



Visual perception of planar orientation: dominance of static depth cues over motion cues

Valérie Cornilleau-Pérès^{a,b,*}, Mark Wexler^c, Jacques Droulez^c, Emmanuel Marin^c,
Christian Miège^d, Bernard Bourdoncle^d

^a IPAL-CNRS, KRDL, 21 Heng Mui Keng Terrace, 119613, Singapore

^b Singapore Eye Research Institute, 11 Third Hospital Avenue, 168751, Singapore

^c LPPA CNRS-Collège de France, 11 pl. M. Berthelot, 75005 Paris Cedex 06, France

^d Service d'Optique Physiologique, Essilor, 57 Avenue de Condé, 94106 St Maur Cedex, France

Received 26 February 2001; received in revised form 27 August 2001

Abstract

We measured the ability to report the tilt (direction of maximal slope) of a plane under monocular viewing conditions, from static depth cues (square grid patterns) and motion parallax (small rotations of the plane about a frontoparallel axis). These two cues were presented separately, or simultaneously. In the latter case they specified tilts that were either collinear (coherent case) or orthogonal (conflict case). The field of view was small (8°) or large (60°). In small field, for motion parallax, the reported tilt depends strongly on the orientation of the plane relative to the rotation axis, being totally ambiguous when tilt is collinear with the rotation axis. In contrast, in large field, the reported tilt depends little on this variable, and is accurately specified by motion cues. In both cases static cues strongly dominated the tilt reports. Hence static grid patterns constitute robust tilt cues, which can dominate contradictory tilt indications from motion parallax, and should be considered as essential for the visual orientation during locomotion, or the immersion in virtual reality environments. © 2002 Published by Elsevier Science Ltd.

Keywords: Motion parallax; Texture; Depth perception; Optic flow; Orientation

1. Introduction

Slant and tilt describe fully the local orientation of a surface relative to the eye. They are defined as the angle relative to a frontoparallel plane, and the direction of maximal slope, respectively (Fig. 1A and B). In this paper, the term “tilt” refers indifferently to the unitary vector τ in the direction of the projected normal vector in the frontoparallel plane, or to the angle of this vector relative to the frontoparallel vertical.

The slant–tilt representation of surface orientation has several advantages (Stevens, 1983). For instance, it does not depend on the choice of specific image (or retinal) directions, contrary to the two gradients of the depth function in a cartesian coordinate system. Tilt is a critical variable for motor control, as it indicates the

direction of the ground slope during locomotion, or the orientation of the fingers and hand for the grasping of a flat object. However, it has received little interest so far, by comparison to the large number of studies dedicated to slant (Braunstein, 1968; Freeman, 1966; Freeman, Harris, & Meese, 1996; Meese & Harris, 1997). Although the final representation of tilt, which is useable for motor control, is probably defined in a body-centered or allocentric reference, we present here an approach to visual tilt perception in a gaze-centered frame of reference.

The human visual system presents the remarkable ability to recover the 3D layout of the environment from 2D retinal images. Gibson (1950) underlined the role of optic flow (the dynamic changes of retinal images) and static perspective cues in this regard. In monocular vision, the optic flow can theoretically reveal plane tilt in multiple frame stimuli (Hoffman, 1982). On the psychophysical side, the perception of tilt from optic flow has already been studied by Norman, Todd, and Phillips (1995) and Domini and Caudek (1999) in orthographic

* Corresponding author. Address: Singapore Eye Research Institute, 11 Third Hospital Avenue, 168751, Singapore.

E-mail addresses: rjizaac@singnet.com.sg, vcp@ccr.jussieu.fr (V. Cornilleau-Pérès).

(or close to orthographic) projection. This restriction prevents from extending the results to large fields of view, for which orthographic projection is an erroneous approximation. Another restriction of these two studies is the used of a fixed direction of motion, which is a very constrained case for tilt recovery because (i) disentangling the tilt from the direction of 3D motion is a critical theoretical problem in the 3D interpretation of visual motion (Longuet-Higgins & Prazdny, 1980) (ii) for a planar surface, there is a special ambiguity for the recovery of orientation from motion, between the 3D translation vector, and the normal to the plane (Longuet-Higgins, 1984), (iii) if the motion direction is fixed, the 2D velocity pattern can be used as a bias for tilt responses (i.e., in the absence of a 3D percept), since the ratio of the vertical and horizontal gradients of the optic flow indicates the tilt direction (see for instance Eq. (4) on p. 428 in Domini and Caudek (1999)). Hence, the first goal of this study was to measure the ability to report the tilt of a moving plane under small and large-field vision, for varied directions of motion.

Our second goal was to examine the perceptual strength of motion cues as compared to static monocular cues. Static depth cues can be easily made contradictory to the true depth pattern, as illustrated in the presentation of hollow masks (Gregory, 1970), of

Necker cubes (Peterson & Shyi, 1988) or, more relevant to our work, in Ames' window (Ittelson, 1952). Using grid patterns, Stevens and Brookes (1988) showed that static depth cues can dominate stereopsis for the perception of plane tilt in small field. Since the perception of depth within surfaces presents high similarities for stereopsis and motion parallax (Rogers & Graham, 1982), this questions whether static depth cues also dominate motion parallax for the perception of tilt.

The Ames' window portrays a trapezoidal window initially in a frontoparallel orientation (Fig. 1C), which is perceived as a rectangle slanted in depth. If this trapezoid rotates around a vertical axis, it is seen as a deforming rectangle. This illustrates the restricted case where static and motion cues portray collinear tilts, but different slant values (for instance in its physically frontoparallel position, the trapezoid has a zero slant, but the static perspective cues conveyed by the trapezoid shape indicate a non-zero slant). Since small-field (under 16° diameter) optic flow conveys little or no slant information (Eagle & Blake, 1995; Cornilleau-Pérès, Wong, Cheong, & Droulez, 2000), the visual system seems to use static cues specifying object orientation to lift the motion ambiguity, as proposed by Graham (1963). In any case, Ames' window clearly demonstrates that a non-rigid 3D shape can be perceived instead of the physically rigid object,

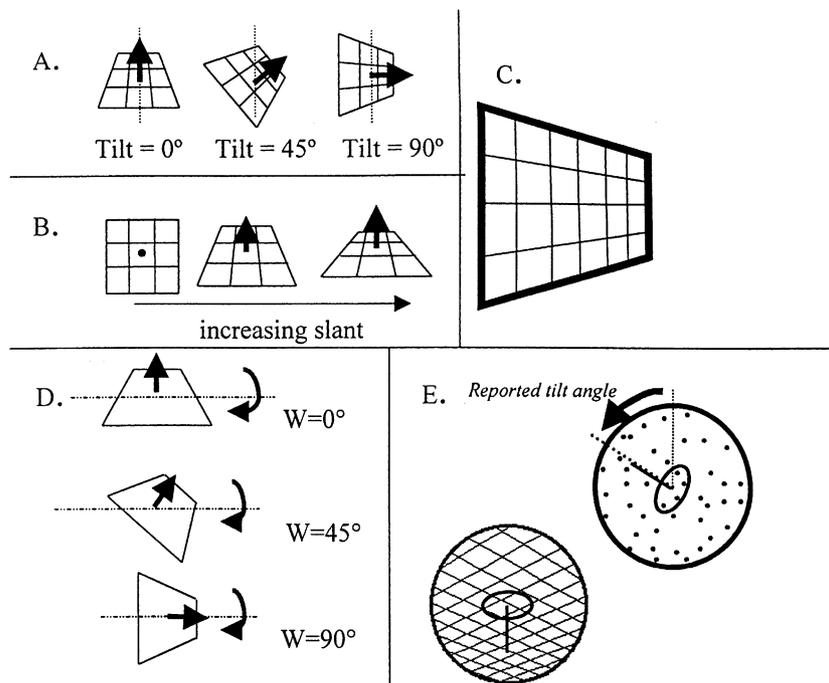


Fig. 1. (A) The tilt angle. The normal to an object plane projects as a vector in the frontoparallel plane. The tilt is the direction of this vector and indicates the direction of maximum slope. (B) The slant angle. The slant of an object plane is the angle between the object plane and the frontoparallel plane. (C) Intersecting lines are usually perceived as orthogonal in space. Grid patterns combine this depth cue with texture cues, as the 2D size of a mesh varies with depth, yielding the percept of a receding surface. (D) Several motion/plane configurations for different values of the winding angle W . W is the angle between the tilt and the direction of frontal translation, which itself is orthogonal to the axis of rotation. Only rotations around a horizontal axis are shown here. The small arrow indicates the surface normal. (E) The stimuli and probe. Dotted planes or grid planes were presented within a circular window. As illustrated here, the angle between the grid lines and the tilt was randomised.

and that static perspective cues can modify the perception of 3D shape from motion. However, the question arises whether the dominance of static depth cues also occurs when (1) the conflict concerns the tilt directions, rather than slant (2) the variable of interest (the tilt) is unambiguously specified by motion parallax, under large-field perspective for instance.

To elicit a perception of tilt from motion, we chose the 3D movement that was found to optimise the perception of 3D shape from motion (Dijkstra, Cornilleau-Pérès, Gielen, & Droulez, 1995), namely an oscillation around a frontoparallel axis. The component of frontoparallel translation is then orthogonal to the rotation axis. To measure the relative orientation between the plane and the 3D motion, we define the winding angle W as the angle between the tilt and of the frontal translation (Fig. 1D). W is unsigned and ranges between 0° and 90° .

Stevens (1983) demonstrated that the visual system tends to perceive intersecting lines as orthogonal in 3D space. He displayed the projection of a tilted plane textured with a square grid (Fig. 1C). In this stimulus, the ‘orthogonal interpretation’ combines with others static monocular cues, such as the gradients of size and density of texture elements, to yield a compelling percept of 3D shape, and accurate tilt responses. Therefore, we chose square grid patterns as stimuli to elicit static perspective cues.

In summary, we compare tilt perception in four conditions:

- when motion cues are isolated by presenting dotted planes oscillating around a frontal axis (condition DOT_MV), with a uniform dot density in the 2D image,
- when static cues are presented using square grids (GRI_ST) “drawn” over a static plane,
- when the two cues are combined coherently (condition GRI_MV), as the grid plane oscillates with the motion used for DOT_MV condition,
- when the two cues compete (condition GRI_CF), as motion cues indicate a tilt τ , and the plane texture (a trapezoidal grid) specifies a tilt orthogonal to τ .

Using a method of visual probe adjustment (Fig. 1E), we measured the perceived tilt of planes for small and large fields of view (FOV), of width 8° or 60° (equal to the angle of perspective projection).

2. Methods

2.1. Stimuli

In condition DOT_MV, a dotted plane (average 350 dots) rotated around a frontoparallel axis which was randomly chosen as vertical, horizontal or along one of

the two 45° oblique directions. The amplitude of the 3D oscillations was $\pm 6^\circ$, and the frequency was 0.5 Hz (with a refresh rate of 72 Hz, there were 72 different views per motion sequence). The dot density was uniform within the display screen, and could not be used as an orientation cue.

In condition GRI_ST, the stimuli were grid patterns with square elementary meshes within the object plane. The mean number of visible intersections was 43 in small field, and 19 in large field. The grids were composed of antialiased lines (width 3 pix), or regularly spaced dots (20 dots for each side of a mesh). Although the dot density was an additional cue to orientation in the latter case, we found that this choice (true lines or dotted lines) did not influence tilt responses. The orientation of the grid on the stimulus plane was randomly chosen between six orientations (from 0° to 75° relative to the vertical, by steps of 15°), and was independent from the tilt direction. The length of the mesh side was chosen randomly in the range $1.23^\circ \pm 20\%$ in small field, or $18.9^\circ \pm 20\%$ in large field.

In the “coherent two cues” condition (GRI_MV), the grid plane oscillated with the motion used for condition DOT_MV. In the “conflicting two cues” stimulus (GRI_CF), motion cues indicated a tilt τ_m , while the plane texture was a trapezoidal grid specifying an instantaneous tilt τ_g orthogonal to τ_m (Fig. 2).

Two additional control stimuli GRI + DOT_MV and GRI + DOT_CF were designed exactly as GRI_MV and GRI_CF, except that the dot distribution of the moving plane DOT_MV was superimposed to the grid pattern on the planar surface.

All conditions involving motion induced an apparent oscillation of the tilt direction in the frontoparallel plane, which increases with the slant and W , and never exceeds 6° (the rotation amplitude). Similarly, the slant would also vary with a maximum of 6° , reaching this maximum when $W = 0$. Therefore, our subjects were asked to report the mean tilt and slant values. Since tilt is our main variable of interest here, we plotted the average tilt change at the bottom of Fig. 4.

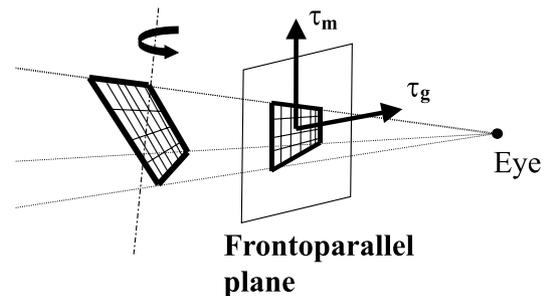


Fig. 2. The conflict case. The tilt direction τ_g which is specified by the static grid cues is orthogonal to the actual tilt of the surface τ_m which is depicted by the motion cues.

All images were computed under perspective projection, with the actual position of the subject's eye as the center of projection. The luminance of lines and dots varied over the image, in the range 9.5–10.5 cd/m² in small field, and 0.2–2 cd/m² in large field.

2.2. Apparatus

The images were presented on a monitor (small-field experiments) or projected by a videoprojector Barco 1208 on a translucent screen (large-field experiments), within a circular luminous window of 8° or 60° diameter, with spatial resolutions of 1.2 and 9 arcmin/pix respectively. The monitor and videoprojector had high resolutions of 1280 × 1024, but we used only the central 400 pixels, in order to avoid the image distortions, which may occur in the periphery of the CRT. The frame rate was 72 Hz and the viewing distance 72 cm.

The subjects viewed the stimuli monocularly, with an eye patch covering the dominant eye. They reported the perceived orientation of the plane by adjusting the orientation of a graphic probe (Fig. 1E) made of one needle and one ellipse, superimposed on the stimulus. The direction of the probe needle indicates the perceived tilt (the needle has one of its extremities fixed in the centre of the image). The small axis of the ellipse can be adjusted until the ellipse appears as a circle within the tilted plane.

The probe adjustments were made with a computer device. Three subjects used a mouse, then two subjects used a joystick, because we found it slightly more comfortable for understanding the task and giving the responses. Subsequently, we verified that this factor did not influence the results, most likely because the important factor was the graphic probe itself. To minimise the interactions between the stimulus and probe, the probe was visible intermittently, when the subject was pushing the mouse or joystick button.

2.3. Design

The tilt of the plane was randomly chosen between 0° and 360°, and its slant was 30° or 45° (the slant was equal for the motion and static cues when both were presented coherently or in conflict). For each slant, FOV, and condition, subjects performed 32 trials, which corresponded to 32 random values of the tilt. For conditions involving movement, they also corresponded to 32 random values of the winding angle W between the tilt and the direction of the frontal translation. In the analysis of the results, we grouped W values into three intervals of 30° width (0–30°, 30–60°, 60–90°).

Five subjects performed the four conditions DOT_MV, GRID_ST, GRID_MV, GRID_CF. Two of them performed also the control conditions GRI + DOT_MV and GRI + DOT_CF. The comparisons be-

tween the unsigned tilt error in different conditions was done either on raw error distributions using the median test (through the computation of χ^2) or on averaged error values with the Wilcoxon matched pair test (through the computation of Z), with threshold at $p < 0.05$. Both tests are non-parametric, which is necessary because of the strongly non-gaussian nature of the unsigned tilt error distributions. The median test is for independent samples, while the Wilcoxon test is applicable for matched data. Both tests were run using the software STATISTICA.

2.4. Subjects

The subjects (1 female and 4 males) were 20–27 years old and had normal vision. They were naive as to the goal of the experiment and paid for their participation. All gave their informed written consent for their participation.

3. Results

3.1. Ambiguities on the tilt sign

We measure the ambiguity in the tilt sign as the percentage of responses where the reported tilt is closer to $-\tau$ than to τ (the stimulus tilt). In small field, this ambiguity is partial for grid cues, but total (median 48.4%) for motion cues (Fig. 3). In large field it is not observed, except for a residual of 3.4% (median value) for motion cues. This supports the view that large-field motion reduces the ambiguity on the depth sign (Dijkstra et al., 1995). It is also compatible with the idea that small-field projection is closed to orthographic projection. Note that under orthographic projection we would expect that the tilt for static and dynamic conditions would be totally ambiguous. Here the ambiguity is much stronger for the motion cues. Hence, our results show

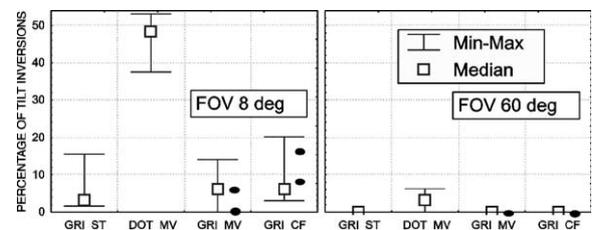


Fig. 3. Percentage of tilt inversions for each stimulus (indicated in abscissae), in small and large field. Since condition GRI_CF is dominated by grid cues, the percentage of depth reversals is calculated using the grid-defined tilt as a baseline. The square and brackets indicate the median and extrema, respectively, of the percentage of inversions across all subjects (each percentage is calculated over 64 trials). The black dots indicate the tilt inversions for two subjects in conditions GRI + DOT_MV (next to GRI_MV) and GRI + DOT_CF (next to GRI_CF).

that, in our experimental conditions, the visual system is more sensitive to the effects of perspective projection for static cues, in terms of foreshortening of grid segments, and tilting of the grid lines, than for motion cues, in terms of the second order of the velocity field (Cornilleau-Pérès et al., 2000).

Due to the presence of this ambiguity in small field, we decided to study the tilt responses, up to the 180° ambiguity. We removed this ambiguity from the results, by reversing the reported tilt when a tilt inversion occurs. Hence, we limited the tilt errors to the range (−90° to 90°), and computed the unsigned tilt error ranging between 0° and 90°. Due to the low amount of tilt reversals in large field, our results address the full tilt perception for the case of large field.

3.2. Tilt errors for motion parallax cues

When motion parallax is the only depth cue, the tilt error increases significantly with W (Table 1, Fig. 4), more strongly so in small field than in large field. In

small field for the large W range (60–90°), the histograms of the unsigned tilt error (Fig. 5) show that the tilt perception is strongly ambiguous, in the sense that the errors are widely spread between 0° and 90°.

The surface motion induces an oscillation of the tilt that increases in amplitude as W grows. The corresponding average angular displacement is plotted on the lower curves of Fig. 4. It amounts to less than 16% of tilt errors in small field, and less than 43% of it in large field. Therefore, the dynamic changes of tilt during the motion seem too weak to account for the strong influence of W on the tilt error, particularly in small field. In addition, the dynamic changes of the tilt are identical in small and large field, which fails to predict the larger influence of W in small field.

3.3. Comparison between coherent cue conditions (one or two cues)

For motion and static cues, the response accuracy increases with the FOV and with the slant (Fig. 4). We

Table 1
Spearman rank correlation between the unsigned tilt error and the winding angle W

Condition	DOT_MV	GRI_MV	GRI + DOT_MV	GRI_CF	GRI + DOT_CF
Number of subjects	5	5	2	5	2
FOV 8°, slant 30°	0.573*	0.038	0.198	−0.218*	−0.022
FOV 8°, slant 45°	0.385*	0.119	0.126	−0.115	−0.209
FOV 60°, slant 30°	0.208*	0.048	−0.036	−0.110	−0.177
FOV 60°, slant 45°	0.252*	−0.013	−0.146	−0.053	−0.116

*Indicates significant values ($p < 0.05$ or less).

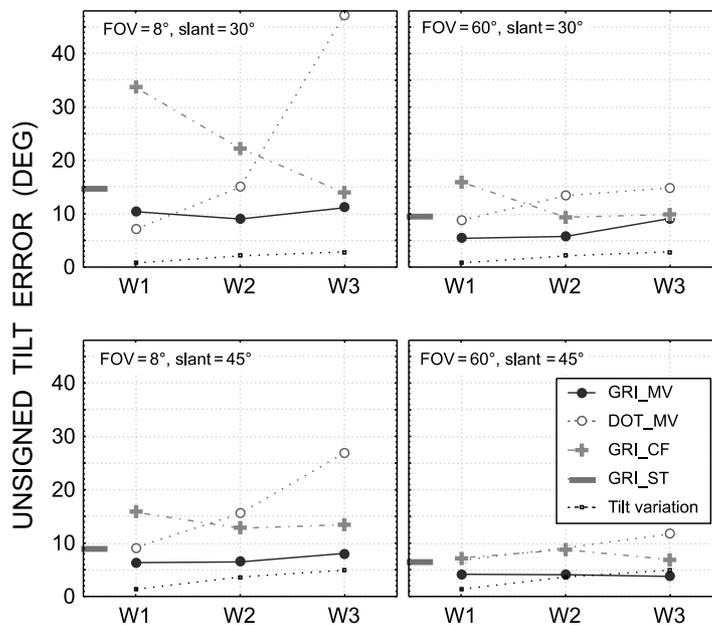


Fig. 4. Median unsigned tilt error in degrees for conditions GRI_ST, DOT_MV, GRI_MV and GRI_CF, for three winding angle ranges ($W1 = 0–30^\circ$, $W2 = 30–60^\circ$, and $W3 = 60–90^\circ$). For GRI_CF the baseline for calculating the tilt error is the grid-defined tilt. The lower curves (small circles) represent the average tilt variation in the stimulus during the motion.

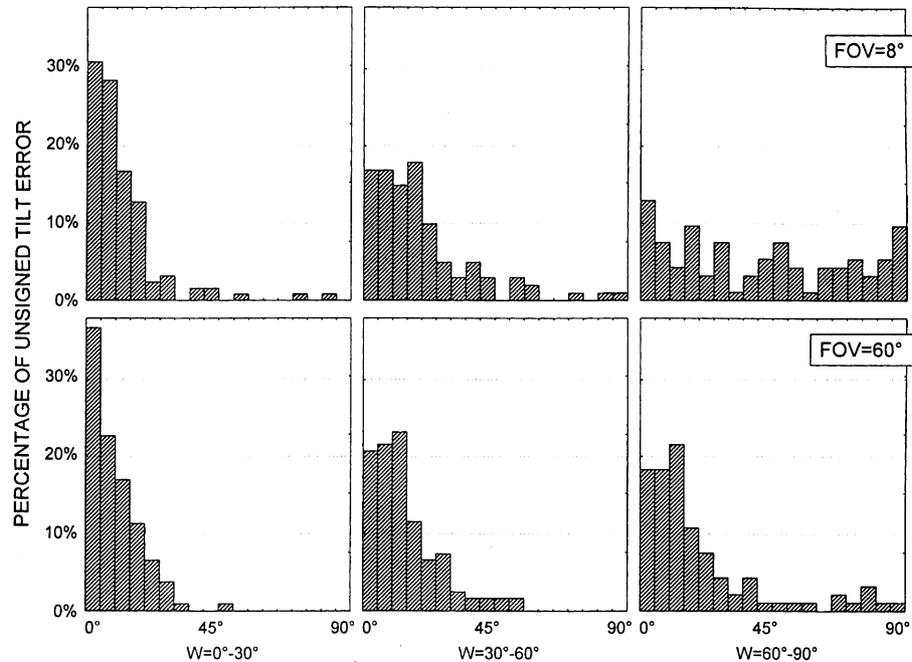


Fig. 5. Histograms of the unsigned tilt error in the condition DOT_MV, for each W range. Bin width: 5° .

grouped the responses of the five subjects, and tested these effects separately (for instance the FOV effect was tested for each slant and each condition). Both effects were significant for conditions GRID_ST and GRID_MV (χ^2 between 5.5 and 11.25, $p < 0.02$, $N = 160$). For motion cues (DOT_MV) the FOV effect was significant only for the large slant ($\chi^2 = 5.84$, $p < 0.016$), and the slant effect was significant only in large field ($\chi^2 = 9.27$, $p < 0.002$).

Fig. 4 also shows that the tilt errors are usually smaller for motion cues than for grid cues for small W ($0\text{--}30^\circ$), while the reverse trend is observed for large W ranges.

Responses were more accurate when the grid was moving, rather than static. In over 30 cases (5 subjects, 3 W ranges, 2 slants), the decrease in the tilt error was statistically significant both in small field (median value 1.43° , $Z = 2.46$, $p < 0.014$) and in large field (median value 3.84° , $Z = 3.84$, $p < 0.001$).

We also compared the performance for the moving grid to the best performance obtained for one cue alone (GRI_ST or DOT_MV) in each of the 30 cases. The tilt error decreased when the two cues are present, rather than one, by a median value of 0.7° for small field and 1.7° for large field, but significantly so only for the large-field condition ($Z = 2.81$, $p < 0.005$). Therefore, there exists a cooperation between static and motion cues, which tends to be stronger in large field than in small field.

It has previously been found that coupling two different depth cues improves the detection of depth differences, or the assessment of the slant or curvature

(Young, Landy, & Maloney, 1993; Cornilleau-Pérès & Droulez, 1993). The present result extends these conclusions to the precision in reporting ordinal depth relationship (tilt).

3.4. Results for the conflicting two cue condition

In the conflict case, the grid-defined tilt served as a baseline to calculate the tilt error. If this error is close to 0° , the reported tilt lies along the “grid tilt”, whereas a tilt error of 90° indicates that it lies along the “motion tilt”. In Fig. 4, the median tilt errors are smaller than 45° , this effect being significant for 11 of the 12 points of the GRI_CF curves. Hence, tilt is clearly reported according to grids, rather than motion cues. Moreover, the error histograms (Fig. 6, right column) indicate that the response peak is close to the grid tilt in condition GRI_CF for both field sizes.

We questioned whether the dominance of grid cues could be due to a difference in the accuracy of each cue alone. For the small W range ($0\text{--}30^\circ$), motion cues elicit a similar or better accuracy in tilt responses than grid cues. Yet the dominance of grid cues is very strong in this case (the reported tilt is only 10° from the grid tilt in average). We also restricted the analysis to the 17 cases out of 60 (5 subjects, 2 field sizes, 2 slants, 3 W -ranges) where motion cues yield smaller tilt errors than grid cues. In 16 of these 17 cases, the median of the unsigned tilt error was smaller than 40° (from 7.3° to 39.1° , median 19.7°), showing that the perceived tilt was much closer to the grid tilt than to the motion tilt. Therefore, grid cues tend to drive the tilt responses, even when

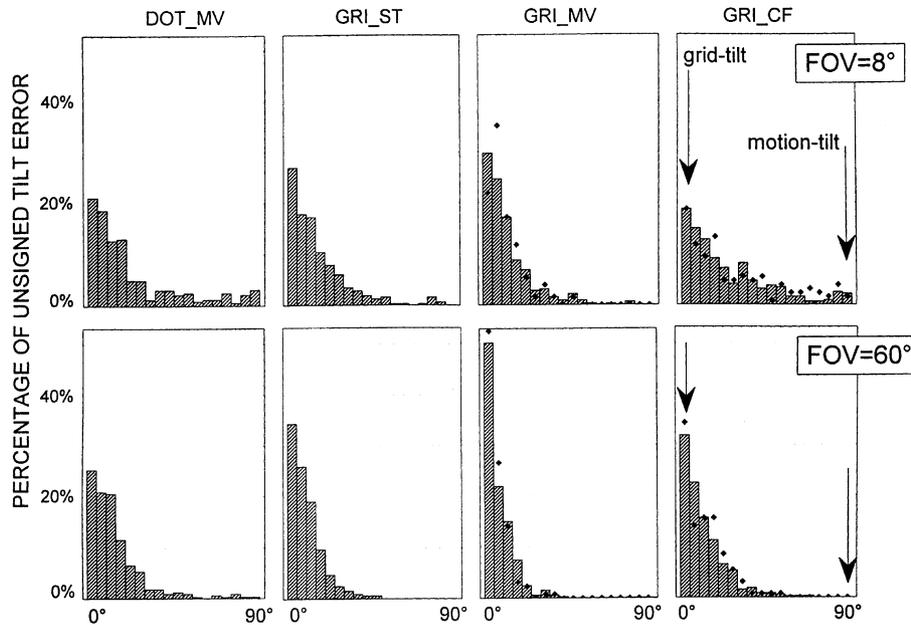


Fig. 6. Histograms of the unsigned tilt errors. Top row: small-field results. Bottom row: large-field results. Each column corresponds to a different condition. Bin width: 5° . The diamond in conditions GRI_MV and GRI_CF indicate the results for conditions GRI + DOT_MV and GRI + DOT_CF respectively. The grid-defined tilt is the baseline for calculating the unsigned tilt error in conditions GRI_CF (right column). Thus, a zero value of the error indicates a tilt reported from grid cues, whereas a 90° value indicates a tilt reported according to motion cues.

conflicting motion cues are more reliable if presented alone.

In spite of this dominance, conflicting motion cues do perturb the use of grid cues, because the error on the reported grid tilt increases significantly from condition GRI_ST to GRI_CF. This effect was significant both in small field (median value 7.9° , $Z = 3.84$, $p < 0.001$) and in large field (median value 1.7° , $Z = 2.62$, $p < 0.009$).

Another sign of this perturbation is the small negative correlation of the tilt error with W (Table 1) for condition GRI_CF. This suggests that, as W decreases, motion cues gain in accuracy and weight, and influence more strongly tilt perception. However, in no condition did we observe a shift of the peak of the tilt distributions toward the motion-defined tilt (Fig. 6), and the general effect of conflicting motion cues is only an increase in the variability of tilt responses.

3.5. Control conditions: superimposing dots and grid patterns

The lines in stimulus GRI_CF could provide a poorer velocity information than the dots in condition DOT_MV, thereby weakening the strength of motion cues. We designed stimuli GRI + DOT_MV and GRI + DOT_CF in the same way as GRI_MV and GRI_CF respectively, except that dot distributions were superimposed to the grids. For two subjects this did not modify the results for the coherent-cue and conflict conditions (diamonds in Fig. 6 right).

4. Discussion

Our results can be summarised as follows:

1. The perception of tilt from motion suffers from a strong 180° ambiguity in small field. For static grid cues this ambiguity is lower. For both cues, large-field perspective information reduces strongly the ambiguity.
2. The perception of plane tilt from motion parallax depends highly on the relative orientation of the plane and of the 3D motion (here a rotation around a frontoparallel axis) in small field. Large values of the winding angle W lead to a strong perceptual indeterminacy on the tilt direction (in addition to the 180° ambiguity).
3. Changing the size of the field of view from small (8°) to large (60°) leads to a general improvement of tilt reports, in particular to a full disambiguation of the tilt vector for large values of W .
4. In all cases the static grid cues clearly dominate over motion parallax when they indicate a tilt which is orthogonal to the motion-defined tilt.
5. The presence of conflicting motion cues tends to increase the variability of tilt reports, but not to shift their mean value away from the grid-defined tilt.

4.1. The perception of tilt from motion

Several computational studies have addressed the question of recovering tilt from optic flow. Hoffman

(1982) finds that tilt can be computed from multiple orthogonal views of a moving plane, but does not mention the possible role of W . Longuet-Higgins (1984) shows that two perspective views of a plane lead to confounding the tilt and frontal translation vector in the computation of surface orientation. However, this ambiguity does not seem to explain our results for two reasons. First, we do not observe an influence of W in large field as expected from Longuet-Higgins' approach. Second, the spurious solution (a tilt aligned with the frontal translation) also corresponds to a 90° slant. Such a configuration, i.e., a plane including the gaze axis, would probably be discarded by the subjects as the uniform dot density rather indicates a frontoparallel plane. Also, the slant responses of our subjects were in the range $3\text{--}64^\circ$, which undermines the possible existence of a 90° slant percept. Therefore, our results point to the discrepancy between existing models of tilt computation from motion, and the psychophysical reports.

When $W = 0^\circ$ the optic flow corresponds to a pure compression (the velocity magnitude varies in the direction of the velocity vector), whereas the case $W = 90^\circ$ corresponds to a shearing motion (the velocity magnitude varies in the direction orthogonal to the velocity direction). Several authors demonstrated that the visual sensitivity to compression and shearing differ (Nakayama, Silverman, MacLeod, & Mulligan, 1985; van Doorn & Koenderink, 1982; Rogers & Graham, 1983), but this effect concerned the detection of relative motion, whereas our task concerns the determination of a specific direction within the optic flow. With our present data it is not yet possible to determine whether the W -related anisotropy in tilt reports is due to the accuracy of the 3D reconstruction process, or whether it is purely a 2D effect, namely in the coding of the direction of the optic flow derivatives.

4.2. *The case of competing depth cues*

The ability of the visual system to use depth cues such as motion parallax or grid cues indicates that it uses some hypotheses to relate 2D visual images to 3D variables. For grid cues, such hypothesis can be that the texture is uniform on the surface, and that crossing lines are orthogonal in 3D space. Similarly, the 3D rigidity assumption is necessary to explain theoretically the perception of 3D structure from motion (Ullman, 1979; Longuet-Higgins & Prazdny, 1980). Yet our experiments show that 3D rigidity is not always the predominant hypothesis. Indeed, in the conflict case, our subjects perceived apparent 3D deformations of the plane, because the perceived grid tilt (indicated by static cues) was not compatible with a rigid interpretation. Hence the visual system favoured a non-rigid interpretation, instead of the rigid configuration corresponding to the

motion-defined tilt. Such non-rigid interpretations have already been reported under small-field orthographic projection (Braunstein & Payne, 1968; Hershberger, Stewart, & Laughlin, 1976; Jansson, 1977; Jansson & Johansson, 1973; Norman & Todd, 1993; Sparrow & Wren Stine, 1998). In our large-field stimuli this is not due to the weakness of motion cues, because (1) the tilt was reported precisely enough from motion parallax only, (2) a non-rigid percept was perceived, indicating that the non-rigidity was detected by the visual system. Rather, the 3D interpretation conveyed by grid cues seems to act as a constraint in a top-down control of the analysis of the 3D structure from motion. This interpretation is in agreement with the view that the perception of 3D structure from motion is influenced by learning and memorising a moving wireframe object (Sinha & Poggio, 1996). Finally, our result agrees with the report by O'Brien and Johnston (2000), who found that the perception of slant is also dominated by texture cues when they contradict motion cues.

In large-field vision, the 3D rigidity hypothesis is probably critical for motor control, because our ability to maintain equilibrium requests the existence of a stable frame of reference (Lee & Lishman, 1975; Dijkstra, Schöner, & Gielen, 1994). Also, small-visual motion is often produced by non-rigid transformations of the visual scene (displacements of objects for instance), whereas large-field motion is usually due to self-motion in a stationary environment. Note, however, that the dominance of static cues is stronger in large field than in small field, and that apparent 3D deformations of the stimulus were particularly visible in large field. Hence, our results do not support the view that 3D rigidity is used in a stricter way for large visual scenes. However, the question remains whether the conclusion would be similar during self-motion. In this case the analysis of 3D structure from motion receives some input from non-visual information regarding the direction of the 3D motion, as shown by the disambiguation of the depth sign through self-motion in small field (Cornilleau-Pérès & Droulez, 1994; Rogers & Rogers, 1992). In small field again, Wexler, Panerai, Lamouret, and Droulez (2001) have demonstrated that under a conflict between the static grid tilt and the motion tilt, there tends to be a bimodal response distribution around the two possible tilt values, the motion tilt being more frequently reported during self motion than during object-motion. Therefore, the weight of motion cues increases during self motion, as compared to object motion.

For object motion, Wexler et al. reported a bimodal distribution in their conflict condition. On the opposite, we find a unimodal, grid tilt centred, distribution. Two factors may explain this discrepancy. First, Wexler et al. used grids that were missing random cells, which they found to reduce the relative weight of perspective. Sec-

ond, they presented a unique motion direction (horizontal) in the stimulus, which is likely to ease the determination of tilt from motion parallax (indeed the computation of tilt from motion is theoretically simplified if the motion direction is known, as shown by Longuet-Higgins, 1984). Combining the results from Wexler et al. and our study yields therefore the tentative conclusion that, for object motion, the conflict case yields:

- two peaks in the tilt responses if the motion direction is fixed, (at the grid tilt, and at the motion tilt),
- one peak at the grid tilt, larger than in the coherent condition, if the motion direction varies randomly.

The mechanisms of depth cue combination have been modelled and tested in cases where the tilt indicated by conflicting cues were collinear, when the question concerns the amount of perceived slant, or curvature. If applied along every image directions, the weak fusion model (Clark & Yuille, 1990) predicts that the perceived tilt should lie between the tilts defined by motion and grid cues, (because the slant in each direction would be intermediate between the slants specified by each cue, being closer to the slant specified by the most reliable cue). Such a shift in the reported tilt is not observed here. Rather, grid cues play a role close to the veto defined by Bühlhoff and Mallot (1987), since the mean perceived tilt is hardly changed by contradictory motion information.

Hence, static monocular cues are critical for the perception of plane orientation. This may explain why people experience sensations of discomfort living in trapezoidal rooms (Crunelle, 1996). The static perception of the room as rectangular would contradict the actual 3D geometry specified by the optic flow during self-motion, hence creating apparent distortions of the room, and generating an instability in the perceived 3D space.

Acknowledgements

This research was partly funded by Essilor International. We thank A. Treffel and M. Ehrette for their invaluable technical help regarding the experimental set up.

References

- Braunstein, M. L. (1968). Motion and texture as sources of slant information. *Journal of Experimental Psychology*, 78, 247–253.
- Braunstein, M. L., & Payne, J. W. (1968). Perspective and the rotating trapezoid. *Journal of the Optical Society of America*, 58, 339–403.
- Bühlhoff, H. H., & Mallot, H. A. (1987). Interaction of different modules in depth perception. In *In 1st International Conference on Computer Vision* (pp. 295–305). New York: IEEE Society Press.
- Clark, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing systems*. Boston, MA: Kluwer.
- Cornilleau-Péres, V., & Droulez, J. (1993). Stereo motion cooperation and the use of motion disparity in the visual perception of 3D structure. *Perception and Psychophysics*, 54, 223–239.
- Cornilleau-Péres, V., & Droulez, J. (1994). The visual perception of 3D shape from self-motion and object-motion. *Vision Research*, 34, 2331–2336.
- Cornilleau-Péres, V., Wong, T. K., Cheong, L. F., & Droulez, J. (2000). Visual perception of slant from optic flow under orthographic projection and perspective projection. *Investigative Ophthalmology and Visual Sciences*, 41 (Suppl.), 3820 (abstract).
- Crunelle, M. (1996). A problem in perception: living in trapezoidal spaces. *Perception* 25 (Suppl.), 60B (abstract).
- Dijkstra, T. M. H., Cornilleau-Péres, V., Gielen, C. C. A. M., & Droulez, J. (1995). Perception of 3D shape from ego- and object-motion: comparison between small and large field stimuli. *Vision Research*, 35, 453–462.
- Dijkstra, T. M. H., Schöner, G., & Gielen, C. C. A. M. (1994). Temporal stability of the action-perception cycle for postural control in a moving visual environment. *Experimental Brain Research*, 97, 477–486.
- Domini, F., & Caudek, C. (1999). Perceiving surface slant from deformation of optic flow. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 426–444.
- van Doorn, A. J., & Koenderink, J. J. (1982). Spatial properties of the visual detectability of moving spatial white noise. *Experimental Brain Research*, 45, 189–195.
- Eagle, R. A., & Blake, A. (1995). Two-dimensional constraints on three-dimensional structure from motion tasks. *Vision Research*, 35, 2927–2941.
- Freeman, R. B. (1966). Absolute threshold for visual slant: the effect of stimulus size and retinal perspective. *Journal of Experimental Psychology*, 71, 170–176.
- Freeman, T. C. A., Harris, M. G., & Meese, T. S. (1996). On the relationship between deformation and perceived surface slant. *Vision Research*, 36, 317–322.
- Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.
- Graham, C. H. (1963). On some aspects of real and apparent visual movement. *Journal of the Optical Society of America*, 53, 1019–1025.
- Gregory, R. L. (1970). *The intelligent eye*. New York: McGraw-Hill.
- Hershberger, W. A., Stewart, M. R., & Laughlin, M. K. (1976). Conflicting motion perspective simulating simultaneous clockwise and counterclockwise rotation in depth. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 174–178.
- Hoffman, D. D. (1982). Inferring local surface orientation from motion fields. *Journal of the Optical Society of America*, 72, 888–892.
- Ittelson, W. H. (1952). *The ames demonstrations in perception*. Princeton, NJ: Princeton University Press.
- Jansson, G. (1977). Perceived bending and stretching motions from a line of points. *Scandinavian Journal of Psychology*, 18, 209–215.
- Jansson, G., & Johansson, G. (1973). Visual perception of bending motion. *Perception*, 2, 321–326.
- Lee, D. N., & Lishman, J. R. (1975). Visual proprioceptive control of stance. *Journal of Human Movement Studies*, 1, 87–95.
- Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London*, 208, 385–397.
- Longuet-Higgins, H. C. (1984). The visual ambiguity of a moving plane. *Proceedings of the Royal Society of London*, B223, 165–175.
- Meese, T. S., & Harris, M. G. (1997). Computation of surface slant from optic flow: orthogonal components of speed gradient can be combined. *Vision Research*, 37, 2369–2379.

- Nakayama, K., Silverman, G., MacLeod, D. I. A., & Mulligan, J. (1985). Sensitivity to shearing and compressive motion in random dots. *Perception*, *14*, 225–238.
- Norman, J. F., & Todd, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception and Psychophysics*, *53*, 279–291.
- Norman, J. F., Todd, J. T., & Phillips, F. (1995). The perception of surface orientation from multiple sources of optical information. *Perception and Psychophysics*, *57*, 629–636.
- O'Brien, J., & Johnston, A. (2000). When texture takes precedence over motion in depth perception. *Perception*, *29*, 437–452.
- Peterson, M. A., & Shyi, G. C.-W. (1988). The detection of real and apparent concomittant rotation in a three-dimensional cube implications for perceptual interactions. *Perception and Psychophysics*, *44*, 31–42.
- Rogers, B. J., & Graham, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research*, *22*, 261–270.
- Rogers, B. J., & Graham, M. (1983). Anisotropies in the perception of three-dimensional surfaces. *Science*, *221*, 1409–1411.
- Rogers, S., & Rogers, B. J. (1992). Visual and non visual information disambiguate surfaces specified by motion parallax. *Perception and Psychophysics*, *52*, 446–452.
- Sinha, P., & Poggio, T. (1996). Role of learning in three-dimensional form perception. *Nature*, *384*, 460–463.
- Sparrow, J. E., & Wren Stine, W. M. (1998). The perceived rigidity of rotating eight-vertex geometric forms: extracting nonrigid structure from rigid motion. *Vision Research*, *38*, 541–556.
- Stevens, K. A. (1983). Surface tilt (the direction of slant): a neglected psychophysical variable. *Perception and Psychophysics*, *33*, 241–250.
- Stevens, K. A., & Brookes, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research*, *28*, 371–386.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.
- Wexler, M., Panerai, F., Lamouret, I., & Droulez, J. (2001). Self-motion and the perception of stationary objects. *Nature*, *409*, 85–88.
- Young, M. J., Landy, M. S., & Maloney, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research*, *33*, 2685–2696.