction of heading and the direction of gaze, although performance decreased to 60% correct with a final angle of 2.5°. The results were interpreted as support for the theory of differential motion, but observers had a great deal of training, and judgments were again relative to a fixation point. In addition, the displays did not actually simulate pursuit eye movements, for rotations did not correspond to the tracking of an element in the scene, and the displays included a component of rotation about the line of sight.

In the present research the hypotheses are tested systematically by using a methodology that avoids these problems, including judgments relative to an environmental target and random placement of the heading and fixation points. The results support Gibson's original notion of a decomposition based on flow-field information and are consistent with the differential-motion theory.

GENERAL METHODS

Displays depicting self-motion relative to random-dot surfaces were generated in real time on a Raster Technologies Model 1/380 graphics terminal hosted by a Vax/780 computer. Unless otherwise specified, each display consisted of 45 images with a 1280 (horizontal) × 1024 (vertical) pixel resolution presented at 15 Hz. The screen was viewed from a distance of 45 cm and subtended 40° (horizontal) × 32° (vertical). Dots were single white pixels with a luminance of 118 nits (1 nit = 1 cd/m^2), on a black background with a luminance of 0.2 nit and did not expand with motion. Display motion corresponded to a fast walking speed of 1.9 m/sec^2 for an assumed eye height of 1.6 m. On each trial, the first frame of the display appeared for 1 sec as a warning signal, and the dots moved for 3 sec. In the last frame a vertical 1° target line appeared and remained visible together with the dots until the observer responded. Observers pushed one of two buttons to indicate whether they appeared to be heading to the left or the right of the target. Unless otherwise specified, the heading angle with respect to the target varied horizontally, with values of ±0.5°, ±1°, ±2°, and ±3°, where positive values indicate a heading to the right of the target and negative values indicate a heading to the left. A fixation point appeared in a random position in the first frame of the display, was constrained so that it would remain on the screen, and did not expand with motion. Observers were familiar with optical flow displays and received 20 practice trials with feedback; no feedback was provided on test trials.

EXPERIMENT 1

Procedure

At this point it is not even known whether observers can in fact perceive their direction of self-motion during a pursuit eye movement.[28] The first experiment answered this question and tested hypotheses based on edge parallax and multiple fixations. Displays depicted self-motion parallel to a random-dot ground plane with no depth edges, satisfying the smoothness constraint. Under the stationary condition, the fixation point remained stationary on the screen, so that the retinal flow pattern contained only a translational component. Under the moving condition, the fixation point moved as though it were a spot on the ground surface and induced a pursuit eye movement, so that the retinal flow pattern contained both translational and rotational components. The two types of trial were intermixed randomly. If observers depend on either depth edges or multiple fixations to resolve the effects of translation and rotation, performance should be at the level of chance in the moving condition.

In this experiment only, displays had 56 frames and lasted 3.7 sec. The ground surface extended 37.3 m in depth and had a dot density of 0.12 dot/m^2, with an average of 63 dots visible at the start of a trial. The target appeared in one of four positions on the pseudohorizon, ±1° or ±3° from the center of the screen, and the heading angle varied with values of ±0.5°, ±1°, ±2°, ±4°, and ±6° from the target. To control for effects of retinal eccentricity, trials were matched so that the stationary fixation point appeared in one of four possible positions along the trajectory of the moving fixation point. The amount of eye rotation (excursion of the fixation point) varied from 2° to 10°, with a mean velocity that varied from 0.5 to 2.7 deg/sec. In this experiment, the screen was surrounded by a black mask and viewed binocularly, while eye movements were recorded with an ASL Model 210 head-mounted binocular limbus tracker accurate to within ±0.5°. The output was superimposed upon a videotaped image of the display, the records were screened, and trials containing sequences of fixations or drifts greater than 1° from the fixation point were excluded from further analysis. Eight observers each underwent 160 trials under each condition, and fewer than 5% of the trials were rejected.

Results

Observers were quite accurate under both the stationary and the moving conditions (Fig. 4), with no difference in percent correct performance between them [$F(1, 7) = 2.99$, not significant]. Heading thresholds were computed by fitting each subject's data with an ogive and adopting the heading angle at the 75%-correct response level as the threshold. Mean thresholds were 1.7° in the stationary condition and 1.8° in the moving condition [$t(7) = 1.96$, not significant]. Thus observers can in fact accurately perceive their direction of self-motion during pursuit eye movements and do not see themselves as heading toward the singularity at the fixation point. Neither edge parallax nor multiple fixations are necessary to resolve the effects of translation and rotation, contrary to the differential-motion-parallax, flow-asymmetry, and ocular-drift hypotheses.

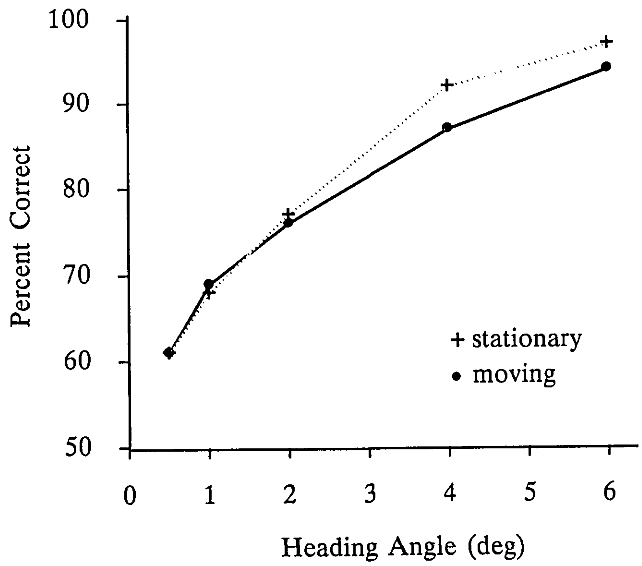

Fig. 4. Percentage of correct responses as a function of heading angle with stationary and moving fixation points (experiment 1, ground surface). Chance-level performance is 50% correct.

EXPERIMENT 2

Procedure

In the second experiment, we tested the oculomotor-signal and frame-of-reference hypotheses, again using displays of self-motion parallel to a ground plane. Under the moving condition, we presented a moving fixation point that induced a pursuit eye movement, as before. Under the simulated condition, matched displays simulated the retinal flow pattern that would occur with such an eye movement, but the fixation point remained stationary on the screen (Fig. 3). Thus the display contained both translational and rotational components without an actual eye movement. To eliminate the visible orbit of the eye and to minimize the frame of reference provided by the display screen, we used a translucent reduction screen that reduced and blurred the luminance difference at the edges of the display. This placed the flow pattern, which specified translation plus eye rotation, in conflict with the oculomotor state and any frame effects, which indicated that no eye rotation was occurring. If observers depend on oculomotor signals or reference frames, performance under the simulated condition would be at the level of chance.

Displays depicted a ground plane, as in the preceding experiment. The target appeared ±1° or ±3° from the center of the screen, and the heading angle varied with respect to the target. The fixation point appeared in a random position on the ground within 5° of the target, and the amount of rotation varied from 1° to 2° with a mean velocity that varied from 0.3 to 0.7 deg/sec. In this and subsequent experiments, displays were viewed monocularly through the reduction screen, so eye movements were not monitored. Six observers underwent 128 trials under each condition.

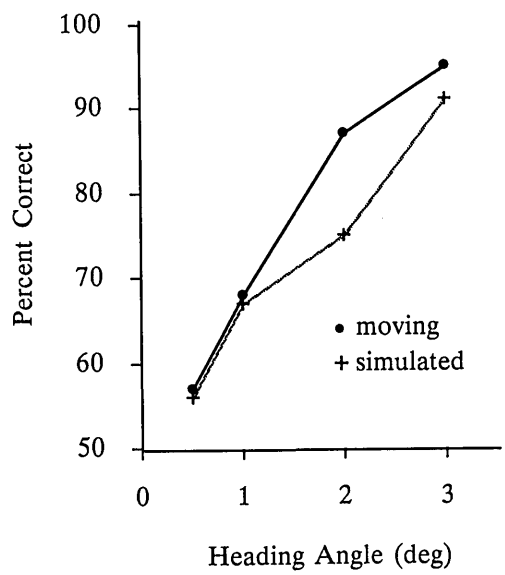

Fig. 5. Percentage of correct responses as a function of heading angle with moving fixation point and simulated eye rotation (experiment 2, ground).

Results

The results again show accurate heading judgments in both conditions (Fig. 5). The mean heading thresholds were 1.3° under the moving condition and 1.5° under the simulated condition, with no difference between them [$t(5) = 1.79$, not significant]. There were marginally significant differences in percent correct performance at heading angles of 2° [$t(5) = 3.46$, $p < 0.05$] and 3° [$t(5) = 3.16$, $p < 0.05$], perhaps because of residual frame effects from the edges of the screen. It is interesting to note that observers typically could not distinguish moving and simulated trials, for in the latter there was a strong illusion that one's eye was actually moving. It is clear that observers can determine their heading under the simulated condition on the basis of flow-field information alone. Thus neither oculomotor signals nor the visible orbit of the eye is necessary to resolve the effects of translation and rotation with a smooth ground plane.

EXPERIMENT 3

Procedure

In the third experiment, we tested the differential, least-squares, and dynamical models and the divergence maximum hypothesis by using displays of movement relative to a 3D cloud of dots. Pure translation through a 3D cloud yields a radial outflow pattern, with vector magnitudes depending on element distances (Fig. 6). With combined translation and rotation, element distance also influences vector direction, yielding a radically discontinuous field in

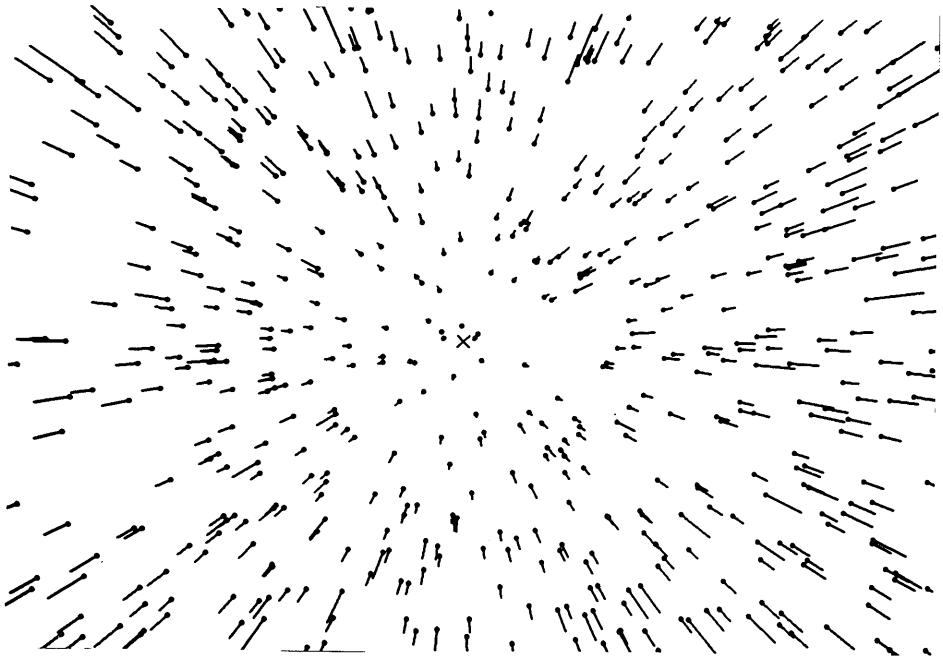

Fig. 6. Velocity field produced by observer translation through a 3D cloud of elements. The × indicates heading.

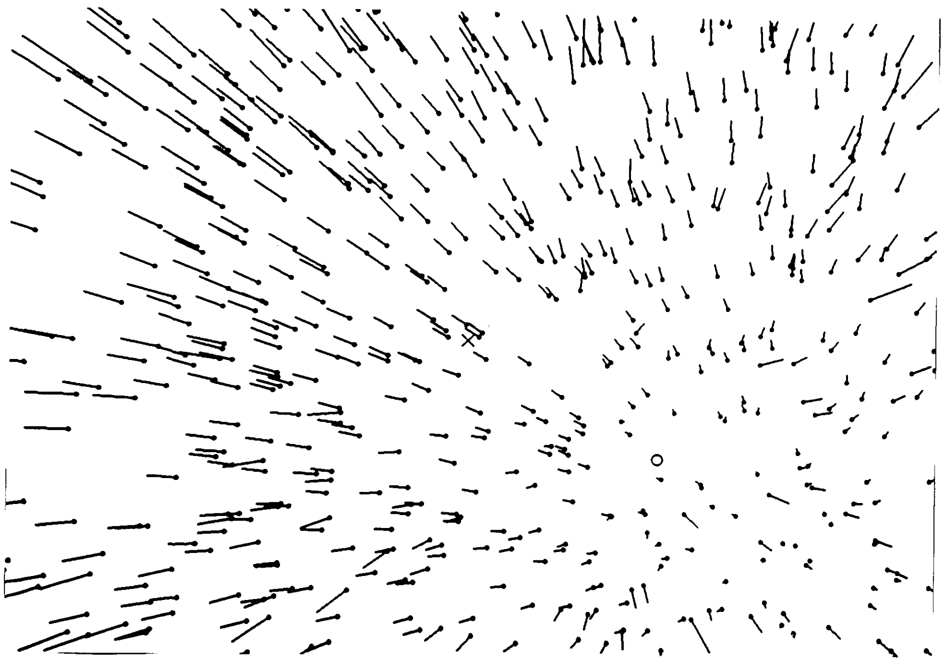

Fig. 7. Velocity field produced by combined observer translation and rotation through a 3D cloud of dots, resulting from translating toward the × while fixating the circle in the middle of the cloud.

which differential motion is particularly evident (Fig. 7). Differential, least-squares, and dynamical models all require locally continuous flow fields, but with a discontinuous field spatial derivatives and the topology of singular points are undefined. Further, the maximum of divergence is described for a plane,[17] but with a complex environmental structure such as a cloud there is no single maximum. Thus, if observers depend on these properties to decompose translation and rotation, performance should be at the level of chance under the simulated condition.

Moving and simulated displays depicted self-motion toward a cubic volume of random dots extending from 6.9 to 37.3 m in depth with its sides off screen, with a speed of 1.9 m/sec. Approximately 50 dots appeared on screen at the start of a trial, so the front and rear surfaces of the cloud were not visibly defined. The observer's heading varied horizontally, with angles of 0°, ±3°, and ±6° from the center of the screen, and the target appeared to the left or to the right of this point. The fixation point appeared in a random position on a horizontal axis in the middle of the cloud (at an initial distance of 21.3 m) within 4° of the target line, generating eye rotation about the vertical axis. The amount of rotation varied from 0.5° to 2.0° with a mean velocity that varied from 0.2 to 0.7 deg/sec. Four observers underwent 120 trials under each condition.

Results

Once again, heading judgments were accurate under both conditions (Fig. 8) with no difference in percent correct performance between them [$F(1, 3) = 2.83$, not significant]. Heading thresholds were 1.2° in the moving condition and 1.4° in the simulated condition [$t(3) = 1.78$, not significant]. Thus observers can perceive heading quite accurately with discontinuous fields, indicating that the visual system does not depend on the divergence maximum, other differential invariants, the topology of flow about singular points, or other methods such as least-squares models that assume a smoothness constraint, in order to decompose translation and rotation.

EXPERIMENT 4

Procedure

Self-motion toward a plane provides a special case that allows us to test general computational models and the differential-motion hypothesis. With linear translation perpendicular to a plane, the standard radial outflow pattern occurs. With translation and eye rotation, radial outflow from the heading point is replaced by nearly radial outflow from the fixation point (Fig. 9). Thus, if observers relied on the radial pattern, they would see themselves as heading toward the fixation point. Because local depth differences between neighboring elements are small in this case, there is little differential motion (although it increases toward the plane's horizon). Thus, if observers depend on differential motion, they should fail to decompose translation and rotation with such displays. On the other hand, the general discrete, differential, least-squares, and dynamical models predict that the observer would successfully recover heading. The maximum of divergence and the gradient are available as well, so heading judgments should also be successful if these differential invariants are sufficient.

Moving and simulated displays of self-motion toward a random-dot wall surface were designed so that if observers saw themselves as heading toward the fixation point, performance would be at the level of chance. The path of approach was always perpendicular to the wall, and the angle of the wall with respect to the screen (and thus the heading direction with respect to the center of the screen) varied, with values of 0°, ±3°, and ±6°, about a vertical axis. Time-to-contact with the surface was 4.9 sec at the beginning and 1.9 sec at the end of a trial, equivalent to initial and final distances of 9.3 and 3.7 m with a speed of 1.9 m/sec. Approximately 140 dots were visible at the start of a trial. The fixation point appeared in a random position on the horizontal midline of the wall such that its final position was within 4° of the target, generating eye rotation about the vertical axis. Total rotation varied from 0.5° to 3.5°, with a mean velocity that varied from 0.2 to 1.2 deg/sec. Six observers underwent 120 trials under each condition.

Results

In contrast to the preceding experiments, performance was accurate under the moving condition but near the level of chance under the simulated condition ($M = 52.9\%$ correct; Fig. 10), with a highly significant difference between them [$F(1, 5) = 98.60$, $p < 0.001$]. The mean threshold in the moving condition was 1.2°, but in the simulated condition subjects reported seeing themselves as always heading toward the fixation point. Thus self-motion toward a plane appears to be a special case. Observers cannot perform the decomposition to determine heading in the absence of local depth differences, contrary to the general computational models and the divergence and gradient hypotheses. (One caveat: least-squares models may also fail in the present case because of the restricted field of view, but other tests of them are provided by experiments 3 and 5.) The results suggest that differential motion is not only sufficient but

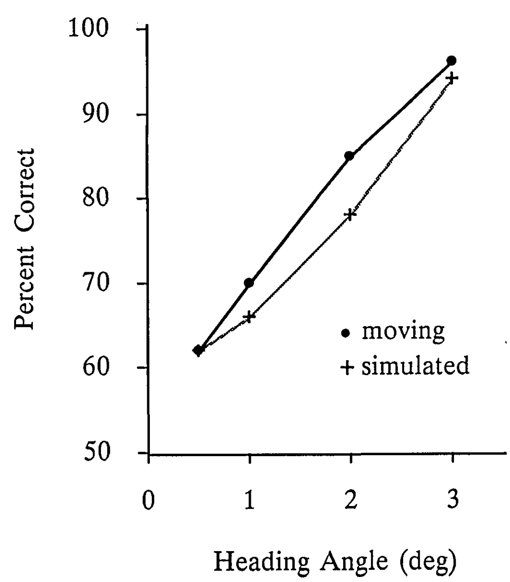

Fig. 8. Percent correct as a function of heading angle with moving fixation point and simulated eye rotation (experiment 3, cloud).

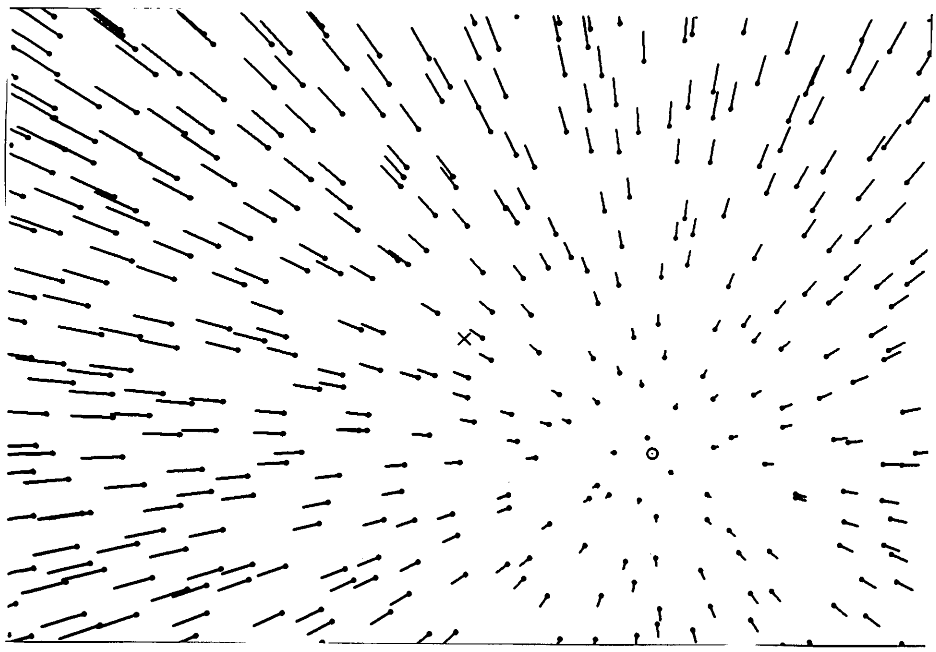

Fig. 9. Velocity field produced by combined observer translation and rotation relative to a wall surface, resulting from translating toward the × while fixating circle on the wall.

necessary for a decomposition based on information in the flow pattern itself. Because of practical limits on the resolution of optical flow,[34] the visual system cannot rely on precise measurements of the velocity field and appears to use more robust motion-parallax information. However, accurate performance in the moving condition indicates that some extra-flow-field information, such as an oculomotor signal or a residual frame of reference, can be used to resolve ambiguity in the optical flow.

EXPERIMENT 5

Procedure

In a final experiment, we examined the neighborhood size and the number of elements necessary to decompose observer translation and rotation. As the density of a 3D cloud of dots decreases, the number of elements decreases and the angular separation of neighboring elements increases. With larger angular separations the resulting difference vectors deviate increasingly from a radial pattern, and with fewer elements there are fewer difference vectors to average over; thus the differential motion theory predicts a decline in performance with decreasing density. On the other hand, the theory can tolerate relatively sparse flow fields that differential, least-squares, dynamical, and differential motion parallax theories cannot. By varying the cloud density, we can determine the angular separation and the number of dots sufficient for an accurate perception of heading.

Moving and simulated displays of self-motion toward a random-dot cloud were similar to those in experiment 3, except that the heading angle was held constant at 2°. The density of the cloud was manipulated to vary the neighborhood size and the number of dots while holding the volume within which dots were placed constant to preserve differential motion and visual angle. The neighborhood size was 1°, 2°, 4°, or 6°, with the corresponding numbers of visible dots at the start of a trial being 50, 25, 12, or 6. A neighborhood of, for example, 2° was defined to mean that there was an average of three or more pairs of dots (generating a total of three or more difference vectors) with angular separations of ≤2°, but not three pairs separated by ≤1°, in the first frame of the display (and thus on average in subsequent frames). Two observers underwent 120 trials under each condition.

Results

Observers could determine their direction of heading with rather sparse flow fields (Fig. 11). There were no overall

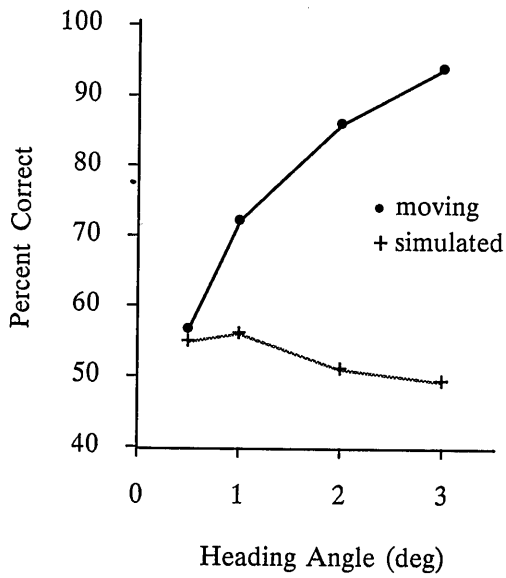

Fig. 10. Percentage of correct responses as a function of heading angle with moving fixation point and simulated eye rotation (experiment 4, wall).

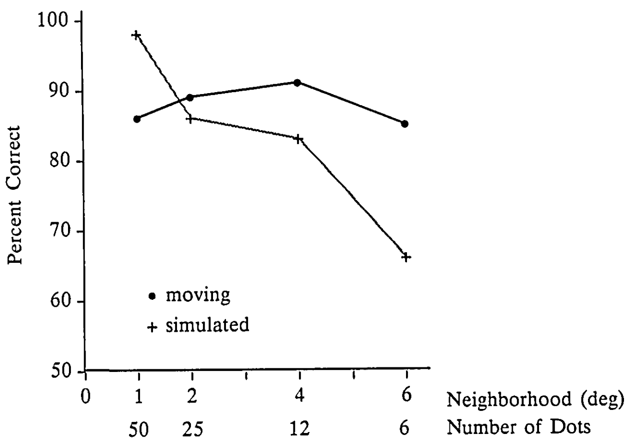

Fig. 11. Percentage of correct responses as a function of neighborhood size and dot density with moving fixation point and simulated eye rotation (experiment 5, cloud).

effects of density or condition on percent correct performance, but there was a significant interaction [$F(3, 3) = 10.15$, $p < 0.05$]; the only significant difference between results for moving and simulated conditions occurred with 6° neighborhoods [$t(2) = 6.50$, $p < 0.05$]. Consistent with the differential-motion hypothesis, performance under the simulated condition declined steadily with decreasing density, although observers could determine their heading quite accurately with as few as 12 elements and a 4° neighborhood. Thus, contrary to the predictions of other models, relatively sparse, discontinuous flow fields can be visually decomposed. The significant drop with a 6° neighborhood could be a result of either neighborhood size or the presence of only six dots, but in either case it contradicts most discrete models, which predict that four to six elements are sufficient. The results are consistent with a hypothesis of reliance on differential motion between rather widely separated elements (as much as 4° apart).

DISCUSSION

The experiments demonstrate that observers can distinguish the direction in which they are heading from the direction in which they are looking on the basis of flow information alone, consistent with Gibson's original notion that the visual system decomposes the flow field. The results question the relevance of the general computational models to biological vision, in particular the assumption of the existence of a smoothness constraint and continuous fields, but are thus far consistent with the narrower hypothesis of differential motion.

Let us evaluate each of the theories in turn. General discrete models are inconsistent with the data because observers fail where the models succeed, in the special case of an approach to a plane. Contrary to predictions, observers also appear unable to decompose a flow field containing only six elements, perhaps reflecting the models' vulnerability to noise. Differential models are contradicted because observers fail when differential invariants are available with the wall, but observers succeed when the differential invariants are undefined with the cloud. This casts doubt on the biological utility of the divergence maximum, the gradient, and other properties based on spatial derivatives of the flow field. This conclusion converges with results of a previous study of optical flow,[2] recent results on structure from motion[35] and stereopsis,[36] and observed limitations of the motion system for determining spatial derivatives.[18] Least-squares models are likewise contradicted by observers' failure with the wall but success with sparse, discontinuous fields with the cloud. Although such models may fail with a restricted view of a wall, observers do not require many elements or a smoothness constraint. The dynamical model is inconsistent for the same reason, as observers fail with the wall but succeed with sparse discontinuous fields in which the flow about singular points is not well behaved. Extra-flow-field variables such as oculomotor signals and visible frames of reference appear to be unnecessary because observers succeed under the simulated condition when flow information is placed in conflict with them. Likewise, ocular drift is unnecessary because observers are successful without multiple fixations.

On the other hand, observers appear to use some type of motion-parallax information to decompose the flow field. Edge parallax is unnecessary because observers succeed in the absence of depth edges. The differential-motion-parallax hypothesis is inconsistent with the data because observers succeed with sparse fields and without multiple fixations. But the results are consistent with a hypothesis of reliance on differential motion produced by depth differences within a neighborhood. Taken together, Experiments 2–5 indicate that differential motion is both necessary and sufficient for a decomposition based on flow-field information alone, because observers under the simulated condition fail when differential motion is removed during an approach to a plane and succeed when it is isolated in discontinuous clouds. In principle, differential motion could be detected by relative-motion-sensitive units, for which there is neurophysiological evidence.[37–39]

Although the results show that oculomotor signals, frames of reference, and multiple fixations are unnecessary, they do not prove that these variables could not contribute under some conditions. When differential motion is placed in conflict with oculomotor signals and reference frames under the simulated condition, observers rely on differential motion. However, successful performance with the wall under the moving condition (experiment 4) indicates that extra-flow-field variables can also be used in the absence of differential motion. We believe that residual frame effects probably account for this success as well as for slight differences between results for moving and simulated conditions in experiments 2 and 3, because the edges of the display were faintly visible despite the reduction screen.

Finally, we should note that the instantaneous velocity field is a limited characterization of optical flow that presents several ambiguities. First, there are two possible interpretations of the velocity field produced by translation and rotation relative to a plane: the veridical one and one in which the axis of translation and the surface normal are interchanged.[7,9] Second, identical velocity fields are produced by translation plus eye rotation and by curvilinear movement about a parallel axis external to the eye.[40,41] Computational models typically assume the eye to be the center of rotation, so that curvilinear movement is interpreted erroneously as translation plus eye rotation. Both of these ambiguities are resolved if the flow field is permitted to evolve over time, as it was in our displays of continuous dot motion. This suggests that the visual system makes use of extended temporal samples of optical flow.[14] The more complex case of eye rotation during curvilinear movement deserves further study.

In sum, the visual system appears to rely primarily on differential motion to decompose observer translation and rotation but, in its absence, may use extra-flow-field variables to resolve ambiguity in the optical flow. Contrary to computational models that assume a smoothness constraint, the visual system performs best with discontinuous flow fields, consistent with the structure of natural environments.

ACKNOWLEDGMENTS

This research was supported by grant AG05223 from the National Institutes of Health. We thank Jan Koenderink and Michael Turvey for their comments. Some of the results were summarized previously.[42]

REFERENCES

1. J. J. Gibson, *Perception of the Visual World* (Houghton Mifflin, Boston, Mass., 1950).
2. W. H. Warren, M. W. Morris, and M. Kalish, "Perception of translational heading from optical flow," J. Exp. Psychol. Hum. Percep. Perf. **14**, 646–660 (1988).
3. J. J. Gibson, *The Senses Considered as Perceptual Systems* (Houghton Mifflin, Boston, Mass., 1966).
4. C. Torrey, "Visual flow and the direction of locomotion," Science **277**, 1064 (1065).
5. J. J. Gibson, "The visual perception of objective motion and subjective movement," Psychol. Rev. **61**, 304–314 (1954).
6. H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," Nature **293**, 133–135 (1981).
7. H. C. Longuet-Higgins, "The visual ambiguity of a moving plane," Proc. R. Soc. London Ser. B **223**, 165–175 (1984).
8. R. Y. Tsai and T. S. Huang, "Estimating three-dimensional motion parameters of a rigid planar patch," IEEE Trans. Acoust. Speech Signal Process. **ASSP-29**, 1147–1152 (1981).
9. R. Y. Tsai and T. S. Huang, "Estimating three-dimensional motion parameters of a rigid planar patch: III. Finite point correspondences and the three-view problem," IEEE Trans. Acoust. Speech Signal Process. **ASSP-32**, 213–220 (1984).
10. R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-6**, 13–27 (1984).
11. K. Prazdny, "Egomotion and relative depth map from optical flow," Biol. Cybern. **36**, 87–102 (1980).
12. H.-H. Nagel, "On the derivation of 3D rigid point configurations from image sequences," in *Proceedings of the IEEE Conference on Pattern Recognition and Image Processing* (IEEE Computer Society Press, New York, 1981).
13. A. Mitiche, "Computation of optical flow and rigid motion," in *Proceedings of the IEEE Workshop on Computer Vision: Representation and Control* (IEEE Computer Society, New York, 1984).
14. W. H. Warren, W. Griesar, A. W. Blackwell, M. Kalish, and N. G. Hatsopoulos, "On the sufficiency of the velocity field for perception of heading," to be submitted to Biol. Cybern.
15. J. J. Koenderink and A. J. van Doorn, "Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer," Opt. Acta **22**, 773–791 (1975).
16. J. J. Koenderink and A. J. van Doorn, "Local structure of movement parallax of the plane," J. Opt. Soc. Am. **66**, 717–723 (1976).
17. J. J. Koenderink and A. J. van Doorn, "Exterospecific component of the motion parallax field," J. Opt. Soc. Am. **71**, 953–957 (1981).
18. K. Nakayama, "Extraction of higher-order derivatives of the optical velocity vector field: limitations imposed by biological hardware," in *Brain Mechanisms and Spatial Vision*, D. Ingle, M. Jeannerod, and D. Lee, eds. (Martinus-Nijhoff, The Hague, The Netherlands, 1985).
19. H. C. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal image," Proc. R. Soc. London Ser. B **208**, 385–397 (1980).
20. A. M. Waxman and S. Ullman, "Surface structure and 3D motion from image flow: a kinematic analysis," Int. J. Robotics Res. **4**, 72–94 (1985).
21. K. Prazdny, "Determining the instantaneous direction of motion from optical flow generated by a curvilinearly moving observer," Comput. Graphics Image Process. **17**, 238–258 (1981).
22. A. R. Bruss and B. K. P. Horn, "Passive navigation," Comput. Graphics Image Process. **21**, 3–20 (1983).
23. D. H. Ballard and O. A. Kimball, "Rigid body motion from depth and optical flow," Comput. Vision Graphics Image Process. **22**, 95–115 (1983).
24. A. Verri, F. Girosi, and V. Torre, "Mathematical properties of the two-dimensional motion field: from singular points to motion parameters," J. Opt. Soc. Am. A **6**, 698–712 (1989).
25. J. H. Rieger and D. T. Lawton, "Processing differential image motion," J. Opt. Soc. Am. A **2**, 354–360 (1985).
26. J. Cutting, *Perception with an Eye for Motion* (MIT Press, Cambridge, Mass., 1986).
27. W. Richards, "Visual space perception," in *Handbook of Perception*, E. C. Carterette and M. P. Friedman, eds. (Academic, New York, 1975), Vol. 5, pp. 351–386.
28. D. Regan, "Visual flow and the direction of locomotion," Science **277**, 1064–1065 (1985).
29. E. von Holst, "Relations between the central nervous system and the peripheral organs," Animal Behav. **2**, 89–94 (1954).
30. I. R. Johnston, G. R. White, and R. W. Cumming, "The role of optical expansion patterns in locomotor control," Am. J. Psychol. **86**, 311–324 (1973).
31. J. J. Gibson, "What gives rise to the perception of motion?" Psychol. Rev. **75**, 335–346 (1968).
32. D. Regan and K. I. Beverley, "How do we avoid confounding the direction we are looking and the direction we are moving?" Science **215**, 194–196 (1982).
33. J. H. Rieger and L. Toet, "Human visual navigation in the presence of 3D rotations," Biol. Cybern. **52**, 377–381 (1985).
34. J. J. Koenderink and A. J. van Doorn, "Facts on optic flow," Biol. Cybern. **56**, 247–254 (1987).
35. J. Lappin and F. Norman, "Detectabilities of the components of optical flow," presented at the Psychonomic Society Annual Meeting, Chicago, Ill., November 10–12 1988.
36. B. Gillam, "Is orientation disparity a stereoscopic primitive?" presented at the Psychonomic Society Annual Meeting, Chicago, Ill., November 10–12 1988.
37. P. Hammond and A. T. Smith, "On the sensitivity of complex cells in feline striate cortex to relative motion," Exp. Brain Res. **47**, 457–460 (1982).
38. B. Frost and K. Nakayama, "Single visual neurons code opposing motion independent of direction," Science **220**, 744–745 (1983).
39. G. A. Orban, B. Gulyas, and R. Vogels, "Influence of a moving textured background on direction selectivity of cat striate neurons," J. Neurophys. **57**, 1792–1812 (1987).
40. J. H. Rieger, "Information in optical flows induced by curved paths of observation," J. Opt. Soc. Am. **73**, 339–344 (1983).
41. W. H. Warren, D. R. Mestre, A. W. Blackwell, and M. W. Morris, "Perception of curvilinear heading from optical flow," submitted to J. Exp. Psychol. Hum. Percep. Perf.
42. W. H. Hannon and D. J. Hannon, "Direction of self-motion is perceived from optical flow," Nature **336**, 162–163 (1988).