Stereo and motion information are not independently processed by the visual system

Article *in* Vision Research · June 2006

DOI: 10.1016/j.visres.2005.11.018 · Source: PubMed

3 authors, including:

Corrado Caudek
University of Florence
90 PUBLICATIONS **655** CITATIONS

SEE PROFILE

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 46 (2006) 1707–1723
Vision Research
www.elsevier.com/locate/visres

Stereo and motion information are not independently processed by the visual system ☆

Fulvio Domini [a,*], Corrado Caudek [b], Hadley Tassinari [c]

[a] Department of Cognitive and Linguistic Sciences, Brown University, USA
[b] Department of Psychology, University of Florence, Italy
[c] Department of Psychology, New York University, USA

Received 15 July 2004; received in revised form 10 November 2005

Abstract

Many visual tasks are carried out by using multiple sources of sensory information to estimate environmental properties. In this paper, we present a model for how the visual system combines disparity and velocity information. We propose that, in a first stage of processing, the best possible estimate of the affine structure is obtained by computing a composite score from the disparity and velocity signals. In a second stage, a maximum likelihood Euclidean interpretation is assigned to the recovered affine structure. In two experiments, we show that human performance is consistent with the predictions of our model. The present results are also discussed in the framework of another theoretical approach of the depth cue combination process termed Modified Weak Fusion.

Keywords: Vision; Cue integration; Disparity; Motion

1. Introduction

The question of how the visual system integrates the information provided by several depth cues is central for vision research. Here, we present a model for how the human visual system combines disparity and velocity information. The model provides a depth interpretation for a sub-space defined by the covariation of the two signals. In two experiments, we show that human performance is consistent with the predictions of the model. We also compare the predictions of our model to those of another theoretical approach, the modified weak fusion model (MWF) (Landy, Maloney, Johnston, & Young, 1995), and discuss the validity of each approach as a model for human perception of three-dimensional (3D) shape from multiple cues to depth.

1.1. Rationale

Binocular disparity and image motion are important sources of information for the perceptual recovery of 3D shape from two-dimensional (2D) retinal projections (Julesz, 1971; Wallach & O'Connell, 1953). Theoretical work has demonstrated that either of these two sources of depth information, when presented in isolation, permits the veridical recovery of the 3D Euclidean structure of the projected objects. The previous statement, however, requires careful qualification.

Let us consider the case of motion information first. In principle, an unbiased metric interpretation of motion information is possible if 2nd-order temporal information of the optic flow (i.e. acceleration) is used (Hildreth, 1984; Koenderink & van Doorn, 1991; Longuet-Higgins & Prazdny, 1980). Even though such information is available when at least three views of a moving object are

☆ We wish to thank Michael Landy, Suzanne P. McKee and James M. Hillis for useful comments on a preliminary version of this manuscript. We also thank two anonymous reviewers for several constructive suggestions. Research supported by Grant NSF BCS 0345763.
* Corresponding author. Tel.: +1 401 863 1356; fax: +1 401 863 2255.
E-mail address: Fulvio_Domini@brown.edu (F. Domini).

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.11.018

available, empirical research has shown that human observers are not sensitive to acceleration information (Hogervorst & Eagle, 2000; Liter, Braunstein, & Hoffman, 1993; Todd & Bressan, 1990). A large number of psychophysical results has shown that, to solve the structure-from-motion problem, human observers make use of only the 1st-order temporal information (Domini & Caudek, 1999, 2003a, 2003b)[1]. This usage of only a subset of the available information has an important consequence: The same 2-frame motion sequence (providing only 1st-order temporal information) can be produced by infinite combinations of local orientation and 3D rotation. The 1st-order temporal information thus identifies the affine structure of the projected shape, but not its "relief", since the angular rotation parameter cannot be recovered from the velocity field (Koenderink & van Doorn, 1991).

Now, let us consider the case of disparity information. As is the case with motion information, the veridical recovery of metric depth is possible. However, this requires usage of vertical disparities and accurate measurements of extra-retinal signals, such as vergence and version of the eyes (Backus, Banks, van Ee, & Crowell, 1999; Longuet-Higgins, 1981; see also Rogers & Bradshaw, 1995). When the size of the display is small, as for the present investigation, vertical disparities are negligible and, under these circumstances, are not used by human observers (Rogers & Bradshaw, 1995)[2]. In these circumstances, for central gaze (fixating straight ahead), the interpretation of horizontal disparities requires an estimate of the distance to the fixation point (e.g. Garding, Porrill, Mayhew, & Frisby, 1995; Landy & Brenner, 2001). What is important to note here, again, is that such information is not provided by optical information. Within some stimulus settings (and for viewing distances smaller than 70 cm), observers use vergence information to scale image data (Mon-Williams, Tresilian, & Roberts, 2000). In other stimulus settings (and larger viewing distances), however, observers do not scale the disparities with an unbiased estimate of distance derived from extra-visual cues (e.g. Glennerster et al., 1996). Moreover, distortions of 3D shape have also been reported in natural full-cues settings (Lind, Bingham, & Forsell, 2002; Todd, Tittle, & Norman, 1995).

These results might be interpreted as follows: for viewing distances over one meter, extra-visual signals might influence the perceptual solution, even though they do not guarantee a veridical interpretation of disparity information. If information about viewing distance is missing, the 1st-order stereo information specifies only the affine structure of the distal object. Analogous to this is the case of an isolated velocity field, where 1st-order temporal information defines the projected shape only up to an affine transformation.

We can thus say that there are two missing parameters necessary for a metric interpretation of motion and disparity information. These parameters are the component of rotation about an axis contained in the image plane (ω) for motion information and the fixation distance (μ) for disparity information.

In the present work, we choose to develop and empirically test a model which ignores the contribution of extra-retinal signals. In such circumstances, the information that is available to the visual system, provided either by an isolated velocity field or by an isolated disparity field, does not suffice for the recovery of full metric structure. However, it has been shown that, in principle, disparity and parallax could be combined to recover veridical information about 3D shape without recourse to additional scaling parameters. This possibility has been called *promotion* by Landy et al. (1995). Richards (1985), in fact, has shown that, given two binocular views of a moving object, there is a family of shapes (parameterized by viewing distance) which is consistent with binocular disparities, and a family of shapes (parameterized by the rotation angle relating the two views) which is consistent with the image velocities. These two one-parameter families of solutions, however, have only one shape in common, which corresponds to the "correct" Euclidean interpretation of the image signals.

Some empirical findings support this model (e.g. Bradshaw, Parton, & Eagle, 1998; Brenner & van Damme, 1998; Johnston, Cumming, & Landy, 1994; Richards & Lieberman, 1985), although other findings suggest that disparity and motion may remain relatively independent in the recovery of 3D shape (e.g. Brenner & Landy, 1999; Brenner & van Damme, 1998; Tittle, Todd, Perotti, & Norman, 1995). Very often, however, distortions of perceived 3D shape have been reported, even for judgments of real objects in fully illuminated natural environments (e.g. Bradshaw, Parton, & Glennerster, 2000; Cuijpers, Kappers, & Koenderink, 2000; Hecht, van Doorn, & Koenderink, 1999; Koenderink, van Doorn, Kappers, & Todd, 2002; Loomis, Da Silva, Fujita, & Fukusima, 1992; Loomis & Philbeck, 1999; Norman, Lappin, & Norman, 1996;

[1] One important aspect of the perception of perceived 3D structure from motion is the distinction between the viewing conditions of a non-moving, passive observers and those of an active moving observer. Wexler, Panerai, Lamouret, and Droulez (2001), for example, found that observers' estimates of surface tilt (i.e., the direction of its normal relative to the frontoparallel plane) improved in the active condition, when motion parallax was due to the observer's head movements, relative to the passive case, when the observer remained still. Such results indicate that active observers' perception of 3D structure depends on extra-visual self-motion information. The results of Wexler et al., however, are not relevant for the present investigation that is concerned solely with the case of stationary viewing conditions.

[2] By using a nulling task (i.e., by asking observers to adjust the pattern of horizontal disparities until the test surface appears to be flat and lie in a frontal plane), Rogers and Bradshaw (1995) found that observers indeed used vergence information for a veridical scaling of disparities. Such a result, however, has not been replicated with other tasks involving metric judgments of 3D shape. We note, moreover, that Glennerster, Rogers, and Bradshaw (1996) showed that the effectiveness of vergence information for scaling disparity depends on the task: They found a higher degree of constancy for matching tasks (in which only the ratio of distances to the two test surfaces needs to be known) than for depth-to-height judgements (requiring an estimate of the viewing distance).

Todd & Norman, 2003; but see also Buckley & Frisby, 1993).

The purpose of the present investigation is to develop and test a model for the perceptual derivation of 3D shape from stereo and motion information. Such model is based on the information that is actually available to the human visual system, and does not assume unbiased estimates of the missing parameters μ and ω.

The proposed model rests on the assumption that the visual system recovers veridical information of the *affine*, but not of the Euclidean, structure of the distal 3D shape (see Todd & Bressan, 1990). Following Koenderink and van Doorn (1991), we will distinguish two phases in the depth-recovery process. In the first phase, a 3D representation of the distal shape will be constructed up to an arbitrary affine transformation; in the second phase, a maximal likelihood decision rule will uniquely determine the rigid structure of the 3D shape.

2. Disparity and motion

Let us consider an observer fixating a point **F** belonging to an object which rotates about the horizontal axis (Fig. 1). Each identifiable feature $\mathbf{P}_i$ on the surface of the object produces a binocular disparity d_i and a velocity v_i. Both quantities (measured relative to the fixation point **F**) depend on the depth difference Δz_i of the point $\mathbf{P}_i$ with respect to the fixation point **F**. Let us consider first disparity information.

Let relative disparity d_i be defined 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, respectively. If the visual angle subtended by the object is small (<8°), then d_i can be approximated by:

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

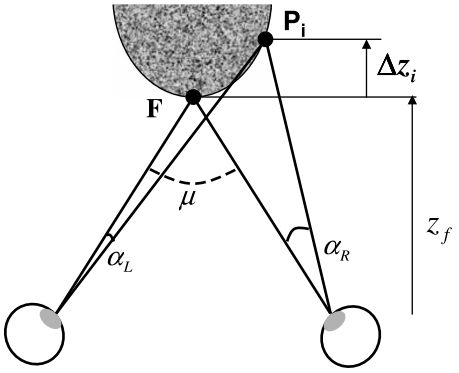

Fig. 1. Schematic illustration of an observer fixating a horizontally oriented hemi-cylinder at a viewing distance z_f. Point **F** represents the fixation point; the vergence angle is represented by μ; α_L and α_R are the angles formed by point **P** relative to the fixation point. Δz is the front-to-back depth of the cylinder (along the line of sight).

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: $d_i \approx \mu z_i + \varepsilon_d$.

Similarly, let the relative velocity v_i be defined as the difference between the visual angles α_2 and α_1 subtended by $\mathbf{P}_i$ and **F** at the cyclopean vantage point in two different instants of time t_1 and t_2 (Fig. 2). If the object rotates by a small amount ω during the time interval $[t_1, t_2]$, then v_i can be approximated by the following equation:

$$v_i \approx \omega z_i + \varepsilon_v, \tag{2}$$

where ε_v is the noise of velocity measurements.

2.1. Dimensionality reduction

In the first stage of processing, the goal of the Intrinsic Constraint (IC) model is to obtain the best possible estimate of the affine structure of the distal depth z_i. We have seen above that the affine structure can be estimated from each signal in isolation. Because of measurement noise, however, the estimates based on the isolated signals are subject to error. The IC model postulates that, in order to reduce measurement noise, the visual system pools together different depth signals so as to obtain a composite estimate that is maximally associated with the distal depth map (i.e., more strongly associated with the distal 3D shape than any of the estimates obtained from the isolated signals).

A composite estimate having the desired properties can be found by means of a dimensionality reduction technique such as, for example, a Principal Component (PC) analysis. For reasons explained below, the PC analysis must be carried out on the scaled signals $(\bar{d}_i, \bar{v}_i)$, that is, on the disparity d_i and velocity v_i signals divided by the standard deviations of their measurement errors[3]:

$$\bar{d}_i \approx \bar{\mu} z_i + \varepsilon, \tag{3}$$

$$\bar{v}_i \approx \bar{\omega} z_i + \varepsilon, \tag{4}$$

where $\bar{\mu}$ and $\bar{\omega}$ are the scaled vergence angle and angular velocity, and $\varepsilon \sim \mathcal{N}(0,1)$ (see Fig. 3).

For purpose of demonstration, we ran a simulation showing that the correlation between the scores on the first principal component and the distal z depth magnitudes is higher than the correlation between each separate signal and the distal z values. In the simulation, the z values were drawn from a uniform distribution varying in the same range as the z depth values of Experiment 1. The disparity and velocity signals were

[3] The measurement errors of the disparity and velocity signals indicate the internal noise of the visual system that, in the process of measurement, corrupts these image signals.

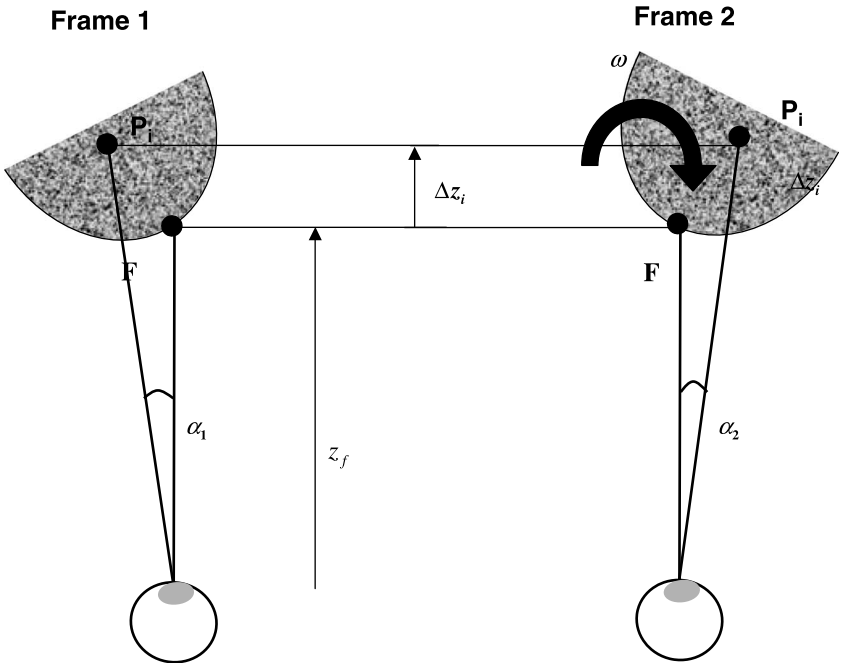

Fig. 2. Schematic illustration of a generic surface rotating through the angle ω. z_f represents the viewing distance; **F** represents the fixation point; α_1 and α_2 represent the angles formed by point $\mathbf{P}_i$ relative to the fixation point, in two moments of the surface's rotation (frame 1 and 2). The angular velocity of the projection of the point $\mathbf{P}_i$ is defined as the difference between the angles α_1 and α_2.

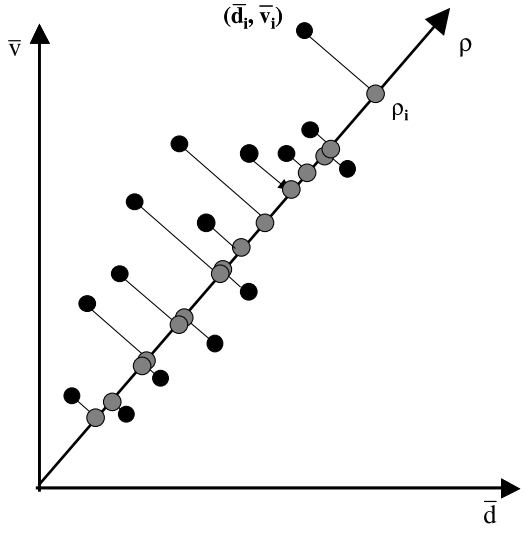

Fig. 3. Intrinsic constraint line plotted in a velocity–disparity space. The norm of the vector identified by the largest velocity–disparity pair is termed ρ.

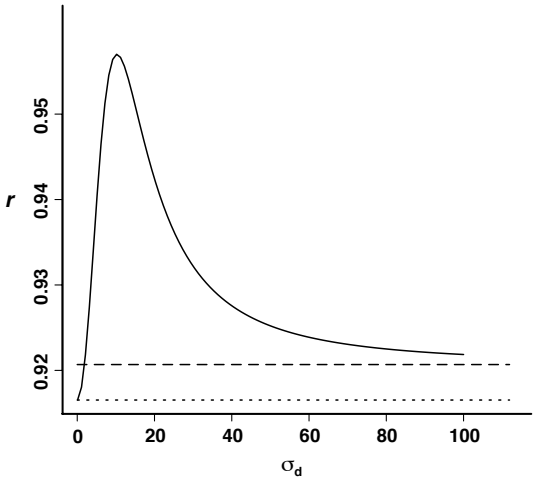

Fig. 4. Results of a simulation computing the correlation between the simulated z values and the scores on the first principal component calculated from the scaled disparity and velocity signals generated by the stimulus used in Experiment 1. In the simulation, the disparity signals are scaled by different amounts and the velocity signals are scaled by the estimated measurement error reported by Lappin and Craft (2000). The correlation between the PC_1 scores and z is maximum when both signals are scaled by their measurement errors. The horizontal lines show the correlation between the two single signals and the z values.

expressed in terms of displacement (by assuming a temporal window of 150 ms). A random component of Gaussian noise with standard deviation σ_{ε_v} or σ_{ε_d} was added to the velocity and disparity signals, respectively. In the simulation, the velocity signals were scaled by $\sigma_{\varepsilon_v} = 20$ arcsec and the disparity signals were scaled by a variable amount. Fig. 4 reports the correlation between the simulated z values and the scores on the first principal component, by letting the disparity signals being scaled by different amounts. Note that the correlation between the PC_1 scores and z is maximum when both the signals are scaled by their measurement error (or by any two numbers preserving the ratio between σ_{ε_v} and σ_{ε_d}). The figure shows, moreover, the correlations

between each single signal (velocity or disparity) and the z values (the two horizontal lines[4]). Note that both these two correlations are lower than the correlation between the PC_1 scores and z.

In conclusion, an important consequence of this preprocessing of the stimulus information is that the projections onto the first principal component (PC_1) have a higher correlation with the scaled depth z_i than any of the signals considered in isolation. In other words, the PC_1 scores provide the best estimate of the affine structure of the projected object, given the available disparity and velocity signals[5].

2.2. Depth interpretation

The goal of the second stage of processing, according to the IC model, is to estimate (up to a scale factor) the Euclidean depth map z_i from the scores ρ_i on PC_1. If the viewing parameters μ and ω are unknown and if they cannot be estimated in an unbiased fashion, then an unbiased estimate of the Euclidean 3D structure is not possible. In these circumstances, we hypothesize that the visual system estimates the Euclidean depth map through the same procedure described by Domini and Caudek (1999, 2003a, 2003b) for the perceptual analysis of the optic flow.

Let us examine such a procedure by considering Eq. (2) above. Note that the term on the left side of the equation (v_i) can be produced by an infinite number of different pairs of depth (z_i) and angular rotations (ω) values. For the case of slant perception, Domini and Caudek (2003b) proposed that the visual system chooses, among these infinite z_i and ω pairs, the one that maximizes the likelihood function $p(v_i|z_i)$:

$$\hat{z}_i = \text{argmax}_{z_i} p(v_i|z_i), \tag{5}$$

where:

$$p(v_i|z_i) \propto \int_{\omega} p(v_i|z_i,\omega)p(\omega)\,\mathrm{d}\omega. \tag{6}$$

In Fig. 5, $p(v_i|z_i, \omega)$ is coded by luminance for each z_i and ω pair. The graph in the top part of the figure shows the marginal distribution $p(v_i|z_i)$. Note that $p(v_i|z_i)$ has a maximum: The value z_i corresponding to the maximum of the marginal distribution $p(v_i|z_i)$ is the ML estimate $\hat{z}_i$. In other words, Domini and Caudek postulated that human structure-from-motion can be described as the best-guess about the 3D property given the (ambiguous) 2D information. In a series of papers (Caudek and Domini, 1998; Caudek and Rubin, 2001; Di Luca et al., 2004; Domini & Caudek, 1999, 2003a, 2003b, 2003c; Domini et al., 1997; Domini et al., 1998; Domini et al., 2003; Domini et al., 1998; Domini et al., 2002), they provided many empirical data in support to their hypothesis.

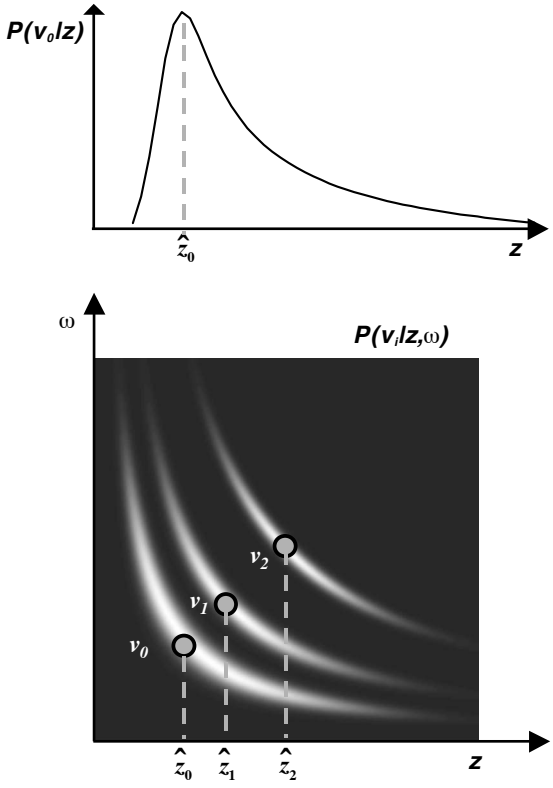

Fig. 5. In Fig. 2, the relative velocity v_i is defined as the difference between the visual angles α_2 and α_1 subtended by $\mathbf{P}_i$ and $\mathbf{F}$ at the cyclopean vantage point in two different instants of time t_1 and t_2. If the object rotates by a small amount ω during this time interval, then v_i can be approximated by $v_{ij} \approx \omega z_j + \varepsilon_{v_i}$. In this figure, three velocity magnitudes, v_0, v_1, v_2, are represented in the ω, z space. The probability $p(v_i|z_j,\omega)$ is coded by luminance for each z_j and ω_i pair. For the velocity v_0, the graph at the top shows the marginal distribution $p(v_0|z_i)$. The maximum of this marginal distribution is the ML estimate $\hat{z}_j$ from the velocity v_0.

Here, we will analyze the composite stereo-motion score ρ_i in a similar manner as we have done above for v_i. In general, if enough points are measured, the measurement noise on the first eigenvector $\mathbf{e}_1$ (i.e. the orientation of the first principal component in the signal space) is negligible. Then

$$\rho_i = z_i\sqrt{\bar{\mu}^2 + \bar{\omega}^2} + \varepsilon, \tag{7}$$

where $\varepsilon \sim \mathcal{N}(0,1)$. The above equation can be re-written as

$$\rho_i = z_i R + \varepsilon, \tag{8}$$

where $R = \sqrt{\bar{\mu}^2 + \bar{\omega}^2}$ can be termed *global rotation*.

Note that Eq. (8) is structurally identical to Eqs. (3) and (4) and, in fact, is equal to Eq. (4) when stereo information specifies a flat surface, and is equal to Eq. (3) when motion information specifies a flat surface. The important point is that ρ_i can be produced by an infinite number of different

[4] The difference between the values of the two correlations represented by the horizontal lines in Fig. 4 is the product of random fluctuations in the simulation.

[5] If the affine structure could be recovered perfectly, the correlation coefficient would be equal to one. A smaller value of the correlation coefficient indicates the presence of noise in the specification of the relative-depth relations of the distal object.

pairs of z_i and R values. The IC model proposes that z_i can be estimated from ρ_i through the same ML procedure described above for the structure-from-motion case. The estimated scaled-depth thus becomes:

$$\hat{z}_i = \text{argmax}_{z_i} p(\rho_i|z_i). \tag{9}$$

Again, the likelihood function $p(\rho_i|z_i)$ can be calculated by integrating over the unknown parameter R:

$$p(\rho_i|z_i) \propto \int_R p(\rho_i|z_i, R)p(R)\,\mathrm{d}R. \tag{10}$$

The Euclidean solution found in this way is not, in general, veridical. But it is not, in any way, arbitrary: it represents *the best possible guess* about the depth z_i, given the input signals and the assumptions that we have introduced in the interpretation process. It remains a goal of the psychophysical research to establish whether the constraints that we have presently used are adequate to describe the actual functioning of the human perceptual system.

The exact function $p(\rho_i|z_i)$ can be described only if the a priori distribution of the global rotation R is known. Even without knowing $p(R)$, however, specific predictions can be made. Our main assumption is that the perceptual solution depends only on ρ_i (i.e. the scores on PC_1), regardless of the orientation of the IC line (i.e. the eigenvector $\mathbf{e}_1$). From this it follows that:

HP$_1$ constant values of ρ produce the same perceived depth value $\hat{z}$, even if ρ is associated with different $\mathbf{e}_1$ magnitudes;
HP$_2$ the increase of ρ is associated with a monotonic increase of $\hat{z}$;
HP$_3$ for constant values of ρ, the variability of observers' performance is not affected by $\mathbf{e}_1$.

In the next section we will describe how the previous three hypotheses can be tested.

3. Empirical test of the IC model

The experimental procedure described in this section allows to test the IC and MWF models in conditions in which they make different predictions. Let the stimuli be random-dot cylinders. In a preliminary phase of Experiment 1, by fixing the velocity of rotation ω, we found the viewing distance z_f for which the two single-cue stimuli (velocity-only and disparity-only) evoked the same amount of perceived depth. It is well known, in fact, that observers either overestimate or underestimate the depth of disparity-defined cylinders, according to whether the viewing distance is shorter or larger than 1 meter (Johnston, 1991). If we indicate with z_0 the simulated front-to-back depth of the cylinder scaled by the viewing distance z_f, then the relative disparity (d_0) and relative velocity signals (v_0) are:

$$d_0 \approx \mu_0 z_0 + \varepsilon_v, \tag{11}$$

$$v_0 \approx \omega_0 z_0 + \varepsilon_d, \tag{12}$$

where ω_0 is the angular rotation in a 150 ms temporal window, and μ_0 is the vergence angle equating perceived depth from stereo and motion signals, as indicated above.

As has been shown in the previous section, in the first stage of processing, the IC model recovers the scores ρ_i by computing the first principal component on the scaled velocity and motion signals. When only the single-cue signals are present, the PC_1 scores basically coincide with the scaled signals (if the noise relative to the measurement of the first eigenvector is negligible; see Fig. 6):

$$\rho_{d_0} = \bar{d}_0 \approx \bar{\mu}_0 z_0 + \varepsilon, \tag{13}$$

$$\rho_{v_0} = \bar{v}_0 \approx \bar{\omega}_0 z_0 + \varepsilon. \tag{14}$$

A first hypothesis of the IC model is that the perceived depth magnitudes z_i are determined by ρ, regardless of the orientation of the eigenvector $\mathbf{e}_1$. Remember that, in the preliminary phase of Experiment 1, μ_0 is chosen in such

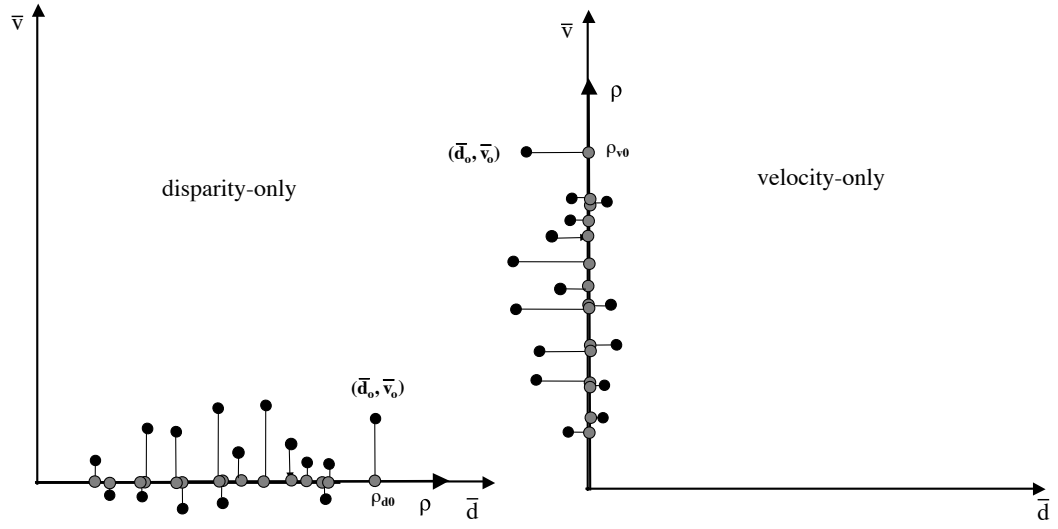

Fig. 6. Schematic representation of two single-cue displays in the space defined by the scaled velocity and disparity signals. In both cases, the eigenvector associated with the first principal component coincides with the axis of the signal specifying the 3D structure of the distal object.

a way that the two single-cue stimuli give rise to the same amount of perceived depth $\hat{z}$. In the disparity–velocity space, then, the vector ρ_{v_0} (associated with the disparity-only signals) will be equal to the vector ρ_{d_0} (associated with the velocity-only signals):

$$\rho_{d_0} = \rho_{v_0} = \rho_0. \tag{15}$$

From the previous equality, it follows that

$$\bar{\mu}_0 = \bar{\omega}_0 = R_0. \tag{16}$$

After this preliminary stage, in the second part of Experiment 1, a two-interval forced-choice scheme was used in which a sequence of two stimuli was presented on each trial. For one stimulus, the simulated cylinder depth was fixed; for the other, it was varied through a staircase procedure. The vergence angle μ_0 and the amount of rotation ω_0 took on the same values as in the calibration phase of the experiment. In each trial of the experiment, two successive computer renderings of cylinders (which were defined by either disparity information, motion information, or both) were shown. Observers were asked to decide which of the two simulated cylinders appeared to be more elongated in depth. Three interleaved staircases were run for each of the two conditions described below.

- ***Condition 1***. In each 2AFC trial, the combined-cues stimulus simulated a constant depth amount. Within three separate staircases, the simulated depth of: (i) a disparity–velocity stimulus (with consistent simulated depth magnitudes from each cue), (ii) a disparity-only stimulus, and (iii) a velocity-only stimulus was varied.
- ***Condition 2***. Within three separate staircases, the simulated depth of the combined disparity–velocity stimulus was varied, while maintaining constant the simulated depth of: (i) a disparity–velocity stimulus (with consistent simulated depth magnitudes from each cue), (ii) a disparity-only stimulus, and (iii) a velocity-only stimulus.

Note that, if we consider one trial at a time, there is no difference between conditions 1 and 2 (within the staircase, in condition 1, the single-cues stimuli were varied and the combined-cues stimulus was kept fixed; in condition 2, the role of the single-cue and combined-cues stimuli was reversed). According to the MWF model, therefore, this experimental manipulation should not affect perceptual performance. The IC model, instead, makes different predictions for the two conditions, concerning both the PSEs and the variability of observers' judgments, as indicated in the following sections.

3.1. Points of subjective equality

According to the first hypothesis stated above, the perceived depth magnitudes $\hat{z}_i$ depend on ρ alone, regardless of the values of $\mathbf{e}_1$.

- ***Condition 1***. First, let us consider the combined-cues stimulus that, in each 2AFC trial of the staircase, simulates a constant amount of depth. The magnitude of ρ_{c_0} for this stimulus is:

$$\rho_{c_0} = z_0\sqrt{\bar{\mu}_0^2 + \bar{\omega}_0^2}. \tag{17}$$

Since $\bar{\mu}_0 = \bar{\omega}_0 = R_0$, it follows that $\rho_{c_0} = z_0 R_0\sqrt{2}$.
Now, let us consider the other stimulus that is displayed in each 2AFC trial. Within each of the three staircases, this stimulus is defined either by disparity and velocity (c), by disparity-only (d), or by motion-only information (v). Within each of the three staircases, the amount of depth simulated by such stimuli is varied. For the three stimuli manipulated by the observers, the magnitudes of ρ are:

$$\rho_{\mathrm{c}} = z_{\mathrm{c}}\sqrt{\bar{\mu}_0^2 + \bar{\omega}_0^2} = R_0 z_{\mathrm{c}}\sqrt{2}, \tag{18}$$

$$\rho_{\mathrm{d}} = \bar{\mu}_0 z_{\mathrm{d}} = R_0 z_{\mathrm{d}}, \tag{19}$$

$$\rho_{\mathrm{v}} = \bar{\omega}_0 z_{\mathrm{v}} = R_0 z_{\mathrm{v}}. \tag{20}$$

Within the staircase, the simulated depth of the above three stimuli was varied until the two displays shown on each trial were perceived as having the same depth. The IC model predicts that the PSE is reached when, within each staircase, observes equate the ρ magnitude of the stimulus that is being manipulated to the ρ_{c_0} value of the fixed stimulus: $\rho_{c_0} = \rho_{\mathrm{c}}$, $\rho_{c_0} = \rho_{\mathrm{d}}$ and $\rho_{c_0} = \rho_{\mathrm{v}}$. Since $\rho_{c_0} = z_0 R_0\sqrt{2}$ (Eq. (17)), the previous three equalities will be met when $z_{\mathrm{c}} = z_0$, $z_{\mathrm{d}} = \sqrt{2}z_0$ and $z_{\mathrm{v}} = \sqrt{2}z_0$.
For the MWF model, conversely, the PSE should be found when $z_{\mathrm{c}} = z_{\mathrm{d}} = z_{\mathrm{v}} = z_0$, since the main assumption of this model is that the depth estimates are unbiased. The previous statement is based on the assumption that no other cues are used in the depth estimation process.
- ***Condition 2***. Relative to condition 1, the role of the fixed and varying stimuli is reversed. Within each of the three staircases, the stimuli simulating a fixed amount of depth are defined either by disparity-only (d), or by motion-only (v), or by disparity and velocity (c) information. For these stimuli, ρ takes on the following values:

$$\rho_{c_0} = z_0\sqrt{\bar{\mu}_0^2 + \bar{\omega}_0^2} = R_0 z_0\sqrt{2}, \tag{21}$$

$$\rho_{d_0} = \bar{\mu}_0 z_0 = R_0 z_0, \tag{22}$$

$$\rho_{v_0} = \bar{\omega}_0 z_0 = R_0 z_0. \tag{23}$$

Within each of the staircases, the depth of the combined-cues stimulus is varied. For the combined-cues stimulus, the magnitude of ρ is:

$$\rho_{\mathrm{c}} = z_{\mathrm{c}}\sqrt{\bar{\mu}_0^2 + \bar{\omega}_0^2} = R_0 z_{\mathrm{c}}\sqrt{2}. \tag{24}$$

Also in this case, the IC model predicts that the PSE is reached when, within each staircase, the ρ magnitudes of the varying and fixed stimuli are equated: $\rho_c = \rho_{c_0}$, $\rho_c = \rho_{d_0}$, $\rho_c = \rho_{v0}$. For the staircase comprising the two combined-cues stimuli, the condition $\rho_c = \rho_{c_0}$ is met when $z_c = z_0$. For the two staircases in which the single-cue stimuli simulate a fixed amount of depth, the conditions $\rho_c = \rho_{d_0}$ and $\rho_c = \rho_{v_0}$ are met when $z_c = \frac{z_0}{\sqrt{2}}$.
For the MWF model, because of the hypothesis of unbiased depth estimates, the PSE should again be found when $z_c = z_d = z_v = z_0$.

3.2. Variability of judgements

In proposing a new model for depth cue combination, we must not only show that the IC model predicts empirical results incompatible with the MWF model, but must also address the previous empirical findings in favor of MWF. The strongest support to the MWF model comes from the findings of "optimal" cue combination: For two or more cues simultaneously present, observers have been found to reduce the variance of their 3D estimates below that which could be achieved from either cue alone (Alais & Burr, 2004; Ernst & Banks, 2002; Gepshtein & Banks, 2003; Hillis, Watt, Landy, & Banks, 2004; Knill & Saunders, 2003; Landy & Kojima, 2001).

To test optimal cue combination in the case of stereo and motion cues, MWF proponents often choose a certain baseline value of simulated depth z_0 and then measure the discrimination accuracy for this depth value in three conditions: disparity-only (d), motion-only (v) and stereo plus motion (c). The goal is to estimate the standard deviations σ_{z_d}, σ_{z_v} and σ_{z_c} of observers' judgments. If the scene variable of interest (i.e. depth) is estimated in a statistically optimal fashion, then σ_{z_c} must be related to σ_{z_d} and σ_{z_v} according to the following equation (Clark & Yuille, 1990):

$$\sigma_{z_c}^2 = \frac{\sigma_{z_d}^2 \sigma_{z_v}^2}{\sigma_{z_d}^2 + \sigma_{z_v}^2}. \tag{25}$$

For the IC model, discrimination accuracy depends on the just-noticeable-difference ($\Delta\rho_{JND}$) of the scores on the IC line, and is independent of the orientation angle of the IC line. In other words, if ρ is kept constant, then according to the IC model, discrimination accuracy should be the same for the single-cue and combined-cues stimuli, since these stimuli differ only in the orientation of the IC line. A difference between the variances of depth judgements σ_z^2, conversely, is expected across the two experimental conditions, as indicated below.

- ***Condition 1***. In condition 1, the combined-cues stimulus was kept constant within the staircase. The simulated depth of the disparity-only, velocity-only or disparity–velocity stimuli were varied so as to match the perceived depth of the two displays presented in each trial. Let us express the variation of ρ as a function of Δz:

$$\Delta\rho_c = \sqrt{\bar{\omega}_0^2 + \bar{\mu}_0^2}\Delta z_c = R_0 \Delta z_c \sqrt{2}, \tag{26}$$

$$\Delta\rho_d = \bar{\mu}_0 \Delta z_d = R_0 \Delta z_d, \tag{27}$$

$$\Delta\rho_v = \bar{\omega}_0 \Delta z_v = R_0 \Delta z_v. \tag{28}$$

According to the IC model, two stimuli can be discriminated when $\Delta\rho = \Delta\rho_{JND}$:

$$\Delta\rho_c = \Delta\rho_d = \Delta\rho_v = \Delta\rho_{JND} \tag{29}$$

and thus

$$\Delta z_d = \frac{\Delta\rho_{JND}}{R_0}, \tag{30}$$

$$\Delta z_v = \frac{\Delta\rho_{JND}}{R_0}, \tag{31}$$

$$\Delta z_c = \frac{\Delta\rho_{JND}}{R_0\sqrt{2}}. \tag{32}$$

It therefore follows that

$$\Delta z_c^2 = \frac{\Delta z_d^2 \Delta z_v^2}{\Delta z_d^2 + \Delta z_v^2}. \tag{33}$$

If Δz_d, Δz_v, Δz_c are considered to be measures of the JNDs in the disparity-only, velocity-only and combined-cues conditions, then Eq. (33) is equivalent to the optimal combination rule postulated by the MWF model Eq. (25). In condition 1, therefore, both MWF and IC models make the same predictions about the variability of observers' judgments.

- ***Condition 2***. In condition 2, the two models generate different predictions concerning the variability of observers' judgments. Remember that, in our model the variability of observers' judgments, for any given ρ, depends only on the just-noticeable-difference $\Delta\rho_{JND}$ of the scores on the IC line. In condition 2, the simulated depth of the combined-cues stimulus was varied until it appeared as deep as the fixed (disparity-only, velocity-only, or disparity–velocity) stimulus. We can assume that the same ρ value is associated with the fixed stimuli, since in the preliminary phase of the experiment the viewing parameters were chosen in such a manner so as to equate the amount of depth evoked by the single-cue stimuli. The observers' task, in terms of the IC model, is to vary the ρ value of the combined-cues stimulus, to match its perceived depth to that of the three fixed stimuli. For each of the three comparisons (combined-cues vs. motion-only, combined-cues vs. disparity-only, combined-cues vs. combined-cues), the IC model predicts that the variability of performance depends only on the JND for the varying combined-cues stimulus. The three psychometric functions of condition 2, therefore, are expected to have the same variability.
The MWF model, conversely, yields the same predictions as for condition 1 since, for that model, it

is inconsequential whether observes vary the single-cue or the combined-cues stimuli.

4. General methods

4.1. 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 (Fig. 7). This equipment follows the same design as 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 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 armatures. Thus changes in eye position can be dissociated from changes in retinal images.

4.2. Stimuli

Hemi-cylinders that were either stretched or compressed along the line of sight were simulated with random dot stereograms on each CRT screen. Cylinders were comprised of 200 dots; the 2D arrangement of dots was randomly defined with each stimulus presentation. The dots were projected on the two eyes such that horizontal disparities specified a protruding elliptical cross section of varying depth, parallel to the line of sight. Cylinders were oriented horizontally and rotated 20° back and forth (up 10°, down 10°) about the horizontal axis perpendicular to the line of sight.

Simulated cylinder height and width were constant and equal to 50 mm, while disparity and motion information specified a cross-section (Δz) that was less deep, more deep, or equal to the height of the cylinder. Vertical subtense was sufficiently small (less than 6° of visual angle) so that vertical disparities were not a reliable cue to distance (Cumming, Johnston, & Parker, 1991; Rogers & Bradshaw, 1993).

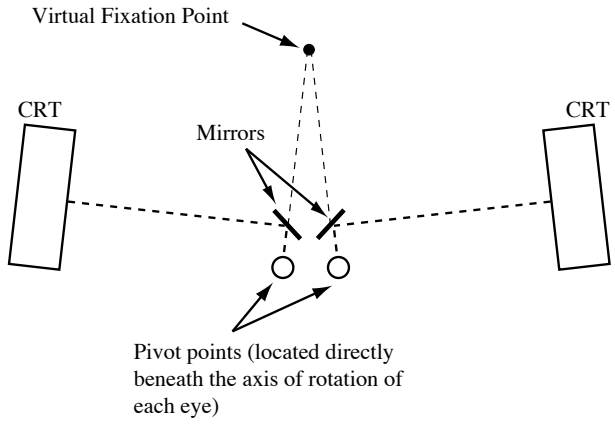

Fig. 7. Schematic illustration of the stereo viewing apparatus (haploscope). CRT monitors project through mirrors to each eye individually and are moved about pivots which coincide with each eye's axis of rotation. The virtual fixation point is therefore specified by vergence angle, which was adjusted to the proper setting for each subject's IOD and viewing distance.

5. Experiment 1

The purpose of this experiment was to establish whether perceptual performance complies with the predictions of the IC model described in Section 3.

5.1. Method

5.1.1. Observers

Four observers with normal or corrected-to-normal vision participated in Experiment 1. Two observers were naïve to the purpose of the experiment, and two were the second and the third authors.

5.1.2. Procedure

In a pre-test phase, the viewing distance and the angle of rotation were chosen for each observer so that, for a constant simulated depth $z_0 = 25$ mm, each cue in isolation gave rise to the same amount of perceived depth. Then, observers' performance was measured in each of the 3 (disparity-only, velocity-only, disparity–velocity stimuli) $\times 2$ (condition 1 and 2) comparisons described above. A 4-interleaved 2AFC staircase procedure was used.

5.2. Results

The six conditions of Experiment 1 were analyzed by means of a generalized linear mixed model with multivariate normal random effects, using penalized quasi-likelihood (R Development Core Team, 2005). The `R` routine `glmmPQL` (Venables and Ripley, 2002; Section 10.4) was used in the work reported here. The resulting psychometric functions are shown in Fig. 8, together with their 95% CI.

5.2.1. Variance of observers' settings

In *condition 1*, the estimated $\hat{\sigma}$ are the following (the 95% CIs are reported in parenthesis): $\hat{\sigma}_v = 7.992(6.662–9.987)$; $\hat{\sigma}_d = 6.872(5.637–8.802)$; $\hat{\sigma}_c = 3.221(2.651–4.102)$. By applying the optimal combination rule, the single-cue data predict a value of 5.211 (4.302–6.603) for the combined-cues condition. Note that the value that we have found is well predicted by the theoretical value computed from Eq. (25) (it is actually slightly smaller). The methodology of condition 1, therefore, is adequate to reproduce the finding of optimal combination, often reported in support of the MWF model. As we have noted above, in condition 1, both the MWF and the IC model make the same prediction.

One possible explanation of the fact that the variance of observers' judgments in the combined condition is smaller than that predicted by Eq. (25) may be formulated in the following way. According to both the IC and the MWF models, the optimal combination rule applies when the motion-only, disparity-only, and motion–disparity simulate the same amount of depth. In the present experiment, instead, the PSE was reached when the amount of simulated depth was about 40% larger for the single-cue than for

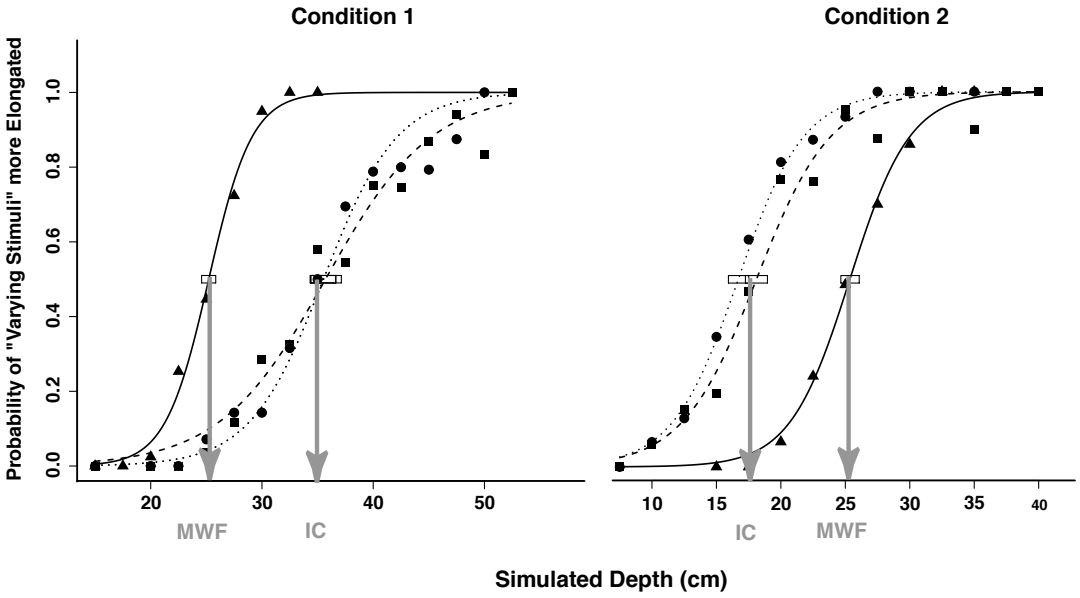

Fig. 8. Estimated probability of an "elongated" response as a function of the simulated cylinder depth (Δz), in conditions 1 and 2 of Experiment 1. The 50% level of the fitted functions are taken to be the points of subjective equality (PSEs). The boxes represent the estimated 95% CIs around the PSEs. The gray arrows indicated the predictions of the IC and MWF models for the comparisons between the single-cue and the combined-cues stimuli (dashed and dotted lines). 'Condition 1: The dashed and dotted lines represent, respectively, the disparity-only and velocity-only stimuli that were varied within the staircase. Condition 2: The dashed and dotted lines represent, respectively, the disparity-only and velocity-only stimuli that were kept constant within the staircase. For both conditions 1 and 2, the continuous lines represent the comparisons between two (consistent) velocity–disparity stimuli (control conditions).

the combined-cues stimuli (see below). If the variability of performance is linked to the absolute value of the stimulus (according to a Weber fraction rule), then it may be expected that the data may slightly deviate from the theoretical values.

In *condition 2*, the estimated standard deviations are the following (the 95% CIs are reported in parenthesis): $\hat{\sigma}_v = 4.895(4.049–6.186)$; $\hat{\sigma}_d = 4.230(3.449–5.470)$; $\hat{\sigma}_c = 3.863(3.161–4.968)$. Note how, in this second condition, the variability of observers' responses is the same in all three conditions, as predicted by the IC model. These results cannot be accounted for by the MWF model, which yields the same predictions as for condition 1.

5.2.2. Points of subjective equality

In *condition 1*, when the combined-cues stimulus was matched with itself, the estimated PSE was equal to 25.210 (95% CI: 24.575–25.845), thus indicating that our procedure was unbiased. When the motion-only stimulus was matched with the fixed combined-cues stimulus, the estimated PSE was equal to $25 \times \sqrt{2} = 35.775$(95% CI : 34.413–37.136); for the stereo-only stimulus, the estimated PSE was equal to 35.491 (95% CI: 34.344–36.638). In both these latter cases, the value of 35.355 predicted by the IC model is contained within the 95% CIs.

Also in *condition 2* performance was unbiased, when the combined-cues stimulus was matched with itself (PSE = 25.367, 95% CI: 24.654–26.080). When the combined-cues stimulus was matched to the motion-only stimulus, the estimated PSE was equal to 18.103 (95% CI: 17.249–18.958). For the stereo-only stimulus, the estimated PSE was equal to 16.721 (95% CI: 15.931–17.511). Also in this case, the value of $\frac{25}{\sqrt{2}} = 17.678$ predicted by the IC model is very closely approximated by the empirical settings.

Note that none of the critical results concerning the PSEs, neither in condition 1 nor in condition 2, can be predicted by the MWF model, which predicts unbiased estimates for all the comparisons. Conversely, a very good fit between the model's predictions and the data by was achieved by the IC model without any free parameters.

6. Experiment 2

In Experiment 2, the IC model was tested by using a "cue-conflict" paradigm (Young, Landy, & Maloney, 1993). For the MWF approach, cues are in conflict when different cues specify unequal amounts of metric depth, a situation which may arise from measurement error and internal noise of the system. The size of cue-conflicts are of limited extent, given that the metric estimates derived from independent depth-processing modules are assumed to be unbiased. No specific predictions are made for "large" amounts of cue conflict, apart from the hypothesis that some operation akin to vetoing may take place (Landy et al., 1995). The characteristics of such a process, however, are left unspecified—in the case of only two cues, for example, which one must be discarded?

For the IC model, conversely, cue conflict arises only when different depth cues specify different affine structures. The situation in which different cues specify the same affine structure, but different amounts of metric depth, is considered compatible with the natural viewing conditions. If stereo and motion specify two cylinders having different front-to-back scaled depth z_d and z_v, then their relative disparities and velocities are:

$$d \approx \mu z_d + \varepsilon_d, \tag{34}$$

$$v \approx \omega z_v + \varepsilon_v, \tag{35}$$

and the scaled signals are:

$$\bar{d} \approx \bar{\mu} z_d + \varepsilon, \tag{36}$$

$$\bar{v} \approx \bar{\omega} z_v + \varepsilon, \tag{37}$$

where $\varepsilon \sim \mathcal{N}(0,1)$.

Since the velocity-specified and disparity-specified cylinders have the same affine structure (but different amounts of metric depths), the two signals are related to each other in a linear fashion. For such a stimulus, the scores on the first principal component PC_1 will be:

$$\rho = \sqrt{\bar{d}^2 + \bar{v}^2} = \sqrt{\bar{\mu}^2 z_d^2 + \bar{\omega}^2 z_v^2}. \tag{38}$$

In this experiment, observers were shown computer renderings of cylinders in which the stereo- and motion-signals could simulate depth extents that were different. In each trial, observers were asked to judge whether the simulated cylinder appeared to be flatter or more elongated in depth relative to a cylinder with circular cross section. Such a task, devised by Johnston (1991), is called the apparently circular cylinder (ACC) task. The ACC was estimated as the 50% point on the resulting psychometric function. Five stereo- and motion-depth pairs were presented over three viewing distances (Fig. 9).

A first test of the IC model concerns HP_1: A constant value of ρ gives rise to the same depth percept, regardless of whether different signals simulate different metric depth magnitudes. According to the IC model, therefore, an ACC should be perceived for all combinations of stereo and motion signals that correspond to a constant value of ρ (see Eq. (5)).

A second test of the IC model concerns the opposite case (HP_2): Variation of ρ is expected to give rise to different depth percepts. This hypothesis was tested by manipulating the viewing distance. The observers were requested to adjust the perceived z-depth ($\Delta z'$) so as to match it to the perceived radius (r'): $\Delta z' = r'$. The perceived radius of the cylinder is equal to the simulated radius r multiplied by the ratio between the perceived and the simulated fixation distance: $r' = r\frac{z'_f}{z_f}$. Upon dividing both terms ($\Delta z'$ and r') by z'_f, we obtain: $\frac{\Delta z'}{z'_f} = \frac{r}{z_f}$. By definition, the ratio $\Delta z'/z'_f$ is equal to the perceived scaled depth z'. Since the simulated radius $r = 25$ mm was kept constant for the three viewing distances, the perceived scaled depth z' must thus vary as an inverse function of the simulated distance z_f. Even though we cannot make a precise quantitative statement about the relation between z' and ρ, since the distribution of R is unknown (see Eq. (10)), we can however say that z' and ρ must be related by a monotonically increasing function. We can thus conclude that ρ must also be an inverse function of the simulated distance z_f. The ρ values estimated from the observers settings' for the three viewing distances, therefore, are expected to vary as an inverse function of z_f.

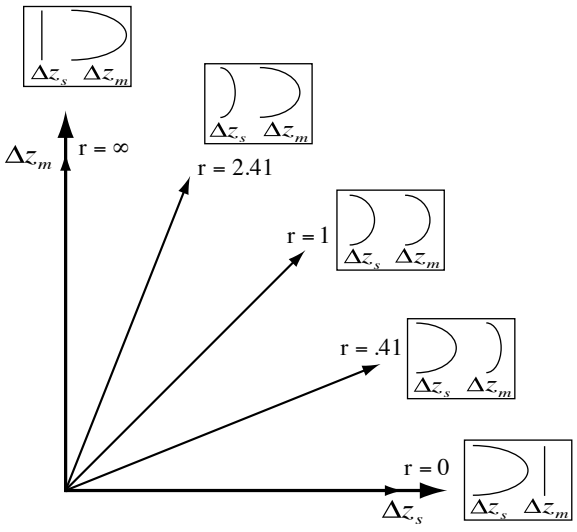

Fig. 9. Experiment 2: Five ratios of simulated depth-from-motion to depth-from-stereo. For the condition $r = 0$, the cylinder is specified by stereo information, while motion information is consistent with a flat surface. In $r = .41$, stereo information specifies a cylinder that is more elongated than does motion information. The condition $r = 1$ is a congruent stimuli situation, where stereo and motion information simulate the same cylinder depth. In $r = 2.41$, stereo information indicates less depth than motion information. Finally, in the condition $r = \infty$, motion specifies the cylinder while stereo information specifies a flat surface.

6.1. Method

6.1.1. Observers

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

6.1.2. Procedure

The haploscope was adjusted for each observer's interocular distance (IOD) and the proper vergence angle was chosen for simulating each viewing distance. Observers performed a two-alternative forced-choice task: they were asked to decide if the cylinder was flatter or elongated in depth than an apparently circular cylinder[6]. Observers viewed each stimulus presentation binocularly for as much time as necessary to make their decision. A response

[6] In Experiment 1, instead, observers performed a 2AFC: they were asked to decide which of two simulated cylinders appeared to be more elongated in depth.

automatically triggered the next presentation. No feedback was provided.

Five ratios $r = \frac{\Delta z_v}{\Delta z_d}$ of depth-from-motion and depth-from-stereo were chosen (Fig. 9). The line designated $r = \infty$ represents the stimulus condition where motion information specifies the cylinder, and depth specified by stereo is flat. In the next condition, the ratio of motion to stereo is still greater than one ($r = 2.41$), thus there is more depth-from-motion than stereo. The central dashed line ($r = 1$) represents the third stimulus condition, where depth-from-stereo and depth-from-motion are congruent. By the fourth condition ($r = .41$), depth-from-stereo is greater than the depth specified by motion (the ratio is less than one). Finally in the fifth condition ($r = 0$), the cylinder is specified by stereo, while motion defines a nearly flat surface. The ACC in each of these five experimental conditions was determined for every subject over three viewing distances of 50, 100, and 200 cm (2D cylinder height subtending 5.73°, 2.86°, and 1.43° of visual angle, respectively). Five cylinder elongations were generated and presented in random order. Overall, there were 100 trials per psychometric function.

In each trial, observers judged whether the depth of a cylinder was greater than or less than half its height ($h = 25$ cm) in a binary forced-choice task. With the method of constant stimuli, the data for one psychometric function were collected for each of the five IC lines represented in Fig. 5. On each IC line, 5 points were chosen. The five Δz_v, Δz_d values of each IC line were determined in a preliminary calibration stage for each observer, so as to optimize the shape of the resulting psychometric function. With three viewing distances and five IC lines, 15 points of subjective equality were measured. The ACC was estimated as the 50% point on the resulting psychometric functions.

The five Δz_v, Δz_d of each psychometric function were presented for 20 times to each observer. The 20 presentations were broken into four blocks as follows: stimulus conditions were codified as A ($r = \infty$), B ($r = 2.41$), C ($r = 1$), D ($r = .41$), and E ($r = 0$). The five conditions were then randomly ordered and presented four times for each of the three viewing distances.

6.2. Results

The present experiment was motivated by what we can call the "metamerism" hypothesis, whereby different combinations of disparity and motion information are expected to give rise to equivalent depth percepts for constant ρ values. For each viewing distance, therefore, we expect that observers, in order to perceive a cylinder with circular cross section, would select (in the different IC lines) disparity and velocity values corresponding to the same ρ magnitude. By definition, $\rho = \sqrt{\bar{d}^2 + \bar{v}^2}$, where $\bar{d}$ and $\bar{v}$ are the disparity and velocity signals scaled by their measurement error.

The present experimental manipulations do not provide a direct measure of the measurement errors for disparity and motion signals. By using hyperacuity tasks, however, previous research has revealed that stereo acuity is in the order of 10 arcsec (Lappin & Craft, 2000; McKee, Levi, & Bowne, 1990; McKee, Welch, Taylor, & Bowne, 1990) and, thus, such a value can be used as an estimate of the measurement error for disparity signals[7]. In the present context, moreover, the ratio between the slopes of the psychometric functions in the motion-only and disparity-only conditions informs us about the ratio between the measurement errors of these two signals. The slopes of the psychometric functions ($\hat{\sigma}_m$ for the motion-only condition and $\hat{\sigma}_v$ for the velocity-only condition), in fact, depend on two sources of variability: (i) the measurement errors for each signal, and (ii) other factors such as task difficulty, "internal noise", lapses of attention and so on. If we assume that the second source of variability is constant across the two conditions, then the ratio between $\hat{\sigma}_v$ and $\hat{\sigma}_d$ estimates the ratio between the measurement errors of the two signals. By using the estimate reported by Lappin and Craft (2000) for the disparity measurement error, the measurement error for the velocity signals was computed for each observer as $10 \times \frac{\hat{\sigma}_v}{\hat{\sigma}_d}$. The values of $\hat{\sigma}_v$ and $\hat{\sigma}_d$ were computed from the single-cue conditions, for each observer, by using the procedure of Wichmann and Hill (2001).

Having scaled the values collected by means of the staircase procedure by the ratio between $\hat{\sigma}_m$ and $\hat{\sigma}_d$, the magnitude of ρ was estimated from the mean ACC settings for each IC line and each viewing distance. The analysis was performed by means of a generalized linear mixed model, as in Experiment 1. The resulting $\hat{\mu}_{PSE}$ are reported in Fig. 10 together with their estimated 95% confidence intervals. As the Figure indicates, the "metamerism" hypothesis was confirmed: At each viewing distance, an ACC was perceived when different combinations of disparity and velocity signals defined the same value of ρ.

A second test of the model concerns the opposite case: A variation of ρ was expected to give rise to different depth percepts. This hypothesis was tested by manipulating the viewing distance. As described in the introduction of the experiment, the ρ values estimated from the observers settings' for the three viewing distances were expected to vary as an inverse function of z_f. The results reported in Fig. 10 are consistent with this prediction. Once more, we stress the fact that the predictions of the IC model are generated with no free parameters.

[7] Note that stereo acuity is not a singular value. It varies with the distance of the target from the fixation plane, as Ogle (1953), Blakemore (1970), Westheimer and McKee (1978), Badcock and Schor (1985) and others have found. Secondly, the disparity noise varies with the separation between the targets. McKee et al. (1990) found that the disparity noise increased proportionately with the disparity separating the targets. Glennerster and McKee (1997) confirmed this finding. The general answer for velocity noise is that it is about 5% of the speed from 1 to 60 deg/s. The estimate of the measurement error used in the present simulation, therefore, is only a first approximation.

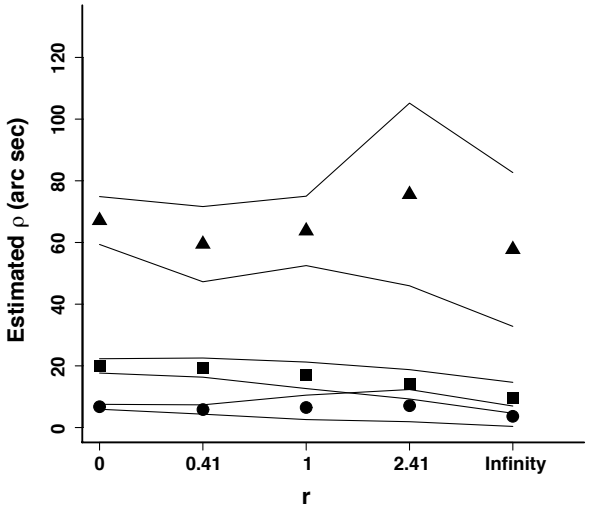

Fig. 10. Experiment 2: Apparently circular cylinders (ACCs) corresponding to five ratios of simulated depth-from-motion and depth-from-stereo, for three viewing distances (50 cm: bottom: 100 cm: middle; 200 cm: top). The lines around the ACCs represent the estimated 95% CIs.

7. General discussion

The cue-combination model that we propose is based on the assumption that the visual system exploits the linear relationship between the disparities and the velocities projected by a rigidly-moving object. As described in Section 1, this linear relationship, identified in a one-dimensional disparity–velocity sub-space, is called the *intrinsic constraint* line. Our main assumption is that the visual system does not process the disparity and velocity signals independently, but rather derives 3D shape directly from the IC manifold.

To test this hypothesis, we run two experiments in which observers made judgments of 3D shape for displays simulating stereo- and motion-specified cylinders that could either be elongated or flattened along the line of sight (z-axis). Experiment 1 addressed the issue of comparing the IC and MWF models in conditions in which they make different predictions. This test was carried out by considering both the amount of depth perceived by observers and the variability of their performance. Concerning the perceived depth magnitudes, the MWF model predicts unbiased performance; the IC model, conversely, predicts systematic biases. Results from Experiment 1 clearly indicate that observers' performance is biased, and are in very good agreement with the predictions of the IC model. Concerning the variability of observers' performance, it is very important to determine whether the IC model can fit the data in the literature as well as the MWF model does. The empirical support to the MWF model, in fact, comes mainly from the finding that observers, in combining different sources of depth information, comply with the optimal combination rule of Eq. (25). In the introduction, we showed analytically that the IC model yields the same predictions (Eq. (33)) as those of Eq. (25), if the experiment is carried out by means of the procedure termed condition 1. The data of condition 1 showed, indeed, that the variability of observers' judgments was consistent with the predictions of Eqs. (25) and (33). In condition 2, however, only the IC model accounts for the variability of observers' performance (the MWF model does not).

In Experiment 2, we determined when stereo information, motion information, or a combination of the two produced the perception of an apparently circular cylinder (the ACC task, see Johnston, 1991). Observers were presented with displays simulating random-dot cylinders in which the z-axis elongation (Δz) specified by binocular disparity could be different from the amount of elongation specified by motion. The purpose of the experiment was to identify the different combinations of stereo-specified depth (Δz_d) and motion-specified depth (Δz_v) that are perceived as apparently circular cylinders. The IC model predicts that perceived depth depends on the length of the vector $\rho = \sqrt{\bar{d}^2 + \bar{v}^2}$, where $\bar{d}$ and $\bar{v}$ are the disparity and velocity signals scaled by their measurement errors. We found that the observers' settings were in good agreement with predictions of the IC model: An ACC was perceived for constant values of ρ, regardless of the orientation of the IC line.

Our proposal provides an alternative to the currently most accepted model of depth cue-integration, termed Modified Weak Fusion (see Landy et al., 1995). The MWF model posits a modular architecture of the visual system, in which specialized independent modules are devoted to estimating 3D depth from different image signals. Two processing stages are hypothesized: in a first stage, termed *promotion*, the depth estimates from separate cues are brought to a common metric representation. Because of the noise in the image measurements and in processing, each module produces a different estimate of 3D shape. In a second stage, therefore, these different 3D estimates must be combined. The problem of finding an optimal combination of such 3D estimates finds its solution in a weighted average of the outputs of the different modules, with weights inversely proportional to the reliability of each cue (Clark & Yuille, 1990). Since the MWF model provides the current framework for depth cue combination, in the following we will discuss such a model in more details.

7.1. Cue promotion

The clearest specification of how the promotion stage may work has been provided by Richards (1985), who showed that the simultaneous presence of disparity and velocity is sufficient, in principle, to provide metrical depth without recourse to extra-retinal scaling information. Even though Richard's proposal is very appealing, the available psychophysical results do not seem to support it. Several reports have shown, for example, that metric judgements do not become more veridical when motion is added to a stereo-specified display (Landy & Brenner, 2001). Such results suggest that the visual system does not provide a veridical interpretation to a 2D projection, even when the

simultaneous presence of motion and disparity information make such an interpretation possible. In other published studies, the addition of motion to a stereo-specified display increased performance (e.g. Johnston et al., 1994). By examining those studies more carefully, however, we can point out several problematic methodological aspects.

(1) In some studies, the presence or absence of motion signals covaried with the presence or absence of other cues, such as texture gradient and object-boundary foreshortening—i.e., texture gradients and deforming occluding boundaries had a consistent depth interpretation with motion (e.g. Brenner & Landy, 1999; Johnston et al., 1994). It is not clear, therefore, to what extent the more veridical estimates found in the combined stereo-motion condition can be ascribed to the specific effect of the motion signals.
(2) In the studies in which other cues-to-depth (such as texture gradient and foreshortening) were controlled, static stereo displays were underestimated (Hibbard & Bradshaw, 2002). It is not clear, therefore, whether the addition of motion to a stereo-specified display makes observers' judgments more veridical, or whether the addition of motion simply increases the apparent depth of the perceived shapes. In this second case, the addition of motion would also make judgments in a combined condition *less* veridical, if the appropriate stimulus settings were chosen so as to induce a veridical solution when stereo-only information is provided. The results of the experiments of Tassinari, Domini, and Caudek (2005) support this second possibility.

Does this mean that a process akin to promotion does not take place? Not necessarily. The present results may still be consistent with the hypothesis that separate modules may mutually constrain each other. What the present results do not support is the idea that each module solves an inverse-geometry problem (in the sense of the computer-vision algorithms), and that the interactions among modules necessarily produce a more veridical estimate of 3D shape. Finally, it must be pointed out that the first stage of processing that we hypothesize is similar, in spirit, to the promotion stage proposed by Landy et al. (1995). The subspace of the *intrinsic constraint line* necessarily brings stereo and motion signals to a single representational format, since points on the IC line define a unique (though not necessarily veridical) 3D map.

7.2. Flatness cues

Proponents of the MWF approach ascribe a great importance to flatness cues. In many experimental situations, vergence, accommodation, prior knowledge, pixelization, and so on, may signal that the stimulus display is flat (which, in fact, it is). The use of computer displays may thus introduce potential artifacts which contribute to the poor performance in depth constancy experiments (Banks, Watt, & Ernst, 2002; Bradshaw et al., 1998; Buckley & Frisby, 1993; Frisby et al., 1995). For example, Frisby, Buckley, and Duke (1996) reported an improvement by almost three times in the Weber fractions for a 3D length discrimination task when they replicated the experiment of Todd and Bressan (1990) by using "real-world" stimuli. Note, however, the comment of Bradshaw et al. (2000): "It is interesting to note in this regard that performance in Frisby et al.'s monocular viewing conditions remained surprisingly good (although no recognized depth cue was present): this would have been unlikely in the experiments of Todd and Bressan." Other investigations reveal that observers often report a larger amount of depth in real world situations with well-illuminated environments than in equated computer-generated displays (Bradshaw et al., 2000; Durgin, Proffitt, Olson, & Reinke, 1995; Tittle & Braunstein, 1993).

How can the IC model account for the effects of flatness cues? Remember that the outcome of the IC model depends solely on the first principal component. The scores on PC_1 have high loadings on the variables that are highly correlated among themselves, and negligible loadings on the variables uncorrelated with the previous ones. For example, in the case of a 3D shape simulated with disparity and velocity, these two cues will be highly correlated (since they are generated by the same 3D structure), while the flatness cues will be uncorrelated with the disparity and velocity signals (since they are generated by a different 3D structure—the flat surface of the monitor). The scores on PC_1 (i.e. the input to the IC model), will therefore be unaffected by all those sources of depth information which signal the flatness of the monitor display.

Let us suppose that two (consistent) cues (say, disparity and velocity) specify a 3D shape, whereas other cues (such as blur, accommodation and so on) specify the fronto-parallel surface of the monitor. In such circumstances, the ρ value associated to the tip-to-bottom depth of the simulated object takes on a given value, let say ρ_1. Let us now consider a real object that is compared with the computer simulation. For such an object, all the available cues—disparity and velocity, on the one side, and blur, accommodation and so on, on the other—specify the same depth. It is easy to understand that, in the multidimensional space of signals (with blur, accommodation and so on being further dimensions with non zero intensities), the length of the resulting ρ_2 vector (associated with the tip-to-bottom depth of the target object) will be larger than ρ_1. In this way, the IC model can explain the increased amount of depth found for "real-world" stimuli.

7.3. Modular architecture or strong fusion?

Does the model that we propose represent a departure from the fundamental assumption of a modular organization to the visual system? Not necessarily. What characterizes our proposal is the assumption that the perceptual analysis relies on the deterministic or statistical relationships present among the retinal signals. For example, retinal

velocities and binocular disparities are related to each other in a linear fashion, because they are both produced by the same 3D rigid object (or to a local patch of a non-rigid object). Points belonging to the IC line, in fact, codify an actual depth map. In a similar fashion, the relationship among first-order gradients (texture-gradient, stereo-gradient, motion-gradient) is related to local surface orientation. We speculate, therefore, that the visual system may still be organized in a modular fashion, but with visual modules specialized for the processing of *specific 3D world properties*, rather than for the processing of separate retinal signals.

7.4. Open questions

We conclude by noting that the IC model, at this stage, is not yet fully specified, and that its psychological plausibility needs to be confirmed by future research. From a theoretical standpoint, several issues need to be addressed.

(1) The IC model, in its present form, does not address the issue of how extra-retinal signals affect perceptual performance. From Eq. (7), for example, we can see that ρ depends on z_f (since it is linearly related to the scaled depth map) and it is reasonable to assume that both retinal (e.g., vertical disparities) and extra-retinal (e.g., vergence angle) signals affect the perceptual estimate of z_f (e.g. Brenner, Smeets, & Landy, 2001; Glennerster, Rogers, & Bradshaw, 1998). Presently, we have not specified in which manner extra-retinal signals may be used to estimate z_f. Once extra-retinal signals are fixed (as for the present experiments), we expect that the perceptual outcome depends only on ρ. In other situations, however, vergence angle may vary and may affect the perceptual outcome.

(2) In its current form, the IC model is limited to stimuli subtending a small visual angle (for which vertical disparity information is ineffective). We are not concerned, therefore, with the increase of veridicality of depth constancy that has been found when vertical disparity information is added to the stimulus displays (e.g. Rogers & Bradshaw, 1995). It will be a task for future research to model the perceptual effects ensuing from vertical disparity and acceleration varying over a larger range of values than those here considered (so as to be perceptually effective).

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*, 257–262.

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.

Badcock, D. R., & Schor, C. M. (1985). Depth-increment detection function for individual spatial channels. *Journal of the Optical Society of America A, 2*, 1211–1216.

Banks, M. S., Watt, S. J., & Ernst, M. O. (2002). Screen cues to flatness affect 3D percepts [Abstract]. *Journal of Vision*, *2*(10), 1a, http://journalofvision.org/2/10/1/, doi:10.1167/2.10.1.

Blakemore, C. (1970). The range and scope of binocular depth discrimination in man. *Journal of Physiology, 211*, 599–622.

Bradshaw, M. F., Parton, A. D., & Eagle, R. A. (1998). The interaction of binocular disparity and motion parallax in determining perceived depth and perceived size. *Perception, 27*, 1317–1333.

Bradshaw, M. F., Parton, A. D., & Glennerster, A. (2000). The task-dependent use of binocular diparity and motion parallax information. *Vision Research, 40*, 3725–3734.

Brenner, E., & Landy, M. S. (1999). Interaction between the perceived shape of two objects. *Vision Research, 39*, 3834–3848.

Brenner, E., Smeets, J. B. J., & Landy, M. S. (2001). How vertical disparities assist judgements of distance. *Vision Research, 41*, 3455–3465.

Brenner, E., & van Damme, W. J. M. (1998). Judging distance from ocular convergence. *Vision Research, 38*, 493–498.

Buckley, D., & Frisby, J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Research, 33*, 919–933.

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

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

Clark, J., & Yuille, A. (1990). *Data fusion for sensory information processing systems*. Boston, MA: Kluwer.

Cuijpers, R. H., Kappers, A. M., & Koenderink, J. J. (2000). Investigation of visual space using anexocentric pointing task. *Perception & Psychophysics, 62*, 1556–1571.

Cumming, B. G., Johnston, E. B., & Parker, A. J. (1991). Vertical disparities and perception of 3-dimensional shape. *Nature, 349*, 411–413.

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

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. (2003a). Perception of slant and angular velocity from a linear velocity field modeling and psychophysics. *Vision Research, 43*, 1753–1764.

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

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

Domini, F., Caudek, C., & Proffitt, D. R. (1997). Misperceptions 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., & Skirko, P. (2003). Temporal integration of motion and stereo cues to depth. *Perception & Psychophysics, 65*, 48–57.

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

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

Durgin, F. H., Proffitt, D. R., Olson, T. J., & Reinke, K. S. (1995). Comparing depth from motion with depth from binocular disparity. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 679–699.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429–433.

Frisby, J. P., Buckley, D., & Duke, P. A. (1996). Evidence for good recovery of lengths of real objects seen with natural stereo viewing. *Perception, 25*, 129–154.

Frisby, J. P., Buckley, D., Wishart, K. A., Porrill, J., Garding, J., & Mayhew, J. E. W. (1995). Interaction of stereo and texture cues in the perception of 3-dimensional steps. *Vision Research, 35*, 1463–1472.

Garding, J., Porrill, J., Mayhew, J. E. W., & Frisby, J. P. (1995). Binocular stereopsis, vertical disparity and relief transformations. *Vision Research, 35*, 703–722.

Gepshtein, S., & Banks, M. S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology, 13*(6), 483–488.

Glennerster, A., & McKee, S. P. (1997). Sensitivity to depth or lateral displacement on a slanted reference plane. *Investigative Ophthalmology and Visual Science, 38*, S907.

Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1996). Stereoscopic depth constancy depends on the subject's task. *Vision Research, 36*, 3441–3456.

Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1998). Cues to viewing distance for stereoscopic depth constancy. *Perception, 27*, 1357–1365.

Hecht, H., van Doorn, A., & Koenderink, J. J. (1999). Compression of visual space in natural scenes and in their photographic counterparts. *Perception & Psychophysics, 61*, 1269–1286.

Hibbard, P. B., & Bradshaw, M. F. (2002). Isotropic integration of binocular disparity and relative motion in the perception of three-dimensional shape. *Spatial Vision, 15*, 205–217.

Hildreth, E. C. (1984). *The measurement of visual motion*. Cambridge, MA: MIT Press.

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.

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Reseach, 31*, 1351–1360.

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

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago, IL: University of Chicago Press.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research 43*(24), 2539–2558.

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

Koenderink, J. J., van Doorn, A. J., Kappers, A. M., & Todd, J. T. (2002). Pappus in optical space. *Perception & Psychophysics, 64*, 380–391.

Landy, M. S., & Brenner, E. (2001). Motion–disparity interaction and the scaling of stereoscopic disparity. In L. R. Harris & M. R. M. Jenkin (Eds.), *Vision and attention* (pp. 129–151). New York: Springer Verlag.

Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A, 18*, 2307–2320.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. J. (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.

Lind, M., Bingham, G. P., & Forsell, C. (2002). The illusion of perceived metric 3D structure, *infovis*. IEEE (p. 51).

Liter, J. C., Braunstein, M. L., & Hoffman, D. D. (1993). Inferring structure from motion in two-view and multiview displays. *Perception, 22*, 1441–1465.

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

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

Loomis, J. M., Da Silva, J. A., Fujita, N., & Fukusima, S. S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 906–921.

Loomis, J. M., & Philbeck, J. W. (1999). Is the anisotropy of perceived 3-D shape invariant across scale? *Perception & Psychophysics, 61*, 397–402.

McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research, 30*, 1763–1779.

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.

Mon-Williams, M., Tresilian, J. R., & Roberts, A. (2000). Vergence provides veridical depth perception from horizontal retinal image disparities. *Experimental Brain Research, 133*, 407–413.

Norman, J. F., Lappin, J. S., & Norman, H. F. (1996). The visual perception of 3D length. *Journal of Experimental Psychology Human Perception and Performance, 22*, 173–186.

Ogle, K. N. (1953). Precision and validity of stereoscopic depth perception from double images. *Journal of the Optical Society of America, 43*, 907–913.

R Development Core Team. (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project. org.

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

Richards, W., & Lieberman, H. (1985). Correlation between stereo ability and the recovery of structure from motion. *American Journal of Optometry and Physiological Optics, 62*, 111–118.

Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature, 361*, 253–255.

Rogers, B. J., & Bradshaw, M. F. (1995). Disparity scaling and the perception of frontoparallel surfaces. *Perception, 24*, 155–179.

Tassinari, H., Domini, F., & Caudek, C. (2005). The intrinsic constraint model for stereo-motion integration (submitted).

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

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). The systematic distortion of perceived 3D structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 663–678.

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.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S.* (fourth ed.). Springer.

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology, 73*, 117–129.

Westheimer, G., & McKee, S. P. (1978). Stereoscopic acuity for moving retinal images. *Journal of the Optical Society of America, 68*, 450–455.

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

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

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.