Available online at www.sciencedirect.com
ScienceDirect
Vision Research 47 (2007) 1335–1349
Vision Research
www.elsevier.com/locate/visres

The relation between disparity and velocity signals of rigidly moving objects constrains depth order perception

Massimiliano Di Luca [a], Fulvio Domini [a,*], Corrado Caudek [b]

[a] Department of Cognitive and Linguistic Sciences, Brown University, 200 Metcalf Research Laboratory, 190 Thayer St., Providence, RI 02912, USA
[b] Max Planck Institute for Biological Cybernetics, Spemannstr. 41, 72076 Tübingen, Germany

Received 16 January 2006; received in revised form 11 October 2006

Abstract

In two experiments, observers were asked to judge the relative depth of a probe and one or two flanker dots. In Experiment 1, we found that such judgments were influenced by the properties of adjacent image regions, that is, by the amount of angular rotation of a surrounding cloud of dots. In Experiment 2, we found that the properties of the adjacent image regions affected the precision of the observers' judgments. With only the probe and the flanker dots presented in isolation, the precision of observers' judgments was much lower than when probe and the flanker dots were surrounded by a rigidly-connected cloud of dots. Conversely, a non-rigid rotation of the surrounding dots was detrimental to the precision of visual performance. These data can be accounted for by the Intrinsic Constraint model [Domini, F., Caudek, C., & Tassinari, H. (2006). Stereo and motion information are not independently processed by the visual system. *Vision Research, 46*, 1707–1723], which incorporates the mutual constraints relating disparity and motion signals. The present investigation does not show that the rigidity constraint affects the visual interpretation of motion information alone. Rather, our results show that perceptual performance is affected by the linear relation between disparity and velocity signals, when both depth-cues are present and the distal object is, in fact, rigid.

Keywords: Vision; Rigidity constraint; Cue integration; Disparity; Motion

1. Introduction

The visual environment contains many cues to the three-dimensional (3-D) properties of a visual scene (e.g., Cutting & Vishton, 1995). The hypothesis of a modular architecture of the visual system (e.g., Marr, 1982) has lead to theoretical studies on how information about 3-D structure may be extracted from each depth source in isolation (e.g., Bülthoff, 1991). Several algorithms have been proposed on shape from stereo, shading, texture, geometric structure, surface contours, motion, accommodation, and occlusion. Such algorithms produce unbiased outcomes when their underlying assumptions hold. These constraints range from the surface smoothness constraint as in Grimson (1981), the rigidity constraint used in motion analysis (Tsai & Huang, 1984; Ullman, 1984; Liu & Huang, 1988), the Lambertian assumption used in shape-from-shading (Horn, 1975; Brown, Ballard, & Rainero, 1983), the directional isotropy assumption (Witkin, 1981) or uniform density (Aloimonos, 1988), or weaker forms thereof (Gårding, 1993) used in shape-from-texture, and the knowledge of the light source in shape-from-shading methods (Pentland, 1982; Brooks & Horn, 1985). If these constraints are not met, however, the analysis of each isolated source of depth information will produce a non-veridical outcome.

Notwithstanding these theoretical analyses, when a single depth cue is shown in isolation, there is little evidence that such constraints are perceptually effective. Whereas some investigations support the hypothesis that visual processing is consistent with the above-mentioned theoretical analyses (e.g., Hildreth, Grzywacz, & Adelson, 1990 for motion; Backus, Banks, van Ee, & Crowell, 1999 for stereo,

* Corresponding author. Fax: +1 401 863 2255.
E-mail address: Fulvio_Domini@brown.edu (F. Domini).

0042-6989/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2006.10.029

Pentland, 1989 for shading), other investigations have cast doubts over such interpretations (Caudek & Proffitt, 1993; Domini, Caudek, & Proffitt, 1997; Caudek & Domini, 1998; Domini, Caudek, & Richman, 1998; Domini, Caudek, Turner, & Favretto, 1998; Domini & Caudek, 1999; Caudek & Rubin, 2001; Caudek, Domini, & Di Luca, 2002; Domini, Caudek, & Skirko, 2003; Domini et al., 2003; Todd & Bressan, 1990; Todd & Norman, 2003; Todd, Tittle, & Norman, 1995; Liter & Braunstein, 1998; Tittle & Braunstein, 1993).

Constraints in the interpretation of optical information acquire a different role when several cues are simultaneously available (see Domini, Caudek, & Tassinari, 2006). Let us consider the case of two depth cues, for example, motion, v, and disparity, d. If the *rigidity constraint* holds (and only in that case), then, necessarily, these two cues must be linearly related to each other (Figs. 1 and 2, see discussion below). Domini et al. (2006) proposed a model that, by relying on such constraint, derives 3-D affine structure from the projected d and v signals, and provided psychophysical evidence supporting the psychological plausibility of such a model.

The motivation of this investigation is to determine whether, in the case of a rigid object rotation, *the visual analysis makes use of the mutual constraint relating disparity and velocity signals*. We investigated such a question by examining perceptual performance in stimulus conditions in which *a small amount of non-ridigity* was added to the stimulus displays. Our hypothesis is that, under the rigidity constraint, the linear relation between the d and v signals affects the perceptual interpretation, so that *the non-rigid dots are perceived, not with their actual d, v values, but with those values that are expected in absence of measurement noise*. We will show that such hypothesis is consistent with our empirical findings, even though such re-interpretation undermines the veridical recovery of the depth order relations.

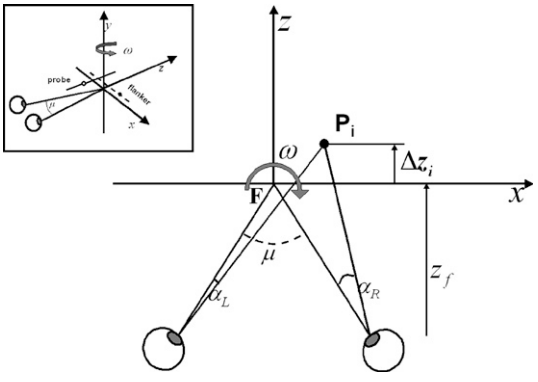

Fig. 1. Schematic representation of the viewing geometry for the stimuli used in the present investigation. The z and x axes represent the line-of-sight and the horizontal dimension, respectively. The vertical axis of rotation of the simulated 3-D configuration is placed in correspondence with the origin of the coordinate system and ω represents the angular velocity. **F** represents the fixation point and μ represents the vergence angle. $\mathbf{P}_i$ is a generic point. α_L and α_R represent the angles subtended by the points **F** and $\mathbf{P}_i$ to the left and right eye, respectively. z_f is the fixation distance; Δz_i is the difference between the fixation distance and the distance between the point $\mathbf{P}_i$ and the observer. The inset depicts the 3D viewing condition of the actual experiments.

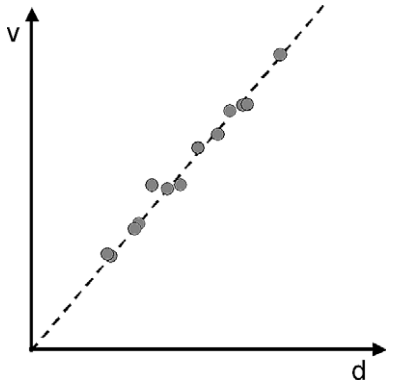

Fig. 2. The gray points represents disparity and velocity measurements corrupted by noise. The dashed line represents the linear relationship between disparities and velocities signals, in absence of measurement noise.

1.1. Perceived depth order from a single depth cue

In Experiment 1, observers were asked to manipulate the disparity and velocity values of one dot (*probe*), so that it appeared to be at the same egocentric distance as another dot (*flanker*, see inset on Fig. 1). In principle, this task can be performed by matching the velocity values, v, or the disparity values, d, of the two dots.

Let $\mathbf{P}_i$ (for $i = 1, 2$) be the two points, and **F** the fixation point (see Fig. 1). Let us define the relative disparity d_i as the difference between the visual angles α_L and α_R subtended by $\mathbf{P}_i$ and **F** at the left and right vantage points. Disparity is coded in such a manner that larger (positive) values correspond to smaller simulated distances from the observer. The amount of relative depth Δz_i between $\mathbf{P}_i$ and **F** is thus related to relative disparity d_i through the following (approximated) equation:

$$d_i \approx \frac{E}{z_f^2}\Delta z_i + \varepsilon_d \tag{1}$$

where z_f is the fixation distance, E is the inter-ocular distance and ε_d is the noise of disparity measurements. We assume that ε_d is Gaussian and independent from the measurement noise of other signals. If we define the *scaled depth* as $z_i = \frac{\Delta z_i}{z_f}$ and the *vergence angle* as $\mu \approx \frac{E}{z_f}$ (this is a good approximation for objects at a distance of at least 50 cm from the observer), then the previous equation becomes:

$$d_i \approx \mu z_i + \varepsilon_d \tag{2}$$

Let us now define the relative velocity v_i as the difference, at two moments in time t_1 and t_2, between the visual angles α_2 and α_1 subtended by $\mathbf{P}_i$ and **F** at the cyclopean vantage

point. Similarly to Eqs. (2) and (3), there is a linear relation between relative velocity v_i and the scaled depth z_i:

$$v_i \approx \omega z_i + \varepsilon_v \quad (3)$$

where ω is the angular rotation during the time interval $[t_1, t_2]$ and ε_v is the noise of velocity measurements. Again, velocity is coded in such a manner that larger (positive) values correspond to smaller simulated distances from the observer.

In conclusion, Eqs. (2) and (3) suggest that the depth-alignment task of Experiment 1 can be performed by simply matching the disparity values, or the velocity values, of the two points $\mathbf{P}_1$ and $\mathbf{P}_2$, so that $d_1 = d_2$, or $v_1 = v_2$. The upper limit of precision in performing such a task depends only on the noise in the disparities and velocities measurements.

1.2. Perceived depth order from combined stereo-motion signals

Even though the information provided by each depth cue in isolation is sufficient for performing the task of Experiment 1, the simultaneous presence of the disparity-cue and of the motion-cue introduces an additional constraint in the depth-interpretation of the image signals. To clarify this point, let us consider again Eqs. (2) and (3), and let us suppose that the display contains n dots which rotate rigidly about, say, the vertical axis. For any *rigid rotation*, the parameters ω and μ of Eqs. (2) and (3) must be invariant for all the points of the 3-D configuration. In such circumstances, by necessity, *the velocity and disparity values will be linearly related to each other* (see Fig. 2).

We argue that, as a consequence of such constraint, the depth interpretation of $\mathbf{P}_1$ and $\mathbf{P}_2$ will be determined by the velocity and disparity values of all the other dots which belong to the same perceived rigid 3-D object. When a rigid interpretation is imposed, in fact, $\mathbf{P}_1$ and $\mathbf{P}_2$ are constrained by the properties of the other dots comprising the 3-D rigid configuration.

The present experiments tested two hypotheses:

(1) the depth-alignment task is performed by separate analyses of each of the two depth cues;
(2) the depth-alignment task is affected by the constraint relating the velocity and disparity values, under the rigidity assumption.

Different mechanisms might be used by the visual system to integrate the disparity and velocity values, so as to give rise to the perception of a rigid 3-D structure. The Modified Weak Fusion model (e.g., Hillis, Watt, Landy, & Banks, 2004; Kham & Blake, 2000; Landy, Maloney, Johnston, & Young, 1995) maintains that disparity and velocity signals are process by independent depth-processing modules, each giving rise to a separate depth estimate. In a second stage of analysis, depth estimates are integrated in a statistically optimal fashion. A different strategy of depth-cue combination has been proposed by Domini et al. (2006). Their model combines the depth-cues at the signals level (not at the level of the 3-D Euclidean depth interpretations), so as to obtain the optimal estimate of the 3-D affine structure.

To distinguish between the two hypotheses, it is necessary to describe in detail one of the possible mechanisms which might be responsible for integrating velocity and disparity signals. In the following, we will speculate that the visual system computes a 3-D shape from multiple depth-cues as indicated by the Intrinsic Constraint (IC) model proposed by Domini et al. (2006). This choice was guided by considerations regarding the observers' task in our experiments. Since observers were asked to perform an affine task, the IC model seemed a natural choice, given that provides a normative model for the optimal recovery of 3-D affine shape.

1.3. Optimal recovery of 3-D affine structure

Let us consider three points on a rigid object $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$, such that $\mathbf{P}_1 < \mathbf{P}_2 < \mathbf{P}_3$. These points generate three retinal velocities v_1, v_2 and v_3, and three disparities d_1, d_2 and d_3 (see Fig. 3a). In absence of noise, the disparity and velocity signals provide equivalent information about the 3D affine structure, since $v_1 < v_2 < v_3$ and $d_1 < d_2 < d_3$. In general, however, disparity and velocity signals are corrupted by measurement noise. The measured d_i and v_i values (with $i = 1,2,3$ indexing the three points) are thus two random variables. For simplicity, d_i and v_i will be considered as Gaussian random variables with means μ_{d_i} and μ_{v_i}, and standard deviations σ_{d_i} and σ_{v_i} (see Fig. 3b).

What is important here is to determine the *relative information-content* of the velocity and disparity signals, by keeping in mind that the model of Domini et al. (2006) is only concerned with the recovery of the 3D affine structure.

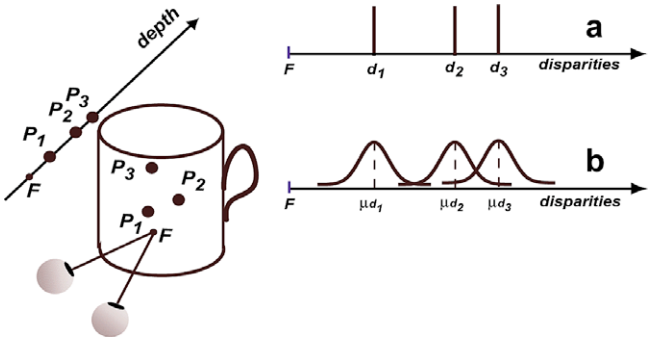

Fig. 3. Left panel: Schematic representation of a cup. P_1, P_2, P_3, three generic points on the surface; f fixation point. Right panel: (a) The intensity of the signals in absence of measurement noise is represented by the spikes d_1, d_2, d_1. (b) The intensity of the disparity signals is represented by the distance among the means of the Gaussian distributions; the spread of the distribution represents the fact that each signal is corrupted by noise. There is a distribution of noisy signal because we assume that multiple measurements are available for each image signal.

We can safely say that the relative information-contents of d_i and v_i differ solely in terms of their uncertainty levels. In order to compare these two signals, thus, a convenient transformation is the scaling by the standard deviations of the random components. After such scaling, the two signals can be directly compared, since both have the same variance ($\sigma = 1$).

In this manner, we can define a Cartesian space in which the axes represent the scaled disparity and velocity signals. Because of repeated measurements, within this space, the scaled velocity and disparity signals define bivariate Gaussian distributions centered at μ_{d_i} and μ_{v_i}, with standard deviations $\sigma_{d_i} = \sigma_{v_i} = 1$. For the three points $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$, the means μ_{d_i} and μ_{v_i} will lie on a straight line (which we will call ρ), since the local disparities and velocities values are linearly related to each other (Fig. 4). It can be shown that the projection of the scaled measurements of the three points $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$ onto the ρ line results in an *optimal encoding of the 3D affine structure* (Domini et al., 2006).

Let us assume that the visual system measures the instantaneous values of disparity and velocity projected by the points belonging to an object. These d_i and v_i values are linearly related to each other, and an estimate of ρ can be found by a Principal Component Analysis (PCA) on the scaled signals (see Fig. 5). The eigenvector of the first principal component (PC_1) provides an estimate of the slope of the ρ line on which the means μ_{d_i} and μ_{v_i} of the three bivariate distributions lie. The orientation of the first eigenvector, moreover, depends on the relative signal-to-noise ratio of each depth cues (see Figs. 4 and 5). It must be pointed out that the orientation of the ρ line depends on the viewing distance and the angular rotation of the object. Moreover, the intercept of this line with the velocity axis is determined by the amount of translation of the object.

With enough measurements available at each moment in time, the variation of estimated orientation of ρ will be negligibly small across repeated observations. The projection onto the ρ line of the measurements of $\mathbf{P}_i$, conversely, will vary considerably across repeated observations. Such variation defines a univariate distribution of ρ_{ij} scores (with $j = 1,\ldots,n$ for n measurements). It should be noted that the univariate distributions of the scaled disparity measurements and of the scaled velocity measurements and the univariate distribution of the ρ_{ij} scores have all unit variance.

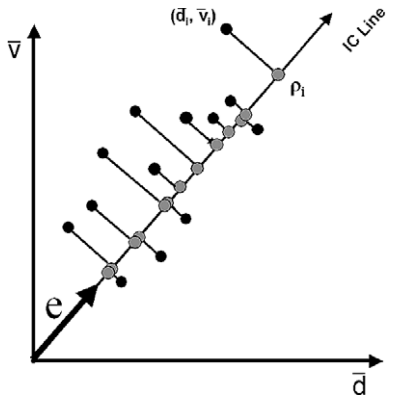

Fig. 5. The black points represent the noisy disparity and velocity signals scaled by their measurement errors ($\bar{d}$, $\bar{v}$), and **e** represents the first eigenvector of the scaled disparity-velocity covariance matrix. The IC line is termed ρ. The gray points represent the projections ρ_i of the scaled signals on the first principal component.

The most important aspect of such encoding of the 3-D affine structure is that *the means of the distributions of the noisy measurements of the three points* $\mathbf{P}_i$ *on the first principal component are necessarily more spaced among each other than the means of the distributions of the single-cue measurements*. In other words, *the scores on the first principal component provide a better estimate of the 3-D affine structure than the single-cue signals considered in isolation* (see Fig. 4).

The scaling by the standard deviation of the measurement errors is just one of the infinite possible scaling which all produce the same end result, if the scaling factor is proportional to the standard deviation of the measurement errors. Since the scaling factor is arbitrary, with the above

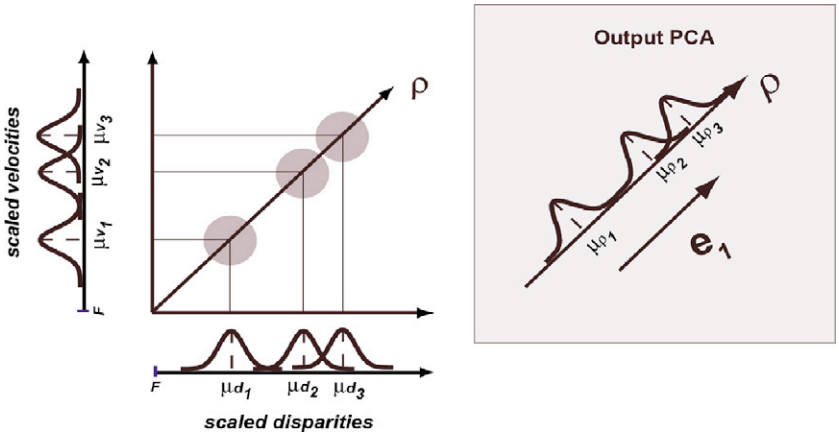

Fig. 4. The optimal encoding of the affine structure is provided by the scores on the first principal component (ρ) computed on the scaled velocity and disparity signals: The means of the distributions on the first PC are spaced apart from one another by a greater amount than those relative to the single-cue scaled-signals.

constraint, then it should be clear that the absolute values of the means and standard deviations of the composite signals on PC_1 are not directly interpretable: Only the ratios of these quantities are informative.

In conclusion, according to Domini et al. (2006), the best estimate of the local 3-D affine structure is provided by the scores on the first principal component computed on the signals scaled by their standard deviations. The scaling of the signals determines the orientation of PC_1. The choice of the standard deviation of the measurement error as the scaling factor optimizes the recovered 3-D affine solution.

1.4. Affine judgments

Let us consider the previous example again, in which $\mathbf{P}_1 < \mathbf{P}_2 < \mathbf{P}_3$, and $\mathbf{P}_3 - \mathbf{P}_2 < \mathbf{P}_2 - \mathbf{P}_1$ (see Fig. 4). Over repeated measurements, the relation among the average scores ρ_i on PC_1 (μ_{ρ_1}, μ_{ρ_2} and μ_{ρ_3}) will be isomorphic to the depth-relations among the points $\mathbf{P}_i$, since $\mu_{\rho_1} < \mu_{\rho_2} < \mu_{\rho_3}$ and $\mu_{\rho_3} - \mu_{\rho_2} < \mu_{\rho_2} - \mu_{\rho_1}$. Given a single measurement for each point $\mathbf{P}_i$, moreover, more reliable relative depth comparisons are obtained by using the ρ_i scores ($\rho_1 < \rho_2 \Rightarrow \mathbf{P}_1 < \mathbf{P}_2$, and $\rho_2 - \rho_1 < \rho_3 - \rho_2 \Rightarrow \mathbf{P}_3 - \mathbf{P}_2 < \mathbf{P}_3 - \mathbf{P}_1$), rather than the raw signals.

This method guarantees an optimal estimation of the local 3D affine structure and leads to the formulation of the main hypothesis of Domini et al. (2006). According to such hypothesis, the visual system combines the disparity and velocity values that are linearly related to each other (i.e., those belonging to a rigid structure) by performing a PCA on the signals scaled by the standard deviation of their measurement noise. As a consequence of such cue-combination strategy, the perceived local 3-D affine structure will depend solely on scores on the first principal component.

For the present purposes, it is important to note that there is a monotonic relation between the ρ_i scores and the relative-distances among the points $\mathbf{P}_i$ with respect to a reference point—this implies that larger ρ values correspond to smaller distances to the observer. Domini et al. (2006) hypothesized that the perceived depth order relations are determined by the values that ρ takes on: Two points $\mathbf{P}_1$ and $\mathbf{P}_2$ are perceived at the same egocentric distance if $\rho_1 = \rho_2$.

Domini et al. (2006) developed the IC model for the specific case of the fixation point on the axis of rotation and, thus, in its present form, the IC model cannot be applied to more generic viewing conditions. For a generic not-fixed axis of rotation, the mathematical description becomes more complex, since image velocities are not parallel to each other and cannot be described by a field of scalars. This problem, however, is easily solved, by discounting the rotational component parallel to the z-axis from the global velocity field (Ullman, 1979). A different problem concerns the fact that, within a temporal measurement window, the linear relationship between disparities and velocities changes over time, for a not-fixed axis of rotation. Even though we plan to extend the IC model to more generic viewing conditions, such theoretical developments cannot be presented here.

A final consideration concerns the choice of a fixed vertical-axis rotation for the stimuli of the present experiments. Such choice was made to generate a velocity field that, in general, is compatible with a 3-D rigid rotation. In such circumstances, to discriminate between rigid and non-rigid motion, is sufficient to determine whether the disparities and velocities signals are linearly related to each other.

2. Experiment 1

The goal of Experiment 1 was to establish whether the rigid shape specified by motion and disparity of a cloud of surrounding dots affects the local interpretation of two dots that are integrated into that interpretation. In Experiment 1, observers binocularly viewed two dots (one *probe dot* and one *flanker*) surrounded by a cloud of dots (the *context dots*), and were asked to judge whether the probe dot appeared in front or behind the flanker (see Figs. 6 and 7). Three dots are thus involved in such a task: the probe dot, the flaker dot and the fixation point. A staircase procedure was used to determine the velocity and disparity values of the probe which satisfy the experimental demand.

In different experimental conditions, we manipulated the simulated 3-D angular velocity, so as to vary the orientation of the PC_1 line. The IC model predicts that observers place the probe "in front" or "behind" of the flanker, according to the position of PC_1. If the surrounding dots can have such a dramatic effect, as to reverse the perceived depth order of the two central dots, then we must conclude that the recovery of the local 3-D affine structure depends on the rigid interpretation that has been provided to all the points that have been integrated into that interpretation.

In Experiment 1, the velocity-disparity of the probe was forced to vary, as a consequence of the adjustments of the

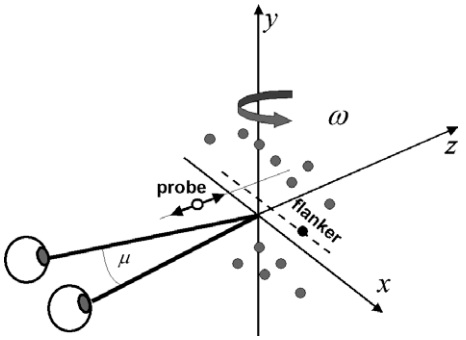

Fig. 6. Schematic representation of the stimulus display for Experiment 1. The x, y plane represents the image plane and z represents the axis coinciding with the line of sight. The gray and black dots are the *context dots* and the *flanker*, respectively (see text for details). Observers adjusted the probe by varying its z-coordinate. ω represents the angular velocity of the simualted 3-D configuration; μ represents the vergence angle.

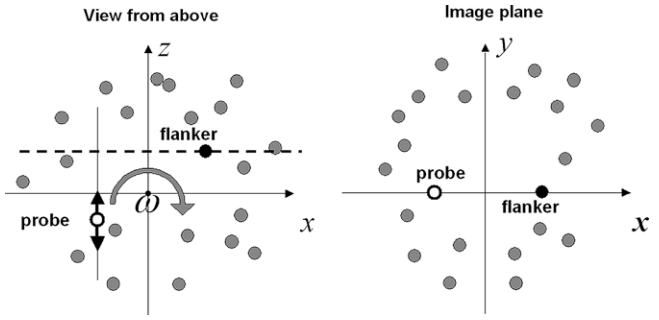

Fig. 7. Schematic representation of the stimulus display for Experiment 1. Left panel: View from above. The surrounding dots are placed within a spherical volume. Right panel: Front view. The probe and the flanker are placed in the central region of the display, separated by an empty region from the surrounding dots.

observer, only within a straight line in the velocity-disparity space (the *''control'' line*). Such subspace, however, did not contain the velocity-disparity values of the flanker (see Fig. 8). In this manner, observers could not perform the task by simply matching the velocity-disparity values of probe and flanker.

Four experimental conditions were created. We start by describing in detail the first two, since they clarify the rationale of the Experiment. Conditions 1 and 2, represented schematically in Fig. 9a, differed in terms of the simulated 3-D angular velocity of the surrounding cloud of dots. In both conditions, all points of the stimulus display but one were the projection of a rigid object-rotation about the vertical axis. The non-rigid point was the flanker in condition 1 and the probe in condition 2.

- In *condition 1*, the surrounding dots were rigidly connected to the probe, but not to the flanker–equivalently, the angular rotation of the surrounding dots was chosen so that their v, d ratio matched that of the probe dot. In these circumstances, the surrounding dots defined a first Principal Component, termed IC_1, passing through the probe (see Fig. 9a, left panel). The control line for the probe coincided with IC_1.

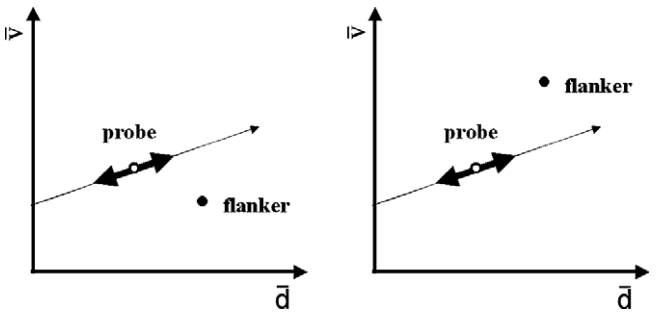

Fig. 8. Experiment 1. Probe and flanker dots are represented in the disparity, velocity space. The observers' adjustments of the probe are constrained in such a way that the probe can take on only the values of the control line (thin arrow). The left and right panels represent the conditions represented in Figs. 9a and b, respectively.

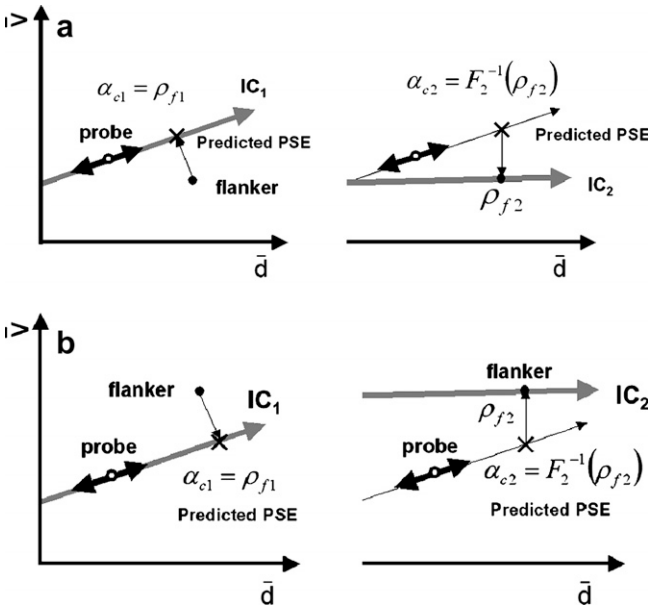

Fig. 9. (a) Experiment 1. Left panel: Condition 1. As indicated in Fig. 5, the probe can take on only the values belonging to the control line (gray arrow). In condition 1, such values coincide with those identified by the first eigenvector of the variance-covariance matrix of the disparity and velocity signals of all dots in the display. In other words, the probe can solely take on the values that belong to the IC_1 line (gray bold arrow). The flanker identify the only point of the stimulus display which does not belong to the IC_1 line. According to the proposed model, the flanker is considered as a random perturbation of the IC_1 line. We hypothesize that the visual system recovers the "unperturbed" value of such point. If the signals are scaled by their measurement error, the recovered value will correspond to the orthogonal projection on the IC_1 line. In terms of the proposed model, the task is to match the position of the probe with the position of the recovered value of the flanker. The cross in the figure indicates the predicted Point of Subjective Equality (PSE). Right panel: In condition 2, the probe was constrained to assume the same values as in condition 1, but, in this case, such values did not correspond to the first eigenvector of the variance-covariance matrix (IC_2, gray bold arrow). According to the proposed model, in condition 2, the position of the probe is considered a random perturbation of the IC_2 line. We hypothesize that the visual system recovers the "unperturbed" value of the probe on the IC_2 line. Since the task is to match the probe with the flanker, we expect that observers will select, among the values that the probe can take on, the value which, by being orthogonally projected on the IC_2 line, corresponds to the flanker. (b) Experiment 1. Left panel: Condition 3. This condition is similar to condition 1, except for the position of the flanker. Right panel: Condition 4. This condition is similar to condition 2, except for the position of the IC_2 line and the flanker.

- In *condition 2*, the surrounding dots were rigidly connected to the flanker, but not to the probe. Hence, the surrounding dots defined a first Principal Component, termed IC_2, passing through the flanker (see Fig. 9a, right panel). The control line for the probe did *not* coincide with IC_2, but rather with the subspace spanned by IC_1 in condition 1.

It is important to note that the disparity-velocity values of the flanker in condition 1, or the probe in condition 2, differed by only a small amount from fitting into the rigid structure. Consequently, *all points of the stimulus displays*

were always perceived as forming a rigid structure. Remember that, for any rigid structure, the velocity-disparity values are linearly related. In other words, if the stimulus' features of a rigidly-rotating object are represented within the disparity-velocity space, they will all lie on a straight line (the IC sub-space).

Because of the perceived rigidity, in condition 1, the measured disparity and velocity values of the flanker can be considered as a noisy perturbation of the "true" value of the flanker within the IC_1 subspace. According to our hypothesis, *observers match the value of the probe laying within IC_1 subspace with the "best guess" of the position that the flanker should take on within the IC_1 subspace*. If the image signals are scaled by their standard deviations, then we hypothesize that *the best guess of the unperturbed flanker's value can be found by orthogonal projection of the flanker onto IC_1*. Let us term ρ_{f_1} such a value. The predicted point of subjective equality in condition 1, α_{c_1}, thus becomes $\alpha_{c_1} = \rho_{f_1}$ (Fig. 9a, left panel).

In condition 2, the first Principal Component, IC_2, is oriented differently than in condition 1, and passes through the flanker point (Fig. 9a, right panel). The control line, conversely, coincides with the subspace spanned by IC_1 in condition 1. In these circumstances, *the v, d values of the probe can be considered as a noisy perturbation of the "true" value of the probe within the IC_2 subspace*. If we indicate with $\mathcal{F}_2$ the function which associates (through orthogonal projection) points belonging to the control line to scores on IC_2, then in condition 2 the predicted Point of Subjective Equality, α_{c_2}, becomes: $\alpha_{c_2} = \mathcal{F}_2^{-1}(\rho_{f_2})$ (Fig. 9a, right panel).

Note that $\mathcal{F}_2^{-1}(\rho_{f_2}) > \rho_{f_1}$ (see Fig. 10, left panel). We thus expect two different PSEs for the two experimental conditions, *even though the v, d values of the flanker are kept constant*.

The predicted PSEs for conditions 1 and 2 are summarized in Fig. 11. In condition 1, the predicted PSE coincides with the point on the control line coinciding with the orthogonal projection of the flanker on IC_1 (left panel, upper-left). In the condition 2, the predicted PSE is the point on the control line coinciding with the back-projection of the probe from IC_2 (left panel, lower-right). In summary, α_{c_1} and α_{c_2} represent the predicted *d*, *v* values of the probe which make the probe dot to appear at the same *z*-depth distance as the flanker for IC_1 and IC_2 subspaces, respectively (Fig. 11, right panel, main diagonal).

Now we can ask: what does it happen if the probe takes on the value α_{c_1}, but the context dots define the IC_2 subspace (Fig. 8, left panel, upper-right)? The projection of α_{c_1} on IC_2 determines a score $\mathcal{F}_2(\alpha_{c_1})$ on IC_2 smaller than ρ_{f_2} (where ρ_{f_2} is the predicted PSE when the IC line is IC_2). As a consequence, we predict that a change in the angular velocity of the context dots will lead the observers to perceive, in one condition, probe and target dots at the same distance, and, in a second condition, the probe as farther away from the observer than the flanker[1] (Fig. 11, left/right panels, upper-right).

A similar prediction can be made for the case in which the probe takes on the value α_{c_2} and the context dots define the IC_1 sub-space (Fig. 11, left panel, lower-left). In this case, the orthographic projection of the flanker on IC_1 determines a score $F_1(\rho_{f_2})$ smaller than α_{c_2}. In these circumstances, therefore, we expect that the probe will be perceived as closer to the observer than the flanker.

Conditions 3 and 4, schematically represented in Fig. 9b, provide a control condition in which the relative positions of probe, flanker, control line, IC_1 and IC_2 take on different values. The stimuli for these two conditions were generated so as to reverse the predictions for the relative positions of the PSEs, with respect to conditions 1 and 2, as indicated in Fig. 12. It is important to note that the stimuli in these conditions only differed from those in conditions 1 and 2 by adding a translational component to the stimulus.

In conclusion, the four stimulus conditions of Experiment 1 were created by crossing two variables: *control-line position* (control line coinciding or not with the IC subspace: conditions 1, 3 vs. 2, 4), and *IC_2 position* (IC_2 "above" or "below" the probe: conditions 1, 2 vs. 3, 4—see Figs. 11 and 12). We predict that *control-line position* will interact with *IC_2 position* in determining the PSEs set by the observers.

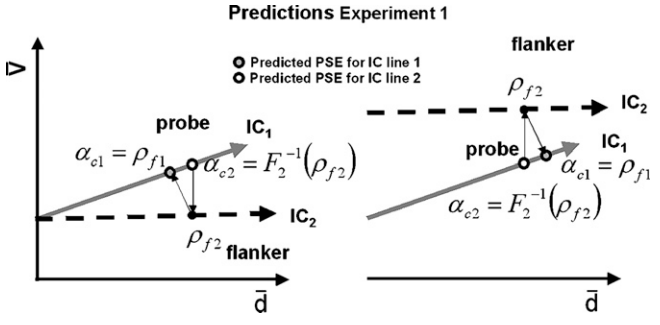

Fig. 10. Experiment 1. Left panel: In both conditions 1 and 2, the probe can take on the values belonging to the control line. In this case, such values coincide with the IC_1 line. However, we expect that the PSE will take on different values in the two conditions: The solid gray dot and the open white dot indicate the predicted PSEs for conditions 1 and 2, respectively. Right panel: The solid gray dot and the open white dot indicate the predicted PSEs for condition 3 and 4, respectively.

2.1. Method

2.1.1. Apparatus

Stereoscopic stimuli were displayed on a haploscope consisting of two CRT monitors (.22 mm dot pitch) located on swing arms pivoting directly beneath the observer's eyes. This equipment follows the design described by Backus et al. (1999). Anti-aliasing and spatial calibrating procedures allow spatial precision of dot location greater than hyper-acuity levels (see Backus et al., 1999). Each monitor

[1] Remember that larger ρ values correspond to smaller distances to the observer.

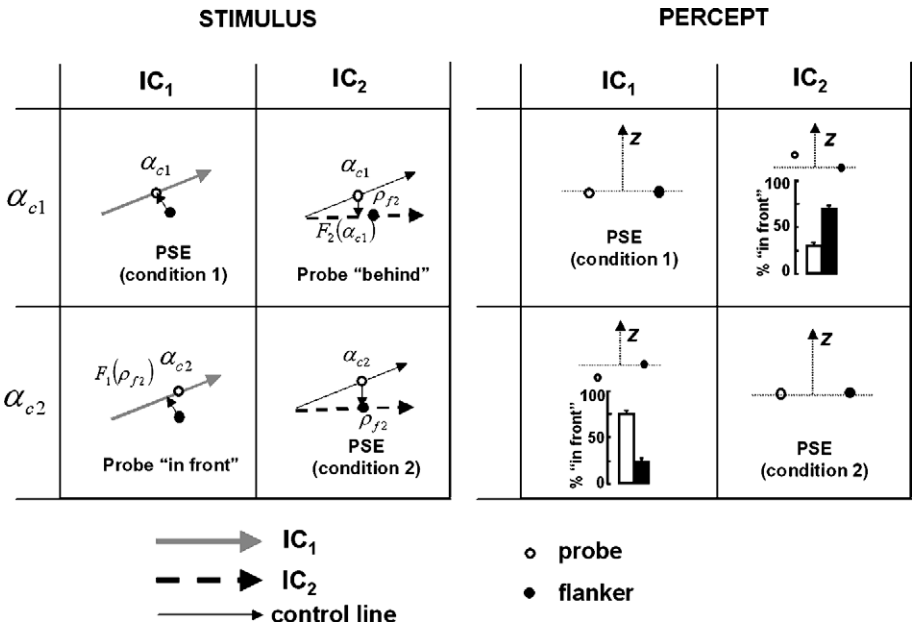

Fig. 11. Schematic representation of the predictions of the IC model for Conditions 1 and 2. The "Stimulus" panel represents the relative positions of the probe, flanker and IC line in the disparity, velocity space. The "Percept" panel represents the corresponding perceptual solution. IC_1 and IC_2 represent the conditions 1 and 2, respectively. α_{c1} and α_{c2} represent the PSEs for the probe. ρ_{f2} is the projection of the PSE on the IC_2 line. $F_2(\alpha_{c1})$ is the back-projection of α_{c1} on IC_2. $F_1(\alpha_{c2})$ is the projection of the flanker on IC_1. The corresponding cells on the "Percept" represent the depth-map on the probe and flanker. In the main diagonal, the probe's PSE and the flanker are at the same z-depth distance from the observer. The cells off the main diagonal represent the relative depth relations between the probe and flanker, if the PSE of the probe in condition 1 would be compared to the flanker in condition 2, and the other way around. The histograms in these two cells represent the results of the second analysis described in Section 2.2. The black bar represent the percentage of judgments "probe in front of the flanker"; the white bar is the complement: "flanker in front of the probe". Vertical bars represent 1 SE.

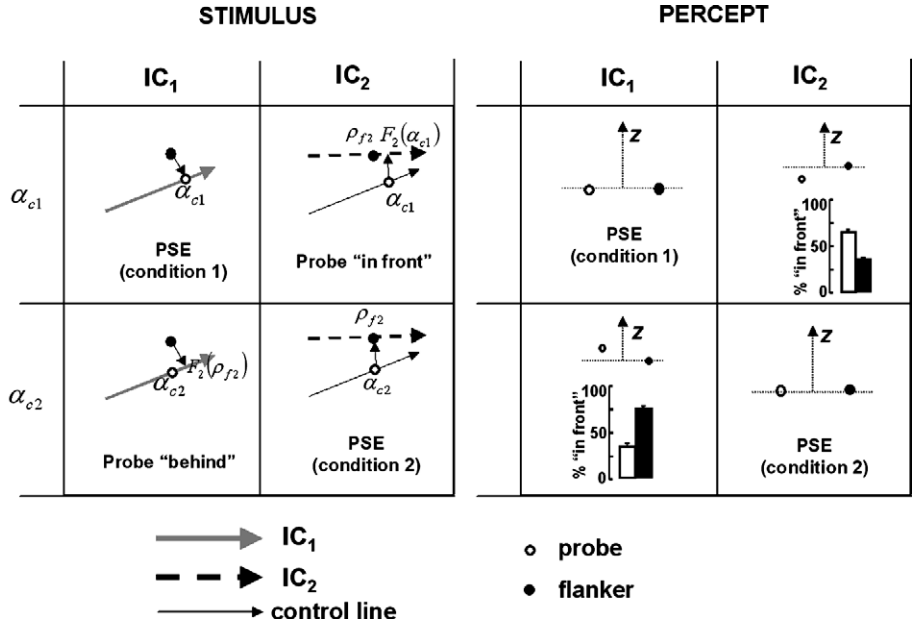

Fig. 12. Schematic representation of the predictions of the IC model for conditions 3 and 4. Apart from this difference, the structure of Fig. 11 is identical to that of Fig. 12.

was seen in a mirror through one eye. Head position was fixed with a chin-and-forehead locating apparatus. The actual distance from each eye to the corresponding monitor was 95 cm. The eyes' vergence was directly manipulated by physically moving the monitors on their swing arms. Since the monitors and mirrors pivot rigidly about the eye's axis of rotation, the retinal images always remain the same for all positions of the two CRT monitors. Thus, changes in eye position can be dissociated from changes in retinal images.

2.1.2. Stimuli

The stimuli were 200 high luminance dots displayed against a low luminance background. Probe and flanker were rendered with different colors and were positioned near the center of the display on a straight line passing through the center of the screen. The distance between each of these two dots and the center was equal to 0.4 cm (or 0.25° of visual angle). The left-right placement of the probe-flanker pair was determined at random on each trial. The x, y orientation of the virtual line comprising probe and flanker dots was determined at random on each trial within a range of ±5° from the horizontal.

The surrounding dots, termed "context dots", simulated the projection of a cloud of dots randomly placed within a spherical volume with a ray of 4 cm (2.4° of visual angle). A cylindrical region with axis orthogonal to the image plane and a ray of 1.33 cm (0.8°) centered in the middle of the screen did not contain any context dots, so as to make the probe and flankers dots clearly visible (see Figs. 6 and 7). The direction of rotation was determined at random in each trial.

The overall stimulus subtended about 4.8° of visual angle. Depth information was provided by the simultaneous presence of disparity and velocity cues. Disparities where calculated so as to simulate a 3-D structure viewed at 1000 cm from the observer. The vergence angle was calculated accordingly for each subject, by taking into account her or his inter-ocular distance. The 2D velocities were computed by simulating a rotation of the dots about a vertical axis. In Experiment 1, the axis of rotation was tangent to the far side of the sphere containing the context dots.

The velocities and disparities of the flanker dots were kept constant throughout the experiment. In condition 1, they were equal to 206 arcsec and 616 arcsec/s, respectively; in condition 2, they were equal to 206 arcsec and 1046 arcsec/s, respectively.

The velocity and disparity of the probe dot varied so as to maintain a constant ratio equal to $\frac{v_p}{d_p} = 1.16$. In condition 1, the context dots specified IC_1 (Fig. 9a, left panel), since their rotation about the vertical axis was equal to 16° within a time window of 500 ms. In condition 2, the context dots rotated about the vertical axis by 4° and, therefore, specified IC_2 (Fig. 6a, right panel).[2]

2.1.2.1. Procedure. The observers' task was to judge which of the two central dots appeared to be closer (along the line-of-sight). On each trial, the stimulus was shown for 500 ms. For the successive 500 ms the screen was dark. This sequence was repeated four times, or until the subject provided his/her response.

2.1.2.2. Design. All the independent variables were studied within observers; in each block, the subject viewed only stimuli belonging to one of the four experimental conditions. The disparity and velocity of the probe were varied by the participants through a 4-interleaved 2AFC staircase procedure. The ratio between the velocity and disparity values that the probe could take on was kept constant throughout the experiment.

2.1.2.3. Observers. Seven observers with normal or corrected-to-normal vision participated in Experiment 1. Six observers were naïve to the purpose of the experiment, and one was the first author.

2.2. Results

Fig. 13 shows the mean PSEs in the four experimental conditions, together with their standard errors. The reported velocity and disparity values are scaled by the estimated standard deviations of their measurement errors. As estimates of the standard deviations of the measurement errors for the two signals, we used the values reported by Lappin and Craft (2000) in their hyper-acuity tasks. The disparity values and the velocity values (computed within a temporal window of 150 ms) were thus scaled by 10 arcsec and 20 arcsec, respectively. Note, however, that several investigations have shown that the internal noise is not constant (McKee, Verghese, & Farell, 2005; Farell, Li, & McKee, 2004a, Farell, Li, & McKee, 2004b; Verghese, McKee, & Grzywacz, 2000; McKee, Welch, Taylor, & Bowne, 1990), being typically proportional to sensory magnitude and depending, for example, on retinal eccentricity and/or the separation between reference and test lines. The estimates of internal noise that we used here, therefore, might be very imprecise. Nevertheless, they proved adequate for the present purposes.

Psychometric functions were fitted by a cumulative Gaussian function, for each participant and each condition, using *psignifit* version 2.5.6 (see http://bootstrap-software.org/psignifit/), a software package which implements the maximum-likelihood method described by Wichmann and Hill (2001a). Confidence intervals were found by the 'BCa' bootstrap method implemented by psignifit, based on 50 simulations (see Wichmann & Hill, 2001b).

The average PSEs are shown again in Fig. 14, and were analyzed by a mixed-effects model with the participants as the grouping factor (R Development Core Team, 2005). The `R` routine `lmer` was used in the work reported here. The interaction between the factors *control-line position* (conditions 1, 3 vs. 2, 4), and *IC_2 position* (conditions 1, 2 vs. 3, 4) was significant [$F(1, 25) = 6.843$, $p < 0.05$], confirming the hypothesis that the surrounding dots affect the local depth order interpretation of the probe and flanker dots.

[2] Note that in this experimental condition the angular rotation is very small (only 4°) and, therefore, the difference between the velocities of the probe and flanker dots is very small. In such circumstances, therefore, we expect that the observer will rely on binocular disparities, for the most part (since the IC line is almost parallel to the disparity axis), and will perform the task by disparity matching.

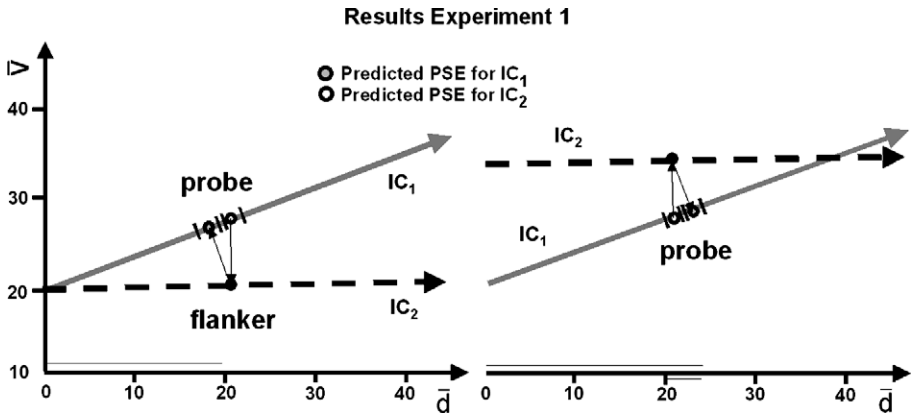

Fig. 13. Left panel: Predicted and observed PSEs for the Experiment 1. The predicted PSEs for conditions 1 and 2 are represented by a solid gray dot and an open white dot, respectively. Right panel: The predicted PSEs for conditions 3 and 4 are represented by a solid gray dot and an open white dot, respectively. The two small bars indicate ±1 SE for the empirical PSEs. The unit of measurement are the scaled disparity (horizontal axis) and velocity values (vertical axis) – see text for details.

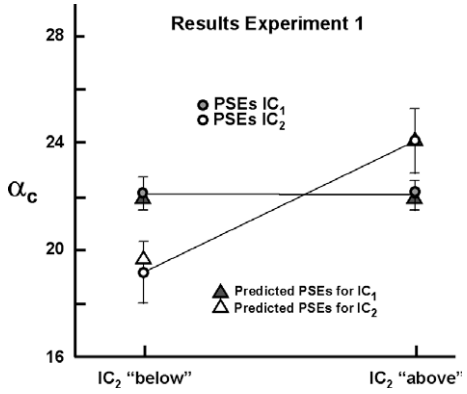

Fig. 14. Average and predicted PSEs of Experiment 1 represented in terms of α_c, for the four experimental conditions (IC_1, IC_2) × (IC_2 "below", IC_2 "above").

In condition 1, when the angular rotation of 16^0 for the surrounding dots defines the IC_1 line, α_{c_1} is the predicted PSE that makes the probe appear at the same distance from the observer as the flanker. What does it happen if we consider *the same point on the control line*, but now with the angular rotation of 14^0 which defines the IC_2 line? According to our hypothesis, α_{c_1} would be considered as a random perturbation of the "true" value $\mathcal{F}_2(\alpha_{c_1})$ (see Fig. 11, left panel, top right). Since $\mathcal{F}_2(\alpha_{c_1})$ corresponds to a ρ score smaller than ρ_{f_2} associated to the flanker, we expect that, *by changing the angular rotation of the context dots*, the probe should appear *farther away* from the observer than the flanker. In fact, if we consider the judgments associated to ρ values smaller than $\mathcal{F}_2(\alpha_{c_1})$, the 76.2% (SE = 4.6) of the times observers judged the probe accordingly (see Fig. 11, right panel).

The opposite result is expected when considering the point α_{c_2} with the angular rotation of 16^0, which defines the IC_1 line. In this case, $\mathcal{F}_2(\rho_{f_2})$ is smaller than α_{c_2}. As a consequence, we expect that the probe should appear as *closer* to the observer than the flanker. For ρ values smaller than $\mathcal{F}_2(\rho_{f_2})$, in fact, observers judged the probe as farther away than the flanker in only the 30.5% (SE = 4.5) of the cases.

In conclusion, with a certain angular velocity for the context dots, the v, d values of the PSE make the probe appear to be at the same distance from the observer as the flanker. When the angular rotation of the context dots changes then, the same v, d values make the probe to appear either closer to the observer (Fig. 11, left panel, bottom left) or farther away than the flanker (Fig. 11, left panel, top right).

For the conditions 3 and 4 represented in Fig. 12, we expected the opposite pattern of results. In condition (α_{c1}, IC_2), in fact, we found that the percentage of judgments "probe closer to the observer than the flanker" was 66.1% (SE = 2.1). In condition (α_{c2}, IC_1), the percentage of judgments "probe closer to the observer than the flanker" was equal to 32.5% (SE = 3.0). This interaction (represented by the histograms of Fig. 12) was statistically significant [$F(1, 18) = 115.843$, $p < 0.0001$]. We can thus conclude that a change in the 3-D angular velocity of the context dots is sufficient to reverse the sign of the perceived depth order relations of the probe and flanker dots.

3. Experiment 2

If the 3-D angular-velocity context dots affects the depth order interpretation of probe and flanker, then also the precision of the depth-alignment task should depend on the "reliability" by which the IC line is estimated. In terms of the PCA, the reliability of the IC line can be described as the ratio between the first eigenvalue and the sum of all eigenvalues associated to the sample variance-covariance matrix.

To test this second hypothesis, we created displays similar to those of the first experiment with the following differences: (i) the ratio between the disparity and velocity values was kept constant for probe and flanker dots (see Fig. 15, upper left panel), (ii) the context dots could

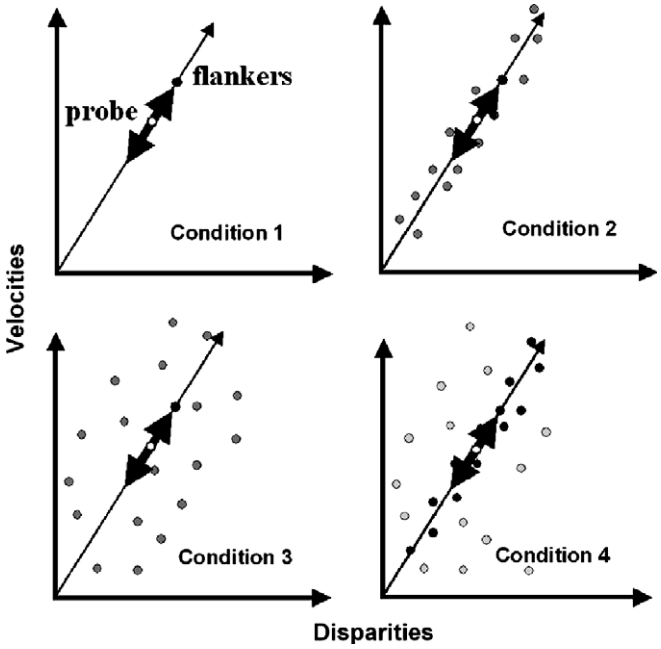

Fig. 15. Schematic representation of the four conditions of Experiment 2. Top left panel: Condition 1. Only the probe (open white dot) and the two flankers (solid black dot) are shown. Top right panel: Condition 2. The probe and the flanker dots are presented together with the context dots (solid gray dots). All dots in the display are rigidly connected and, thus, they line up on the IC line. Bottom left panel: Condition 3. Each of the surrounding dots oscillates about the same (vertical) axis at a different speed (non rigid motion). As a consequence, the surrounding dots do not lie on the IC line. Bottom right panel: Condition 4. Half the dots rotate at different speeds about the same (vertical) axis and half are rigidly connected to the probe and flanker dots.

undergo rigid or non-rigid motion, and (iii) a second flanker was added in the center of the coordinate system.

The first change was made to allow observers to perform the depth-alignment task by matching the disparity and velocity values of the probe with those of the flankers. We asked whether the context affects the precision of performance, even when the task can be performed by following this simple strategy. The second change was introduced to investigate the effects of the "reliability" of the IC line on the variance of observers' judgments. If the IC line does not affect the perceptual interpretation of the d, v values of probe and flankers, then the properties of the context dots should not affect the accuracy of judgments. According to our hypothesis, conversely, the precision in the depth-alignment task should be modulated by the reliability of the IC line. In other words, as the reliability of the IC line increases, the precision of observers' judgments should improve.

Note that, in Fig. 15, the two flankers are represented with a single point in the signal space. This happens because they were both simulated with the same velocity and disparity values—i.e., they were positioned at the same distance, along the z-axis, with respect to the plane containing the axis of rotation.

Four experimental conditions were created: (1) no context-dots (Fig. 15, panel 1); (2) context-dots rigidly connected to the probe and flankers—this produces an unperturbed IC line (Fig. 15, panel 2); (3) nonrigid context dots—this corresponds to a noisy IC line (Fig. 15, panel 3); (4) a surrounding cloud of dots in which half of the dots are rigidly connected to the probe and the flankers, whereas the other half is nonrigid random noise—-this produces a partly perturbed IC line (Fig. 15, panel 4). In the fourth condition the amount of noise perturbation added to the stimulus was twice the amount of noise perturbation added to the stimulus in condition 2.

Note, again, that any model that integrates probe or flanker dots into an estimated rigid 3-D structure (not just the one which characterizes the information in the display as we have proposed) will predict the same pattern of results as the ones that we have put forward before. In order to run an ideal-observer analysis, however, it is necessary to propose a specific hypothesis about the mechanism by which the visual system integrates different features into a rigid 3-D shape. According to our proposal, prior to integrating the disparity and velocity values into a composite score, their values must be scaled by their estimated measurement errors. In the simulation described below, such estimates were provided by the standard deviations of the disparity and velocity judgments for the hyper-acuity tasks reported by Lappin and Craft (2000). Disparity values were thus scaled by 10 arcsec; velocity values, computed within a temporal window of 150 ms, were scaled by 20 arcsec.

In each trial of the simulation, the orientation of the IC line was calculated by PCA on the scaled disparity and velocity values perturbed by standardized Gaussian noise. Due to the variability of the estimation of the IC line, across trials, probe and flankers did not lie on the IC line. We hypothesize that, in these circumstances, the system chooses, as the most likely value, their orthogonal projections on the IC line. In different trials of the simulation, therefore, the probe was orthogonally projected on different positions of the IC line. The variance of the probe's projections on the IC line was hence used to estimate the variance of the psychometric functions in the four experimental conditions. Fig. 16 shows the results of this simulation.

The purpose of Experiment 2 was to establish whether the variability of observers' performance in the depth-alignment task is affected by the context dots, as suggested by the ideal-observer analysis described above.

3.1. Method

3.1.1. Observers

Six observers with normal or corrected-to-normal vision participated in Experiment 2. Five observers were naïve to the purpose of the experiment, and one was the first author.

3.1.2. Stimuli

In the second experiment, a second flanker dot was added and positioned at the center of the screen. The surrounding dots, termed "context dots", simulated the projection of a cloud of dots randomly placed within a

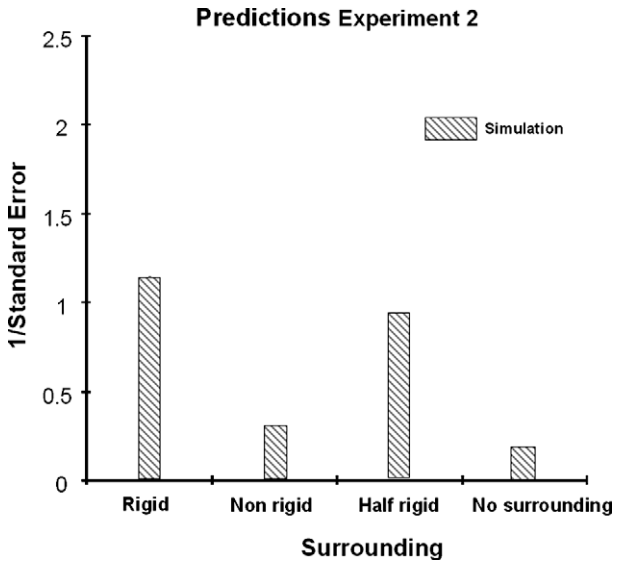

Fig. 16. Predictions for Experiment 2. In each experimental condition, the IC line was calculated by PCA on the scaled disparity and velocity values perturbed by standardized gaussian noise. The variability $\hat{\sigma}^2$ of the orthogonal projection of the probe on the IC line across trials was computed. The vertical bars show $1/\hat{\sigma}$, which provides the prediction for the precision of observers' judgments.

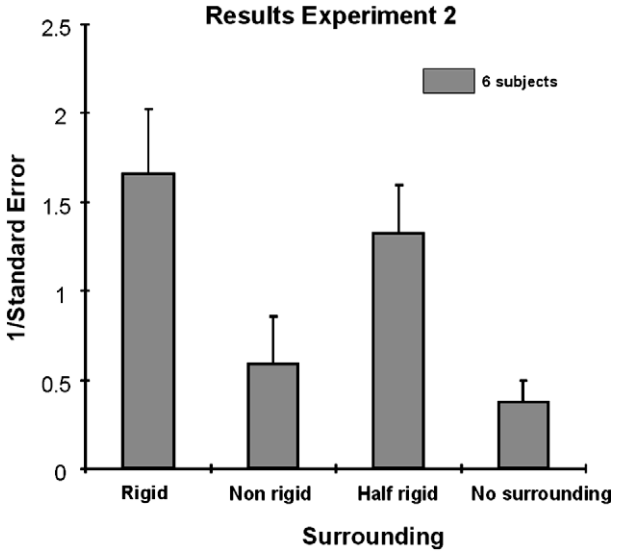

Fig. 17. Results of Experiment 2. The vertical bars represent the precision of observers' judgments $(1/\hat{\sigma})$ in the four experimental conditions of Experiment 2. Vertical bars represent 1 standard error. Note how the pattern of results closely matches the predictions shown in Fig. 16.

spherical volume with a ray of 4 cm (2.4° of visual angle). A cylindrical region with axis orthogonal to the image plane and a ray of 1.33 cm (0.8°) centered in the middle of the screen did not contain any context dots, so as to make the probe and flankers dots clearly visible.

In Experiment 2, the axis of rotation passed through the center of the sphere containing the context dots. The velocities and disparities of the flanker dots were kept constant and were equal to 90 arcsec/s and 280 arcsec, respectively. Velocity and disparity of the probe varied in such a manner as to maintain a constant ratio equal to $\frac{v_p}{d_p} = 1.38$.

In condition 1, only the probe and the flankers were presented (no context dots). Condition 2 simulated a rigid rotation of the structure comprising the context dots, the probe and the flankers. In condition 3, the depth used to calculate the disparity values of the context dots was chosen randomly, within a ray of 4 cm from the center of the coordinate system. As a result, the amounts of angular rotation and disparity of the dots were not consistent with those of a rigid structure. In condition 4, half the dots were perturbed as in condition 3, while the other half simulated a rigid rotation.

In condition 3, a rotational velocity component of 430 arcsec/s, consistent with the direction of rotation, was added to all dots in the display; otherwise, the stimulus was identical to condition 1. In condition 4, the same velocity component was added to the flanker dot; otherwise, the stimulus was identical to condition 2.

3.2. Results

Psychometric functions were fitted to the individual data, as in Experiment 1. Fig. 17 shows the means of the inverse standard errors estimated through the staircase procedure in the four conditions. These data were analyzed by a mixed-effect model having participant as the grouping factor. The effect of the experimental manipulation was significant [$F(3, 15) = 12.063$, $p < 0.001$]. The mean of the condition "no context dots" was significantly different from both the means of the conditions "rigid context" [$t(15) = 5.210$, $p < 0.001$] and "half-rigid context" [$t(15) = 3.876$, $p < 0.002$]. The conditions "no context dots" and "non-rigid context" did not differ significantly from each other [$t(15) = 0.885$, *n.s.*]. Note how the estimated means match the predictions of the simulation represented on Fig. 16.

A second analysis compared the mean PSEs in the four conditions. As expected, the average PSEs did not differ significantly across conditions [$F(3,15) = 1.747$, *n.s.*].

4. General discussion

The goal of the present investigation was to study the visual interpretation of a small number of dots which are not rigidly connected to an otherwise rigid, binocularly-viewed, rotating structure. Behind this question lies the bigger problem of whether the visual analysis incorporates the mutual constraints relating different depth-cues present in a retinal projection.

Observers were asked to judge the relative depth of one probe, and one or two flanker dots. In Experiment 1, we found that the perceived depth order relations were influenced by the amount of rigid rotation of a surrounding cloud of dots. In Experiment 2, we found that the properties of the surrounding dots affected the precision of observers' judgments: With only probe and flanker dots in isolation, the precision of observers' judgments was

much lower than when probe and flanker dots were surrounded by a rigidly-connected cloud of dots. When the surrounding dots underwent a non-rigid rotation, perceptual performance was the same as with the isolated probe and flanker.

Our stimulus displays were devised so as to mimic a generic condition in which disparity, d, and velocity, v, measurements are corrupted by noise. Since the d, v signals are affected by measurement error, the deviation from rigidity of a single feature point cannot be interpreted unambiguously: Such deviation could be due to actual non-rigidity, or to measurement's noise. With the present investigation, we tried to determine whether such an ambiguity is solved perceptually by relying on the constraint relating the disparity and velocity values of the other dots in the stimulus display.

So far, Richards (1985) has provided the most important theoretical analysis on the mutual constraint relating disparity and velocity signals. He proposed a model for recovering 3-D rigid shape through the integration of disparity and motion information. Richards' analysis start from the consideration that disparity and velocity signals, by themselves, are not sufficient for recovering 3-D shape correctly. The interpretation of binocular disparities requires the knowledge of the fixation distance in units of interpupillary separation. The correct configuration and disposition of a 3-D object, moreover, cannot be recovered from retinal velocities, since measurements of retinal accelerations are also needed (Koenderink et al., 1991). Visual processing of optic acceleration is very poor and highly sensitive to measurement noise (e.g., Hogervorst & Eagle, 2000).[3]

Richards (1985) solves these problems by analyzing two or more *binocular* views of a set of feature points. The minimal inputs to the algorithm are (i) the ratio between the relative disparities of a feature point with respect to a reference point in two instants of time, and (ii) the ratio between the horizontal separation of the same point with respect to a reference point in two instants of time. Even though two binocular views of two points are sufficient for 3-D Euclidean reconstruction, they are not sufficient for a rigidity test. For solving such a problem, two views of three points, or three views of two points, are needed.

Even though Richards' algorithm is sound from a computer-vision standpoint, it may not be appropriate as a model for human visual processing. Two major problems can be pointed out.

First, Richards' model has not been devised to deal with image noise, when discriminating between rigid and non-rigid structures. Richards' algorithm is deterministic, and does not show the graceful degradation which is proper of perceptual performance (e.g., Caudek & Rubin, 2001).

Second, Richards' algorithm does not account for the specific properties of perceptual performance. In Experiment 1, we found that the perceived depth order relations among the three central dots changed according to the motion properties of the surrounding dots, even though the 2-D signals of the central dots did not vary. Such a result is incompatible with Richards' model, since the 3-D interpretation of the three central dots should not be influenced by the surrounding dots. In Experiment 2, we found that perceptual performance was affected by the number of dots in the stimulus display. The output of Richards' algorithms, conversely, does not vary as a function of number of points (after meeting the minimal conditions). In Experiment 2, moreover, we found that perceptual performance deteriorated when half of the surrounding dots underwent non-rigid motion. Such manipulation, however, would not affect the output of Richards' model, since the non-rigid dots would be discarded after the rigidity test.

Other empirical evidence, beside the present experiments, reveals that perceived 3-D structure from stereo-motion is not veridical. Tassinari, Domini, and Caudek (in press), for example, found greater perceptual biases for combined stereo-motion cues than for each cue in isolation. In that study, observers provided depth judgments from (i) stereo-only, (ii) motion-only, and (iii) stereo-motion cues. The observers' task was to judge whether the apparent elongation of a cylinder was more or less deep than an Apparently Circular Cylinder (ACC, Johnston, Cumming, & Landy, 1994). In general, observers reported a larger depth elongation for the stereo-motion stimuli than for the single-cue stimuli (see also Domini et al., 2006). However, the depth elongation of the single-cue stimuli was veridical and, thus, the combined-cue judgments were less accurate than the single-cue judgments.

In the present paper, we propose an alternative model for integrating stereo and motion information. Such alternative model does not show the above limitation and, therefore, is more psychologically plausible than Richards' algorithm. It should be noted, however, that also other models, which integrate disparity and motion information into a rigid structure, would make predictions similar to ours. Nevertheless, our model seems to be particularly suited to account for perceptual performance in the present experiments, because has been devised to maximize accuracy in the derivation of 3-D affine structure.

In this respect, it is worth mentioning here the work of Fernández, Watson, and Qian (2002), who also proposed that disparity and motion information are processed by the visual system so as to take into account their mutual constraints. Fernandez et al. proposed a physiologically plausible model for computing 3-D shape from disparity and motion information, in which MT neurons tuned to

[3] Hogervorst and Eagle asked participants to match the dihedral angle of a hinged plane (probe) defined by multiple-depth cues to one defined by motion only. Their results show effects of perspective information and also effects of image accelerations. However, in most of their experimental conditions, metric judgments revealed large biases. Even though these biases are well accounted for by the Bayesian model proposed by the authors, they also confirm the poor sensitivity of the visual system to higher-order temporal information of the optic flow.

both velocity and disparity are connected to each other laterally to form modulatory interactions. Such a model computes the correct 3-D shape under a wide range of conditions, but can also reproduce the structure-from-motion illusions involving coaxial cylinders.

In summary, the results of our experiments reveal systematic distortions in the perceptual analysis, when a small amount of non-rigidity is added to an otherwise rigid stimulus display. Such systematic distortions concerns the perceived depth order relations, and can be attributed to "errors of measurement" for the velocity and disparity values of the non-rigid dots. These biases in the perceived depth order relations are well predicted by the IC model, which is a normative model for the optimal recovery of the 3-D affine structure.

The present investigation does not show that the rigidity constraint affects the visual interpretation of motion information alone. Rather, our results show that perceptual performance is affected by the linear relation between disparity and velocity signals, when both depth-cues are present and the distal object is, in fact, rigid.

References

Aloimonos, J. (1988). Shape from texture. *Biological Cybernetics, 58*, 345–360.

Backus, B. T., Banks, M. S., van Ee, R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research, 39*, 1143–1170.

Brooks, M. J., & Horn, B. K. P. (1985). Shape and source from shading. *Proceedings of the Joint Conference in Artificial Intelligence*, 932–936.

Brown, C., Ballard, D., Rainero, E. (1983). Constraint propagation in shape from shading — Proc. Image Understanding Workshop, 1983.

Bülthoff, H. H. (1991). Shape from X: Psychophysics and computation. In Michael S. Landy & J. Anthony Movshon (Eds.), *Computational models of visual processing* (pp. 305–330). Cambridge, MA, US: The MIT Press.

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.

Caudek, C., Domini, F., & Di Luca, M. (2002). Short-term temporal recruitment in structure from motion. *Vision Research, 42*, 1213–1223.

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

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

Cutting, J. E., Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth IN Perception of space and motion. Handbook of perception and cognition (2nd ed.). Epstein, W. and Rogers, S. Eds. San Diego, CA, US, Academic Press, Inc., 69–117.

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

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.

Domini, F., Caudek, C., & Richman, S. (1998). Distortions of depth-order relations and parallelism in structure from motion. *Perception & Psychophysics, 60*, 1164–1174.

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

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

Domini, F., Caudek, C., & Tassinari, H. (2006). Stereo and motion information are not independently processed by the visual system. *Vision Research, 46*, 1707–1723.

Farell, B., Li, S., & McKee, S. P. (2004a). Coarse scales, fine scales, and their interactions in stereo vision. *Journal of Vision, 4*, 488–499.

Farell, B., Li, S., & McKee, S. P. (2004b). Disparity increment thresholds for gratings. *Journal of Vision, 4*, 156–168.

Fernández, J. M., Watson, B., & Qian, N. (2002). Computing relief structure from motion with a distributed velocity and disparity representation. *Vision Research, 42*, 883–898.

Gårding, J. (1993). Direct estimation of shape from texture. *IEEE Trans Pattern Analysis and Machine Intelligence, 15*, 1202–1208.

Grimson, W. E. L. (1981). A computer implementation of a theory of human stereo vision, Philosophical transactions of the royal society of london. *Series B, Biological sciences, v292*, 217–253.

Hildreth, E. C., Grzywacz, N. M., & Adelson, E. H. (1990). The perceptual buildup of three-dimensional structure from motion. *Perception & Psychophysics, 48*, 19–36.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optional cue combination. *Journal of Vision, 4*, 1–3.

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.

Horn, B. K. P. (1975). *The psychology of computer vision, chap. obtaining shape from shading information*. New York: McGraw-Hill.

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research, 34*, 2259–2275.

Kham, K., & Blake, R. (2000). Depth capture by kinetic depth and by stereopsis. *Perception, 29*, 211–220.

Koenderink, J. J., & van Doorn, A. (1991). Affine structure from motion. *Journal of the Optical Society of America A, 8*, 377–385.

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.

Lappin, J. S., & Craft, W. D. (2000). Foundations of spatial vision: From retinal images to perceived shapes. *Psychological Review, 107*, 6–38.

Liter, J. C., & Braunstein, M. L. (1998). The relationship of vertical and horizontal velocity gradients in the perception of shape, rotation, and rigidity. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1257–1272.

Liu, Y., & Huang, T. S. (1988). Estimation of rigid body motion using straight line correspondences. *Computer Vision, Graphics and Image Processing, 43*, 37–52.

Marr, D. (1982). *Vision*. W.H. Freeman Co.

McKee, S. P., Verghese, P., & Farell, B. (2005). Stereo sensitivity depends on stereo matching. *Journal of Vision, 5*, 783–792.

McKee, S. P., Welch, L., Taylor, D. G., & Bowne, S. F. (1990). Finding the common bond: Stereoacuity and the other hyperacuities. *Vision Research, 30*, 879–891.

Pentland, A. P. (1982). Finding the illuminant direction. *Journal of the Optical Society of America, 72*, 448–455.

Pentland, A. (1989). Shape information from shading: a theory about human perception. *Spatial Vision, 4*, 165–182.

Richards, W. (1985). Structure from stereo and motion. *Journal of the Optical Society of America A: Optics and Image Science, 2*, 343–349.

Tassinari, H., Domini, F., Caudek, C. (in press). The Intrinsic Constraint Model for Stereo-Motion Integration, submitted to *Perception*.

Tittle, J. S., & Braunstein, M. L. (1993). Recovery of 3-D shape from binocular disparity and structure from motion. *Perception & Psychophysics, 54*, 157–169.

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

Todd, J. T., & Norman, J. F. (2003). The visual perception of 3-D shape from multiple cues: are observers capable of perceiving metric structure? *Perception & Psychophysics, 65*, 31–47.

Todd, J. T., Tittle, J. S., & Norman, J. F. (1995). Distortions of three-dimensional space in the perceptual analysis of motion and stereo. *Perception, 24*, 75–86.

Tsai, T. S., & Huang, R. Y. (1984). Uniqueness and estimation of 3D motion parameters of rigid objects with curved surfaces. *IEEE Trans Pattern Analysis and Machine Intelligence, 6*, 13–27.

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

Ullman, S. (1984). Maximizing rigidity: The incremental recovery of 3-D structure from rigid and nonrigid motion. *Perception, 13*, 255–274.

Verghese, P., McKee, S. P., & Grzywacz, N. M. (2000). Stimulus configuration determines the detectability of motion signals in noise. *Journal of the Optical Society of America A, 17*, 1525–1534.

Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function. I. Fitting, sampling and goodness-of-fit. *Perception & Psychophysics, 63*, 1293–1313.

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function. II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics, 63*, 1314–1329.

Witkin (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence, 7*, 17–45.