Systematic distortions of perceived planar surface motion in active vision

Carlo Fantoni
Center for Neuroscience and Cognitive Systems, Italian Institute of Technology, Genova, Italy

Corrado Caudek
Department of Psychology, University of Florence, Firenze, Italy

Fulvio Domini
Center for Neuroscience and Cognitive Systems, Italian Institute of Technology, Genova, Italy, & Department of Cognitive and Linguistic Sciences, Brown University, Providence, RI, USA

Recent studies suggest that the active observer combines optic flow information with extra-retinal signals resulting from head motion. Such a combination allows, in principle, a correct discrimination of the presence or absence of surface rotation. In Experiments 1 and 2, observers were asked to perform such discrimination task while performing a lateral head shift. In Experiment 3, observers were shown the optic flow generated by their own movement with respect to a stationary planar slanted surface and were asked to classify perceived surface rotation as being small or large. We found that the perception of surface motion was systematically biased. We found that, in active, as well as in passive vision, perceived surface rotation was affected by the deformation component of the first-order optic flow, regardless of the actual surface rotation. We also found that the addition of a null disparity field increased the likelihood of perceiving surface rotation in active, but not in passive vision. Both these results suggest that vestibular information, provided by active vision, is not sufficient for veridical 3D shape and motion recovery from the optic flow.

Keywords: active vision, 3D shape and surface perception, motion—3D, optic flow

Introduction

In the present investigation, we explored the effects of extra-retinal signals produced by head motion on the presence–absence discrimination of surface rotation. We compared the predictions of two classes of models for the perceptual interpretation of the optic flow. In the first approach, the perceptual analysis of the optic flow relies on information provided by the extra-retinal signals (Colas, Droulez, Wexler, & Bessière, 2007; Dijkstra, Cornilleau-Pérès, Gielen, & Droulez, 1995; Dyde & Harris, 2008; Jaekl, Jenkin, & Harris, 2005; Ono & Steinbach, 1990; Panerai, Cornilleau-Pérès, & Droulez, 2002; Peh, Panerai, Droulez, Cornilleau-Pérès, & Cheong, 2002; Rogers & Rogers, 1992; van Boxtel, Wexler, & Droulez, 2003; Wexler, 2003; Wexler, Lamouret, & Droulez, 2001; Wexler, Panerai, Lamouret, & Droulez, 2001; Wexler & van Boxtel, 2005). In the second approach, the perceptual analysis of the optic flow is mainly driven by retinal information, even if extra-retinal signals are available (Rogers & Graham, 1979; van Damme & van de Grind, 1996; Wallach & O'Connell, 1953; Wallach, Stanton, & Becker, 1974).

The structure-from-motion problem

The Structure-from-Motion (SfM) problem has been mostly studied under the *rigidity assumption*. Under this assumption, the Euclidean three-dimensional (3D) structure of the distal object can in principle be recovered from the image transformations produced by an orthographic projection, if second-order temporal information is available (Longuet-Higgins, 1984; Longuet-Higgins & Prazdny, 1980; Ullman, 1979). However, a large number of psychophysical studies have shown that human observers exhibit a very limited sensitivity to the second-order temporal properties of the optic flow. It has been found, in fact, that perceived SfM depends only on the specific properties of the first-order optic flow, that is, on the gradients of the velocity field (Domini, Caudek, & Proffitt, 1997; Todd & Bressan, 1990).

Under polar (as opposed to orthographic) projection, the first-order temporal properties of the optic flow provide sufficient information, in principle, for veridical perception of Euclidean metric structure (Longuet-Higgins, 1981; Mayhew & Longuet-Higgins, 1982). Perspective information in motion displays, however, is perceptually effective only for stimuli subtending large

visual angles (Eagle & Hogervorst, 1999; Hogervorst & Eagle, 2000).

Optic flow and active vision

In active vision, the optic flow is actively generated (not passively observed) by the observer who receives extra-retinal signals (e.g., efference copies of motor commands, proprioceptive and vestibular information) and couples them with retinal motion signals in order to produce a 3D percept (Colas et al., 2007; Wexler & van Boxtel, 2005). In the present study, the stimulus displays subtended relatively small visual angles. Hence, perspective information was negligible and the perceptual analysis of the local optic flow can be approximated by
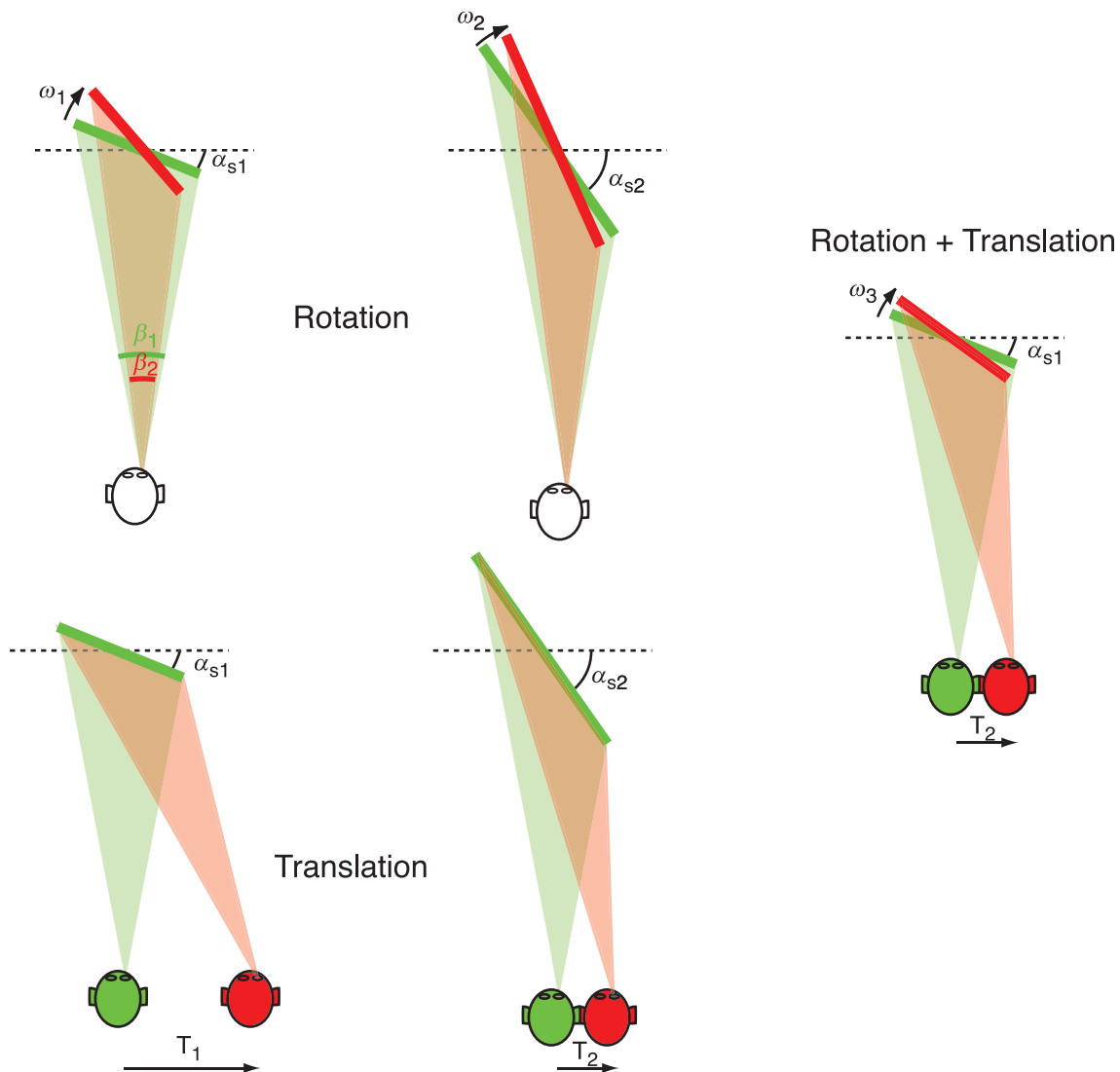

Figure 1. Ambiguity of first-order temporal information in active and passive visions. Each sketch shows two successive bird's-eye views of a planar surface slanted about the vertical axis with the colors (green and red) coding for the temporal ordering of the views (first and second, respectively). The rate of change of the visual angle subtended by a surface approximates a local property of the velocity that is informative about 3D shape: *def*. Here, the average *def* can be visualized by the difference between two subsequent visual angles subtended by the surface on the right eye (β_1 in green; β_2 in red). The four sketches on the left panel depict four specific cases in which the same optical angles are produced for two differently slanted surfaces (α_{s1} or α_{s2}) either by two different clockwise rotation angles (top row: ω_1 and ω_2) of the surface relative to an immobile observer or by two different amount of observer rightward translation (bottom row: T_1 and T_2) relative to an immobile surface. A surface with a small slant (α_{s1}) can produce the same amount of image deformation as a surface with a larger slant (α_{s2}) if its rotation around the vertical is larger (to some well-defined extent) or if it is viewed while performing a larger head movement. The right panel shows that the same visual angles can be produced even in a general case in which both the observer and the surface are moving. Without adding further assumptions about the motion of the surface or the motion of the observer, it is not possible to extract the veridical motion of the plane.

the relations described in Appendix A. Koenderink and van Doorn (1975, 1978; Koenderink, 1986) have shown how the distortion of an object's two-dimensional (2D) image can be decomposed into components of divergence (isotropic expansion or dilatation), curl (2D rotation or vorticity), and deformation (symmetric and anti-symmetric shearing). The deformation (*def*) is the only component that provides information about the surface orientation and motion in the 3D scene.

The relation between *def*, head's motion, surface rotation, and surface slant (see Figure A1) is shown in the following equation (see Appendix A for a derivation):

$$def = (T_{\alpha x} + \omega)\tan(\alpha_s + \alpha_0), \qquad (1)$$

where α_0 is the visual direction, $T_{\alpha x}$ is the horizontal translatory motion component of the observer's head (expressed in terms of angular velocity), α_s is the surface's slant about the vertical axis, and ω is the angular rotation velocity of the surface about the vertical axis (see Figure A2).

Discrimination of presence or absence of surface rotation

Does *def* provide sufficient information to allow veridical discrimination of the presence or absence of surface rotation? In any single moment in time, *def* does not specify the presence or absence of surface rotation even if $T_{\alpha x}$ and α_0 are known.[1] The contribution of ω to the total deformation, in fact, is confused with the contribution of surface slant (i.e., α_s). As a consequence, the same *def* can be produced by a stationary or by a rotating surface, depending on α_s (see Figure 1).

A veridical discrimination of the presence or absence of surface rotation, however, is possible if we combine extra-retinal information resulting from head motion with deformation change over time (Caudek, Domini, & Di Luca, 2002; Domini, Caudek, & Skirko, 2003; Domini, Vuong, & Caudek, 2002). Let us assume that $T_{\alpha x}$ and α_0 are specified by proprioceptive information. If the distal surface is stationary, then $\omega = 0$ and an unbiased estimate of α_s (surface slant) can be found from Equation 1. This estimate of α_s remains constant in successive moments of the motion sequence.

If the distal surface undergoes a rotation during head translation, solving Equation 1 for α_s with $\omega = 0$ will produce a *biased* estimate of α_s. Such biased estimate will take on different values in different moments in time.

We can thus distinguish between two classes of events: those in which $\hat{\alpha}_s$ remains constant and those in which $\hat{\alpha}_s$ varies in time. These two classes of events set apart the optic flow fields produced by stationary and rotating surfaces, respectively.

In conclusion, veridical discrimination of the presence or absence of surface rotation is possible only if extra-retinal information is taken into account, together with deformation change over time. Moreover, the manipulation of surface tilt should have no effect. In fact, surface tilt does not enter in Equation 1.
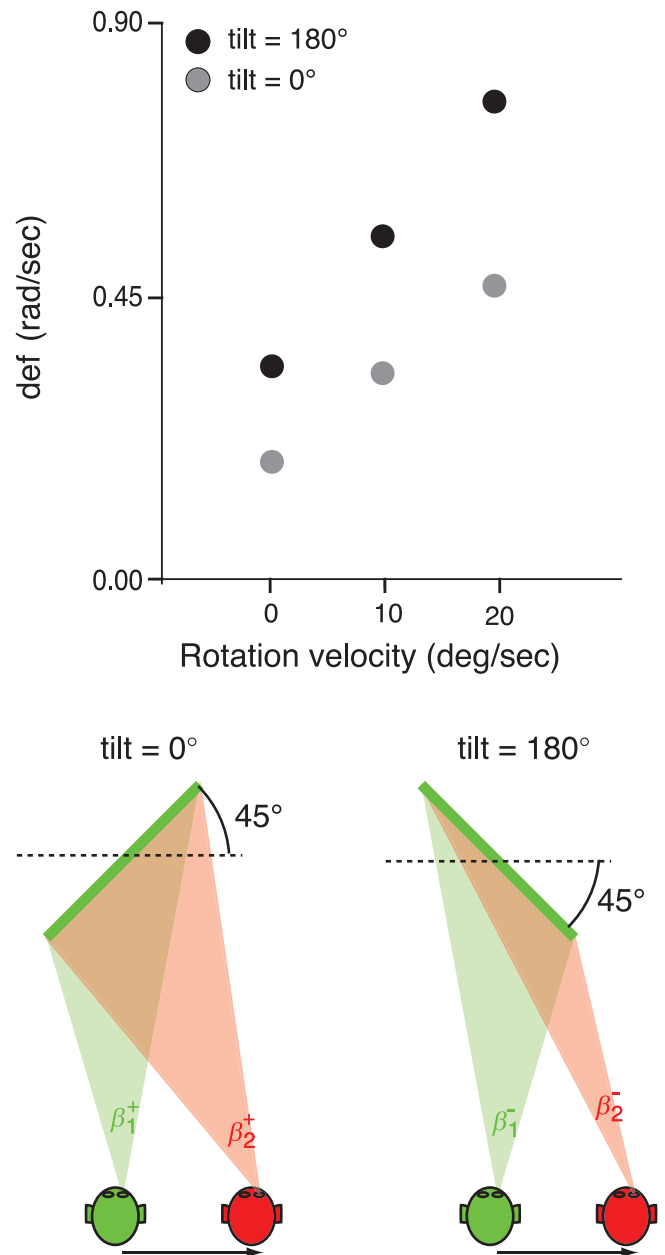

Figure 2. (Top) Average *def* as a function of rotation velocity and tilt for the stimuli of Experiment 1. Values have been calculated by entering into Equation A4 the viewing parameters characterizing Experiment 1 (x_0 ranging from 30 to 65 mm, z_f kept constant at 480 mm, and T_x = 125 mm/s). (Bottom) Different *def*s are produced when viewing surfaces with equal slant magnitude (i.e., 45°) but opposite tilt directions (0° on the left and 180° on the right) while performing the same rightward lateral head shift. The difference between the visual angles subtended by the 180° tilted surface ($\beta_2^- - \beta_1^-$) is indeed larger than the difference between the visual angles subtended by the 0° tilted surface ($\beta_2^+ - \beta_1^+$).

Heuristic interpretation of the optic flow

Suppose that extra-retinal signals are ignored. In these circumstances, $T_{\alpha x}$ and α_0 are left unspecified and *def* remains ambiguous, not only in each single moment in time but also across an extended time window. Veridical discrimination of the presence or absence of surface rotation is not possible. Therefore, in active vision we should expect to find the same systematic biases reported in passive SfM: perceived rotation should be a positive function of *def*, *regardless of actual surface rotation* (Caudek & Domini, 1998; Caudek & Proffitt, 1993; Caudek & Rubin, 2001; Domini et al., 1997; Domini, Caudek, Turner, & Favretto, 1998; Todd & Bressan, 1990; Todd & Perotti, 1999).

For the stimuli of Experiments 1a and 1b, *def* as a function of angular rotation and surface tilt is plotted in Figure 2. Note that in Experiments 1a and 1b, *def* covaried with tilt. If the perceptual analysis relies exclusively on retinal signals, then we should expect a larger number of "Rotation" responses for surfaces with 180° tilt, regardless of actual surface rotation.

Experiment 1a

The experimental setting provided the following depth cues: (i) motion-parallax produced by the lateral head shifts of the observer's head and (in some trials) by the concurrent rotation of the simulated surface about the vertical axis, and (ii) proprioceptive information of self-generated eye and head motion. Observers viewed monocularly the moving image of a planar surface yoked to the movements of their head (Figure 3, left). The

simulated planar surface was always slanted 45°, but in different trials tilt was varied (see Figure 2).

Methods

Participants

Eleven undergraduates at the University of Parma participated in the experiment. All had normal or corrected-to-normal vision and were naive to the purpose of the experiment.

Apparatus

The translational displacements and orientation of the participant's head were recorded on-time by an Optotrak Certus system with two position sensors (0.01-mm resolution). The two position sensors recovered the 3D positions of three markers from three infrared emitting diodes (8.0 V, 2500 Hz). The markers were attached on the frontal part of a Sensic PiSight Head Mount. Position sensors were placed at optimal distance from the observer's helmet. The two sensors were arranged 225 cm apart, with the observer's resting position centered between them at a distance of 265 cm. Sampling of the head tracker was set at 2500/3 Hz. The tracker's latency was lower than the sample interval.

A Dell Precision T3400 525W (using an Intel Core 2 Extreme 5252W, QX9650, 3.00 GHz, 1333 MHz FSB, 12 MB L2 Cache) controlled the stimulus display and sampled the tracker (using a standard PCI card). The positions of Random Dots (RDs) forming our stimuli were updated in real time on a ViewSonic 9613, 19″ CRT monitor. The monitor was set at 1024 × 768 pixel resolution (0.24-mm diagonal dot pitch: 1.68 arcmin at the observer's distance of 480 mm) and was driven by an

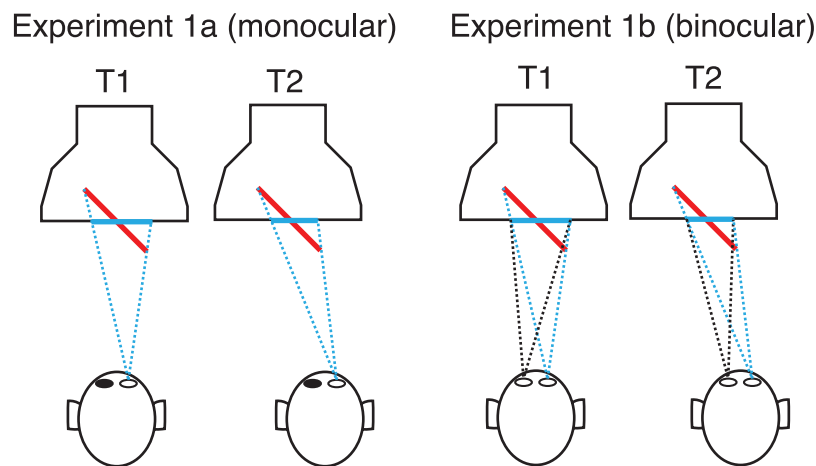Experiment 1a (monocular)   Experiment 1b (binocular)

Figure 3. We yoked the movement of the image of a planar surface (cyan continuous line) to the movement of the head. In Experiment 1, we manipulated the mode of viewing: monocular to the left (Experiment 1a), binocular to the right (Experiment 1b). Both illustrations depict a bird's-eye view of a rightward lateral head shift from an aligned head position (T1) to an eccentric head position (T2). When viewing was binocular, a null disparity field was presented together with optic flow information.

nVidia Quadro FX 4600 with 768 Mb. A custom Visual C++ program supported by OpenGL Libraries and combined with Optotrak API routines was used for stimulus presentation/response recording.

Displays were viewed through a high-quality front-silvered mirror (150 × 150 mm) placed in front of the observer's central viewing position and slanted 45° away from the monitor and the observer's inter-ocular axis. The effective distance from the pupil to the center of the screen was 480 mm (Figure 4).

Displays were viewed through liquid-crystal-diode (LCD) shutter glasses (FE-1 Goggles, Cambridge Research System) synchronized to the monitor. Depending on the viewing mode, monocular (Experiment 1a) or binocular (Experiment 1b), shutter over the dominant eye was opened or closed electronically. The electronically driven shutters allowed us to randomly switch between the two viewing modes. Follow-up interviews revealed that none of the participants realized that (binocular vs. monocular) viewing was manipulated. Moreover, all participants reported perceiving the RD stimuli as either a stationary or a rotating rigid surface, in both monocularly and binocularly viewed displays. As a consequence of using the shutter glasses, the effective CRT refresh rate was halved (60 Hz).

Displays

The stimulus displays will be described by using a reference frame with the xy-plane coplanar with the
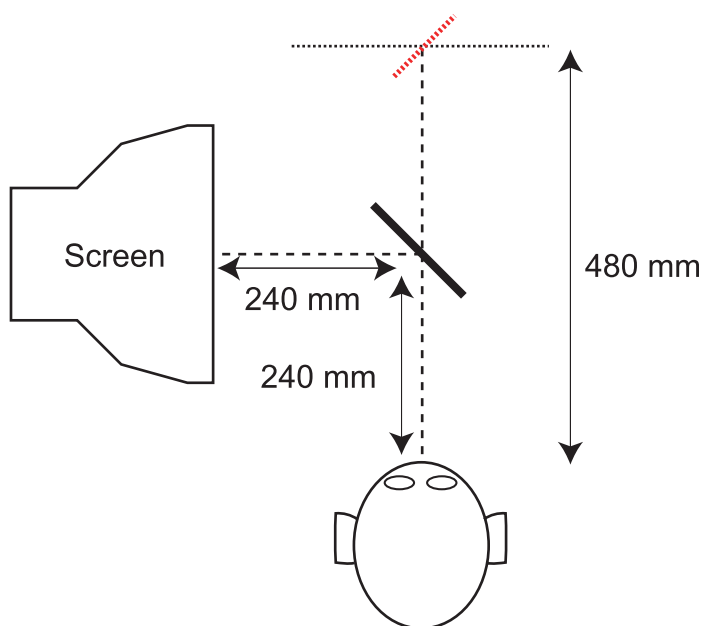

Figure 4. A diagram of the viewing apparatus and setting, including the mirror, the CRT screen, the observer, and the virtual image of the screen (dashed black line) plus the simulated slanted plane (dashed red line). Dashed lines show the light path, from the CRT to the lumen of the eye, for a standard observer at rest.

monitor screen, the x-axis pointing to the subject's right, the y-axis upward, and the z-axis away from the subject. The origin of the reference frame was set at the center of the monitor's screen.

The stimulus displays were arrangements of 300 random anti-aliased red dots (red color was used to eliminate cross-talk) simulating the projection of a squared RD planar surface (5.9° × 5.9°) centered on the image screen and slanted ±45° around the y-axis. The dot arrangement was varied by taking into account the observer's head position and his/her orientation with respect to the simulated surface. We set the dots to the maximal electron-gun value of 82 cd/m^2; the black background was 3 cd/m^2.

To remove texture (nonmotion) cues, the dots were randomly distributed in the projected image (not on the simulated surface) by imposing $z_0 = \tan(g1)x_0 + \tan(g2)y_0$, with x_0 and y_0 randomly selected in the range between ±25 mm from the screen center, and g1 and g2 representing the amount of surface rotation around the y- and x-axes, respectively. In terms of slant σ and tilt τ of a planar surface, g1 corresponds to $\sin(\tau)\tan(\sigma)$, and g2 to $\cos(\tau)\tan(\sigma)$, so that $\sigma = \arctan\sqrt{\tan(g1)^2 + \tan(g2)^2}$ and $\tau = \text{atan}(\tan(g2)/\tan(g1))$. In different trials, we set g1 = ±45° and g2 = 0, thus producing two planar surfaces with equal slant (45°) but opposite tilt angles (represented by the sign of g1; see Figure 2).

The stimulus displays were centered at the observer's cyclopean eye at rest. For each stimulus frame, the dots of the simulated planar surface were projected onto the screen by using a generalized perspective pinhole model with the observer's right eye position (measured with almost no latency) used as the Center Of Projection (COP).

The real-time stimulus update, as a function of observers' position, produced a relative rotation of the simulated planar surface of about ±4.13° about the vertical axis, regardless of tilt (viewing distance: 480 mm; maximum lateral head shift: 35 mm). The motion of the dots generated an approximately linear optic flow with velocity vectors almost parallel.

At the beginning of each trial, surface slant was 41.4° or 48.6° for the 180° tilted and 0° tilted surfaces, respectively (calculated for an average inter-ocular distance of 60 mm and a right dominant eye as the COP). At the extreme lateral head positions (35 mm rightward head shift), surface slant was 37.3° or 52.7° for the 180° and 0° tilted surfaces, respectively.

In half of the trials, the simulated surface was stationary, as described above. In the remaining trials, we simulated a planar surface undergoing a vertical-axis rotation of either 10 deg/s or 20 deg/s (constant angular rotation velocity). For the displays in which we simulated a surface rotation, the optic flow produced by surface rotation was added to the velocity field generated by the observer's rightward lateral head movement. At the average head shift velocity of 15 deg/s (determined during a preliminary training phase of the experiment), the
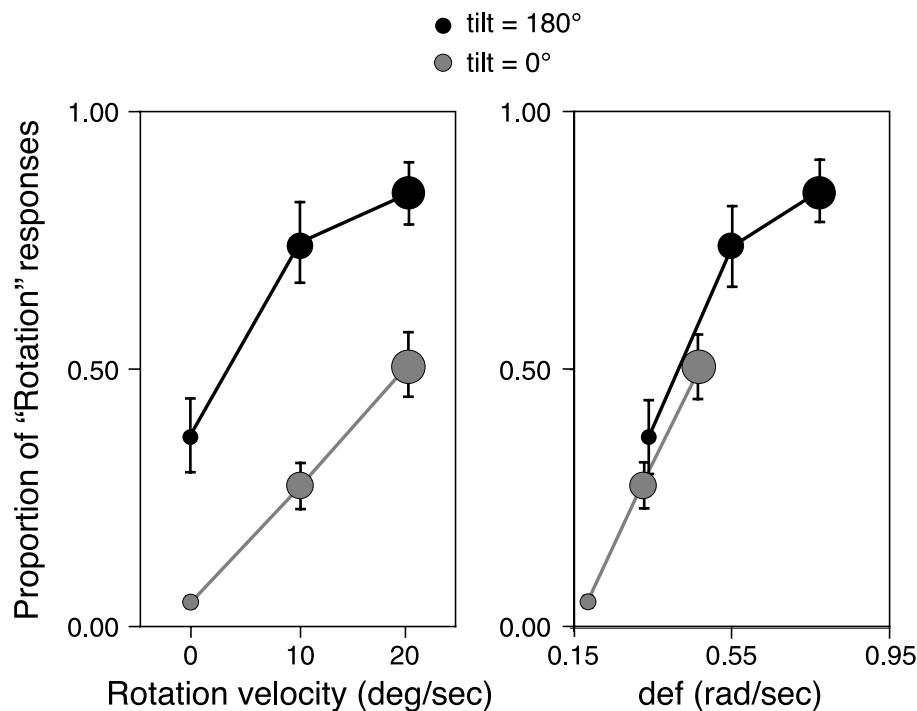
stimulus was presented on the monitor's screen for 0.28 s. For the present stimuli, a plot of *def* as a function of Rotation Velocity and Tilt is provided in Figure 2 (top panel).

Design

A 2 × 3 within-subjects experimental design was used, with two Tilt angles (180° and 0°), and three angular Rotation Velocities (0, 10, and 20 deg/s).

Procedure

Participants were tested individually in a dark room. After 5 min of dark adaptation, the observer's head was positioned on a horizontally extended chin rest allowing 80 mm of lateral head shift. The chin rest, parallel to the horizontal dimension of the monitor's screen, was adjusted in height to position the participant's cyclopean eye at the screen center.

At the beginning of each trial, a red fixation mark was shown in the center of the screen and the observer was required to move his head rightward (Figure 5). The observer was required to reverse the direction of head motion after hearing a beep signaling a head shift of 35 mm relative to the center of the screen and after hearing a beep signaling a −35 mm shift in the opposite direction.

After two cycles, when the observer's head passed through the center of the screen moving rightward, the fixation mark was replaced by the stimulus display. The simulated random-dot planar surface remained visible for half of a cycle, at which time it disappeared and was replaced by a blank screen. The observer provided his or her judgment after stopping head motion. The observer's task was to classify, with a press of a mouse button, the surface as being rotating or static.

The experimental session lasted for about 90 min and was divided into two blocks of 160 trials each. Each block consisted of 10 random sequences of 12 experimental conditions: 2 Experiments (Experiment 1a/monocular, Experiment 1b/binocular) × 2 Tilt angles (180°, 0°) × 3 angular Rotation Velocities (0, 10, and 20 deg/s). To balance the number of static and rotating surfaces, 40 more displays were generated (20 displays representing a static surface with a 180° tilt and 20 displays representing a static surface with a 0° tilt).

The experimental session was preceded by a session in which observers were trained to maintain a constant velocity during lateral head movement (∼125 mm/s corresponding to 15 deg/s). Observers were then presented with 46 trials randomly selected from the different experimental conditions. Only participants with more than 60% correct responses during training were admitted to the experimental session.

Results and discussion

The left panel of Figure 6 shows the average proportion of valid "Rotation" responses as a function of rotation velocity and tilt angle. We considered "valid" only the trials in which head translational velocity was included in the interval from 44 mm/s (meaning that the observer interrupted the head motion) to 350 mm/s (meaning that the observer performed a sudden movement of the head). By this criterion, less than 1% of the trials were eliminated. The average head translational velocity was pretty stable across subjects (129 mm/s ± 3.6 mm/s): the 90% of head translational velocities were in the interval between 120.3 mm/s and 138.1 mm/s.

If we consider each tilt condition separately, we see that observers reported a larger number of "Rotation"

Figure 5. Temporal sequence used in Experiment 1. The illustration refers to a 180° tilted surface trial.

Figure 6. Rotation detection performance in Experiment 1a. Panels on the left show mean proportions of "Rotation" responses as a function of rotation velocity for the two levels of tilt (coded by color). The same data are replotted as a function of average *def* in the panels on the right. The size of the circles codes for the rotation velocity (small for static, medium for 10 deg/s, and large for 20 deg/s). Error bars represent ±1 *SE*.

responses for rotating (10 deg/s and 20 deg/s rotation) rather than for stationary surfaces. Moreover, Figure 6 shows that the likelihood of a "Rotating" response increases with rotation velocity.

If we consider both tilt conditions together, however, we realize that the present results do not provide evidence of a correct discrimination between stationary and rotating surfaces. First, observers reported a larger number of "Rotation" responses for stationary surfaces with a 180° tilt than for surfaces rotating by 10 deg/s with a 0° tilt. Second, the likelihood of a "Rotation" response was larger for surfaces with 180° rather than 0° tilt. These results are consistent with a 3D interpretation based exclusively on the *def* component of the first-order optic flow (see Figure 2, top panel), which does not allow a veridical discrimination of presence or absence of surface rotation. In fact if we replot the proportion of "Rotation" responses as a function of average *def* (right panels of Figure 6), we see that, in active, as well as in passive vision, proportion of "Rotation" responses increased monotonically with *def* (*x*-axis), not with Rotation Velocity (circle's size), regardless of the fact that the simulated surface is stationary or rotating.

In order to test the discrimination ability of our observers, once the effect of *def* is controlled, we computed *d′* by following the procedure indicated by Wright and London (2009) and Wright, Horry, and Skagerberg (2009). One of the advantages of computing

d′ by means of a linear mixed effects (*lme*) analysis involves the possibility of adding a continuous variable to the model (in our case, *def*). In a first *lme* model, disregarding the effect of *def*, we asked whether observers can discriminate stationary from rotating surfaces. In this model, *d′* took on the value of 1.123 ($z = 8.738$, $p < 0.001$), indicating a veridical performance. More interesting, however, was to repeat the same analysis when the effect of *def* was statistically controlled. In a second *lme* model, by adding *def* as a covariate, *d′* becomes statistically equal to zero ($d′ = -0.100$, $z = -0.576$, $p = 0.565$): there was no evidence that observers could veridically discriminate stationary surfaces from rotating ones *when def was kept constant*. In this second model, the likelihood of a "Rotating" response was completely explained by *def* ($\beta_{\text{def}} = 5.292$, $z = 15.566$, $p < 0.001$).

As indicated in the Introduction section, our displays provided sufficient information for a correct discrimination between stationary and rotating surfaces. Vestibular information can be used to estimate the parameters $T_{\alpha x}$ and α_0 of Equation 1, and we know that vestibulo-ocular reflex is more effective than pursue eye movements for image stabilization (Bennur & Gold, 2008; Buizza, Leger, Droulez, Bertoz, & Schmid, 1980; Ferman, Collewijn, Jansen, & van den Berg, 1987; Gu, Angelaki, & DeAngeis, 2008; Gu, DeAngeis, & Angelaki, 2007; Liu & Angelaki, 2009). Therefore, we might have expected an advantage of active over passive vision. Nevertheless, the present data

suggest that in active vision, perceived surface rotation is a direct function of *def*, rather than of actual surface rotation. For our stimuli, the additional information provided by vestibular information was not sufficient for the veridical estimation of planar surface motion from the optic flow.

Experiment 1b

In Experiment 1b, the optic flow was generated in the same manner as in Experiment 1a, but viewing was binocular (Figure 3, right). The same optic flow was shown to both eyes, thus creating a null disparity field. The purpose was to provide conflicting information about the viewing distance.

The pairing of a velocity gradient with a null disparity field does not necessarily correspond to a conflict of cues. Such a stimulus, in fact, can be generated by a target object positioned at a large viewing distance from the observer.[2] For the stimuli of Experiment 1b, the velocity/disparity pairing was compatible with a viewing distance of at least 3 m (see Fantoni, 2008). Such a large viewing distance, however, was at odds with the fact that both vergence and accommodation were modulated by a much smaller viewing distance of only 480 mm (i.e., screen distance). We reasoned that this conflict between the cues

to viewing distance could be resolved either by vetoing extra-retinal information (vergence and accommodation), or by disregarding the retinal information (null disparity field).

1. If vergence and accommodation are disregarded, then the 3D interpretation of the optic flow must take into consideration the large viewing distance compatible with the null disparity field. A stationary surface positioned very far from the observer can generate only a negligible motion parallax, if the observer's head moves by a small amount. In our experiments, head's motion was small but motion parallax was far from negligible. This stimulus situation is only compatible with a rotation of the distal surface—see the motion-distance invariance principle (Gogel & Tietz, 1973; Hay & Sawyer, 1969; Tyler, 1974; Wallach, Yablick, & Smith, 1972). If vergence and accommodation are vetoed, therefore, we expect a larger likelihood of "Rotation" responses in Experiment 1b than in Experiment 1a.
2. If in Experiment 1b the information provided by the null disparity field is disregarded, then we should find the same results as in Experiment 1a.

We intermixed the trials of Experiments 1a and 1b to allow direct comparison between them, as this avoided
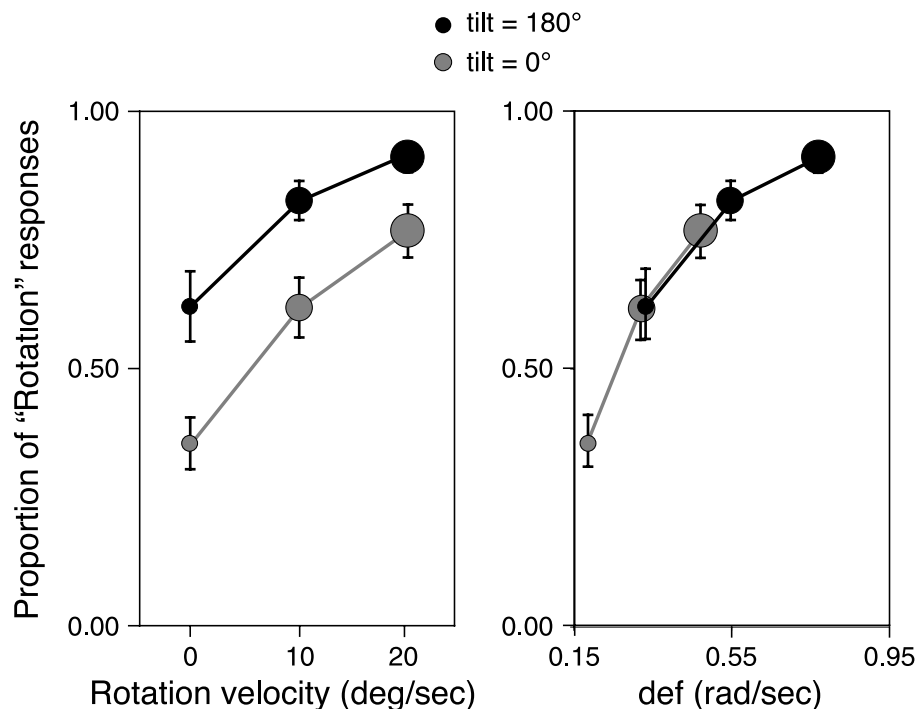

Figure 7. Rotation detection performance in Experiment 1b. Panels on the left show mean proportions of "Rotation" responses as a function of rotation velocity for the two levels of tilt (coded by color). The same data are replotted as a function of average *def* in the panels on the right. The size of the circles codes for the rotation velocity (small for static, medium for 10 deg/s, and large for 20 deg/s). Error bars represent ±1 *SE*.

the possibility of different response biases in the two cases.

Methods and procedure

Methods and procedure were the same as in Experiment 1a, except that viewing was binocular, with the same (monocular) images presented to both eyes (null disparity field).

Results and discussion

The left panel of Figure 7 shows the average proportion of valid "Rotation" responses as a function of rotation velocity and tilt angle. The same criterion as in Experiment 1a was used for identifying the "Valid" responses.

In Experiment 1b, the results were similar to those in Experiment 1a. Specifically, if *def* is not taken into consideration, d' was significant ($d' = 0.849$, $z = 6.768$, $p < 0.001$). When *def* was controlled, d' become statistically equal to zero ($d' = 0.102$, $z = 0.700$, $p = 0.484$). Likewise, in this case, the likelihood of a "Rotation" response was only a function of *def* ($\beta_{def} = 3.172$, $z = 10.103$, $p < 0.001$; Figure 7, right panel).

Comparison of Experiments 1a and 1b

Experiment 1b was motivated by the following hypothesis. Suppose that the visual system must choose between two different 3D interpretations that can be provided to the stimulus displays. This choice can be done by vetoing either a retinal or an extra-retinal signal. If information from vergence and accommodation is vetoed, then the number of "Rotation" responses must be larger in Experiment 1b than in Experiment 1a.

To test this hypothesis, in a first *lme* analysis, we considered only the trials in which the simulated surface was stationary. For these trials, the proportion of "Rotation" responses was significantly larger in Experiment 1b than in Experiment 1a (0.49 vs. 0.21; $z = 8.644$, $p < 0.001$). In a second analysis, we considered only the trials simulating a surface rotation. Likewise, in this case, in Experiment 1b, we found a larger proportion of "Rotation" responses than in Experiment 1a (0.78 vs. 0.59; $z = 5.016$, $p < 0.001$).

In conclusion, the present results are consistent with the hypothesis that the 3D interpretation provided to the stimuli of Experiment 1b was dependent only on the (retinal) information provided by the velocity gradients and the null disparity field. The (extra-retinal) information provided by vergence and accommodation appeared to be disregarded. As a cautionary note, we must add that the stimulus conflict of Experiment 1b can be resolved according to a cue combination rule different than the veto, as indicated in the General discussion section.

Experiments 2a and 2b

The results of Experiments 1a and 1b can be attributed to the manipulation of tilt or to the manipulation of *def*. In Experiments 1a and 1b, in fact, the two variables covaried. The purpose of Experiments 2a and 2b was to disentangle the effects of these two variables. This was done by simulating two surfaces slanted around the horizontal (rather than vertical) axis with a gradient of velocity in a direction orthogonal to the direction of lateral head motion. These two surfaces differed for their tilt directions (as in Experiment 1) but generated similar *def* components. Experiments 2a and 2b thus replicated the design of Experiments 1a and 1b, with the difference that tilt did not covaried with *def*. In these circumstances, we hypothesized that the tilt variation would not affect the perceptual discrimination of the presence or absence of surface rotation. In Experiment 2a, viewing was monocular; in Experiment 2b, viewing was binocular, with the same (monocular) optic flow shown to both eyes (null disparity field).

Methods

Participants

Nine undergraduates at the University of Parma participated to the experiment. All had normal or corrected-to-normal vision and were naive to the purpose of the experiment.

Apparatus, displays, and procedure

The only difference between the stimuli of Experiments 1 and 2 concerns the orientation of simulated 3D surfaces. In Experiments 2a and 2b, the displays were generated by rotating the planar surfaces employed in Experiments 1a and 1b by 90 deg around the z-axis. With such surface orientation, the variation of the horizontal shear induced by the lateral head shift is independent of tilt. Instantaneous *def* was then the same for both tilt conditions (90° and 270°) in each moment of the motion sequence. Across the three Rotation Velocities that had been simulated (0, 10, 20 deg/s), average *def* was equal to 0.26, 0.44, and 0.62 rad/s, respectively. The apparatus, display, and procedure were otherwise identical to those of Experiment 1.

Results

The proportions of "Rotation" responses as a function of angular Rotation Velocity (0, 10, 20 deg/s) and Tilt angle (90° and 270°) are shown in Figure 8 for both Experiments 2a (top panel, monocular viewing) and 2b (bottom panel, binocular viewing). In both experiments,
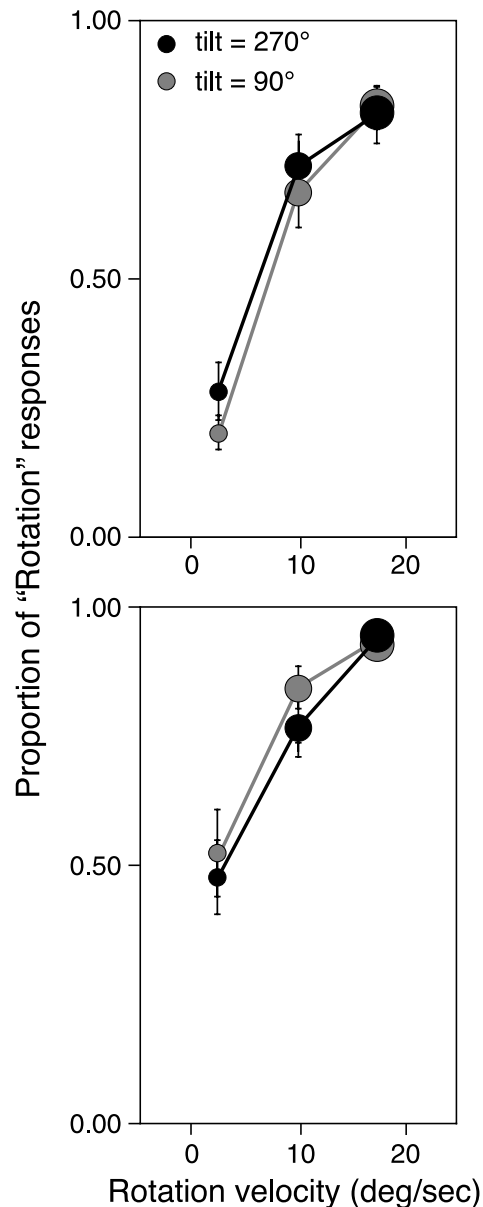
Figure 8. Rotation detection performance in Experiment 2 where the plane was slanted around the horizontal axis rather than around the vertical. Mean proportion of rotation responses across rotation velocity are shown for the two levels of tilt angle (coded by color) in (top) Experiment 2a and (bottom) Experiment 2b. Error bars represent ±1 *SE*.

def covaried perfectly with Rotation Velocity, so the relative effects of these two variables cannot be distinguished.

Experiment 2a

An *lme* model with response as the dependent variable (stationary vs. rotating surface), participants as random factor, and angular Rotation Velocity (0, 10°, 20°/s) and Tilt (90° and 270°) as fixed effects revealed significant main effects for both variables (Rotation Velocity: $z =$

5.153, $p < 0.001$; Tilt: $z = 3.089$, $p < 0.005$) and a not significant interaction ($\chi^2_1 = 1.725$, n.s.).

As shown in Figure 8 (top panel), the likelihood of a "Rotation" response increased with Rotation Velocity (remember that rotation covaries perfectly with *def*) and was larger for surfaces with a 270° tilt.

Even though Rotation Velocity and Tilt had a significant effect on the likelihood of a "Rotation" response, the effect size was very different in the two cases. When Rotation Velocity varied in the interval between 0 deg/s and 20 deg/s (and Tilt was kept constant), the predicted probabilities of a "Rotation" response increased (approximately) between 0.2 and 0.9. On the other hand, as Tilt varied in the interval between 90° and 270° (and Rotation Velocity was kept constant), the predicted probabilities of a "Rotation" response increased from 0.62 to 0.72.

Experiment 2b

An *lme* model with response (stationary vs. rotating surface) as the dependent variable, participants as random factor, and Rotation Velocity (0, 10, 20 deg/s) and Tilt (90° and 270°) as fixed effects revealed significant main effects for Rotation Velocity ($z = 12.330$, $p < 0.001$). Neither the effect of Tilt ($z = 0.953$, n.s.) nor the interaction between Tilt and Rotation Velocity ($\chi^2_1 = 0.911$, n.s.) was significant.

Comparison of Experiments 2a and 2b

An *lme* analysis with response (stationary vs. rotating surface) as the dependent variable, participants as random factor, and angular Rotation Velocity (0, 10, 20 deg/s), Tilt angle (90° and 270°), and Experiment (Experiment 2a/ monocular versus Experiment 2b/binocular) as fixed effects showed that the likelihood of a "Rotation" response was larger for Experiment 2b (Figure 8, bottom panel) than for Experiment 2a (Figure 8, top panel). This result replicates what we found in Experiment 1 ($z = 10.016$, $p < 0.001$).

It is also instructive to compare the overall effect size of the variable Tilt in Experiments 1a and 1b, on the one hand, and Experiments 2a and 2b, on the other. In Experiments 1a and 1b, the odds of a "Rotation" response increased by 139% as tilt changed between 0° and 180°. In Experiments 2a and 2b, the odds of a "Rotation" response increased only by 17% as tilt changed between 90° and 270°. In Experiments 2a and 2b, therefore, the effect of Tilt was extremely small when compared to that found in Experiments 1a and 1b.

Discussion

In conclusion, the results of Experiments 2a and 2b show that, when the tilt did not covary with *def*, the effect of tilt on the perceptual discrimination between stationary

and rotating surfaces disappears or was greatly reduced. The research on passive SfM has shown that *def* is not the only determinant of perceived angular rotation. Domini and Caudek (1999), for example, found that surface tilt accounts for a small component of perceived surface rotation (see also Todd & Bressan, 1990; Todd & Perotti, 1999). In Experiment 2a, we replicated this finding in active SfM: perceived surface rotation was indeed affected by surface tilt, even though this effect was very small if compared to the effect of *def*. For an interpretation of tilt effects in the spatial domain, see Fantoni (2008).

A final consideration concerns the effect of the null disparity field. Likewise, in Experiment 2b, we replicated what found in Experiment 1b: if a null disparity field was added to the optic flow generated by the active observer, the likelihood of a "Rotation" response increased.

Experiment 3

The purpose of Experiment 3 was to directly compare the perceptual interpretation provided to the same optic flow field by an active and a passive observer. We also studied whether the addition of a null disparity field had the same perceptual effects in active and passive visions.

Experiment 3 comprised two viewing conditions. In the Active viewing (Act) condition, observers were shown the optic flow generated (online) by their own movement with respect to a *stationary* planar slanted surface. As detailed in Appendix B, the optic flow produced by the active observer was recorded, and subsequently, it was shown to a stationary observer (passive-viewing condition). In both conditions, observers were asked to classify the apparent rotation of the simulated (stationary) surface as being "small" or "large".

The experimental design of Experiment 3 was similar to Experiments 1a and 1b. The manipulation of tilt covaried with *def*, but the simulated surface was always stationary. The stimulus displays were viewed either monocularly or binocularly.

In both active and passive viewing conditions, we expected that: (1) the likelihood of a "Large Rotation" response would be a function of *def*, even if the simulated surface was always stationary; (2) the addition of a null disparity field would increase the likelihood of a "Large Rotation" response (see Rogers & Collett, 1989).

Methods

Participants

Fourteen undergraduates at the University of Parma with normal or corrected-to-normal vision participated in the experiment. All participants were naive to the purpose of the experiment.

Apparatus and displays

The apparatus and the general properties of the stimulus displays were the same as in Experiments 1a and 1b. The displays generated by the active observer were identical to the displays simulating the static planar surfaces in Experiment 1. The stimulus displays for the passive-viewing conditions were generated as indicated in Appendix B. To isolate the effects of the *def* component of the optic flow for the passive observer, we created two different kinds of displays (see Braunstein & Tittle, 1988; Naji & Freeman, 2004; Rogers & Collett, 1989):

1. both Translational and Rotational (TR) components of the optic flow generated during active vision trials were provided to the passive observer;
2. only the Rotational (Rot) component of the optic flow generated during active vision trials was provided to the passive observer (not the horizontal translational component).

Design

A $2 \times 2 \times 3$ within-subjects design was used, with two Viewing Modes (monocular and binocular), two Tilt angles (180° and 0°), and three Viewing Conditions (Act, TR, Rot).

Procedure

Each observer participated in one active-vision block (Act) and two passive-vision blocks (TR and Rot). The experiment started with the Act block; the ordering of the passive blocks was counterbalanced across the subjects. The procedure of the Act block was the same as in Experiment 1, with the exception that the display was visible on the screen for two and half cycles of head translation (Figure 9). In the passive blocks, participants were instructed to seat in front of the CRT screen with their head on a chin rest. The task was to classify perceived surface rotation as being "Small" or "Large".

Each participant was presented with three blocks of 80 trials. Each block resulted from 20 repetitions of our 4 experimental conditions: 2 Viewing Modes (monocular vs. binocular) × 2 Tilt angles (0° and 180°). The ordering of trials in the TR and Rot blocks differed from the ordering of the trials in the Act block. Each experimental session lasted for about 30 min. A training session (20 trials) preceded each of the three experimental blocks.

Results and discussion

Figure 10 shows the proportions of "Large Rotation" responses in each experimental condition. Let us consider first the Act condition (Figure 10, left panel): the results
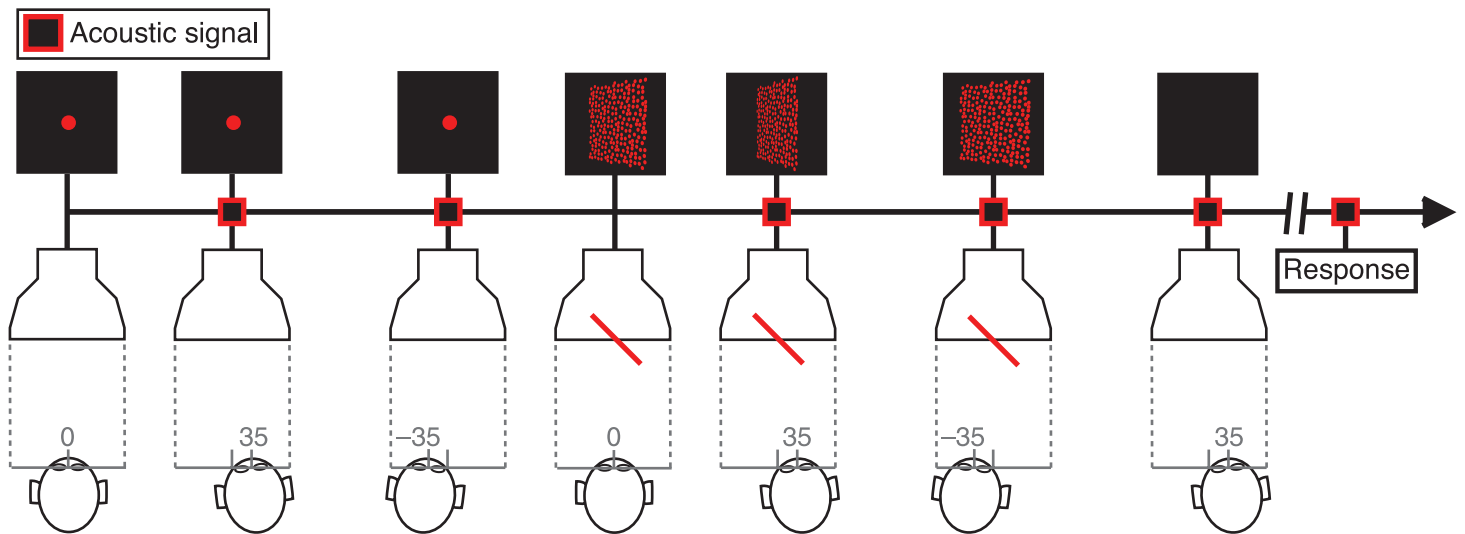
Figure 9. Temporal sequence used in Experiment 3. The illustration refers to an Act 180° tilted surface.

replicated those of Experiment 1. Likewise, in Experiment 3, the manipulation of Tilt covaried with *def* (with *def* equal to 0.19 and 0.34 for 0° tilt and 180° tilt, respectively). Correspondingly, a larger proportion, of "Large Rotation" responses, was associated with the larger *def* magnitude.

For the binocular condition, observers reported "Large Rotation" responses more often for 180° tilted surfaces than for 0° tilted ones ($z = 3.350$, $p < 0.001$). The same effect, but stronger, was found also for the monocular

condition, as revealed by the significant interaction between Viewing Mode and Tilt ($z = 4.282$, $p < 0.001$). From Figure 10, we also see that the likelihood of a "Large Rotation" response was higher in the binocular than in the monocular condition ($z = 8.165$, $p < 0.001$). This replicates the results found in Experiments 1 and 2.

Now, let us consider the TR and Rot conditions (Figure 10, mid and right panels, respectively). The same pattern of results was present for both the TR and Rot conditions, as
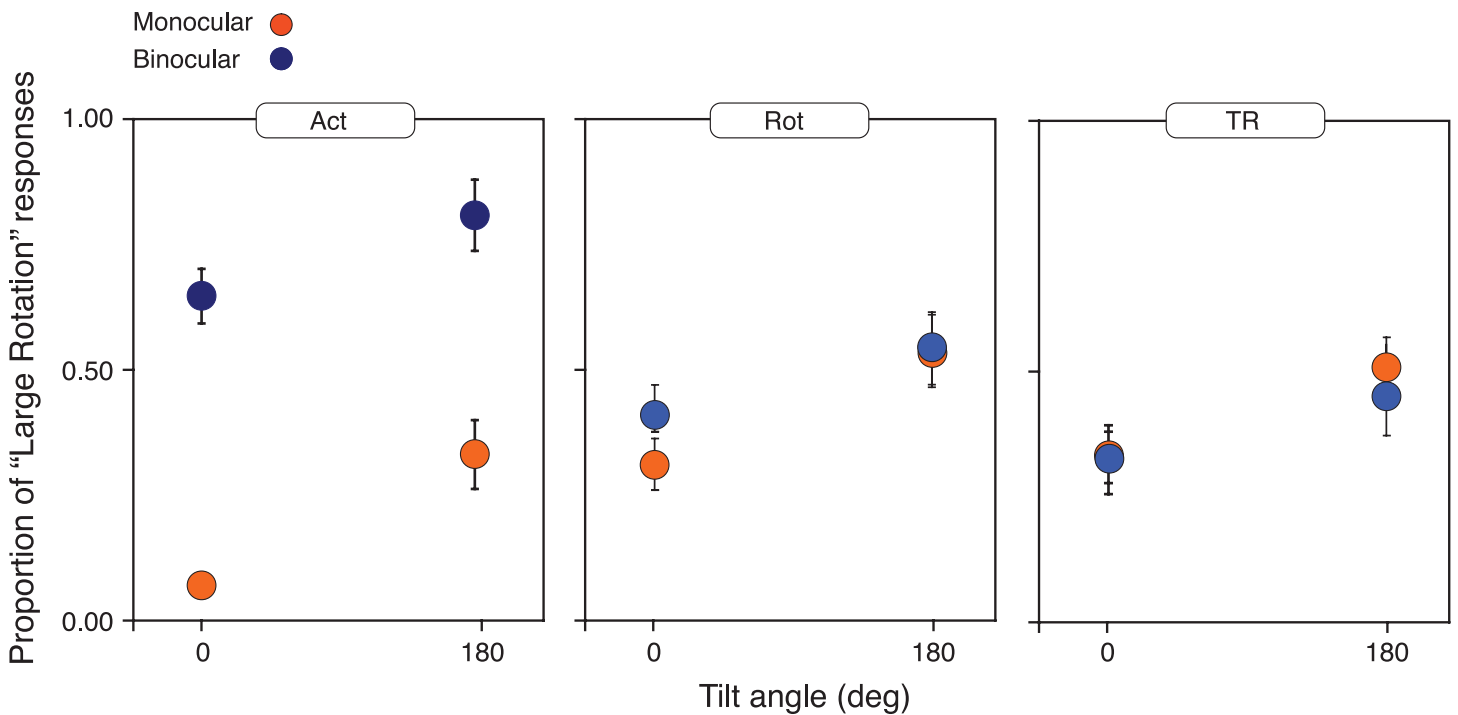
Figure 10. Classification performance in Experiment 3. Mean proportions of "Large Rotation" responses are shown as a function of Viewing condition (Act, left; Rot, middle; TR, right), Tilt angle, and Mode of Viewing (Monocular, red; Binocular, blue). Error bars represent ±1 *SE*.

revealed by the absence of any significant interaction between the variables Passive Optic Flow (TR vs. REL), Tilt (0°, 180°), and Viewing Mode (monocular vs. binocular). The effect of the variable Tilt was significant in the monocular TR condition ($z = 3.427$, $p < 0.001$): a higher likelihood of a "Large Rotation" response was associated with the larger *def* (180° tilt). Figure 10 indicates that a similar effect was also found in the monocular Rot condition. In both TR and Rot conditions, the effect of *def* was stronger in the monocular condition, as revealed by the significant interaction between Passive Optic Flow and Tilt ($z = 2.195$, $p < 0.05$).

When the optic flow was generated by the active observer, the planar surface appeared to undergo a larger amount of rotation for binocular (with null disparity field) than for monocular viewing. This result replicates the findings of Rogers and Collett (1989). However, we found no difference between binocular and monocular viewing for the passive observer, thus suggesting that our passive displays were "less effective" than those used by Rogers and Collett (1989). In fact, they were much smaller in size, they were presented for a very short time, and they had a much smaller translational component. Nevertheless, the same (conflicting) visual information was interpreted differently, depending on the presence of vestibular information consistent with the optic flow.

In active SfM, all available visual information were taken into account, thus increasing the likelihood of perceiving surface rotation; in passive SfM, conversely, the presence of a null disparity field did not change the perceptual interpretation of the optic flow. This difference suggests that, within our stimulus setting, inconsistent visual cues had not been integrated together in the absence of vestibular information consistent with the optic flow.

In conclusion, the results of Experiment 3 suggest that *def* had similar effects on the perception of surface rotation in both active and passive SfMs. This is apparent from Figure 10 (monocular) when we compare the proportions of "Large Rotation" responses for 0° and 180° tilt, across the Act, TR, and Rot conditions. The addition of a null disparity field affected the response in the active but not in the passive viewing condition.

General discussion

Several lines of evidence indicate an advantage of active over passive vision for the 3D perceptual interpretation of the optic flow (Colas et al., 2007; Dijkstra et al., 1995; Jaekl et al., 2005; Ono & Steinbach, 1990; Panerai et al., 2002; Peh et al., 2002; Rogers & Rogers, 1992; but see also Rogers & Graham, 1979; van Damme & van de Grind, 1996; Wallach & O'Connell, 1953; Wallach et al., 1974). In the present investigation, we studied the contribution of vestibular information (and

other extra-retinal signals) on the discrimination of the presence or absence of surface rotation. We also studied discrimination performance when retinal and extra-retinal cues provided conflicting information about viewing distance.

In the Introduction section, we demonstrated that correct discrimination of the presence or absence of surface rotation is possible, in principle, if observers use both retinal and extra-retinal signals, and if they analyze the change of the deformation component of the optic flow over time. The first result of the present research shows that perceived surface rotation was affected by systematic biases, even when sufficient information was provided in active SfM.

In Experiments 1a and 1b, discrimination performance was strongly affected by surface tilt. In Experiments 2a and 2b, conversely, the manipulation of tilt had no effect. Importantly, tilt covaried with *def* in Experiments 1a and 1b, but not in Experiment 2. Likewise, in active vision, therefore, perceived surface rotation seems to depend on the analysis of the first-order optic flow (e.g., Domini & Caudek, 2003a, 2003b). In Experiment 3, we replayed to the passive observers the optic flow previously generated by the active observers. We found a similar response pattern in both active and passive SfMs.

A second result of the present investigation concerns the stimulus conflict of Experiments 1b, 2b, and 3. By hypothesizing that large cue conflicts are resolved by a veto mechanism (a process akin to outlier rejection in statistic; see Landy, Maloney, Johnston, & Young, 1995), we asked how this veto mechanism may operate. Specifically, we asked whether the visual system would veto a retinal or an extra-retinal signal. Different consequences ensue from this choice: with the veto of the null disparity field, veridical discrimination is still possible; with the veto of vergence and accommodation, veridical discrimination is not possible anymore, if the information provided by the null disparity field is taken into account. Our results indicate that observers reliably choose the second interpretation (i.e., the veto of extra-retinal signals), requiring a nonveridical surface rotation.

Other cue combination rules for the stimuli of Experiments 1b, 2b, and 3 are possible, besides the veto mechanism discussed above. Cue combination can be obtained, for example, through a weighted average rather than veto, as indicated by the linear model for cue integration (Landy et al., 1995). According to linear cue combination, different depth-processing modules provide independent estimates of 3D information. In Experiment 1b, disparity, vergence, and accommodation all specified zero depth and no 3D rotation; motion parallax specified nonzero depth and 3D rotation. These cues, therefore, provided largely discrepant estimates of the amount of 3D rotation. In such circumstances, linear cue combination is not appropriate (Oruç, Maloney, & Landy, 2003). However, nonlinear, robust cue integration

models have also been proposed, which could explain the present results (e.g., Knill, 2007).

Our findings seem to be at odds with previous results indicating the importance of extra-retinal signals for perceived SfM. In a first group of studies, ambiguous stimuli were generated by creating a conflict between pictorial (perspective) and motion information (e.g., Wexler, 2003; Wexler, Lamouret et al., 2001; Wexler, Panerai et al., 2001). The perceptual interpretation of such displays clearly showed that observers used extra-retinal signals to resolve stimulus ambiguity. In a second group of studies, errors in tilt judgments were measured (e.g., Cornilleau-Pérès et al., 2002; Dijkstra et al., 1995; van Boxtel et al., 2003), showing that tilt judgments are more precise in active than in passive vision. In a third group of studies, errors in both tilt and slant judgments were investigated (e.g., van Boxtel et al., 2003), revealing that the correlation between simulated and perceived slant is stronger in active vision than in passive vision.

These studies show that vestibular information plays an important role in the perceptual interpretation of the optic flow. Our experiments suggest, instead, that the discrimination of the presence or absence of surface rotation in active SfM is not necessarily veridical, even if the stimulus information is sufficient for veridical performance. We propose that this apparent contradiction may be explained by the different roles that extra-retinal signals play in different perceptual tasks.

Ono, Rivest, and Ono (1986) proposed that extra-retinal information is used in active vision to calibrate motion parallax to absolute-distance information (see also Panerai et al., 2002; Peh et al., 2002). Instead, Cornilleau-Pérès and Droulez (1994) proposed that nonvisual information about self-motion is used mainly as a retinal stabilization factor and that it does not directly improve the processing of depth from motion (see also Oosterhoff, van Damme, & van de Grind, 1993). Our data are consistent with the latter; there is no evidence that nonvisual information is used for veridical 3D shape recovery, even though it may be used for a better measurement of the optic flow (Domini & Caudek, 2010a, 2010b).

A better measurement of the optic flow may help to disambiguate a cue-conflict stimulus (Wexler, 2003; Wexler, Panerai et al., 2001) and to better estimate the tilt (Cornilleau-Pérès et al., 2002; Dijkstra et al., 1995; van Boxtel et al., 2003) and the slant of a planar surface up to a scaling factor (van Boxtel et al., 2003). These tasks do not require the knowledge of Euclidean depth. On the other hand, the discrimination of the presence or absence of surface rotation does require the knowledge of Euclidean 3D properties. An affine analysis alone, in fact, cannot distinguish between the optic flows generated by (i) a surface rotating about a vertical axis while the observer undergoes a horizontal translation, or (ii) a stationary planar surface and a horizontal translation by the observer.

If knowledge of 3D Euclidean properties is not available, how can surface rotation be estimated? In our previous research on passive SfM, we have hypothesized that perceived surface rotation (ω) is a function of *def*, regardless of the actual 3D surface rotation (Domini & Caudek, 1999, 2003a, 2003b). As indicated in Figure 1, for both the active and passive observers, *def* is ambiguous, in the sense that the same *def* can be produced by different slant (σ) and angular rotations (ω) values. For the recovery of surface rotation, Domini and Caudek (2003b) proposed that the visual system chooses, among these infinite σ and ω pairs, the one that maximizes the likelihood function $p(def|\sigma)$:

$$\hat{\omega} = \arg \max_{\omega} p(def \mid \sigma), \qquad (2)$$

where

$$p(def \mid \sigma) \propto \int_{\omega} p(def \mid \sigma, \omega)p(\omega)d\omega. \qquad (3)$$

Note that $p(def|\omega)$ has a maximum: The value ω_i corresponding to the maximum of the marginal distribution $p(def|\omega)$ is the maximum likelihood estimate $\hat{\omega}$. In a series of papers (Caudek & Domini, 1998; Caudek & Rubin, 2001; Di Luca, Domini, & Caudek, 2004; Domini & Caudek, 1999, 2003a, 2003b; Domini et al., 1998), we provided empirical evidence in support to this hypothesis. The present data suggest that, for the present task and stimulus conditions, a similar analysis can be applied to active SfM.

Conclusions

Active SfM provides advantages over passive SfM: the optic flow created by the active movements of the observer generates a vivid percept that is as compelling as that provided by binocular disparity information (Rogers & Graham, 1982). The perceptual interpretation is more reliable in active than in passive SfM, with less variability within and between observers. This does not mean, however, that active SfM is always veridical. In the present investigation, we have shown that the discrimination of the presence or absence of surface rotation in active SfM is strongly affected by the orientation (i.e., tilt) of the distal surface and that these biases can be accounted for by the deformation component of the first-order optic flow. We have also shown that, when presented with conflicting information generated by a null disparity field, the visual system chooses a perceptual interpretation that favors perceived surface rotation, even when the distal

surface is stationary. Overall, the present results suggest that the deformation component of the first-order optic flow elicits similar biases in both active and passive SfMs.

Appendix A

Deformation of a translating surface

The optic flow produced by the translation of a point of view with respect to a slanted surface (Figure A1) is equivalent to the optic flow produced by the translation of a surface with respect to a static point of view (Figure A2). Here, we will consider only the simpler case of a planar surface with a null vertical gradient.

The temporal derivative of the projection of a point $P(x, y, z)$ at a distance z_f from the image plane

$$x_P = \frac{x z_f}{z + z_f},$$ (A1)

is given by

$$\dot{x}_P = \frac{\dot{x} z_f}{z + z_f} - x_P \frac{\dot{z}}{z + z_f}.$$ (A2)

Let us consider a planar surface defined by $z = (x - x_0)g_x$, where x_0 is the horizontal coordinate of the point of the
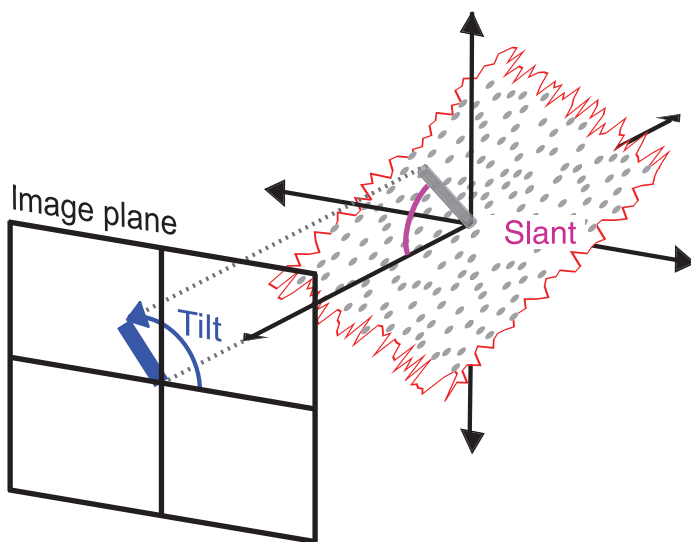
Figure A1. The orientation of a planar surface patch is specified by the slant and the tilt angles. The slant is the angle between the surface normal and the line of sight. The tilt is the projected orientation of the surface normal in the image plane relative to the horizontal.
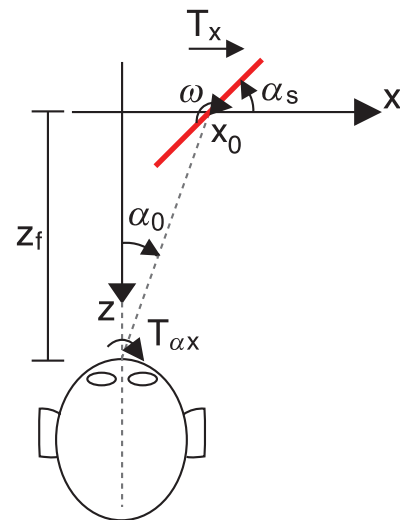

Figure A2. Viewing parameters and reference frame used to characterize the local gradient of the optic flow generated by a surface that is: (1) slanted at an angle α_s about the vertical axis, (2) translating to the right of the observer at T_x image velocity (or $T_{\alpha x}$ in terms of angular velocity), and (3) rotating at an angular velocity ω. The reference frame has: (1) the origin in the center of the surface before it begins translating (at a distance z_f from the point of view), and (2) a z-axis that is centered and aligned with the cyclopean line of sight. The horizontal coordinate of the point where the surface intersects the x-axis is x_0, and α_0 is the angle between the cyclopean line of sight and the horizontal visual direction through x_0.

surface that intersects the image plane and g_x is the horizontal depth gradient (slant). Suppose that the planar surface translates horizontally with speed T_x and rotates with angular velocity ω about a vertical axis passing through the point $(x_0, 0, 0)$. In this case, $\dot{x} = -(x - x_0)g_x\omega + T_x$ and $\dot{z} = (x - x_0)\omega$. If we (1) solve Equation A1 for x after substituting z for the equation of the plane, (2) substitute x, now function of x_P, in the equations for $\dot{x}$, $\dot{z}$, and z, and (3) substitute $\dot{x}$, $\dot{z}$, and z in Equation A2, then we obtain the equation for the image plane velocity field $\dot{x}_P(x_P)$:

$$\dot{x}_P = \frac{T_x(z_f - g_x x_P)}{(z_f - g_x x_0)} - \frac{\omega}{(z_f - g_x x_0)}\left(g_x z_f x_P + x_P^2 - g_x z_f x_0 - x_P x_0\right).$$ (A3)

The gradient of the image optic flow (*def*) calculated at x_0 can be obtained by deriving the previous equation (Equation A3) with respect to x_P:

$$def = \frac{d\dot{x}_P}{dx_P} = \frac{-g_X T_x}{(z_f - g_x x_0)} - \frac{\omega}{(z_f - g_x x_0)}\left(g_x z_f + x_0\right).$$ (A4)
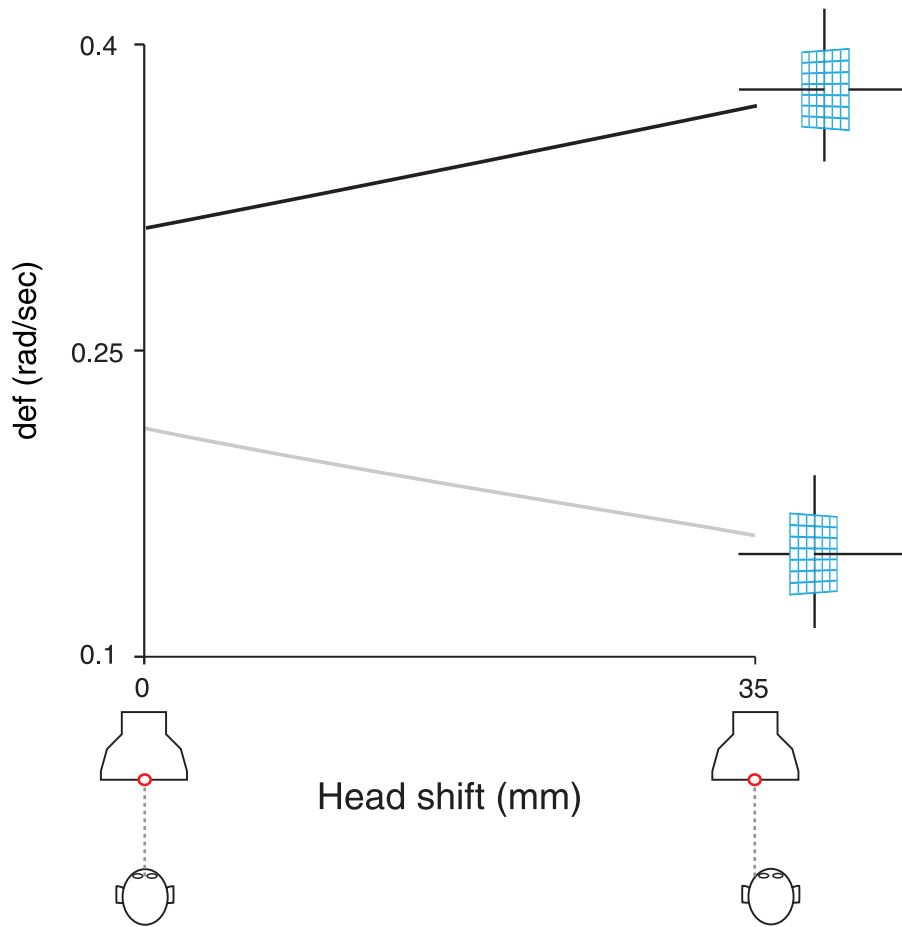
Figure A3. *Def* (*y*-axis) as a function of the amount of lateral head shift from the screen center (*x*-axis) for a static surface that is 45° slanted about the vertical axis with either 0° tilt (gray) or 180° tilt (black; T_x = 125 mm/s, z_f = 480, and x_0 varied from 30 to 65 mm). Depending on the slant direction, the *def* is either a monotonically increasing or a monotonically decreasing function of lateral head shift.

Note that equation is the sum of two terms. The first term is the velocity gradient produced by the translation of the surface, and the second term is the velocity gradient produced by the rotation of the surface.

Equation A4 can also be expressed in terms of visual angle. If we denote with α the horizontal visual direction of a generic point P belonging to the planar surface, then *def* is defined as $\frac{d\dot{\alpha}}{d\alpha}$ calculated at the visual direction $\alpha_0 = \tan^{-1}\left(\frac{x_0}{z_f}\right)$ (Figure A2). This result is found by substituting $\tan(\alpha) = \frac{x_P}{z_f}$, $\tan(\alpha_0) = \frac{x_0}{z_f}$, $T_{\alpha x} = \frac{T_x}{z_f}$ and by observing that $\dot{\alpha} = \frac{1}{1+\tan(\alpha)^2}\frac{\dot{x}_P}{z_f}$. The expression for $\dot{\alpha}$ can be simplified since the values of α are very small for the range of movements relevant to the present study, $\tan(\alpha) \approx \alpha$ and $\tan(\alpha)^2 \ll \alpha$. By this approximation, *def* becomes

$$def = \frac{d\dot{\alpha}}{d\alpha} = (T_{\alpha x} + \omega)\tan(\alpha_s + \alpha_0). \tag{A5}$$

Note that instantaneous *def* varies if the observer translates and the surface is static ($\omega = 0$). In fact,

α_0 increases or decreases with the rightward horizontal position of the viewing point, depending on the sign of α_s (defining the tilt of the surface). Therefore, as shown in Figure A3, the absolute value of the second term of Equation A5 increases if α_s = +45° (i.e., tilt = 180°) and decreases if α_s = −45° (i.e., tilt = 0°). Consequently, the average *def* is larger for α_s = +45° than for α_s = −45°. That is, *def* covaries with the tilt angle of a vertically slanted surface, if the observer undergoes a lateral head translation.

Appendix B

Passive displays

The displays for the passive-viewing condition of Experiment 3 were generated by projecting the points of the actively viewed plane in a new coordinate system

defined as follows: z' corresponds to the cyclopean line of sight in the corresponding actively viewed display; x' is parallel to the inter-ocular axis in the corresponding actively viewed display; y' passes through the intersection of the x'- and z'-axes and is orthogonal to both of them (Figure B1). The center of the new coordinate system was fixed at the distance (D) of 480 mm from the cyclopean
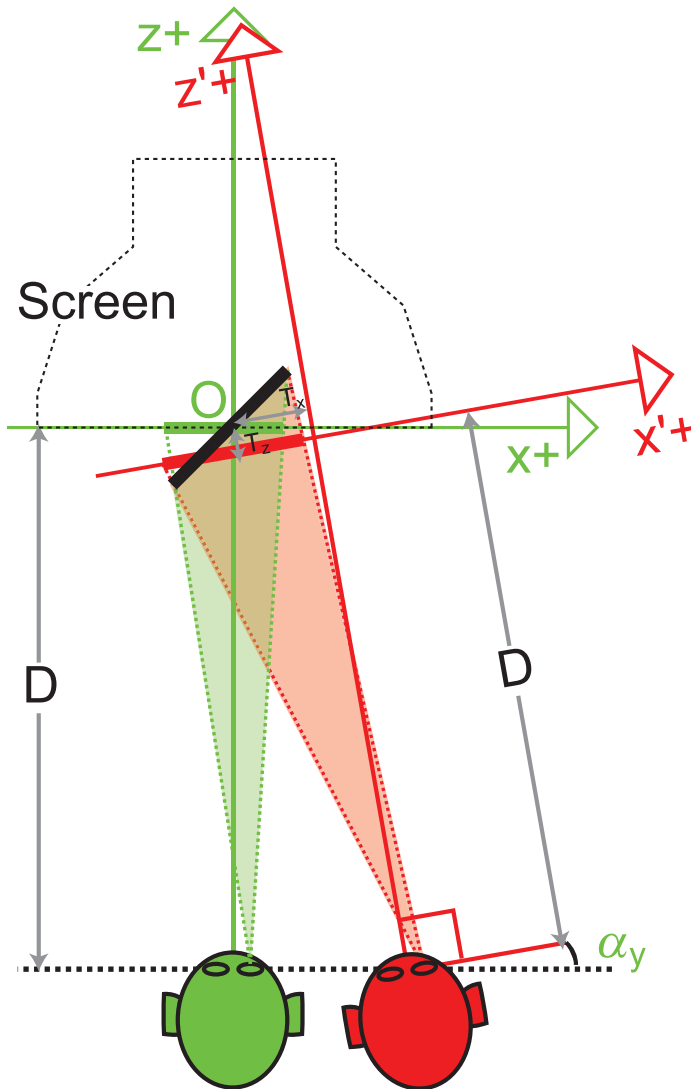
eye and the simulated planar surface was projected onto the x'–y' plane.

The display presented to the passive observer consisted of the transformation of a display generated by the active observer into the frame of reference described above according to the linear equation $P' = P \cdot R_{yz} + T_{xyz}$, where R_{yz} and T_{xyz} identify the rotational and translational components of the following Transformation Matrix:

$$x' \rightarrow a_{11}x + a_{12}y + a_{13}z + T_x$$

$$y' \rightarrow a_{21}x + a_{22}y + a_{23}z + T_y \qquad \text{(B1)}$$

$$z' \rightarrow a_{31}x + a_{32}y + a_{33}z + T_z.$$

For each sequence of frames generated by the active observer, the entries of the Rotation Matrix and the Translation Vector were extracted and stored in a text file as a function of the observers' eye position and head orientation. The Rotation Matrix was obtained by multiplying the two rotational components R_z (about the z-axis) and R_y (about the y-axis). Consistent with Listing's Law, R_x was neglected (the vertical extension of the eyes is null and any rotation around inter-ocular axis of the head leaves the image unchanged). The $R_y R_z$ multiplication resulted in the following Rotation Matrix:

$$R_{yz} = \begin{bmatrix} \cos\alpha_z \cos\alpha_y & \sin\alpha_z \cos\alpha_y & \sin\alpha_y \\ -\sin\alpha_z & \cos\alpha_z & 0 \\ -\cos\alpha_z \sin\alpha_y & \sin\alpha_z \sin\alpha_y & \cos\alpha_y \end{bmatrix}, \quad \text{(B2)}$$

where α_y is the rotation angle of the inter-ocular axis around the y-axis in the actively viewed display and α_z is the rotation angle of the inter-ocular axis around the z-axis. The two rotation angles were calculated according to the left (x_{el}, y_{el}, z_{el}) and right eye (x_{er}, y_{er}, z_{er}) positions during active vision.

Two active-viewing behaviors were defined, which involved different α_y, but not $\alpha_z = \arctan\left(\frac{y_{er} - y_{el}}{\sqrt{(x_{er} - x_{el})^2 + (z_{er} - z_{el})^2}}\right)$, thus defining the two passive-viewing conditions of Experiment 3:

1. TR, in which the observers' fixation was assumed to be straight ahead, regardless of object position and $\alpha_y = \arctan\left(\frac{z_{er} - z_{el}}{x_{er} - x_{el}}\right)$;

2. Rot, in which the observers' fixation was assumed to be centered on the planar surface regardless of actual head position and $\alpha_y = \arctan\left(\frac{(x_{er} + x_{el})/2}{(z_{er} + z_{el})/2}\right)$.

Figure B1. Coordinate systems utilized to generate the same optic flow in the passive condition (TR) as in the active condition (Act). In the passive condition, the coordinates of the actively viewed planar surface (black bold line) are recoded in a rotated and translated coordinate system (red) defined by x', y', and z' and projected onto the x'–y' projection plane assuming the right eye of the observer at rest as the center of projection (lying at a distance D from the center of the coordinate system O'). In the case of an active observer rotating the head counterclockwise by angle α_y while translating horizontally to the right, the screen image of the surface translates in depth of T_z and along the horizontal of T_x (gray arrows).

The Translation Vector resulted in the following three entry vectors:

$$T_{xyz} = \begin{bmatrix} \dfrac{x_{er}+x_{el}}{2}\cos(\alpha_y) + \dfrac{z_{er}+z_{el}}{2}\sin(\alpha_y) \\[2ex] \dfrac{y_r+y_l}{2} \\[2ex] -\dfrac{x_{er}+x_{el}}{2}\sin(\alpha_y) + \dfrac{z_{er}+z_{el}}{2}\cos(\alpha_y) + 480 \end{bmatrix}.$$

(B3)

Acknowledgments

Footnotes

[1] For a passive observer, $T_{\alpha x}$ is null and α_0 is specified by the rotation of the eye; for an active observer, $T_{\alpha x}$ and α_0 are specified by vestibular information and by vestibular-ocular reflex.

[2] We are not claiming here that a null disparity field associated with a velocity field actually elicits the perception of a far surface viewed at large viewing distance from the observer. Instead, we claim that extra-retinal signals may not be used for estimating the object's 3D structure and motion (Cornilleau-Pérès & Droulez, 1994), but only for estimating the egocentric distance: the retinal cues (null disparity and velocity gradient) specify a rotating surface, whereas extra-retinal cues specify a small egocentric distance.

References

Bennur, S., & Gold, J. I. (2008). Right way neurons. *Nature Neuroscience, 11,* 1121–1122. [PubMed]

Braunstein, M. L., & Tittle, J. S. (1988). The observer-relative velocity field as the basis for effective motion parallax. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 582–590. [PubMed]

Buizza, A., Leger, A, Droulez, J., Bertoz, A., & Schmid, R. (1980). Influence of otolithic stimulation by horizontal linear head acceleration in optokinetic nystagmus and visual motion perception. *Experimental Brain Research, 71,* 406–410.

Caudek, C., & Domini, F. (1998). Perceived orientation of axis of rotation in structure-from-motion. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 609–621. [PubMed]

Caudek, C., Domini, F., & Di Luca, M. (2002). Short-term temporal recruitment in structure from motion. *Vision Research, 10,* 1213–1233. [PubMed]

Caudek, C., & Proffitt, D.R. (1993). Depth perception in motion parallax and stereokinesis. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 32–47. [PubMed]

Caudek, C., & Rubin, N. (2001). Segmentation in structure from motion: Modelling and psychophysics. *Vision Research, 41,* 2715–2732. [PubMed]

Colas, F., Droulez, J., Wexler, M., & Bessière, P. (2007) A unified probabilistic model of the perception of three-dimensional structure from optic flow. *Biological Cybernetics, 97,* 461–477. [PubMed]

Cornilleau-Pérès, V., & Droulez, J. (1994). The visual perception of three-dimensional shape from self-motion and object motion. *Vision Research, 34,* 2331–2336. [PubMed]

Cornilleau-Pérès, V., Wexler, M., Droulez, J., Marin, E., Miege, C., & Bourdoncle, B. (2002). Visual perception of planar orientation: Dominance of static depth cues over motion cues. *Vision Research, 42,* 1403–1412. [PubMed]

Dijkstra, T. M., Cornilleau-Pérès, V., Gielen, C. C., & Droulez, J. (1995). Perception of three-dimensional shape from ego- and object-motion: Comparison between small- and large-field stimuli. *Vision Research, 35,* 453–462. [PubMed]

Di Luca, M., Domini, F., & Caudek, C. (2004). Spatial integration in structure from motion. *Vision Research, 44,* 3001–3013. [PubMed]

Domini, F., & Caudek, C. (1999). Perceiving surface slant from deformation of optic flow. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 426–444. [PubMed]

Domini, F., & Caudek, C. (2003a). 3D Structure perceived from dynamic information: A new theory. *Trends in Cognitive Sciences, 7,* 444–449.

Domini, F., & Caudek, C. (2003b). Recovering slant and angular velocity from a linear velocity field: Modeling and psychophysics. *Vision Research, 43,* 1753–1764. [PubMed]

Domini, F., & Caudek, C. (2010a). Matching perceived depth from disparity and from velocity: Modeling and psychophysics. *Acta Psychologica,* doi:10.1016/j.actpsy.2009.10.003 [PubMed]

Domini, F., & Caudek, C. (2010b). Combining image signals before 3D reconstruction: The intrinsic constraint model of cue integration. In J. Trommershäuser, M. S. Landy, & K. Körding (Eds.), *Sensory cue integration*. New York: Oxford University Press.

Domini, F., Caudek, C., & Proffitt, D. R. (1997). Misperceptions of angular velocities influence the perception of rigidity in the kinetic depth effect. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1111–1129. [PubMed]

Domini, F., Caudek, C., & Shirko, P. (2003). Temporal integration of motion and stereo cues to depth. *Perception & Psychophysics, 65,* 48–57. [PubMed]

Domini, F., Caudek, C., Turner, J., & Favretto, A. (1998). Discriminating constant from variable angular velocities in structure from motion. *Perception & Psychophysics, 60,* 747–760. [PubMed]

Domini, F., Vuong, Q. C., & Caudek, C.(2002). Temporal integration in structure from motion. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 816–838. [PubMed]

Dyde, R. T., & Harris, L. R. (2008). The influence of retinal and extra-retinal motion cues on perceived object motion during self-motion. *Journal of Vision, 8*(14):5, 1–10, http://journalofvision.org/content/8/14/5, doi:10.1167/8.14.5. [PubMed] [Article]

Eagle, R. A., & Hogervorst, M. A. (1999). The role of perspective information in the recovery of 3D structure-from-motion. *Vision Research, 39,* 1713–1722. [PubMed]

Fantoni, C. (2008). 3D surface orientation based on a novel representation of the orientation disparity field. *Vision Research, 48,* 2509–2522. [PubMed]

Ferman, L., Collewijn, H., Jansen, T. C., & van den Berg, B. (1987). Human gaze stability in the horizontal, vertical and torsional direction during voluntary head movements, evaluated with a three dimensional scleral induction coil technique. *Vision Research, 27,* 811–828. [PubMed]

Gogel, W. C., & Tietz, J. D. (1973). Absolute motion parallax and the specific distance tendency. *Perception & Psychophysics, 13,* 284–292. [PubMed]

Gu, Y., Angelaki, D. E., & DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience, 11,* 1201–1210. [PubMed]

Gu, Y., DeAngelis, G. C., & Angelaki, D. E. (2007). A functional link between area MSTd and heading perception based on vestibular signals. *Nature Neuroscience, 10,* 1038–1047. [PubMed]

Hay, J. C., & Sawyer, S. (1969). Position constancy and binocular convergence. *Perception & Psychophysics, 5,* 310–312.

Hogervorst, M. A., & Eagle, R. A. (2000). The role of perspective effects and accelerations in perceived three-dimensional structure-from-motion. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 934–955. [PubMed]

Jaekl, P. M., Jenkin, M. R., & Harris, L. R. (2005). Perceiving a stable world during active rotational and translational head movements. *Experimental Brain Research, 163,* 388–399. [PubMed]

Knill, D. C. (2007). Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision, 7*(7):5, 1–24, http://journalofvision.org/content/7/7/5, doi:10.1167/7.7.5. [PubMed] [Article]

Koenderink, J. J. (1986). Optic flow. *Vision Research, 26,* 161–180. [PubMed]

Koenderink, J. J., & van Doorn, A. J. (1975). Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta, 22,* 773–791.

Koenderink, J. J., & van Doorn, A. J. (1978). How an ambulant observer can construct a model of the environment from the geometrical structure of the visual inflow. *Kybernetik,* 224–247.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research, 35,* 389–412. [PubMed] [Article]

Liu, S., & Angelaki, D. E. (2009). Vestibular signals in macaque extrastriate visual cortex are functionally appropriate for heading perception. *Journal of Neuroscience, 29,* 8936–8945. [PubMed]

Longuet-Higgins, H. C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature, 293,* 133–135.

Longuet-Higgins, H. C. (1984). The visual ambiguity of a moving plane. *Proceedings of the Royal Society of London, 223,* 165–175. [PubMed]

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

Mayhew, J. E. W., & Longuet-Higgins, H. C. (1982). A computational model of binocular depth perception. *Nature, 297,* 376–379. [PubMed]

Naji, J., & Freeman, T. (2004). Perceiving depth order during pursuit eye movement. *Vision Research, 44,* 3025–3034. [PubMed]

Ono, H., & Steinbach, M. J. (1990). Monocular stereopsis with and without head movement. *Perception and Psychophysics, 48,* 179–187. [PubMed]

Ono, M. E., Rivest, J., & Ono, H. (1986). Depth perception as a function of motion parallax and

absolute-distance information. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 331–337. [PubMed]

Oosterhoff, F. H., van Damme, W. M., & van de Grind, W. A. (1993). Active exploration of three-dimensional objects is more reliable than passive observation. *Perception, 22,* 99.

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research, 43,* 2451–2468. [PubMed]

Panerai, F., Cornilleau-Pérès, V., & Droulez, J. (2002). Contribution of extraretinal signals to the scaling of object distance during self-motion. *Perception and Psychophysics, 64,* 717–731. [PubMed] [Article]

Peh, C.-H., Panerai, F., Droulez, J., Cornilleau-Pérès, V., & Cheong, L.-F. (2002). Absolute distance perception during in-depth head movement: Calibrating optic flow with extra-retinal information. *Vision Research, 42,* 1991–2003. [PubMed] [Article]

Rogers, B., & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception, 8,* 125–134. [PubMed]

Rogers, B., & Graham, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research, 22,* 261–270. [PubMed]

Rogers, B. J., & Collett, T. S. (1989). The appearance of surfaces specified by motion parallax and binocular disparity. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 41,* 697–717. [PubMed]

Rogers, S., & Rogers, B. J. (1992). Visual and nonvisual information disambiguate surfaces specified by motion parallax. *Perception & Psychophysics, 52,* 446–452. [PubMed]

Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics, 48,* 419–430. [PubMed]

Todd, J. T., & Perotti, V. J. (1999). The visual perception of surface orientation from optical motion. *Perception & Psychophysics, 61,* 1577–1589. [PubMed]

Tyler, C. W. (1974). Induced stereomovement. *Vision Research, 14,* 609–613. [PubMed]

Ullman, S. (1979). *The interpretation of visual motion.* Cambridge, MA: MIT Press.

van Boxtel, J. J. A., Wexler, M., & Droulez, J. (2003). Perception of plane orientation from self-generated and passively observed optic flow. *Journal of Vision, 3*(5):1, 318–332, http://journalofvision.org/content/3/5/1, doi:10.1167/3.5.1. [PubMed] [Article]

van Damme, W. J., & van de Grind, W. A. (1996). Non-visual information in structure-from-motion. *Vision Research, 36,* 3119–3127. [PubMed]

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology, 45,* 205–217.

Wallach, H., Stanton, L., & Becker, D. (1974). The compensation for movement-produced changes in object orientation. *Perception & Psychophysics, 15,* 339–343.

Wallach, H., Yablick, G. S., & Smith, A. (1972). Target distance and adaptation in distance perception in the constancy of visual direction. *Perception & Psychophysics, 12,* 139–145.

Wexler, M. (2003). Voluntary head movement and allocentric perception of space. *Psychological Science, 14,* 340–346. [PubMed]

Wexler, M., Lamouret, I., & Droulez, J. (2001). The stationarity hypothesis: An allocentric criterion in visual perception. *Vision Research, 41,* 3023–3037. [PubMed]

Wexler, M., Panerai, F., Lamouret, I., & Droulez, J. (2001). Self-motion and the perception of stationary objects. *Nature, 409,* 85–88. [PubMed]

Wexler, M., & van Boxtel, J. J. (2005). Depth perception by the active observer. *Trends in Cognitive Science, 9,* 431–438. [PubMed]

Wright, D. B., Horry, R., & Skagerberg, E. M. (2009). Functions for traditional and multilevel approaches to signal detection theory. *Behavior Research Methods, 41,* 257–267. [PubMed]

Wright, D. B., & London, K. (2009). Multilevel modeling: Beyond the basic applications. *British Journal of Mathematical and Statistical Psychology, 41,* 257–267. [PubMed]